Multi-scale Dynamic Temporal Network with Graph Matching Domain Adaptation for Cross-Subject EEG Emotion Recognition

Rongtao Chen, Zhepei Hong, Liting Li, Zhuoyi Huang, Jingcong Li, Jiahui Pan, *Member, IEEE*

***Abstract*—Electroencephalography (EEG) is a powerful and objective tool for detecting emotions, with broad applications across various domains. However, EEG-based emotion recognition faces two major challenges: (1) extracting domain-invariant features that maintain emotion-related information across variations induced by individual differences, and (2) aligning the marginal and conditional distributions of data from different individuals in the feature space. To address these challenges, we propose a novel method: the Multi-scale Dynamic Temporal Network with Graph Matching Domain Adaptation, designed for cross-subject EEG emotion recognition. Our method employs a multi-scale dynamic temporal attention module to extract robust domain-invariant features. In addition, it utilizes a domain adaptation network based on Chebyshev graph representations, which reformulates the emotion recognition task as a graph matching problem. This approach effectively aligns cross-subject data distribution. To validate the effectiveness of our method, we conducted extensive experiments on two benchmark databases (SEED and DEAP) under three different cross-validation protocols. The results demonstrate consistent reliability in subjects and sessions. Specifically, in the cross-subject single-session cross-validation task, our method achieved an accuracy of 94.69%±5.16% on the SEED dataset. On the DEAP dataset, our approach achieved accuracies of 69.45%±7.26% and 67.52%±8.28% for the Valence and Arousal dimensions, respectively. These results demonstrate that our method outperforms existing state-of-the-art approaches. Source code is available at: https://github.com/seizeall/MDTN-GMDA.**

***Index Terms*—Electroencephalography (EEG), emotion recognition, cross-subject, dynamic temporal attention, graph matching domain adaptation.**

I. Introduction

EMOTION recognition is a cornerstone of affective computing, with wide applications in marketing, medical rehabilitation and traffic safety [1]. Its crucial role has driven continuous advancements in decoding complex human states [2], [3]. It typically relies on behavioral and physiological representations, which capture emotional and cognitive aspects through two primary modalities: non-physiological and physiological signals. Non-physiological signals, including facial expressions, speech, and gestures, are easily accessible but less reliable for emotion analysis due to their susceptibility to subjective influences. In contrast, physiological signals are increasingly recognized as more robust and reliable for emotion recognition. Among them, electroencephalography (EEG) provides a direct reflection of emotional states, as it is closely linked to brain activity and the neural processes underlying emotion.

EEG emotion recognition can be categorized into intra-subject and inter-subject models, depending on whether the training and test sets come from the same individual [4]. Intra-subject models, which use data from the same subject for both training and testing, have consistently achieved high accuracy. However, inter-subject models, which train and test on data from different subjects, often exhibit significantly poorer performance [5]. Enhancing the accuracy of inter-subject emotion recognition is therefore a critical issue, as it directly affects the precision and robustness of emotion detection across diverse subjects.

The significant cross-domain variability inherent in EEG signals presents a substantial obstacle to developing stable models that can extract domain-invariant features in cross-subject scenarios [6], [7]. Conventional deep learning approaches, which derive features directly from raw EEG data, are constrained by limitations such as inadequate spatial resolution and vulnerability to physiological artifacts [8], [9]. To mitigate these challenges, researchers have increasingly adopted multi-scale feature extraction frameworks for EEG emotion recognition. For instance, Pang et al. [10] proposed Multi-scale Masked Autoencoders (MSMAE), which leverage multi-scale masked autoencoders to extract noise-robust features. This approach enhances spatial resolution through channel-spatial fusion, improves signal-to-noise ratio (SNR) with robust masking, and reduces artifact sensitivity by aligning correlations. Similarly, Dong et al. [11] introduced Multi-scale 3D-CRU, which employs 3D-CNN for frequency-spatial extraction and GRU for temporal dynamics. This method improves spatial resolution via electrode mapping, boosts SNR through noise discrimination in joint domains, and mitigates artifacts through fusion. Pei et al. [12] presented DTP-Net, which utilizes densely-connected temporal pyramids for multi-scale reuse in the time-frequency domain. This framework enhances spatial resolution using dilated convolutions, elevates SNR through artifact separation, and reduces artifact sensitivity with learnable transforms. Despite these advances, current multi-scale feature extraction methods still face limitations,

This work was supported in part by Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project under Grant 2022ZD0208900, in part by the National Natural Science Foundation of China under Grant 62576142, and in part by the Major Projects of Colleges and Universities in Guangdong Province under Grant 2023ZDZX2021. (Rongtao Chen and Zhepei Hong contributed equally to this work.) (Corresponding authors: Jiahui Pan and Jingcong Li.)

Jiahui Pan and Jingcong Li are with the School of Artificial intelligence, South China Normal University, Guangzhou 510631, China, and also with the Research Center for BrainComputer Interface, Pazhou Laboratory, Guangzhou 510330, China (email: panjiahui@m.scnu.edu.cn; lijingcong@hotmail.com).

particularly in cross-subject emotion recognition. These methods often fail to capture sufficient domain-invariant features, leading to suboptimal accuracy and poor generalization across subjects in real-world applications.

Moreover, EEG signals are nonstationary time series, and each emotion experiment can be considered a unique task. Even the same user may exhibit different neural responses to the same stimuli due to physiological, environmental, and instrumental variations. These individual-specific variations further complicate the task when considering cross-subject experiments. Consequently, another major challenge is aligning the marginal and conditional distributions of different individuals in the same feature space. Traditional domain adaptation methods typically compare extracted features and train the feature extraction module to enhance domain invariance, thereby aligning different individual EEG data in the feature space. For example, Li et al. [13] proposed the BiDANN, which employs global and local domain discriminators adversarially with a classifier to extract emotional features from both brain hemispheres, achieving 84.14% cross-subject accuracy on SEED. Pan et al. [14] presented the GLADA, which balances global and local adversarial adaptations and adjusts weights for marginal and conditional distributions, achieving 90.44% on SEED and 73.89% on SEED-IV. Li et al. [15] presented the MS-FRAN, integrating a wide feature extractor, a random matching operation, and a top-ranked domain classifier selection module. Experiments on SEED and DEAP showed that MS-FRAN outperforms existing methods. Despite these advances, most domain adaptation methods mainly focus on capturing domain differences to update the model. However, they lack structured and dynamic representations that capture fine-grained marginal and conditional distributions across individuals, which limits the model's generalization ability.

To address these challenges, we propose a novel domain adaptation model. It extracts complementary multi-scale domain-invariant features and aligns the marginal and conditional distributions of subject-specific features within a shared feature space using a structured representation. Our contributions can be summarized as follows:

- **Dynamic Temporal Attention (DTA) Module:** We introduce a novel DTA module tailored for cross-subject EEG emotion recognition. It captures both long-term and short-term temporal dependencies through multi-scale convolutions and enhances them with a multi-head attention mechanism. Additionally, a contextual gate layer is incorporated to improve cross-domain transferability, enabling the extraction of robust, domain-invariant emotional features.
- **Chebyshev Graph Matching Domain Adaptation:** We propose a new domain adaptation framework that reformulates cross-subject EEG emotion recognition as a graph matching problem. By dynamically learning a Chebyshev polynomial-based adjacency matrix, our method aligns both marginal and conditional distributions between source and target domains, achieving fine-grained alignment and effective knowledge transfer.
- **Outstanding Performance in Emotion Recognition:** We conduct extensive cross-subject experiments on two widely-used emotion recognition datasets, SEED and DEAP. The results demonstrate that our proposed method outperforms state-of-the-art (SOTA) approaches in terms of various classification metrics, providing strong evidence of the efficacy of our novel methods.

II. Related-work

1) Multi-scale Feature Extraction Network: Capturing multi-scale information is essential for robust EEG emotion recognition, but separate models per scale are computationally expensive. Multi-scale networks capture features across scales, improving pattern identification and performance. Liu et al. [16] proposed a dynamic collaborative evolutionary network using graph-aware enhancement and global convolutional transformer for spatio-temporal features. Zhai et al. [17] introduced a lightweight adaptive inference network with partial convolution and batch interactive attention for efficient task-relevant feature processing. Ding et al. [8] presented LGGNet, using multiscale attentive fusion and graph filtering to model local-global brain activities in emotion recognition. Yan et al. [9] proposed BGAGCN-MT, combining bridge graph attention GCN for over-smoothing mitigation with multi-scale transformer for spatial dependencies via multi-head attention post-1D convolutions.

Despite advances, current methods often overlook finer-grained temporal characteristics in complex EEG time-series. In this work, we introduce a temporal multi-scale module that refines EEG temporal representations at several resolutions while reusing a shared backbone, which enables more fine-grained temporal modeling without incurring the cost of training separate models for each scale. In contrast to traditional multi-scale feature extractors that simultaneously model both spatial and temporal dimensions, our proposed DTA module focuses exclusively on the temporal dimension, enabling efficient fine-grained representation refinement via a shared backbone with significantly lower computational overhead. This focused approach mitigates the risk of overlooking hidden temporal or spatial features due to insufficient modeling depth in joint spatio-temporal processing.

2) Domain Adaptation Network: Efficiently aligning distributions between source and target domains remains a key challenge in cross-subject EEG emotion recognition. Recently, domain adaptation networks have emerged as a powerful solution, aiming to reduce distributional shifts and identify transferable features that enhance model performance. Li et al. [18] incorporated the transposition multi-layer perceptron and sample-reweighted domain adaptation neural network, which prioritize source samples with higher transferability to extract domain-invariant features, significantly improving the performance of EEG-based emotion recognition. Liu et al. [19] proposed the EEG driven spatio-temporal hybrid network based on domain adaptation and dynamic graph attention, which extracts discriminative features through multi-scale residual convolutional learning and temporal self-attention representation. Zhou et al. [20] proposed a prototypical representation-based pairwise learning framework that employs adaptive pseudo-labeling and domain-invariant feature extraction to

improve resilience against individual differences and noisy labels. Wang et al. [21] introduced the cerebral asymmetry representation learning-based deep subdomain adaptation network, which extracts discriminative features through intra-hemisphere spatial learning and asymmetry representation, effectively addressing domain shift.

Despite these achievements, many existing domain adaptation methods demand significant computational resources and often lack dynamic, structured representations for precise data alignment. This limitation can lead to misalignment of the data. Therefore, there is a need to develop efficient dynamic representations to enhance model correction and precise data alignment. The proposed domain adaptation network employed a relatively lightweight alignment module, aiming to balance alignment performance and computational efficiency compared with the aforementioned graph-based or sample-reweighting strategies.

III. Method

A. Data Processing and Enhancement

EEG-based cross-subject emotion recognition is a challenging task due to the inherent variability in EEG signals across different individuals. In this context, we consider a labeled source domain dataset $\{X_s, Y_s\} = \{(x_{s,1}, y_{s,1}), \ldots, (x_{s,n_s}, y_{s,n_s})\}$, where each EEG sample $x_{s,i} \in \mathbb{R}^{D_{EEG}}$ is paired with a one-hot encoded label vector $y_{s,i} \in \mathbb{R}^q$. Here, $D_{EEG} = c \times f \times t$, with c representing the number of EEG channels, each capturing information across f frequency bands, t representing the time dimension of the EEG signal, and q denoting the number of distinct emotional states. Furthermore, we have an unlabeled target domain dataset $X_t = [x_{t,1}, \ldots, x_{t,n_t}] \in \mathbb{R}^{D_{EEG} \times n_t}$, where n_t indicates the number of target samples. The overall architecture of the proposed model is illustrated in Fig. 1.

