AC-CDCN:A Cross-Subject EEG Emotion Recognition Model with Anti-Collapse Domain Generalization

Yubin Sun (23031212428@stu.xidian.edu.cn)
Liying Yang[1] (yangliying1208@163.com)
Huanyu He (23031212455@stu.xidian.edu.cn)
Jingtao Du (jtdu@stu.xidian.edu.cn)
Department of Computer Science and Technology, Xidian University, Xi'an, China

Abstract

Emotion recognition is a critical area in brain-computer interfaces, with electroencephalography (EEG) shown to be effective for emotional analysis. In domain generalization, cross-subject emotion recognition encounters significant generalization challenges, including excessive feature collapse and insufficient capture of EEG features. To tackle these issues, we propose an Anti-Collapse Cross-Domain Consistency Network (AC-CDCN), which leverages Maximum Mean Discrepancy (MMD) to reduce distribution discrepancies between source domains, facilitating the capture of domain-invariant features, and innovatively introduces an Anti-Feature Collapse Strategy (AFCS), which incorporates an Anti-Collapse Domain Discriminator (ACDD) and the code rate loss function, effectively preventing excessive feature collapse. Furthermore, we propose a Flexible Feature Rebalance Module (FlexiReMod), a plug-and-play component that enhances generalization and dynamic feature capture through feature fusion and attention mechanisms. Experimental results indicate AC-CDCN achieved 87.14% (±5.60) and 71.77% (±12.92) accuracy on SEED and SEED-IV datasets, underscoring its significant generalization advantage.

Keywords: Emotion recognition; EEG; Anti-Collapse; Cross-subject; Domain generalization

Introduction

Emotions play an essential role in everyday life, influencing not only communication styles but also personal growth (Picard & Klein, 2002). Affective computing, a branch of AI, enables computers to respond to emotions by identifying, analyzing, and modeling them (Devillers & Cowie, 2023). This technology has broad applications in healthcare, education, and multimedia, enhancing human-computer interaction (Katsigiannis & Ramzan, 2017). Consequently, the field has attracted extensive research interest.

In the domain of emotion recognition research, non-physiological signal-based methods, which focus on behavioral cues such as facial expressions and vocal intonations, have garnered significant attention (Vatcharaphrueksadee, Maliyaem, Viboonpanich, & Phuangkamnerd, 2022). These methods are straightforward and do not require specialized equipment for analyzing facial expressions or vocal modulations. However, their accuracy may be limited due to individuals masking their emotions (Tian, 2023). In contrast, emotion recognition based on physiological signals, such as heart rate, respiration, and skin conductivity, provides a more reliable approach (Picard, Vyzas, & Healey, 2001). Emotion recognition based on the central nervous system is increasingly recognized for its low artifact sensitivity and high accuracy. Among these methods, electroencephalography (EEG) is particularly preferred due to its excellent temporal resolution and cost-effectiveness (Sun, Yang, Wang, Liu, & Ni, 2023).

EEG signals have been demonstrated to correlate with specific stimuli in certain channels and frequency bands, forming the basis for accurate emotion recognition and the detection of subtle emotional patterns (Zheng & Lu, 2015). However, individual differences in skull structures and cognitive processes result in varying sensitivities to the same emotion, leading to substantial distributional disparities across subjects. This variability complicates cross-subject emotion classification and hinders the transferability of models trained on source-domain data to target-domain subjects. Consequently, enhancing model generalization remains a critical challenge.

To address this issue, researchers have explored various strategies, and transfer learning methods have been widely adopted to reduce individual variations in EEG-based emotion recognition (Xu, Dang, Wang, & Zhou, 2023). Early studies suggested that acquiring large volumes of unlabeled data or limited labeled data from target subjects could help mitigate distributional discrepancies between source and target domains. This goal was pursued through unsupervised and semi-supervised domain adaptation techniques (Luo & Lu, 2021; J. Li, Qiu, Shen, Liu, & He, 2019). However, these methods are constrained by their reliance on target subject data, which is often inaccessible or unavailable for training.

Domain generalization has attracted considerable attention in recent years. In contrast to semi-supervised learning and domain adaptation, the objective of domain generalization is to develop a model capable of effectively generalizing across both observed and unobserved data distributions. Song, Zheng, Song, and Cui (2018) proposed DGCNN, and Zhong, Wang, and Miao (2020) introduced RGNN, both of which optimized the adjacency matrix to model EEG channel relationships and achieve good generalization. Recently, adversarial domain generalization (ADG) methods have become a popular choice. Inspired by GAN (Goodfellow et al., 2014), ADG methods optimize the feature space via adversarial learning, thereby improving model generalization. Ma, Li, Zheng, and Lu (2019) employed the gradient reversal technique (Ganin et al., 2016) to confuse the domain dis-

[1]Corresponding author.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

criminator, enabling the extraction of domain-invariant feature representations. Drawing on Khosla, Zhou, Malisiewicz, Efros, and Torralba (2012) and integrating DG-DANN (Ganin et al., 2016), they introduced DResNet, which markedly improved model generalizability. Y. Li et al. (2020) developed a methodology with subnetworks and an adversarial discriminator to extract emotion-related domain-invariant features, effectively improving the generalization performance of EEG emotion recognition.