1) Data Normalization: Individual variability often causes distributional differences between source and target domains, challenging the assumption that samples are independent and identically distributed (i.i.d.) across domains. To address this issue, we propose a channel normalization method applied to both the source and target domains. Specifically, we normalize EEG data along the channel dimension using the mean and standard deviation calculated from the source domain. This approach preserves the temporal properties of EEG signals while standardizing their feature distributions, thereby enhancing data stability and aligning with i.i.d. assumptions. For each channel c in the sample $x \in \mathbb{R}^{D_{EEG}}$, the normalized value is computed as follows:

$$\tilde{x}_c = \frac{x_c - \mu_c}{\sigma_c}, \tag{1}$$

where μ_c and σ_c are the mean and standard deviation of channel c across the dataset.

2) Feature Stacking: In feature stacking, we first compute the Differential Entropy (DE) for each 1-second EEG segment. This has been validated as a highly effective feature for EEG-based emotion recognition [22], [23]. DE provides several advantages: it captures signal complexity with high resolution, reduces noise, characterizes hemispheric asymmetry, and quantifies signal uncertainty to reveal neural dynamics [24], [25]. The value of DE for a segment x_i is calculated as:

$$\mathrm{DE} = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}\right) d_x. \tag{2}$$

After calculating the DE features for each 1-second segment, these features are then stacked across continuous n-second segments to create a new, enhanced feature vector. If we denote the DE feature for each 1-second segment DE_t at time step t, the stacked feature $\mathbf{F}$ for a K-second window is given by:

$$\mathbf{F} = [\mathrm{DE}_1, \mathrm{DE}_2, \ldots, \mathrm{DE}_K], \tag{3}$$

thus, we generate a K-dimensional feature for each K-second EEG segment, which is then used in downstream tasks. By stacking these 1-second DE features, we capture both the temporal evolution of the EEG signal and its frequency-based characteristics, improving the model's ability to recognize emotions effectively.

3) Gaussian Augmentation: During the data augmentation phase, a random subset of the original samples is selected, and Gaussian noise is introduced to generate synthetic data, which is then incorporated into the training set. This augmentation technique leverages a controlled level of variability through Gaussian-distributed noise, effectively serving as a form of regularization that improves the model's robustness and generalization capabilities. The added noise improves the model's ability to adapt to domain shifts and diverse conditions, such as noisy inputs or inputs with outliers. Mathematically, for each feature vector $\mathbf{T}_i$ in the dataset, the noisy version $\mathbf{T}_i^{\text{noisy}}$ is computed as:

$$\mathbf{T}_i^{\text{noisy}} = \mathbf{T}_i + \mathcal{N}(0, \sigma^2), \tag{4}$$

where $\mathbf{T}_i \in \mathbb{R}^{D_{EEG}}$ represents the i-th feature vector, and $\mathcal{N}(0, \sigma^2)$ denotes Gaussian noise with zero mean and variance σ^2. The noise is added to a random subset of the original data, and the modified samples, $\mathbf{T}_i^{\text{noisy}}$, are then used to improve the model's transfer performance.

B. Multi-scale Dynamic Temporal Attention

In traditional domain-adaptive neural network (DANN), the core challenge is to develop feature extractors capable of capturing emotional characteristics. These features are subsequently used to align probability distributions between source and target domains, enabling the extraction of domain-invariant representations through domain-adversarial training. Nonetheless, traditional domain adaptation approaches often rely on single-scale feature extractors, which limit the model's ability to capture comprehensive temporal patterns across varying durations. To address these limitations, we propose a DTA module that employs parallel multi-scale convolutions

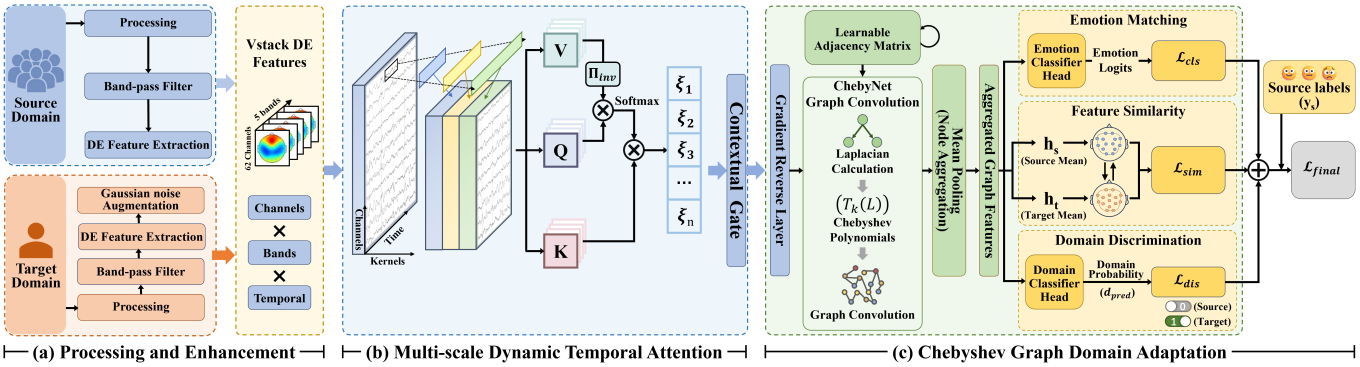


Fig. 1. **Framework of Chebyshev Graph Domain-Adaptive Multi-Scale Dynamic Temporal Attention Network for Cross-Subject EEG Emotion Recognition.** This framework comprises three main components: (a) **Data Processing and Enhancement**: Data are preprocessed using channel normalization, followed by augmentation through stacking features from 1s to Ks and Gaussian generalization; (b) **Multi-Scale Dynamic Temporal Attention Module**: This module captures dynamic temporal features from EEG signals using a multi-head attention mechanism and contextual gating to effectively model temporal and spectral variations; (c) **Chebyshev Graph Domain Adaptation**: Domain adaptation is achieved by aligning the source and target domains through graph matching and adversarial training. Features are projected into a shared space with dynamic threshold labels applied to enhance EEG emotion recognition capabilities.

to extract temporal features of different lengths, and subsequently utilizes a self-attention mechanism to adaptively fuse these multi-scale features. This design focuses on the dynamic integration of temporal information, isolating subject-independent characteristics that explain affective fluctuations in EEG signals.

1) Temporal Feature Extraction: To comprehensively capture temporal characteristics across diverse resolutions, we employ a set of parallel multi-scale convolutional kernels (denoted as T kernels) explicitly designed to extract features at varying temporal granularities. By configuring the length of each kernel as a specific ratio of the input length L, we enable the extraction of a multi-scale representation that spans different temporal scopes. These ratios are denoted as $h^i \in \mathbb{R}$, where i represents the level within the multi-scale temporal extraction layer. Consequently, the size of the temporal convolution kernel at the i-th level, S_i^T, is defined as:

$$S_i^T = \left(1, \frac{L}{h_i}\right), \quad i \in [1, 2, \ldots, T], \tag{5}$$

where $h_i = 2^i$ represents the pre-defined scaling factor at level i. This formulation allows for additional flexibility, where other scaling factors h_i could be applied based on specific temporal dependencies or model requirements.

Multi-scale 1-D convolutional kernels are employed to expand the temporal receptive field, thereby facilitating the extraction of subject-invariant features and the capture of comprehensive temporal patterns across different granularities. Longer kernels are tailored to capture broad, low-frequency patterns associated with long-term dependencies, while shorter kernels focus on high-frequency, short-term dynamics.

Let X represent the EEG input samples, where $X = [X^0, X^1, \ldots, X^n]$ with $X^n \in \mathbb{R}^{D_{EEG} \times L}$. Here, $D_{EEG} \times L = c \times f \times t \times L$, n denotes the total number of EEG samples and L is the length of each sample. To obtain a rich set of multi-scale temporal features, we apply multi-scale temporal convolutional kernels in parallel on the EEG samples.

Let R_i^T denote the output from the i-th level of the temporal kernel, where $R_i^T \in \mathbb{R}^{t \times c \times L}$, with t representing the number of kernels at each level T, c encompassing the spatial and frequency channels, and L is the feature length post-convolution at the i-th level. Formally, R_i^T can be expressed as:

$$R_i^T = AP(\Phi_{L-\text{ReLU}}(\text{Conv1D}(X, S_i^T))), \tag{6}$$

where S_i^T denotes the kernel size at the i-th level T, X is the input EEG sample array, Conv1D$(\cdot)$ represents the 1D convolution operation with kernel size S_i^T and stride $(1, 1)$, and $\Phi_{L-\text{ReLU}}(\cdot)$ is the LeakyReLU $(\cdot)$ activation function. The features are then passed through the LeakyReLU activation function, followed by average pooling (AP), which reduces noise and lowers the feature dimensionality—beneficial for handling the high dimensionality and low signal-to-noise ratio typical of EEG signals.

To integrate features across different temporal scales, R is constructed by stacking the outputs R_i^T from each convolutional level. This aggregation of multi-scale R_i^T outputs allows the model to synthesize information from various temporal resolutions, providing a diverse feature basis for the subsequent dynamic fusion process.

2) Dynamic Temporal Representation: While the previous subsection extracts fixed-scale features, this subsection focuses on the dynamic fusion of these representations. Multi-scale temporal features $\{R_1^T, \ldots, R_T^T\}$, extracted via convolutional kernels, are stacked into a combined representation $R = \oplus_{i=1}^T R_i^T$. A multi-head attention mechanism is then employed to adaptively integrate these features based on their relevance, thereby constructing a genuine dynamic representation. To achieve this, we adopt a comprehensive cross-scale attention strategy: the feature set from each temporal scale, R_i^T, takes a turn as the Query (Q), while the features from other scale R_j^T serve as the corresponding Key (K) and Value (V). This allows

the model to systematically learn the dynamic interactions and relative importance of one temporal scale regarding another. To adaptively process this multi-scale information and promote the domain invariance essential for cross-subject tasks, the internal computation within each attention head j can be further conceptualized as incorporating scale-specific transformations $(\mathcal{F}^i_{Q,j}, \mathcal{F}^i_{K,j})$ for queries and keys, and a domain-invariant projection $(\mathbf{\Pi}_{inv})$ applied to values:

$$Z_{\text{head}_j} = \text{Attention}\Bigg(\oplus_{i=1}^{T} \mathcal{F}^i_{Q,j}(R_i^T) W_{Q,j}, \oplus_{i=1}^{T} \mathcal{F}^i_{K,j}(R_i^T) W_{K,j}, \left(\oplus_{i=1}^{T} R_i^T\right) W_{V,j} \mathbf{\Pi}_{\text{inv}} \Bigg). \tag{7}$$