Although theoretically sound, recent benchmarks reveal practical challenges. Yu, Chan, You, Song, and Ma (2020) observed that excessive collapse of the feature space by extractors can mislead the discriminator. Furthermore, Zhu et al. (2022) highlighted that this collapse can lead to overfitting, thus constraining the model's generalizability.

In this paper, we propose an Anti-Collapse Cross-Domain Consistency Network (AC-CDCN) to address the aforementioned issue. This model leverages adversarial learning at the macro level and draws on maximum coding rate reduction (Yu et al., 2020) at the micro level to prevent feature collapse, thereby ensuring diversity and enhancing generalization. Additionally, inspired by the SOFTS method (Han, Chen, Ye, & Zhan, 2024), we propose a novel module, FlexiReMod, which seamlessly integrates dynamic feature selection with attention mechanisms to enhance dynamic feature capture. The main contributions of this paper are as follows:

- We propose a Flexible Feature Rebalance Module (FlexiReMod), which enhances dynamic feature capture as a plug-and-play module.
- We innovatively propose an Anti-Collapse Domain Discriminator (ACDD), which ensures feature diversity and significantly enhances model generalization.
- We propose the AC-CDCN, which effectively addresses feature collapse, ensures feature diversity, and demonstrates strong generalization capability.
- Experiments on the SEED (Zheng & Lu, 2015) and SEED-IV (Zheng, Liu, Lu, Lu, & Cichocki, 2018) datasets achieved accuracies of 87.14% (±5.60) and 71.77% (±12.92), demonstrating the superiority of AC-CDCN.

Method

This section focuses on the modeling of the domain generalization problem, the proposed Anti-Collapse Cross-Domain Consistency Network (AC-CDCN), and its three core components: the MMD-based Autoencoder, the Anti-Feature Collapse Strategy, and the Flexible Feature Reallocation Module.

Problem Definition and Method Overview

In the EEG domain generalization problem, the dataset of each subject can be considered as a source domain $\mathcal{D}_s^n$, defined as: $\mathcal{D}_s^n = \{(\mathbf{x}_j^n, \mathbf{y}_j^n)\}_{j=1}^{n_n}$ where $\mathbf{x}_j^n$ represents the j-th EEG sample of the n-th subject, $n \in 1,2,\ldots,N$, y_j^n denotes the corresponding emotion label, and n_n is the total number of samples for the n-th subject. The goal of domain generalization is to utilize data from multiple subjects $\mathcal{D}_s = \{\mathcal{D}_s^1, \mathcal{D}_s^2, \ldots, \mathcal{D}_s^N\}$ for training to learn a universal model $f : \mathcal{X} \to \mathcal{Y}$ that performs well on an unseen target domain subject $\mathcal{D}_t = \{(\mathbf{x}_j^t, \mathbf{y}_j^t)\}_{j=1}^{n_t}$ with an unknown data distribution (Blanchard, Lee, & Scott, 2011). Unlike conventional EEG classification tasks, which assume identical training and testing data distributions $\mathcal{P}_{\text{train}} = \mathcal{P}_{\text{test}}$, EEG domain generalization faces the following challenges:

1. There are significant distributional differences between source domains (subjects) $\mathcal{P}_s^i \neq \mathcal{P}_s^j, \forall i \neq j$.
2. The distribution of the target domain (test subjects) is unknown and different from that of the source domains $\mathcal{P}_t \neq \mathcal{P}_s^i$, which is known as a domain shift.

To mitigate overfitting attributable to feature space collapse and bolster generalizability in extant cross-domain generalization approaches, we introduce the Anti-Collapse Cross-Domain Consistency Network (AC-CDCN).