Here, $\mathcal{F}^i$ represents learnable functions adapting to features from scale i, while $W_{Q,j}, W_{K,j}, W_{V,j}$ are standard projection matrices. The term $\mathbf{\Pi}_{inv}$ is a learnable domain-invariant projection matrix. It explicitly transform the value features into a latent subspace where subject-specific variations are suppressed. Let $V_{\text{raw}} = (\oplus_{i=1}^{T} R_i^T) W_{V,j}$ be the value representation before projection. The projection is performed via a simple linear transformation:

$$V_{\text{proj}} = V_{\text{raw}} \mathbf{\Pi}_{\text{inv}}. \tag{8}$$

The matrix $\mathbf{\Pi}_{\text{inv}}$ is not pre-defined but is trained end-to-end with the entire model. This process encourages the projection to retain emotionally relevant information while discarding domain-specific noise. Specifically, $\mathbf{\Pi}_{\text{inv}}$ functions as a learnable subspace projector driven by the adversarial minimax game. It learns to collapse feature dimensions associated with subject-specific noise while preserving orthogonal directions corresponding to shared emotional semantics. This ensures that the Value content is purified of domain bias prior to attention-based integration. As a result, the attention mechanism can focus on dynamically weighing temporal dependencies from a more robust and generalizable feature representation. The overall dynamic temporal representation $Z \in \mathbb{R}^{L_T \times d}$ is then constructed by aggregating the outputs Z_{head_j} from all n heads, followed by a final projection W_O, summarized by the standard multi-head attention equation:

$$Z = \text{Concat}(Z_{head_1}, \ldots, Z_{head_n}) W_O, \tag{9}$$

this resulting representation Z, which encapsulates adaptively fused multi-scale temporal patterns, is subsequently passed through a fully connected layer. This layer scales the features to the final dimensionality $C_{total} = c \times f$, enhancing the model's ability to generalize across subjects and supporting robust domain adaptation in EEG-based emotion recognition.

3) Contextual Gate Layer: In EEG-based emotion recognition, individual variability can introduce noise that degrades the performance of domain adaptation networks. To address this issue, we propose a Contextual Gate Layer that leverages global context to suppress individual-specific noise and highlight emotional features. This layer enhances the ability of the Dynamic Temporal Attention module to capture robust temporal dynamics. Specifically, a context vector $\mathbf{c}$ is computed as the average of all feature vectors in the batch, encapsulating shared and emotional features:

$$\mathbf{c} = \frac{1}{n_s} \sum_{i=1}^{n_s} Z_i, \tag{10}$$

where n_s is the batch size. This context vector $\mathbf{c}$ represents the batch's central tendency, smoothing individual variations via averaging. It captures shared patterns potentially relevant to the task and acts as a reference against which individual features are evaluated. Then the context vector is concatenated with each individual feature vector Z_i and processed through a gating layer to generate adaptive gate weights $\mathbf{w}_i$:

$$\mathbf{w}_i = \sigma(\text{fc}_g([Z_i, \mathbf{c}])), \tag{11}$$

the gate weights $\mathbf{w}_i$ are used to selectively modulate features, dynamically emphasizing relevant input aspects while suppressing individual-specific noise. This modulation is achieved by passing the concatenated input $[Z_i, \mathbf{c}]$ through the fully connected layer fc_g, followed by a non-linear activation function. Here, the fc_g layer learns to compare Z_i against the context $\mathbf{c}$, identifying dimensions in Z_i that align with shared patterns versus those that deviate significantly. The sigmoid function then produces weights between 0 and 1, where higher values preserve features deemed relevant based on context alignment, while lower values suppress deviating or potentially noisy features. The gated feature vector Z'_i is then calculated as:

$$Z'_i = Z_i \cdot \mathbf{w}_i, \tag{12}$$

this element-wise multiplication applies the learned weights, effectively filtering Z_i by attenuating features identified as less relevant or potentially noisy. To encourage sparsity, L1 regularization is applied to $\mathbf{w}_i$. L1 regularization promotes sparsity in $\mathbf{w}_i$, pushing weights of less important features towards zero. This enforces a form of feature selection, further enhancing noise reduction by discarding irrelevant dimensions. The refined features are subsequently passed through another fully connected layer to produce the final output y_i.

By adapting feature representations based on batch-wide context, the Contextual Layer effectively reduces individual-specific noise and enhances the DTA module's ability to extract dynamic temporal features. We extract temporal features with rich sentiment information through the DTA module. It is used for subsequent domain adversarial training to constrain the DTA module to extract domain-invariant sentiment features.

C. Chebyshev Graph Domain Adaptation

The ChebyDiscriminator, a ChebyNet-based domain discriminator, is developed to enhance EEG domain adaptation by leveraging the spatial and functional connectivity of EEG signals through graph matching.

1) Domain Discrimination Task: The domain adversarial task in DANN is reformulated as a cross-domain graph matching problem by the ChebyDiscriminator. To guide the Chebyshev matrix updates, a Gradient Reversal Layer (GRL) is applied before feeding the features from the shared feature extractor into the discriminator. The GRL reverses the gradient during backpropagation, enabling adversarial training that

encourages domain-invariant features. For source and target domain features H_s and H_t, the loss is defined as:

$$\mathcal{L}_{\text{dis}} = -\sum_{i=1}^{N_s} \log \mathbf{y}(H_s) - \sum_{i=1}^{N_t} \log(1 - \mathbf{y}(H_t)), \tag{13}$$

where $\mathbf{y}(H_s)$ and $\mathbf{y}(H_t)$ are the discriminator outputs for source and target samples, and N_s and N_t are the sample counts. This loss encourages the model to produce features that facilitate the update of the Chebyshev matrix across domains.

2) Chebyshev Graph Matching Task: To model EEG signals as graph data, the ChebyDiscriminator constructs a graph structure where EEG channels serve as nodes representing neural activity, and edges capture spatial or functional relationships. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where N is the number of EEG channels, is dynamically learned and symmetrized with self-loops added for stability during feature propagation. It is constructed as:

$$\mathbf{A} = \text{ReLU}\left(\mathbf{A}_{[0]} + \mathbf{A}_{[0]}^{\top}\right) + \mathbf{I}. \tag{14}$$

where $\mathbf{A}_{[0]} \in \mathbb{R}^{N \times N}$ is a learnable parameter matrix used to capture the initial latent spatial dependencies among EEG channels. To ensure that the graph structure is optimized purely from the data during training, $\mathbf{A}_{[0]}$ is initialized using a uniform distribution $\mathcal{U}(0, 1)$, rather than relying on a fixed pre-defined topology.

The adjacency matrix $\mathbf{A}$ defines the graph's connectivity, and the normalized graph Laplacian $\tilde{\mathbf{L}}_G$ facilitates spectral filtering. The degree matrix $\mathbf{D}$ is computed as:

$$D_{ii} = \sum_j \mathbf{A}_{ij}, \quad \mathbf{D}_{\text{norm}} = \text{diag}(D_{ii}^{-\frac{1}{2}}). \tag{15}$$

Chebyshev polynomials $T_k(\tilde{\mathbf{L}}_G)$ approximate graph convolutions efficiently, defined as:

$$T_k(\tilde{\mathbf{L}}_G) = \begin{cases} \mathbf{X}, & k = 0 \\ \tilde{\mathbf{L}}_G \mathbf{X}, & k = 1 \\ 2\tilde{\mathbf{L}}_G T_{k-1}(\tilde{\mathbf{L}}_G) - T_{k-2}(\tilde{\mathbf{L}}_G), & k \geq 2 \end{cases} \tag{16}$$

The Chebyshev convolution for an input signal $\mathbf{X}$ is:

$$\mathbf{Z} = \sum_{k=0}^{K-1} T_k(\tilde{\mathbf{L}}_G)\mathbf{W}_k, \tag{17}$$

where $\mathbf{W}_k \in \mathbb{R}^{d \times d'}$ are learnable weights, and K is the order of the polynomial approximation. Node features $\mathbf{Z}$ are processed with a ReLU activation and dropout:

$$\mathbf{H}_{\text{node}} = \text{Dropout}\left(\text{ReLU}(\mathbf{Z} + \mathbf{b})\right). \tag{18}$$

Node embeddings are aggregated into a graph-level representation $\mathbf{h}$ by mean pooling:

$$\mathbf{h} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_{\text{node}}^{(i)}. \tag{19}$$

Finally, $\mathbf{h}$ is processed through a fully connected layer and sigmoid activation to produce $\mathbf{y}$, which guides the update of the Chebyshev matrix by providing feedback for adjusting the adjacency matrix based on the learned graph representations:

$$\mathbf{y} = \sigma\left(\mathbf{h}\mathbf{W}_{\text{DP}} + b_{\text{DP}}\right) \tag{20}$$

The graph structure is dynamically optimized during training to align the source and target domains. At iteration t, the update process for aligning the edge and conditional distributions is defined as:

$$\mathbf{A}^{(t)} \leftarrow \mathbf{A}^{(t-1)} - \eta \nabla_{\mathbf{A}} \Big[\mathcal{L}_{\text{match}} + \alpha \mathcal{D}_{\text{marginal}}(\mathcal{P}_s, \mathcal{P}_t) + \beta \mathcal{D}_{\text{conditional}}(\mathcal{P}_s, \mathcal{P}_t)\Big], \tag{21}$$

where η is the learning rate, and $\mathcal{L}_{\text{match}}$ is the supervised matching loss. $\mathcal{D}_{\text{marginal}}$ and $\mathcal{D}_{\text{conditional}}$ quantify the discrepancies between domain distributions. Specifically, we employ the Maximum Mean Discrepancy to calculate these terms based on the graph-level representations $\mathbf{h}$. The marginal distribution divergence is calculated as:

$$\mathcal{D}_{\text{marginal}}(\mathcal{P}_s, \mathcal{P}_t) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{h}_{s,i} - \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{h}_{t,j} \right\|_{\mathcal{H}}^2, \tag{22}$$

where $\mathbf{h}_{s,i}$ and $\mathbf{h}_{t,j}$ denote the high-level graph representations of the source and target samples, respectively, and $\mathcal{H}$ represents the Reproducing Kernel Hilbert Space. Similarly, the conditional distribution divergence considers the class-specific alignment using pseudo-labels $\hat{y}_t$ for the target domain:

$$\mathcal{D}_{\text{conditional}}(\mathcal{P}_s, \mathcal{P}_t) = \sum_{c=1}^{C} \left\| \frac{1}{N_s^{(c)}} \sum_{\mathbf{h}_s \in \mathcal{D}_s^{(c)}} \mathbf{h}_s - \frac{1}{N_t^{(c)}} \sum_{\mathbf{h}_t \in \mathcal{D}_t^{(c)}} \mathbf{h}_t \right\|_{\mathcal{H}}^2, \tag{23}$$

where C is the number of emotion classes, and $\mathcal{D}_s^{(c)}$ and $\mathcal{D}_t^{(c)}$ represent the subsets of samples belonging to class c in the source and target domains, respectively.

The matching loss $\mathcal{L}_{\text{match}}$, used in the update rule above, combines two core objectives. First, it involves a classification loss to determine the associated emotion of the source domain graph:

$$\mathcal{L}_{\text{cls}} = \sum_{i=1}^{N_s} \mathcal{L}_{\text{CE}}(f(\mathbf{A}_s^i), y_s^i), \tag{24}$$

where $f(\mathbf{A}_s^i)$ is the predicted emotion class probability and y_s^i is the true emotion label. Here, f denotes the emotion classification function that maps the source domain graph adjacency matrix to predicted class probabilities using the network's classifier.

Second, it includes a similarity loss to match the connectivity patterns between source and target domain graphs:

$$\mathcal{L}_{\text{sim}} = \|\mathbf{A}_s - \mathbf{A}_t\|_F^2, \tag{25}$$

The overall graph matching loss is thus defined as:

$$\mathcal{L}_{\text{match}} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{sim}}, \tag{26}$$

where λ is a hyperparameter balancing the contributions of $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{sim}}$. It avoids bias in the graph matching task and prevents it from over-emphasizing a sub-task.

3) Overall Loss: The ChebyDiscriminator transforms the data into graph-based representations and captures high-order spatial dependencies. The resulting embeddings from the source and target domains are processed to extract domain-invariant features. The final loss combines the domain discrimination and graph matching objectives as follows:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{dis}} + \mathcal{L}_{\text{match}}, \tag{27}$$

where $\mathcal{L}_{\text{match}}$ aligns the connectivity patterns and conditional distributions between source and target domain graphs, targeting the alignment of conditional probability distributions for emotion-specific features, and $\mathcal{L}_{\text{dis}}$ encourages features that facilitate the update of the Chebyshev matrix across domains by aligning marginal probability distributions. This loss fosters domain-invariant, structurally consistent features.

IV. Experiments

A. Emotional Datasets and Preprocessing

To demonstrate the effectiveness of our method for cross-subject EEG-based emotion recognition, we conduct experiments on two public benchmark datasets: SEED [24] and DEAP [26].

SEED. The SEED dataset records EEG data from 15 Chinese subjects (7 males and 8 females, mean age 23.27 $\pm$ 2.37 years). It contains three different emotion categories: positive, negative, and neutral. Emotional states were elicited using 15 carefully selected and edited Chinese film clips, each lasting approximately 4 minutes, designed to effectively elicit the corresponding emotions. For each subject, the data collection process was repeated on three different dates to acquire data from three sessions, each comprising 15 trials (corresponding to the 15 film clips) of EEG signals.

DEAP. The DEAP dataset is a multimodal emotion recognition dataset that collected multimodal physiological signals from 32 healthy subjects (age range 19-37 years, mean age 26.9, 50% female) using a 32-channel electrode cap while they watched videos [26]. The emotion induction process involved 40 one-minute long excerpts of music videos selected from a larger pool based on affective tags from the last.fm website and online subjective assessments, aiming to elicit varied emotional responses. After watching the videos, subjects rated their emotions on valence, arousal, dominance, and liking using the Self-Assessment Manikin with a continuous scale from 1 to 9. For our experiments, the valence and arousal scores were categorized into two groups: high valence (valence $>$5) and low valence (valence $\leq$ 5); high arousal (arousal $>$5) and low arousal (arousal $\leq$ 5).

To mitigate potential noise, we first downsampled the SEED dataset from 1000 Hz to 200 Hz and the DEAP dataset from 512 Hz to 200 Hz. Subsequently, we applied a 1-50 Hz bandpass filter to both datasets to retain frequency components most relevant to emotional states. Through this filtering, the signals were decomposed into five key frequency bands for further analysis: Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-14 Hz), Beta (14-30 Hz), and Gamma (30-50 Hz). To eliminate noise and artifacts, we employed a linear dynamic system approach to refine the signals, ensuring the purity of the extracted features. Finally, after applying the proposed channel normalization method, we computed DE features using a 1-second time window with a step size of 1 second.

B. Experimental Setups

To ensure the reproducibility of the results, we provide a detailed description of the model parameters. We employ the RMSprop optimizer with an initial learning rate of 0.001. The batch size is set to 48, the maximum number of iterations is set to 1000, and the Gaussian generation coefficient is set to 0.2. These parameter choices are designed to optimize model performance and maintain consistency across experiments. Specific parameterizations are presented in Table I.

TABLE I
Experimental hyperparameter settings.

Parameters	Value
Batch Size	48
Learning Rate	0.001
Weight Decay	1e-5
Max Iters	1000
Optimizer	RMSprop
Noise Ratio	0.2
Chebyshev Degree K	3
Multiscale Layers T	3
Input Length L	16
Convolution Kernel Size1	(1, 8)
Convolution Kernel Size2	(1, 4)
Convolution Kernel Size3	(1, 2)
Hidden Size1	64
Hidden Size2	64

C. Experiment Protocols

To rigorously evaluate the cross-subject performance of the proposed method, we employed three distinct evaluation protocols. The detailed performance of our method under these protocols is presented in Table II , providing a basis for comparison with SOTA approaches. For a more obvious comparison, we use * to mark out domain adaptation methods.

1) One-to-One Transfer Validation. The one-to-one transfer paradigm (Protocol-1) involves using data from a single individual as the source domain, with each remaining individual sequentially serving as the target domain for testing [27]. Specifically, each iteration of training and testing exclusively incorporates data from two subjects. Initially, the first subject is assigned as the target domain, while the subsequent subject serves as the source domain. Upon reaching the final subject as the target domain, the first subject is employed as the source domain.

2) Cross-Subject Single-Session Leave-One-Subject-Out Cross-Validation. The cross-subject single-session leave-one-subject-out cross-validation protocol (Protocol-2) is widely recognized as a standard validation approach in EEG-based emotion recognition research [28], [29]. In this Protocol, data from a single session of one subject serves as the target set, while data from single sessions of the remaining subjects constitutes the source set. For example, in the SEED dataset,

TABLE II
PERFORMANCE METRICS ACROSS DIFFERENT EXPERIMENTAL PROTOCOLS FOR SEED AND DEAP DATASETS

Protocols	Datasets	P_{acc}	F1 score	Recall	Precision	PR-AUC
Protocol-1	SEED	83.12±09.63	83.08±09.52	83.12±09.61	83.17±09.67	84.71±08.93
	DEAP (Valence)	63.04±07.82	63.09±07.80	63.04±07.78	63.11±07.81	65.42±07.52
	DEAP (Arousal)	66.58±10.12	66.59±10.13	66.58±10.16	66.61±10.11	68.11±10.06
Protocol-2	SEED	94.69±05.16	94.57±05.11	94.61±05.14	94.63±05.10	94.19±05.09
	DEAP (Valence)	69.45±07.26	69.49±07.21	69.45±07.23	69.47±07.22	72.13±07.18
	DEAP (Arousal)	67.45±08.28	67.51±08.24	67.54±08.27	67.56±08.21	70.22±08.18
Protocol-3	SEED	88.12±04.04	88.09±04.01	88.12±04.09	88.17±04.01	90.31±04.03

Note: P_{acc}: Overall correctness; F1: F1 Score; Recall: True Positive Rate; Precision: Positive Predictive Value; PR-AUC: Area Under the Precision-Recall Curve.

each subject's single session contains 3394 samples and we stacked 3394 samples to get 212 samples, resulting in a source domain of 14 subjects × 1 session × 212 samples and a target domain of 1 subject × 1 session × 212 samples. Specifically, data from a single session of one subject is designated as the test set, while data from the corresponding single sessions of all remaining subjects are used as the training set, iterating until each individual has been designated as the test set once.

3) Cross-Subject Cross-Session Leave-One-Subject-Out Cross-Validation. To evaluate the model's effectiveness on unseen subjects and sessions, all sessions from one subject are designated as the target set, while all sessions from the remaining subjects serve as the source set [20]. For instance, in the SEED dataset, each session comprises 212 samples per subject. In the cross-subject cross-session protocol (Protocol-3), the test set consists of 1 subject × 3 sessions × 212 samples, while the training set is composed of the remaining 14 subjects × 3 sessions × 212 samples. This procedure is iteratively repeated until each subject has been designated as the test set at least once.

D. One-to-One Transfer Paradigm Results

To evaluate the generalizability of our model, we conducted an one-to-one transfer validation on the SEED and DEAP datasets. We present the experimental results of one-to-one transfer validation for SEED and DEAP datasets in Table III and Table IV. Our model achieved (83.12%±9.63%) accuracy in the triple classification task on SEED dataset. It achieved (63.04%±7.82%) and (66.58%±10.12%) accuracy on the Valence and Arousal dimensions of the DEAP dataset, respectively. Compared to other models that have conducted correlation-based experiments, our model achieved the best performance on both datasets. This demonstrates that our model is capable of achieving high performance even in scenarios with limited data. The ability to generalize effectively in the presence of sparse or insufficient data is a key strength of our model, as it maintains robustness across different individuals and is not overly reliant on a large amount of training data from a single source.

E. Cross-Subject Single-Session Leave-One-Subject-Out Cross-Validation Results

The cross-subject single-session leave-one-subject-out cross-validation is currently the most popular evaluation protocol in a BCI systems. In Table V and Table VI, we summarize the experimental results on the SEED database and DEAP database. We also compare them with the related literature. The results show that our proposed model achieves the best performance on SEED dataset (94.69%±5.16%), and the performance on the DEAP dataset for Valence and Arousal classification achieved accuracies of (69.45%±7.26%) and (67.52%±8.28%), respectively. Compared to the recently proposed DANN-based domain adaptation networks, in which both source and target domain data are used for modeling (e.g., GDAKF [5], MASA-TCN [4], LResCapsule [35]) and some traditional machine learning methods [13], [36], [37], [38], our model transforming the emotion recognition task into a graph matching task can avoid the inherent shortcomings of the DANN design (e.g., only considering feature separability on the source domain, pattern collapse). Fig. 2 presents the confusion matrices for the SEED and DEAP datasets alongside several comparative methods, illustrating the recognition accuracy for each emotion. On the SEED dataset, "positive" and "negative" emotions are the

TABLE III
MODEL PERFORMANCE ON SEED USING ONE-TO-ONE TRANSFER VALIDATION

Methods	P_{acc} (%)	F1 score	P value
DANN* [30]	61.57±14.01	62.41±12.97	<0.01
DASER* [31]	65.94±15.99	66.71±14.69	<0.01
PLRSA* [32]	67.51±08.19	67.54±08.08	<0.01
JTSR* [33]	75.72±02.81	74.51±02.93	<0.01
CSDS* [34]	65.80±06.74	64.75±06.51	<0.01
EPNNE [27]	81.30±03.50	81.49±03.40	0.49
Ours	**83.12±09.63**	**83.08±09.52**	–

Note: P values are computed using a two-tailed paired t-test between the proposed and each baseline based on P_{acc}.

TABLE IV
MODEL PERFORMANCE ON DEAP USING ONE-TO-ONE TRANSFER VALIDATION

Methods	Valence	Arousal	P value
DANN* [30]	50.50±09.79	51.01±10.06	<0.01
DASER* [31]	59.39±15.55	60.17±13.58	0.06
PLRSA* [32]	57.60±12.53	64.10±17.69	0.13
GDAKF* [5]	63.04±07.82	62.87±10.37	0.25
Ours	**64.81±08.97**	**66.58±10.12**	–

Note: P values are computed using a two-tailed paired t-test between the proposed method and each baseline based on classification accuracy.