Figure 1 shows the data processing phases of the proposed framework AC-CDCN. When the input data, represented by $\mathbf{X}_{\text{input}}$ is processed by the AC-CDCN, it is first passed through the FlexiReMod module. This module dynamically selects and integrates key features, focusing on important information in the input signals. This ensures that the subsequent feature representations are more stable and representative, providing high-quality feature for the following processing stages. Upon undergoing this process, the data $\mathbf{X}_{\text{Enhanced}}$ is enhanced in both diversity and discriminative power. Subsequently, the enhanced data is fed into an LSTM-based encoder-decoder network, where temporal modeling is employed to capture the dynamic characteristics of the signals, thereby constructing a feature space $\mathcal{FS}$. Specifically, $\mathcal{FS}$ is defined as $\mathcal{FS} = \bigcup_{j=1}^{N} \mathcal{H}^n$, where $\mathcal{H}^n = \{h_j^n \mid j = 1,2,\ldots,n_n\}$ represents the feature distribution corresponding to the n-th subject. The feature space $\mathcal{FS}$ retains critical fine-grained information required for accurate emotion recognition, thereby providing a solid foundation for subsequent classification tasks. To reinforce the domain-invariant properties of features $\mathcal{H}^n$ within $\mathcal{FS}$, we draw inspiration from Gretton, Borgwardt, Rasch, Schölkopf, and Smola (2006) and incorporate the Maximum Mean Discrepancy (MMD) loss $\mathcal{L}_{\text{mmd}}$. This loss function aligns the distributions across multiple source domains, effectively minimizing inter-domain discrepancies within $\mathcal{D}_s^n$ and promoting model consistency across diverse subject datasets. On the macro level, feature distribution guidance is implemented by facilitating interactions between the feature representations $\mathcal{H}^n$ within $\mathcal{FS}$ and the Anti-Collapse Domain Discriminator (ACDD). The ACDD employs adversarial learning to refine the feature space $\mathcal{FS}$, effectively expanding the feature distribution to mitigate over-collapse and enhance diversity. On the micro level, a coding rate loss $\mathcal{L}_{\text{cr}}$ is calculated for each $\mathcal{H}^n$. Optimizing this loss mitigates feature collapse and enhances the cross-domain generalization ability of the model. Finally, the feature representations, denoted by $\mathcal{H}^{1,\ldots,N}$, within the

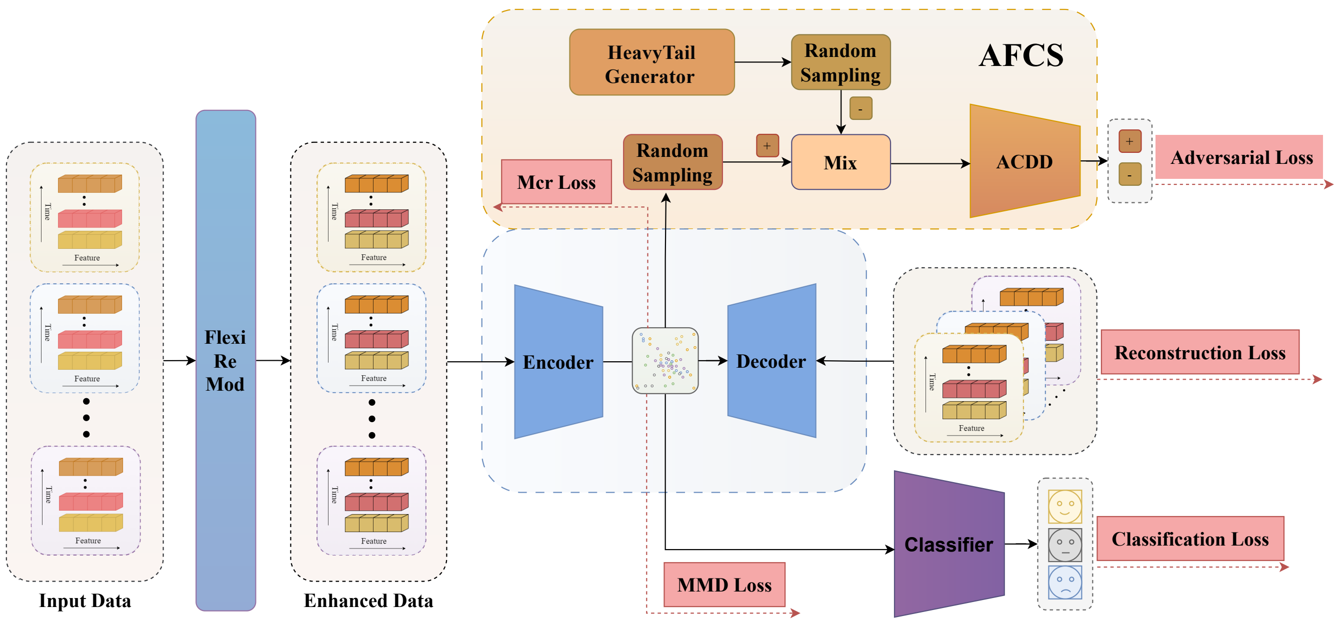

Figure 1: The AC-CDCN model diagram. The solid arrows represent data flow, while the dashed arrows indicate the losses to be computed.

AC-CDCN framework are processed by the classifier *Cls*, to accurately categorise emotional states. Within this framework, FlexiReMod, the classifier, MMD-based Autoencoder, and ACDD function synergistically to maintain feature space consistency and mitigate feature collapse, significantly enhancing the model's robustness, generalization capacity, and classification performance in domain generalization tasks.

Flexible Feature Reallocation Module

To enhance the model's capacity for generalization and dynamic feature extraction in complex electroencephalogram (EEG) signals, we propose a novel module, named as the Flexible Feature Rebalance Module (FlexiReMod). This plug-and-play module employs dynamic feature selection and attention mechanisms to efficiently integrate and amplify relevant information in the input data.

Specifically, let $\mathbf{X}_{input} \in \mathbb{R}^{B\times T\times D}$ represent the input data, where B, D, and T denote the batch size, feature dimension, and time steps, respectively. The sequence representation is projected from the hidden dimension D to the core dimension D_{core} through the feedforward neural network FFN1, yielding the initial core **IC**:

$$\mathbf{IC} = \text{FFN1}(\mathbf{X}_{input}) \in \mathbb{R}^{B\times T\times D_{\text{core}}} \tag{1}$$

Subsequently, the module applies a stochastic pooling mechanism (Zeiler & Fergus, 2013) to perform weighted sampling on **IC**, yielding the selected initial core **SIC**, which is then replicated to obtain the core representation $\mathbf{C} \in \mathbb{R}^{B\times C\times D_{\text{core}}}$. This mechanism dynamically selects features with significant impact on the output. The formulas are as follows:

$$\mathbf{SIC} = \text{Stoch\_Pool}(\text{FFN1}(\mathbf{X}_{\text{input}})) \tag{2}$$

$$\mathbf{C} = \text{Repeat}(\mathbf{SIC}) \tag{3}$$

Building upon this, the module incorporates the attention mechanism from Zhao, Yan, and Lu (2021), which computes the weights on **IC** using a linear transformation and the Softmax function. These weights are then applied to the core representation of the data to obtain $\mathbf{C}_{\text{enhanced}}$, enabling the model to focus on key features and enhance the effectiveness of feature representation. The corresponding formulas are as follows:

$$\alpha = \text{softmax}(\mathbf{WIC} + \mathbf{u}) \tag{4}$$

$$\mathbf{M}_{\text{aug}} = \alpha \odot \mathbf{C} \tag{5}$$

where $\mathbf{W} \in \mathbb{R}^{D_{\text{core}}\times D_{\text{core}}}$ and $\mathbf{u} \in \mathbb{R}^{D_{\text{core}}}$ are the weights and bias for the linear transformation, and $\odot$ denotes element-wise multiplication.

In the feature fusion stage, the original input $\mathbf{X}_{\text{input}}$ is concatenated with the enhanced feature $\mathbf{C}_{\text{enhanced}}$ obtained through the attention mechanism. This concatenated feature is then processed by FFN2 through a linear mapping to perform feature reconstruction. The operations are as follows:

$$\mathbf{X}_{\text{concat}} = \text{Concat}(\mathbf{C}_{\text{enhanced}}, \mathbf{X}_{\text{input}}) \tag{6}$$

$$\mathbf{X}_{\text{enhanced}} = \text{FFN2}(\mathbf{X}_{\text{concat}}) \tag{7}$$

The operation Concat refers to the concatenation of features, and $\text{FFN2} : \mathbb{R}^{D+D_{\text{core}}} \to \mathbb{R}^{D}$ projects the concatenated sequence into the final feature representation, $\mathbf{X}_{\text{enhanced}} \in \mathbb{R}^{B\times T\times D}$, thereby enhancing the model's ability to capture complex data patterns.