TABLE V
MODEL PERFORMANCE ON SEED USING CROSS-SUBJECT SINGLE-SESSION LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION

Methods	P_{acc} (%)	F1 score	P value
Traditional Machine Learning Methods			
TKL [13]	63.54±15.47	65.41±15.14	<0.01
T-SVM [13]	72.53±14.00	74.53±14.24	<0.01
KPCA [36]	61.28±14.62	62.35±14.53	<0.01
TPT [36]	75.17±12.83	77.87±14.86	<0.01
DNN [36]	61.01±12.38	63.65±11.54	<0.01
SVM [36]	58.18±13.85	59.75±13.68	<0.01
Deep Learning Methods			
PPDA* [6]	86.70±07.10	86.78±07.16	<0.01
CL-PSR-TL* [29]	86.80±07.20	86.71±07.29	<0.01
PR-PL* [20]	93.06±05.12	93.04±06.17	0.38
C-PDDM* [28]	92.19±04.70	92.41±04.79	0.16
TMLP-SRDANN* [18]	81.04±06.28	81.62±06.29	<0.01
MS-MWA* [22]	91.36±07.00	91.94±07.11	0.14
MSGDAN* [23]	89.80±03.40	90.49±03.24	<0.01
DANN* [30]	81.65±09.12	82.53±08.55	<0.01
GCPL [39]	80.74±06.05	88.00±06.23	<0.01
MECAM [40]	91.29±03.27	91.27±03.16	0.09
DAEST* [41]	88.10±03.60	89.76±07.54	<0.01
TT-CDAN* [42]	93.62±05.80	93.41±06.16	0.59
Ours	**94.69±05.16**	**94.59±05.11**	–

Note: P values are estimated using t-tests between the proposed method and each baseline based on P_{acc}.

most distinguishable. On the DEAP dataset, high valence and high arousal states are more readily identified compared to low valence and low arousal states. Fig. 3 visually illustrates the model's classification performance through AUC (Area Under Curve) and ROC (Receiver Operating Characteristic). Notably, our model exhibits rapid learning efficiency, with AUC scores converging to high values during training. The final test-set ROC curves further validate this, yielding macro-average scores of 0.971 on SEED, 0.759 on DEAP (Valence), and 0.778 on DEAP (Arousal). Furthermore, compared to SOTA methods, our proposed approach achieves the highest performance across various emotion recognition tasks, demonstrating its strong generalization capability.

TABLE VI
MODEL PERFORMANCE ON DEAP USING CROSS-SUBJECT SINGLE-SESSION LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION

Methods	Valence	Arousal	P value†
Traditional Machine Learning Methods			
SVM [37]	64.30	66.20	<0.01
NB [26]	57.60	62.00	<0.01
ANN [38]	66.35	65.21	<0.01
Deep Learning Methods			
CAEM [43]	60.20	62.09	<0.01
MASA-TCN [4]	62.71	63.51	<0.01
DT-EEGNet [44]	63.44	63.87	<0.01
LResCapsule [35]	66.84	61.62	<0.01
SECT* [45]	66.83	65.31	<0.01
TMLP-SRDANN* [18]	61.88	57.70	<0.01
CARL-DSAN* [21]	67.11	68.67	0.64
Ours	**69.45**	**67.52**	–

Note: P values are computed based on the average accuracy of Valence and Arousal dimensions.

TABLE VII
PARAMETERS OF THE PROPOSED AND BASELINE MODELS.

Model	Parameters	SEED	DEAP
		Test time (ms)	Test time (ms)
DNN [36]	0.01M	3.348	2.478
SVM [46]	0.01M	3.944	2.541
Adaboost [47]	0.01M	3.265	2.221
RF [48]	0.01M	3.416	2.465
DANN* [30]	0.15M	4.324	4.247
PR-PL* [20]	0.51M	6.616	5.598
LResCapsule [35]	4.15M	28.324	25.376
LGGNet [8]	3.28M	22.235	19.531
RSCG-PMDA [49]	3.01M	17.425	16.313
EPNNE [27]	0.67M	7.932	6.842
Ours	**0.45M**	**4.932**	**4.768**

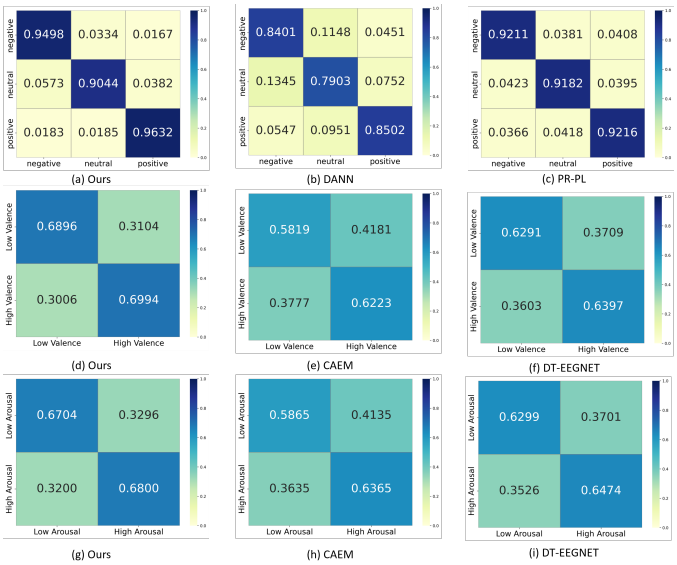


Fig. 2. Confusion Matrices for Cross-Subject Single-Session Leave-One-Subject-Out Experiments on SEED and DEAP with different methods.

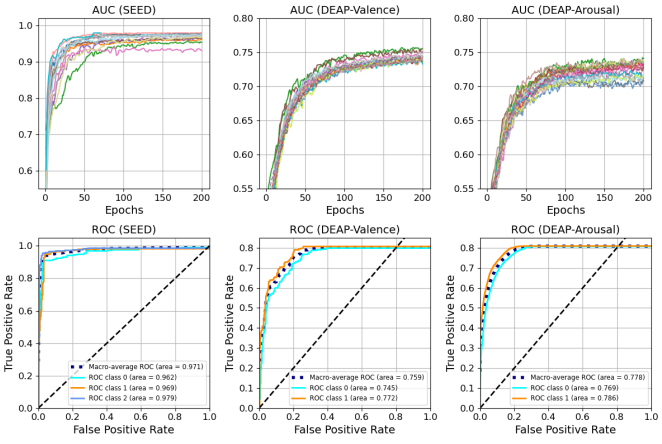


Fig. 3. AUC curves and ROC curves for SEED and DEAP datasets. Different colored lines in the AUC curves represent different individuals, while in the ROC curves represent different categories.

TABLE VIII
MODEL PERFORMANCE ON SEED USING CROSS-SUBJECT CROSS-SESSION LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION

Methods	P_{acc} (%)	F1 score
Traditional Machine Learning Methods		
RF [48]	69.60±07.64	70.69±07.45
KNN [50]	60.66±07.93	60.91±08.42
SVM [46]	68.15±07.38	70.54±07.21
Adaboost [47]	71.87±05.70	72.86±05.76
TCA [51]	64.02±07.96	64.02±07.32
CORAL [52]	68.15±07.83	69.34±08.43
SA [53]	61.41±09.75	62.43±10.32
GFK [54]	66.02±07.59	66.65±07.81
Deep Learning Methods		
DCENet [16]	87.55±07.62	–
DANN* [30]	81.08±05.88	82.43±05.31
ST-DADGAT* [19]	84.68±05.52	–
CL-PSR-TL* [29]	85.80±05.30	85.53±05.64
PR-PL* [20]	86.54±05.17	87.54±05.37
C-PDDM* [28]	86.49±05.20	87.45±05.43
Ours	**88.12±04.04**	**88.09±04.01**

F. Cross-Subject Cross-Session Leave-One-Subject-Out Cross-Validation Results

To evaluate the model's generalization ability, we conducted additional validation using data collected from different subjects across multiple sessions. Since the DEAP dataset only provides single-session recordings, cross-session and cross-subject experiments were performed exclusively on the SEED dataset. This evaluation approach accounts for both subject and session variability, thereby ensuring the reliability and stability of the results while rigorously testing the model's generalization to unseen conditions. We show the results of cross-subject cross-session leave-one-subject-out cross-validation in Table VIII. And our proposed model achieves an accuracy of (88.12%±4.04%) in the three-class classification task of SEED. This assessment scheme, which incorporates subject and session variability, is the most difficult of the three cross-validation schemes for EEG-based emotion recognition. Compared to existing Domain Adaptation methods and traditional machine learning methods, our model achieves a 1.58% improvement in classification accuracy in the SEED dataset and a reduction in standard deviation. Meanwhile, table VII proves that our proposed method achieves better results with less overhead when compared to the latest depth methods. These findings highlight the effectiveness and robustness of our model in dealing with inter-subject and inter-session variability in EEG-based emotion recognition.

G. Ablation Study

We conducted the ablation study to evaluate the contribution of each key module to emotion recognition performance. Table IX shows the detailed metrics for various model combinations, highlighting that each module significantly improves overall performance.

In the first ablation experiment, we validated the following combinations of models: (1) baseline only (Model A), (2) baseline + Gaussian Augmentation (Model B), (3) baseline + Gaussian Augmentation + DTA (Model C), (4) baseline + Gaussian Augmentation + DTA + Contextual Gate Layer (Model D), (5) baseline + Gaussian Augmentation + DTA + ChebyDiscriminator (Model E), (6) baseline + Gaussian Augmentation + Contextual Gate Layer + ChebyDiscriminator (Model F-H), (7) baseline + DTA + Contextual Gate Layer + ChebyDiscriminator (Model I), (8) The proposed (Model J). These results were obtained using the SEED dataset under a cross-subject single-session leave-one-subject-out cross-validation strategy, ensuring robust and generalizable evaluation of the proposed framework's components.

V. DISCUSSIONS

A. Gaussian Augmentation Mitigates Noise

Integrating Gaussian noise into the training data enhances the model's capacity to extract domain-invariant features within a cross-domain context. This strategy replicates the variability seen across recording equipment, subject states, and environmental conditions, thereby bolstering the model's robustness to cross-domain data. By aligning the marginal distributions across subjects, the model prioritizes learning stable, shared emotion-related features rather than overfitting to domain-specific nuances. Consequently, the model can more effectively capture the salient signals required for accurate emotion classification.

From a theoretical perspective, injecting zero-mean Gaussian noise can be regarded as a form of Tikhonov regularization that smooths the decision boundary with respect to small perturbations of the EEG samples, thereby reducing the model's sensitivity to session-dependent artefacts and encouraging the learning of domain-invariant representations. Concretely, we scale the variance of the added Gaussian noise according to the average channel-wise power of the preprocessed EEG signals, so that the noise ratio r is defined as the ratio between the noise standard deviation and the standard deviation of each EEG channel. A moderate ratio ($r \approx 0.2$) leads to a controlled decrease of the effective SNR, which is consistent with realistic fluctuations in EEG caused by changes in electrode impedance and mild muscular and ocular artefacts, without severely distorting the underlying neurophysiological activity. To further investigate this effect, we conducted an ablation study with noise ratios $r \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ applied to the training dataset. Fig. 4 show that $r = 0.2$ achieves the best cross-domain generalization: when r is too small, the perturbed samples remain too close to the original manifold and do not significantly regularize the classifier, whereas for $r \geq 0.4$ the added noise substantially degrades the SNR of EEG waveforms and the model tends to overfit spurious noisy patterns, which in turn harms emotion-related feature learning.