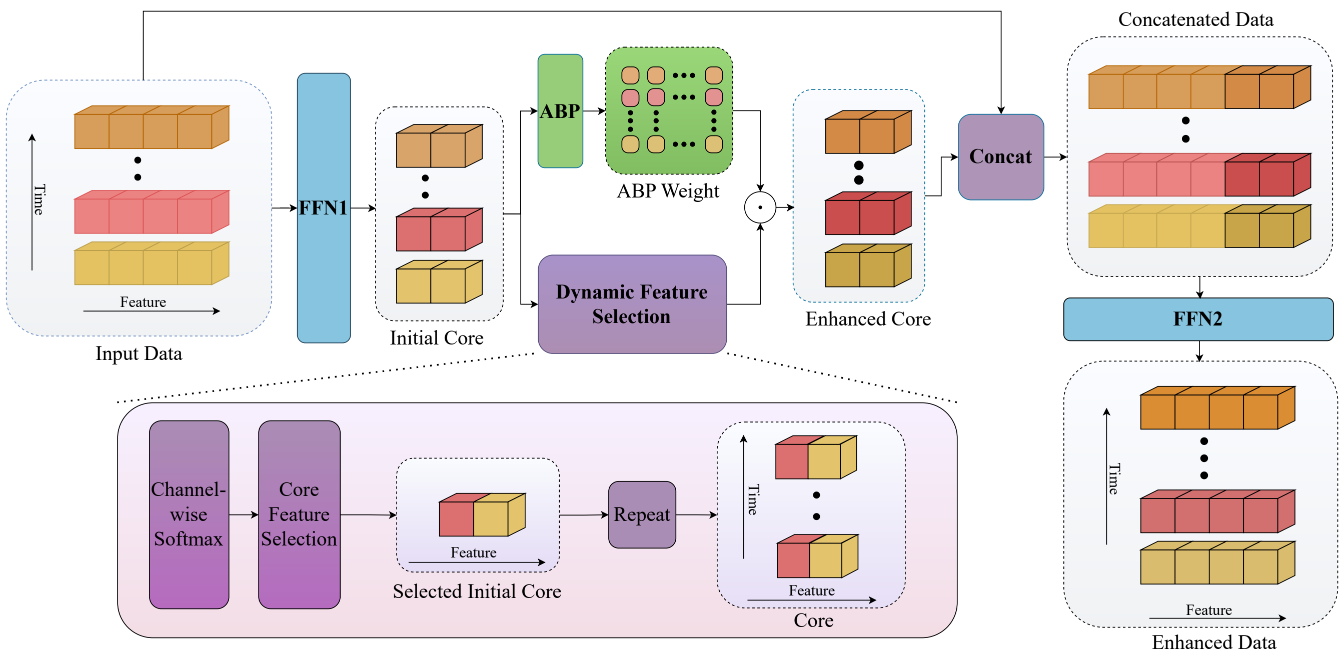

Figure 2: The FlexiReMod module operates as follows: the input data $\mathbf{X}_{\text{input}}$ is first processed by FFN1 to obtain the core representation **IC**, followed by stochastic pooling to aggregate features into **SIC**, which is then replicated to form **C**. Attention weights are applied to **C**, producing the enhanced feature representation $\mathbf{C}_{\text{enhanced}}$. Finally, the original input is concatenated with the $\mathbf{C}_{\text{enhanced}}$ and processed by FFN2 to yield the final enhanced output $\mathbf{X}_{\text{enhanced}}$.

MMD-based Autoencoder

To enhance domain-invariant feature extraction, we adopt the MMD-based Autoencoder approach from Cheng et al. (2021), which uses the MMD loss $\mathcal{L}_{\text{mmd}}$ to align distributions across source domains, reducing differences and improving consistency across subjects.

The MMD-based Autoencoder takes the enhanced source domain dataset $\{\mathcal{D}^n_{enhanced}\}^N_{n=1}$ as input. The encoder maps each input $\mathbf{x_{enhanced}}^n_j$ to the latent space using $Q(\cdot)$, producing the latent representation $\mathbf{h}^n_j = Q(\mathbf{x_{enhanced}}^n_j)$. The decoder reconstructs the latent representation, yielding $\hat{\mathbf{x_{enhanced}}}^n_j = P(\mathbf{h}^n_j)$. The $\mathbf{h}^n_j$ are then mapped to the Reproducing Kernel Hilbert Space (RKHS) through the kernel mapping $\varphi(\cdot)$, resulting in the feature mapping $\hat{\mathcal{H}}^n = \varphi(\mathbf{h}^n_j)$. The mean of the features in $\hat{\mathcal{H}}^n$ and $\mathcal{L}_{\text{mmd}}$ are computed as follows:

$$\mu_{\hat{\mathcal{H}}^n} = \mathbb{E}_{\hat{h}\sim\hat{\mathcal{H}}^n}[\varphi(h^n)] \tag{8}$$

$$\mathcal{L}_{\text{mmd}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \|\mu_{\hat{\mathcal{H}}_i} - \mu_{\hat{\mathcal{H}}_j}\|_2^2. \tag{9}$$

At the same time, the reconstruction loss L_{rec} is calculated as follows:

$$\mathcal{L}_{rec} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{n_n}\sum_{j=1}^{n_n}\|\mathbf{x_{enhanced}}^n_j - \hat{\mathbf{x_{enhanced}}}^n_j\|_2^2 \tag{10}$$

Anti-Feature Collapse Strategy (AFCS)

In domain generalization, although MMD captures distribution differences, it can lead to feature collapse, resulting in the loss of critical information and diminished discriminative ability. To preserve feature diversity and enhance adaptability and generalization in complex tasks, we introduce the Anti-Feature Collapse Strategy (AFCS), which operates at both macro and micro levels to improve feature space diversity and model generalization.

On the macro level, AFCS expands the feature distribution by generating heavy-tailed data through the HeavyTail Generator (HTG), which serves as input for the Anti-Collapse Domain Discriminator (ACDD). Utilizing adversarial learning, ACDD steers the feature space distribution towards a heavy-tailed structure, mitigating over-compression and enhancing feature diversity. Features in $\mathcal{FS}$, represented by $\mathcal{H}^n$, from the real domain are labeled 1, while pseudo-samples from HTG, $\mathcal{H}_{\text{fake}} \in \mathbb{R}^{B\times D_{\text{core}}}$, are labeled 0. The adversarial loss L_{adv} is:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}[\log(d_{\text{real}})] + \mathbb{E}[\log(1-d_{\text{fake}})] \tag{11}$$

where $d_{\text{real}} = \sigma(\text{ACDD}(\mathcal{H}))$, $d_{\text{fake}} = \sigma(\text{ACDD}(\mathcal{H}_{\text{fake}}))$, $\sigma(\cdot)$ denotes the Sigmoid activation function, and $\mathbb{E}$ represents the expectation.