In addition, we systematically compare Gaussian noise augmentation with several widely used EEG data augmentation strategies, including time warping, frequency masking, and channel swapping. While these techniques also provide performance gains over the non-augmented baseline, Gaussian noise yields more consistent improvements in cross-subject and cross-session settings with negligible computational overhead, demonstrating that it is a simple yet effective regularization method for cross-domain emotion recognition.

TABLE IX
ABLATION EXPERIMENTS OF DIFFERENT MODEL COMBINATIONS

Models	Components				Performance				
	GA	DTA	CG	CD	P_{acc}	F1 score	Recall	Precision	PR-AUC
Model A	×	×	×	×	86.71±06.48	86.66±06.51	86.73±06.44	86.64±06.46	87.75±06.54
Model B	✓	×	×	×	88.46±06.31	88.49±06.26	88.40±06.33	88.47±06.24	89.42±06.35
Model C	✓	✓	×	×	90.54±05.76	90.47±05.77	90.59±05.73	90.52±05.82	91.58±05.71
Model D	✓	✓	✓	×	91.62±05.43	91.68±05.41	91.59±05.48	91.67±05.39	92.61±05.50
Model E	✓	✓	×	✓	91.81±05.76	91.77±05.79	91.88±05.75	91.76±05.80	92.83±05.70
Model F	✓	×	✓	✓	91.12±05.61	91.13±05.57	91.07±05.67	91.20±05.58	92.10±05.66
Model G	✓	×	✓	✓	91.24±05.75	91.43±05.76	91.56±05.28	91.31±05.42	92.42±05.53
Model H	✓	×	✓	✓	91.55±05.53	91.56±05.84	91.64±05.36	91.35±05.64	92.12±05.13
Model I	×	✓	✓	✓	93.03±05.43	93.01±05.47	93.09±05.38	93.96±05.44	93.06±05.41
Model J	✓	✓	✓	✓	**94.69±05.16**	**94.59±05.11**	**94.60±05.14**	**94.63±05.10**	**94.19±05.09**

Note: GA: Gaussian Augmentation; DTA: Dynamic Temporal Attention; CG: Contextual Gate Layer; CD: Chebyshev Discriminator. Models F-H extract short-term, medium-term, and long-term temporal features using single-scale convolutional kernel and self-attention mechanisms.

TABLE X
PERFORMANCE OF DIFFERENT GRAPH REPRESENTATIONS

Graph Representations	P_{acc}	F1 Score	Recall	Precision	PR-AUC
GCN	89.45±05.32	89.12±05.28	88.97±05.34	89.23±05.30	90.15±05.31
GAT	90.82±05.18	90.54±05.15	90.67±05.20	90.32±05.17	91.89±05.19
GraphSAGE	89.93±05.40	89.67±05.37	89.88±05.42	89.55±05.38	90.72±05.41
GIN	91.12±05.25	90.89±05.22	91.03±05.26	90.76±05.23	92.34±05.24
Chebyshev	**94.69±05.16**	**94.59±05.11**	**94.60±05.14**	**94.63±05.10**	**94.19±05.09**

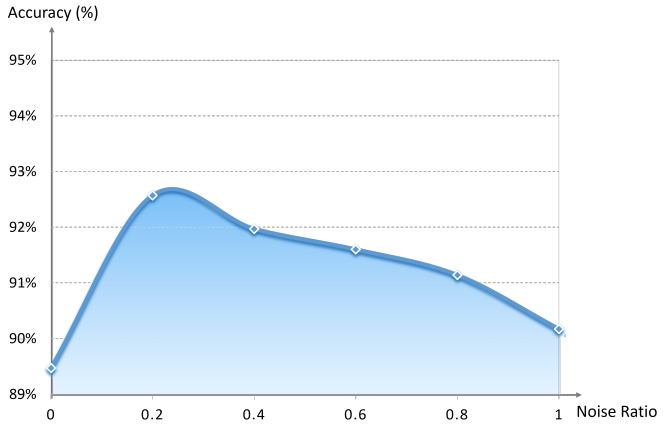


Fig. 4. Effect of Different Gaussian Noise on Model Accuracy.

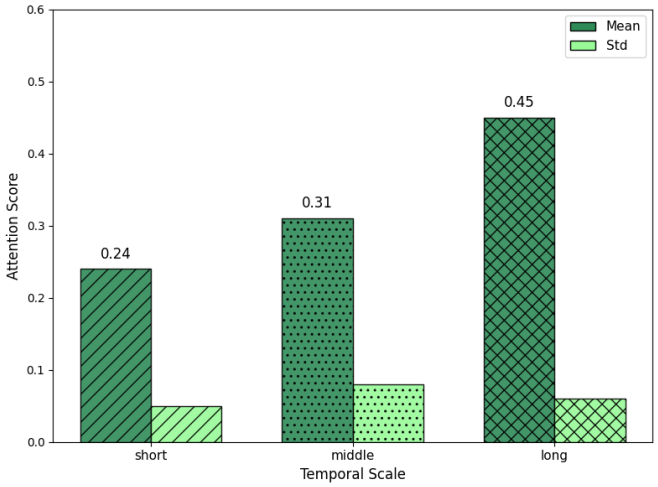


Fig. 5. Average attentional weights for all subjects in the SEED dataset.

B. DTA Extracts Domain-Invariant Features

The DTA module leverages multi-scale convolution and self-attention mechanisms to excel in emotion recognition tasks. Multi-scale convolution kernels are designed to extract emotion-related features from EEG signals across different time scales, capturing both short-term and long-term mood fluctuations. Three convolutional kernels corresponding to different time windows enable the model to capture macroscopic patterns of emotion changes over larger time spans while focusing on faster fluctuations over shorter time scales. This design allows the model to account for both long-term and short-term dependencies when dealing with the complex dynamics of mood signaling. As shown in Table IX, the single-scale variants in Models F–H achieve 91.12%, 91.24%, and 91.55% for short-, medium-, and long-term kernels, respectively, all of which are lower than the full multi-scale model that reaches 94.69%. Moreover, long-term features perform best, followed by medium-term and short-term features, suggesting that longer temporal context captures more stable emotion-related patterns. Fig. 5 illustrates that the average attention scores are 0.24 for short-term, 0.31 for medium-term, and 0.45 for long-term features. This indicates that the model relies more on longer-range temporal cues while still benefiting from short-term fluctuations.

The self-attention mechanism further enhances the DTA module by dynamically weighting features along the temporal dimension. This enables the model to dynamically prioritize emotionally salient moments while filtering out noise and irrelevant information, thereby improving the extraction of emotion-related signals. Additionally, the multi-head attention mechanism integrated into the module enhances its capacity

to process dynamic temporal representations by attending to different aspects of the data in parallel. We further vary the number of attention heads $h \in \{1, 2, 4, 8\}$ and obtain average accuracies of 94.05%, 94.42%, 94.69%, and 94.65%, respectively, showing that performance improves as h increases to 4 and then saturates. therefore, we set $h = 4$ to balance accuracy and efficiency. Crucially, although the convolutional receptive fields are fixed, the self-attention mechanism empowers the model to dynamically adapt to inter-subject and inter-emotion variability. By assigning data-driven weights to these scales for each input sample, the network avoids a rigid fusion strategy. Instead, it automatically prioritizes the temporal resolution that best characterizes the specific physiological manifestation of the current emotion, capturing both transient and sustained patterns.

The contextual gate layer serves as an augmentation module that further improves the model's context-awareness. By computing and integrating global contextual information, the contextual gate layer helps the model optimize feature selection and improve its ability to learn emotionally relevant features. The gating mechanism adaptively adjusts feature weights according to the global context, ensuring that emotionally relevant features are given higher importance. This improves the model's accuracy across individuals and contexts and boosts cross-domain transfer performance. Crucially, these two components play distinct yet complementary roles: while the DTA attention mechanism focuses on intra-sample temporal dynamics to capture fine-grained fluctuations over time, the Contextual Gate Layer operates at the inter-sample level, utilizing batch-wide statistics to calibrate individual deviations against the global distribution. The $\mathbf{\Pi}_{inv}$ projection operates not in isolation, but as a learnable subspace projector driven by adversarial training. It explicitly learns to orthogonally partition features into domain-private (subject-specific) and domain-shared (emotion-specific) subspaces. As shown in Fig. 7, while the global distribution appears similar, the projection significantly alleviates boundary confusion—pulling samples away from decision margins and toward emotion centroids, which directly contributes to the 1.66% accuracy improvement observed in Table IX.

Compared to joint spatio-temporal modeling approaches, using the DTA module as the primary feature extraction backbone to focus exclusively on the temporal dimension, followed by Chebyshev graph matching for spatial alignment, achieves a more thorough decoupling of spatio-temporal features. This avoids the feature entanglement and high computational cost associated with simultaneous spatio-temporal graph convolutions. As evidenced in Table VII, this design yields significant computational advantages while maintaining state-of-the-art accuracy. Finally, we visualize the domain-invariant features extracted by the DTA module during training and observe that the distribution of feature points for different emotions becomes increasingly distinguishable with increasing adversarial training rounds. This indicates that the DTA module continuously adjusts the edges of the data in the feature space, efficiently extracting features with emotional information. Fig. 6 shows the distribution of different categories of emotional feature points with increasing training rounds. Furthermore, we visualize the data distributions with and without the projection head $\mathbf{\Pi}_{\text{inv}}$ in Fig. 7. This demonstrates the capability of the projector to map Value features into a latent subspace where subject-specific variations are effectively suppressed.

C. Chebyshev Graph Matching Aligns Distributions

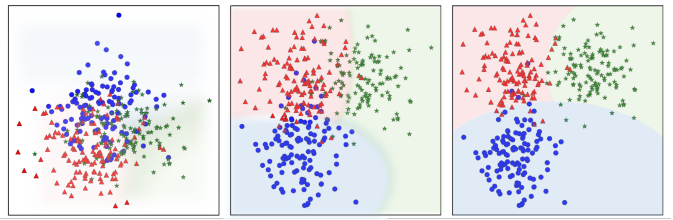
Compared to other graph neural networks (GNNs), Chebyshev graph convolutions offer several key advantages that make them particularly suitable for EEG signal processing. Specifically, Chebyshev polynomials enable an efficient approximation of graph convolutions without requiring eigenvalue decomposition, which is computationally expensive, especially for large graphs. This property makes Chebyshev convolutions more scalable and suitable for real-time applications.

Moreover, Chebyshev convolutions provide a flexible and adaptive approach to handling dynamic graphs. Unlike traditional Graph Convolutional Networks (GCNs) that rely on fixed graph structures, Chebyshev convolutions can dynamically adapt to changes in graph topology. This adaptability is crucial for EEG data, where individual-specific topological changes are common. For example, GCNs struggle with cross-subject variability in EEG data due to their reliance on static adjacency matrices, which fail to capture individual-specific topological changes. Graph Attention Networks (GATs) introduce high computational complexity through attention mechanisms, and their attention weights can be unstable due to EEG signal noise, limiting their effectiveness in cross-subject feature alignment. GraphSAGE, while efficient through sampling, often misses global spatial dependencies critical for EEG data, and Graph Isomorphism Networks (GINs) lack adaptability to dynamic graph structures, making them less effective for handling subject-specific variations in EEG connectivity.

As demonstrated in Table X, the Chebyshev-based approach achieves superior performance with an accuracy of 94.69%±5.16%, outperforming GCN (89.45%±5.32%), GAT (90.82%±5.18%), GraphSAGE (89.93%±5.40%), and GIN (91.12%±5.25%) across multiple metrics, including F1 Score, Recall, Precision, and PR-AUC. These results highlight the computational efficiency and adaptability of Chebyshev convolutions in the context of EEG signal processing.

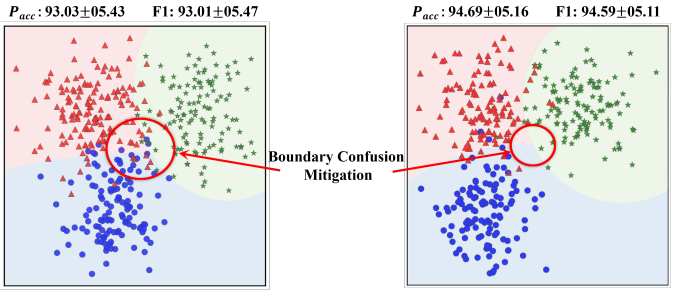
By reframing the emotion recognition task as a graph matching problem using Chebyshev graph convolutions, we enhance the alignment of domain-invariant features across different individuals in cross-domain environments. The Chebyshev graph framework effectively models the spatial and functional dependencies inherent in EEG signals. This approach enables the extraction of higher-order spatial dependencies, reducing domain discrepancies in both marginal and conditional distributions. As a result, the model learns more generalized features, improving its robustness to cross-domain EEG data.

The domain adaptation capability is markedly improved by dynamically adjusting the Chebyshev adjacency matrix A, which provides a structured representation to effectively align data distributions between source and target domains. This matrix is not predefined but is learned during training to



(a) Initial Distribution (b) Intermediate Distribution (c) Final Distribution

Fig. 6. The progression of the emotional feature space on the SEED dataset at different stages of training, illustrating the feature extractor's learning process. (a) **Initial Distribution (Round 0),** represents a disordered feature space with significant overlap among the negative (blue), neutral (green), and positive (red) categories, indicating minimal separation. (b) **Intermediate Distribution (Round 500),** demonstrates partial clustering as the feature extractor begins to differentiate emotional categories, though notable overlaps persist. (c) **Final Distribution (Round 1000),** reveals well-defined and distinct clusters for each emotion, signifying the feature extractor's effective optimization and learning.



(a) non-projected (b) projected

Fig. 7. Visualization of feature distributions with and without the domain-invariant projection head Π_{inv}. (a) Non-projected features exhibit significant class confusion and overlap at the decision boundaries due to domain-specific noise. (b) After projection, the subject-specific variability is suppressed, pulling samples closer to their respective emotion centers. The class confusion at the boundaries is significantly alleviated, resulting in highly distinguishable margins between different emotional states.

represent the functional connectivity between EEG channels, thereby capturing the spatial characteristics of brain activity. Each element A_{ij} quantifies the strength of the learned relationship between channel i and channel j; a higher value signifies a stronger functional connection relevant to the emotion recognition task. By learning A, the model adaptively identifies key brain regions and their interaction patterns that are most discriminative for emotional states, rather than depending on a fixed, universal brain topology. This structural alignment is visually evident in Fig. 8, where the adjacency matrix of the target domain progressively evolves from a disordered state to a structured pattern that closely mirrors the source domain's connectivity. This convergence confirms that our method effectively bridges the topological gap between subjects. Notably, our graph matching fundamentally differs from conventional tasks that seek exact node-to-node combinatorial correspondences between fixed topologies [23][54]. In traditional graph matching (e.g., graph isomorphism or subgraph matching), the goal is to find a permutation matrix mapping nodes between structurally rigid graphs. In contrast, since EEG electrodes are spatially aligned across subjects, we treat matching as a distributional alignment of learnable connectivity. We align domains by minimizing the Maximum Mean Discrepancy (MMD) between graph-level embeddings ($\mathbf{h_s}$, $\mathbf{h_t}$) and the Frobenius distance between dynamically learned adjacency matrices ($\mathbf{A_s}$, $\mathbf{A_t}$). As shown in Fig. 8, the target adjacency matrix evolves from a disordered state to a structured pattern mirroring the source, confirming that our method bridges the topological gap by adapting functional connectivity rather than enforcing rigid structural correspondence. This self-evolving representation accommodates inter-subject neural variability through Chebyshev spectral filtering, ensuring robust knowledge transfer. This dynamic, self-evolving representation accommodates inherent inter-subject neural variability and ensures more effective knowledge transfer, leading to better generalization across diverse conditions. Chebyshev polynomials are then utilized to approximate graph convolutions efficiently, avoiding the computational complexity of eigenvalue decomposition. This ensures effective propagation of information across broader spatial ranges within the graph, aligning source and target domains at the feature level. Theoretically, Chebyshev polynomials capture multi-scale spatial information, enabling the model to learn invariant features that mitigate the impact of inter-subject variability in EEG data. This multi-scale modeling directly addresses the challenge of domain shifts caused by subject-specific variations, ensuring robust feature alignment and improved generalization.

The adversarial training framework further strengthens this alignment, as the Gradient Reversal Layer promotes the extraction of domain-invariant features by minimizing the distinguishability between source and target data. These enhancements collectively contribute to the model's ability to generalize across subjects in EEG-based emotion recognition tasks.

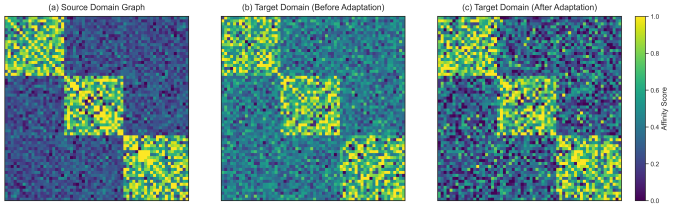


Fig. 8. Visualization of graph structure alignment on the SEED dataset.

D. Parameter Sensitivity Analysis

To investigate the impact of the hyperparameter λ in Eq. (26), which balances the classification loss and the graph matching similarity loss, we conducted a sensitivity analysis on the SEED dataset. We varied λ within the range $\{0.001, 0.01, 0.1, 0.3, 0.5, 0.8\}$ while keeping other parameters fixed. As shown in Fig. 9, the performance initially improves as λ increases, peaking at $\lambda = 0.1$. This indicates that an appropriate weight effectively bridges the domain gap by aligning graph structures. However, performance degrades when λ becomes excessively large, as the optimization becomes dominated by the structural matching objective at the expense of discriminative feature learning. Consequently, we set $\lambda = 0.1$ for all subsequent experiments to achieve an optimal trade-off.

TABLE XI
PERFORMANCE ON HBUED, MPED, AND SEED-VII DATASETS

Datasets	P_{acc}	F1 Score	Recall	Precision	PR-AUC
HBUED [55]	84.63±05.35	84.81±05.51	84.91±05.61	84.23±05.47	85.91±05.96
MPED [56]	25.13±04.83	25.92±04.64	25.28±04.43	25.48±04.39	25.24±04.32
SEED-VII [57]	46.14±08.91	46.32±08.53	46.73±08.89	46.78±08.91	46.48±08.87

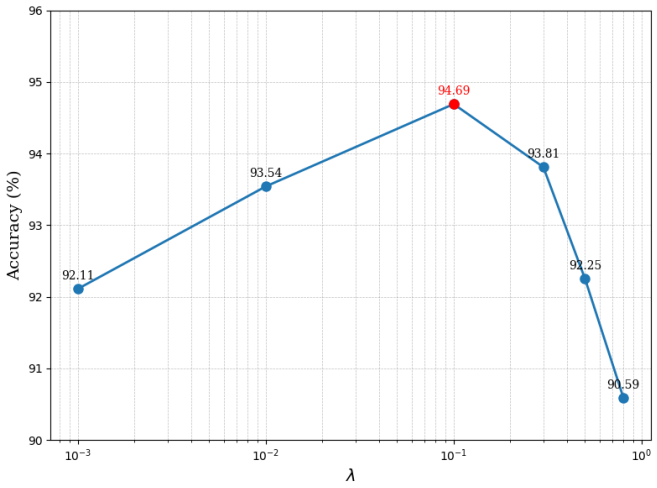


Fig. 9. Sensitivity analysis of the hyperparameter λ on the SEED dataset.

E. Limitations

The proposed model excels in cross-subject EEG emotion recognition, yet several limitations exist. First, the DTA module and Chebyshev Graph Domain Adaptation may struggle to align data distributions when individuals differ greatly from the training data in extreme variability cases, which can degrade performance. Second, EEG signals are inherently noisy and easily contaminated by muscular activity or environmental interference. Although we employ down-sampling, band-pass filtering, and a linear dynamic system for denoising, the extracted features may still be affected in low signal-to-noise conditions. Third, when we extend the evaluation from SEED and DEAP to three additional public datasets (HBUED [55], MPED [56], and SEED-VII [57]), the performance on fine-grained emotion recognition tasks remains limited. As shown in Table XI, the proposed method achieves an accuracy of $84.63\% \pm 5.35\%$ on the HBUED dataset for the four-class classification task. However, on the more challenging MPED and SEED-VII datasets, the accuracy drops to $25.13\%\pm4.83\%$ and $46.14\% \pm 8.91\%$, respectively. These results indicate that distinguishing subtle emotional states across a large number of classes remains a significant challenge for the current architecture.

VI. CONCLUSION

In this study, a Multi-scale Dynamic Temporal Network with Graph Matching Domain Adaptation is proposed. The dynamic temporal representations of EEG signals are efficiently extracted by the Dynamic Temporal Attention module. Additionally, the ChebyDiscriminator is designed to transform the emotion recognition task into a graph matching problem, thereby accomplishing the efficient alignment of EEG data across different individuals. The experimental results demonstrate that the proposed model achieves state-of-the-art cross-subject performance on two widely used open EEG emotion recognition databases, SEED and DEAP, highlighting the potential of the model.

REFERENCES

[1] Y. Wang, W. Tian, J. Xu, Y. Tian, C. Xu, B. Ma, Q. Hao, C. Zhao, and H. Liu, "Wearable wireless dual-channel EEG system for emotion recognition based on machine learning," *IEEE Sensors Journal*, vol. 23, no. 18, pp. 21767–21775, 2023.

[2] C. Fan, K. Zhu, J. Tao, G. Yi, J. Xue, and Z. Lv, "Multi-level contrastive learning: Hierarchical alleviation of heterogeneity in multimodal sentiment analysis," *IEEE Transactions on Affective Computing*, vol. 16, no. 1, pp. 207–222, 2024.

[3] C. Fan, H. Zhang, Q. Ni, J. Zhang, J. Tao, J. Zhou, J. Yi, Z. Lv, and X. Wu, "Seeing helps hearing: A multi-modal dataset and a mamba-based dual branch parallel network for auditory attention decoding," *Information Fusion*, vol. 118, p. 102946, 2025.