On the micro level, considering the relationships between individual features, AFCS introduces the encoding rate $R(\mathcal{H})$, which quantifies the number of bits needed to encode $\mathcal{H}$ at a precision ε, serving as an indicator of feature space compactness.

$$R(\mathcal{H}) = \frac{1}{2}\log\det\left(I + \frac{d}{k\varepsilon^2}\mathcal{H}^\top\mathcal{H}\right) \tag{12}$$

Where k is the batch size, d is the feature dimension, ε is the encoding precision, and I is the $d \times d$ identity matrix. Inspired by Zhu et al. (2022), we introduce the anti-feature collapse loss function $\mathcal{L}_{\text{cr}}$ based on the concept of encoding rate:

$$\mathcal{L}_{\text{cr}} = \frac{1}{\rho} \log\cosh\left(\rho \left(R(\mathcal{H}) - \bar{R}(\mathcal{H})\right)\right) \tag{13}$$

$\bar{R}(\mathcal{H})$ is the moving average of $R(\mathcal{H})$, updated as $\bar{R}(\mathcal{H}) \leftarrow \eta\bar{R}(\mathcal{H}) + (1-\eta)R(\mathcal{H})$, with $\eta = 0.99$. The parameter ρ, set to 0.2, controls the smoothness of the loss function. The Log-Cosh loss reduces when errors exceed the tolerance, focusing on significant collapse issues.

Learning Loss in The training Phase

Following GAN (Goodfellow et al., 2014), the AC-CDCN can be written as the following minimax optimization problem:

$$\min_{\text{AC-CDCN}} \max_{\text{ACDD}} \mathbb{E}_{\mathcal{D}_s}\left[\mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{mmd}} + \lambda_3 \mathcal{L}_{\text{adv}} + \lambda_4 \mathcal{L}_{\text{cr}}\right] \tag{14}$$

Where λ_1, λ_2, λ_3, and λ_4 are hyperparameters controlling the loss weights. Following Ganin et al. (2016), who introduced the gradient reversal layer in adversarial learning, the total loss objective of AC-CDCN can be simplified as follows:

$$\mathcal{L}_{Total} = \min_{\text{AC-CDCN}} \mathbb{E}_{\mathcal{D}_s}\left[\mathcal{L}_{cls} + \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{mmd}}\right] - \lambda_3 \mathcal{L}_{\text{adv}} + \lambda_4 \mathcal{L}_{\text{cr}} \tag{15}$$

As shown in Algorithm 1, the process involves several steps to optimize the parameters.

Algorithm 1 AC-CDCN Training Algorithm

Input: Epoch T, $\mathcal{D}_s^n = (\mathbf{x}_j^n, y_j^n)_{j=1}^{n_n}$ for $i \in \{1, 2, \ldots, N\}$, initialized parameters Φ for AC-CDCN.
Output: Optimized parameters for AC-CDCN Φ^*

```
1:  while The iteration limit has not been exhausted. do
2:      sample a minibatch X_s, Y_s from D_s^n.
3:      FlexiReMod(X_s) and get X_Enhanced.
4:      Q(X_Enhanced) and FS.
5:      Compute L_mmd by (9)
6:      Compute L_cr by (13)
7:      Sample H^n and H_fake.
8:      Compute adversarial loss L_adv by (11)
9:      Compute L_cls using the classifier Cls.
10:     Take a gradient step to update Φ^t by minimizing: (15)
11: end while
12: return Φ* = {θ*_FlexiReMod, θ*_Enc-Dec, θ*_ACDD, θ*_Cls}
```

Experiments

Datasets

The AC-CDCN was validated on two EEG datasets, SEED and SEED-IV, widely used in emotion recognition. SEED includes EEG signals from 15 participants (62 channels) watching 15 Chinese movie clips to evoke positive, negative, and neutral emotions, totaling 675 trials (15 participants × 3 sessions × 15 trials). SEED-IV also involves 15 participants, each undergoing three sessions with 24 trials, viewing clips to elicit happiness, sadness, fear, and neutrality, totaling 1,080 trials. EEG signals were recorded using a 62-channel ESI NeuroScan System at 1000 Hz, downsampled to 200 Hz, and bandpass filtered (1–75 Hz). Features were extracted using Differential Entropy (DE) across five frequency bands: delta (0.4–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz), yielding 310 features per sample (62 channels × 5 frequency bands).

Implementation Details

DE features from the first session of all subjects were utilised as the input data. For the purpose of preprocessing, the overlapping sliding window method with a step size of T was employed, as outlined in the study by Zhao, Yan, and Lu (2021). The data samples are represented as follows: $\mathbf{X}_{\text{window}} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t) \in \mathbb{R}^{T \times D}$, The evaluation used leave-one-out cross-validation, with one subject as the target and the remaining 14 as sources to simulate domain shifts. After each subject was used as the target, the average accuracy and standard deviation were calculated. For both datasets, the input feature dimension was 310, and the encoder's hidden layer dimension was 64. Hyperparameters λ_1, λ_2, λ_3, and λ_4 were set to 0.25, 0.25, 0.4, and 0.1, respectively. We used Adam optimizer with a learning rate of 1.2×10^{-4} for ACDD and 6.5×10^{-3} for other parts, and a weight decay of 5×10^{-2}. The batch size was 256, with time steps of 30 for SEED and 10 for SEED-IV, and the SEED dataset was trained for 150 epochs, while SEED-IV was trained for 200 epochs. To ensure reproducibility, the random seed was fixed at 1137 for all experiments. All experiments were implemented in PyTorch on an NVIDIA GeForce RTX 4060 Ti.

Experiment Results

Performance Comparison We compared the proposed AC-CDCN model with state-of-the-art methods, as shown in Table 1. (Y. Li et al., 2020) captures inter-hemispheric differences via paired subnetworks. Clisa (Shen, Liu, Hu, Zhang, & Song, 2022) uses contrastive learning to generate robust differential entropy (DE) features. GMSS (Y. Li et al., 2022) integrates self-supervised and spatial puzzle tasks to reduce overfitting. RGNN (Zhong et al., 2020) is domain adaptation models adapted for domain generalization (DG) by removing the gradient reversal layer. PPDA (Zhao et al., 2021) can also be adapted for DG tasks after removing the calibration component. However, these methods struggle with dynamic feature capture, limiting their effectiveness on time-varying EEG signals. DGCNN (Song et al., 2018) uses graph convolutional networks to model inter-channel relationships for generalized emotion recognition. DG-DANN (Ganin et al., 2016) and DResNet (Ma et al., 2019) are adversarial DG methods, with DResNet extending DG-DANN by separating weight parameters for unbiased and biased features. However, these methods suffer from feature space collapse, limiting generalization

Table 1: Average Accuracy and Standard Deviation (%) of Different Models on SEED and SEED-IV

Method	SEED Avg.	SEED Std.	SEED-IV Avg.	SEED-IV Std.
BiHDM(DG)	81.55	9.74	67.47	8.22
RGNN(DG)	81.92	9.35	71.65	9.34
PPDA(DG)	85.40	7.10	-	-
BiHDN	85.40	7.53	69.03	8.66
Clisa	86.40	6.40	-	-
GMSS	86.52	6.22	**73.48**	**7.41**
DGCNN	79.95	9.02	-	-
DG-DANN	84.30	8.32	-	-
DResNet	85.30	7.97	-	-
AC-CDCN(ours)	**87.14**	**5.60**	71.77	12.92

Table 2: Impact of Different Components on Performance through Ablation Studies on SEED.

Method	**Avg.**	**Std.**
AC-CDCN (Full Model)	87.14	5.60
w/o ACDD	86.73	4.87
w/o L_{cr}	85.94	4.39
w/o ACDD & L_{cr}	85.40	5.02
w/o FlexiReMod	85.18	4.97
w/o FlexiReMod & ACDD & L_{cr}	84.66	4.84

and leading to lower accuracy compared to AC-CDCN. On the SEED dataset, AC-CDCN achieved 87.14% (±5.60) accuracy. On SEED-IV, our model achieved 71.77% (±12.92), slightly below GMSS, likely due to the smaller dataset size limiting data augmentation. Overall, AC-CDCN outperforms other models, demonstrating superior cross-subject generalization by extracting domain-invariant features and leveraging AFCS to address individual variability.

Ablation studies To assess the effectiveness of the proposed anti-feature collapse technique and data augmentation strategy, ablation experiments were conducted on SEED datasets to evaluate the contribution of each component of the proposed method. The results are presented in Table 2.

The ablation experiments adopt the same hyperparameter settings as AC-CDCN and include the following cases: "w/o ACDD" removes ACDD, relying solely on L_{cr} to prevent feature collapse; "w/o L_{cr}" removes L_{cr}, relying solely on ACDD for feature space optimization; "w/o FlexiReMod" removes FlexiReMod to evaluate its role in dynamic feature selection; "w/o ACDD & L_{cr}" removes both ACDD and L_{cr}; and "w/o ACDD & L_{cr} & FlexiReMod" removes all modules to assess the baseline model's performance. The results underscore the crucial roles of ACDD and L_{cr} in mitigating feature over-compression, enhancing generalization, and improving emotion recognition accuracy. FlexiReMod excels in dynamic feature capture and fusion, enabling effective selection and redistribution of critical information. Removing these modules led to drops in accuracy and stability, confirming their importance in efficient feature extraction, adaptability, and robustness. Together, these modules synergistically enhance AC-CDCN's cross-subject generalization, particularly for complex, noisy EEG signals.

Conclusion and Future Work

This paper proposes AC-CDCN, an innovative cross-subject domain generalization model for EEG-based emotion recognition. The model addresses feature collapse and dynamic feature capture issues by introducing ACDD, $\mathcal{L}_{cr}$, and FlexiReMod. FlexiReMod optimizes feature extraction adaptively across tasks and datasets, while ACDD introduces significant innovation, enhancing feature diversity, model robustness, and cross-task transferability. Experimental results demonstrate that AC-CDCN achieves 87.17% (±5.60) accuracy on SEED and 71.77% (±12.92) on SEED-IV, significantly improving emotion recognition and robustness, especially in cross-subject tasks, showing excellent generalization and stability. Future work will focus on investigating the application conditions of ACDD, exploring its suitability for adversarial learning-guided feature distribution, and optimizing adversarial training strategies to further enhance generalization for complex high-dimensional data and cross-domain tasks.