[4] Y. Ding, S. Zhang, C. Tang, and C. Guan, "MASA-TCN: Multi-anchor space-aware temporal convolutional neural networks for continuous and discrete EEG emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[5] D. Huang, S. Zhou, and D. Jiang, "Generator-based domain adaptation method with knowledge free for cross-subject EEG emotion recognition," *Cognitive Computation*, vol. 14, no. 4, pp. 1316–1327, 2022.

[6] L.-M. Zhao, X. Yan, and B.-L. Lu, "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, 2021.

[7] C. Tang, L. Yang, G. Cao, J. Du, and Q. Zhang, "St-gcn: EEG emotion recognition via spectral graph and temporal analysis with graph convolutional networks," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3748–3751, IEEE, 2024.

[8] Y. Ding, N. Robinson, C. Tong, Q. Zeng, and C. Guan, "LGGNet: Learning from local-global-graph representations for brain–computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[9] H. Yan, K. Guo, X. Xing, and X. Xu, "Bridge graph attention based graph convolution network with multi-scale transformer for EEG emotion recognition," *IEEE Transactions on Affective Computing*, 2024.

[10] M. Pang, H. Wang, J. Huang, C.-M. Vong, Z. Zeng, and C. Chen, "Multi-scale masked autoencoders for cross-session emotion recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1637–1646, 2024.

[11] H. Dong, J. Zhou, C. Fan, W. Zheng, L. Tao, and H. K. Kwan, "Multi-scale 3d-cru for EEG emotion recognition," *Biomedical Physics & Engineering Express*, vol. 10, p. 045018, may 2024.

[12] Y. Pei, J. Xu, Q. Chen, C. Wang, F. Yu, L. Zhang, and W. Luo, "Dtp-net: Learning to reconstruct EEG signals in time-frequency domain by multi-scale feature reuse," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 5, pp. 2662–2673, 2024.

[13] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bi-hemisphere domain adversarial neural network model for EEG emotion recognition," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 494–504, 2018.

[14] T. Pan, N. Su, J. Shan, Y. Tang, G. Zhong, T. Jiang, and N. Zuo, "GLADA: Global and local associative domain adaptation for EEG-based emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 17, pp. 167–178, 2024.

[15] W. Li, W. Huan, S. Shao, B. Hou, and A. Song, "MS-FRAN: A novel multi-source domain adaptation method for EEG-based emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[16] S. Liu, Z. Gu, Y. Zhang, Y. An, S. Zhao, B. Li, and Y. Zhang, "Dynamic collaborative evolutionary network: A novel spatio-temporal feature extraction framework for EEG emotion recognition," *Expert Systems with Applications*, vol. 298, p. 129654, 2026.

[17] S. Zhai and X. Guo, "A novel lightweight dynamic focusing convolutional neural network LAND-FCNN for EEG emotion recognition," *Measurement*, vol. 234, p. 114862, 2024.

[18] W. Li, B. Hou, X. Li, Z. Qiu, B. Peng, and Y. Tian, "TMLP + SRDANN: A domain adaptation method for EEG-based emotion recognition," *Measurement*, vol. 207, p. 112379, 2023.

[19] S. Liu, X. Wang, Z. Gu, Y. An, S. Zhao, B. Li, and Y. Zhang, "Cross-subject emotion recognition by EEG driven spatio-temporal hybrid network based on domain adaptation and dynamic graph attention," *Biomedical Signal Processing and Control*, vol. 110, p. 108231, 2025.

[20] R. Zhou, Z. Zhang, H. Fu, L. Zhang, L. Li, G. Huang, F. Li, X. Yang, Y. Dong, Y.-T. Zhang, *et al.*, "PR-PL: A novel prototypical representation based pairwise learning framework for emotion recognition using EEG signals," *IEEE Transactions on Affective Computing*, vol. 15, no. 2, pp. 657–670, 2023.

[21] Z. Wang, Y. Wang, X. Wan, and Y. Tang, "Cerebral asymmetry representation learning-based deep subdomain adaptation network for electroencephalogram-based emotion recognition," *Physiological Measurement*, vol. 45, no. 3, p. 035004, 2024.

[22] Y. Zhu, Y. Ma, M. Meng, Z. Luo, Y. Gao, and M. Sun, "Application of a multi-source multi-task weight adaptation framework for cross-domain EEG emotion recognition (MS-MWA)," *Signal, Image and Video Processing*, pp. 1–14, 2024.

[23] J. Wang, X. Ning, W. Xu, Y. Li, Z. Jia, and Y. Lin, "Multi-source selective graph domain adaptation network for cross-subject EEG emotion recognition," *Neural Networks*, vol. 180, p. 106742, 2024.

[24] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *2013 6th international IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 81–84, IEEE, 2013.

[25] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[26] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.

[27] H. Zhang, T. Zuo, Z. Chen, X. Wang, and P. Z. Sun, "Evolutionary ensemble learning for EEG-based cross-subject emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, pp. 3872–3881, 2024.

[28] J. Tao, Y. Dan, and D. Zhou, "Possibilistic distribution distance metric: a robust domain adaptation learning method," *Frontiers in Neuroscience*, vol. 17, p. 1247082, 2023.

[29] Y. Wang, Q. Li, J. Jia, and R. Zhang, "A novel transfer learning model for cross-subject emotion recognition using EEGs," in *Proceedings of the 2022 6th International Conference on Computer Science and Artificial Intelligence*, pp. 217–223, 2022.

[30] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[31] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 2, pp. 344–353, 2019.

[32] J. Luo, M. Wu, Z. Wang, Y. Chen, and Y. Yang, "Progressive low-rank subspace alignment based on semi-supervised joint domain adaption for personalized emotion recognition," *Neurocomputing*, vol. 456, pp. 312–326, 2021.

[33] Y. Peng, H. Liu, W. Kong, F. Nie, B.-L. Lu, and A. Cichocki, "Joint EEG feature transfer and semisupervised cross-subject emotion recognition," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 7, pp. 8104–8115, 2022.

[34] Y. Ma, W. Zhao, M. Meng, Q. Zhang, Q. She, and J. Zhang, "Cross-subject emotion recognition based on domain similarity of EEG signal transfer learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 936–943, 2023.

[35] C. Fan, J. Wang, W. Huang, X. Yang, G. Pei, T. Li, and Z. Lv, "Lightweight residual convolution-based capsule network for EEG emotion recognition," *Advanced Engineering Informatics*, vol. 61, p. 102522, 2024.

[36] H. Li, Y.-M. Jin, W.-L. Zheng, and 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*, pp. 403–413, Springer, 2018.

[37] D. S. Naser and G. Saha, "Recognition of emotions induced by music videos using dt-cwpt," in *2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT)*, pp. 53–57, IEEE, 2013.

[38] W. Zhang, Z. Yin, Z. Sun, Y. Tian, and Y. Wang, "Selecting transferrable neurophysiological features for inter-individual emotion recognition via a shared-subspace feature elimination approach," *Computers in Biology and Medicine*, vol. 123, p. 103875, 2020.

[39] W. Li, L. Fan, S. Shao, and A. Song, "Generalized contrastive partial label learning for cross-subject EEG-based emotion recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–11, 2024.

[40] F. Hou, W. Meng, L. Ma, K. Chen, Q. Ai, Q. Liu, and S. Q. Xie, "Mecam: A novel multi-axis EEG channel attention model for emotion recognition," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[41] X. Shen, R. Gan, K. Wang, S. Yang, Q. Zhang, Q. Liu, D. Zhang, and S. Song, "Dynamic-attention-based EEG state transition modeling for emotion recognition," *IEEE Transactions on Affective Computing*, pp. 1–17, 2025.

[42] H. Huang, X. Si, Y. Han, and D. Ming, "A novel conditional adversarial domain adaptation network for EEG cross-subject emotion recognition," *IEEE Transactions on Affective Computing*, pp. 1–13, 2025.

[43] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao, and Y.-D. Zhang, "EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network," *Knowledge-Based Systems*, vol. 265, p. 110372, 2023.

[44] Z. Jia, Y. Ouyang, X. Kong, Y. Guo, Z. Li, and H. Zeng, "A novel dual-task model for EEG-based emotion and cognition recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–14, 2025.

[45] Z. Bai, F. Hou, K. Sun, Q. Wu, M. Zhu, Z. Mao, Y. Song, and Q. Gao, "Sect: A method of shifted EEG channel transformer for emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[46] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, pp. 293–300, 1999.

[47] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[48] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[49] R. Chen, C. Xie, J. Zhang, Q. You, and J. Pan, "A progressive multi-domain adaptation network with reinforced self-constructed graphs for cross-subject EEG-based emotion and consciousness recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 3498–3510, 2025.

[50] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules," *Analytica Chimica Acta*, vol. 136, pp. 15–27, 1982.

[51] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.

[52] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, 2016.

[53] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2960–2967, 2013.

[54] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 2, pp. 354–367, 2020.

[55] S. Liu, X. Wang, Y. An, Z. Wang, Z. Gu, Y. Zhang, and S. Zhao, "HBUED: An EEG dataset for emotion recognition," *Journal of Affective Disorders*, p. 119397, 2025.

[56] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.

[57] W.-B. Jiang, X.-H. Liu, W.-L. Zheng, and B.-L. Lu, "Seed-vii: A multimodal dataset of six basic emotions with continuous labels for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 16, pp. 969–985, 2024.

Chen Rongtao. He is currently an undergraduate student majoring in Software Engineering at South China Normal University. His research interests focus on the interdisciplinary areas of affective computing and transfer learning, with a specialized emphasis on EEG signal processing and its applications across these fields.

Hong Zhepei. He is currently an undergraduate student majoring in Software Engineering at South China Normal University. His research interests focus on the interdisciplinary areas of affective computing and disorder of consciousness detection, with a specialized emphasis on EEG signal processing and its applications across these fields.

Liting Li. She is currently an undergraduate student majoring in Software Engineering at South China Normal University. Her research interests lie in the interdisciplinary field of affective computing, encompassing sentiment analysis, EEG signal processing, and human pose estimation.

Zhuoyi Huang. She is an undergraduate student at the School of Artificial Intelligence, South China Normal University. Her research focuses on affective computing based on multimodal signals. She serves as the principal investigator of a National Innovation Training Program project, leading research on emotion recognition algorithms.

Jingcong Li. He received his Ph.D. in pattern recognition and intelligent systems in 2019 from South China University of Technology, Guangzhou, China. In 2019, he joined the School of Software, South China Normal University, China, where he is now an Associate Professor. His current research interests include the fields of signal processing, neural network, hybrid brain- computer interface, and their clinical applications.

Jiahui Pan (Member, IEEE). He received a B.S. degree in computer science and technology in 2005 and an M.S. degree in computer application technology in 2008 from South China Normal University, Guangzhou, China. In 2014, he received his Ph.D. in pattern recognition and intelligent systems from South China University of Technology, Guangzhou, China. In 2008, he joined the School of Artificial Intelligence, South China Normal University, China, where he is now a Full Professor. His current research interests include the fields of signal processing, hybrid brain-computer interface, and their clinical applications.