Acknowledgments

This work was supported by the Natural Science Foundation of China (62374121,61974109).

References

Blanchard, G., Lee, G., & Scott, C. (2011). Generalizing from several related classification tasks to a new unlabeled sample. *Advances in neural information processing systems*, *24*.

Cheng, J., Gao, M., Liu, J., Yue, H., Kuang, H., Liu, J., & Wang, J. (2021). Multimodal disentangled variational autoencoder with game theoretic interpretability for glioma grading. *IEEE Journal of Biomedical and Health Informatics*, *26*(2), 673–684.

Devillers, L., & Cowie, R. (2023). Ethical considerations on affective computing: an overview. *Proceedings of the IEEE*.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., … Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of machine learning research*, *17*(59), 1–35.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B., & Smola, A. (2006). A kernel method for the two-sample-problem. *Advances in neural information processing systems*, *19*.

Han, L., Chen, X.-Y., Ye, H.-J., & Zhan, D.-C. (2024). Softs: Efficient multivariate time series forecasting with series-core fusion. *arXiv preprint arXiv:2404.14197*.

Katsigiannis, S., & Ramzan, N. (2017). Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, *22*(1), 98–107.

Khosla, A., Zhou, T., Malisiewicz, T., Efros, A. A., & Torralba, A. (2012). Undoing the damage of dataset bias. In *Computer vision–eccv 2012: 12th european conference on computer vision, florence, italy, october 7-13, 2012, proceedings, part i 12* (pp. 158–171).

Li, J., Qiu, S., Shen, Y.-Y., Liu, C.-L., & He, H. (2019). Multisource transfer learning for cross-subject eeg emotion recognition. *IEEE transactions on cybernetics*, *50*(7), 3281–3293.

Li, Y., Chen, J., Li, F., Fu, B., Wu, H., Ji, Y., … Zheng, W. (2022). Gmss: Graph-based multi-task self-supervised learning for eeg emotion recognition. *IEEE Transactions on Affective Computing*, *14*(3), 2512–2525.

Li, Y., Wang, L., Zheng, W., Zong, Y., Qi, L., Cui, Z., … Song, T. (2020). A novel bi-hemispheric discrepancy model for eeg emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(2), 354–367.

Luo, Y., & Lu, B.-L. (2021). Wasserstein-distance-based multi-source adversarial domain adaptation for emotion recognition and vigilance estimation. In *2021 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 1424–1428).

Ma, B.-Q., Li, H., Zheng, W.-L., & Lu, B.-L. (2019). Reducing the subject variability of eeg signals with adversarial domain generalization. In *Neural information processing: 26th international conference, iconip 2019, sydney, nsw, australia, december 12–15, 2019, proceedings, part i 26* (pp. 30–42).

Picard, R. W., & Klein, J. (2002). Computers that recognise and respond to user emotion: theoretical and practical implications. *Interacting with computers*, *14*(2), 141–169.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, *23*(10), 1175–1191.

Shen, X., Liu, X., Hu, X., Zhang, D., & Song, S. (2022). Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition. *IEEE Transactions on Affective Computing*, *14*(3), 2496–2511.

Song, T., Zheng, W., Song, P., & Cui, Z. (2018). Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, *11*(3), 532–541.

Sun, H., Yang, L., Wang, Q., Liu, D., & Ni, P. (2023). Meeg-transformer: transformer network based on multi-domain eeg for emotion recognition. In *2023 ieee international conference on bioinformatics and biomedicine (bibm)* (pp. 3332–3339).

Tian, M. (2023). Emotion recognition based on physiological signals.

Vatcharaphrueksadee, A., Maliyaem, M., Viboonpanich, R., & Phuangkamnerd, S. (2022). Facial emotion classification of multi-type datasets based on svm classifier. In *2022 international conference on computer science and software engineering (csase)* (pp. 71–76).

Xu, T., Dang, W., Wang, J., & Zhou, Y. (2023). Dagam: a domain adversarial graph attention model for subject-independent eeg-based emotion recognition. *Journal of Neural Engineering*, *20*(1), 016022.

Yu, Y., Chan, K. H. R., You, C., Song, C., & Ma, Y. (2020). Learning diverse and discriminative representations via the principle of maximal coding rate reduction. *Advances in neural information processing systems*, *33*, 9422–9434.

Zeiler, M. D., & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv:1301.3557*.

Zhao, L.-M., Yan, X., & Lu, B.-L. (2021). Plug-and-play domain adaptation for cross-subject eeg-based emotion recognition. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 863–870).

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., & Cichocki, A. (2018). Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, *49*(3), 1110–1122.

Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, *7*(3), 162–175.

Zhong, P., Wang, D., & Miao, C. (2020). Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, *13*(3), 1290–1301.

Zhu, W., Lu, L., Xiao, J., Han, M., Luo, J., & Harrison, A. P. (2022). Localized adversarial domain generalization. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 7108–7118).