Cross-Subject EEG Emotion Recognition with Periodic Evolution Representation and Prototype Boundary Learning

Jianing Wu (2023511249@link.tyut.edu.cn)[1], Yanrong Hao[*] (haoyanrong@tyut.edu.cn)[1], Chenchen Zhang (2022511314@link.tyut.edu.cn)[1], Jing Bian (bianjing01@tyut.edu.cn)[1], Xin Wen (xwen@tyut.edu.cn)[1], Mengni Zhou (zhoumengni@tyut.edu.cn)[1] & Rui Cao[*] (caorui@tyut.edu.cn)[1]

[1]School of Software, Taiyuan University of Technology, Taiyuan, China [*] Corresponding author

Abstract

EEG emotion recognition is important for affective computing and for understanding the neural encoding of emotion. However, EEG signals are highly non-stationary and susceptible to task-irrelevant interference. In cross-subject settings, such variability causes intra-class distribution shifts and increases inter-class overlap, undermining stable discriminative learning. We propose PEPBNet, a cross-subject EEG emotion recognition model with periodic evolution representation and prototype boundary learning. PEPBNet captures stable emotion-related rhythms by modeling multi-granularity intra-period patterns and evolutionary relations between adjacent periods. We further decompose the learned representations into mutually orthogonal emotion and interference factors to suppress non-emotional components. Prototype learning is integrated with decision boundary learning by jointly optimizing classification and prototype losses, improving intra-class compactness and inter-class separability. Experiments show that PEPBNet achieves accuracies of 96.30% on SEED and 87.04% on SEED-IV, and delivers competitive performance on DEAP for both the valence and arousal dimensions, demonstrating improved robustness and generalization.

Keywords: Electroencephalography (EEG); Emotion Recognition; Graph Neural Networks; Cross-subject; Prototypes

Introduction

Emotions play a crucial role in memory, decision-making, and social interaction. Consequently, emotion recognition is important not only for applications such as affective computing and human-computer interaction, but also for uncovering how emotional states are represented in neural activity (Meng et al., 2025). As a research field at the intersection of computer science, psychology, neuroscience, and cognitive science, physiological-signal-based emotion recognition has attracted substantial attention (Erat et al., 2024). Among various modalities, EEG is widely used due to its non-invasive acquisition, low cost, and millisecond-level temporal resolution (Mishra & Ranjan, 2026). Furthermore, EEG reflects neural electrical activity and is highly sensitive to changes in neural oscillatory patterns and inter-regional interactions associated with different emotional states.

Nevertheless, the low signal-to-noise ratio, pronounced non-stationarity, and inter-subject variability of EEG continue to hinder representation learning that is simultaneously stable and discriminative (Mohammadi Foumani et al., 2024). Consequently, recent studies have increasingly leveraged deep learning for EEG representation learning. However, as a multi-channel spatial signal, EEG may lose critical spatial information when its channels are modeled as independent Euclidean dimensions, since such modeling ignores the non-Euclidean relationships among electrodes (Pan et al., 2023). Graph Neural Networks (GNNs) address this issue by modeling electrodes as nodes and inter-channel dependencies as edges. For instance, Cao et al. (2025) learned cross-channel spatial structure with the Graph Convolutional Network (GCN), while DGAT used the multi-head Graph Attention Network (GAT) to learn inter-channel attention weights for adaptive information aggregation (Ding et al., 2025). For graph construction, integrating physiological priors as structural constraints with emotion-state-aware adaptive connectivity learning offers a balance between interpretability and adaptability to time-varying dynamics.

In terms of feature extraction, many approaches focus on learning time-domain or frequency-domain representations from EEG signals. To fully exploit multi-band information, W. Tang et al. (2025) combined an efficient capsule network with convolutional attention to improve recognition accuracy. In addition, Garg et al. (2024) integrated a lightweight MobileNet with Recurrent Neural Network (RNN) to extract time-frequency features, achieving a favorable trade-off between efficiency and performance. However, EEG exhibits pronounced oscillatory rhythms; if such periodic evolution patterns are not taken into account, models may struggle to capture the intrinsic rhythmic structure embedded in neural oscillations. This motivates explicitly representing and modeling EEG from the perspective of periodic evolution.

Moreover, in practical acquisition settings, EEG signals are inevitably affected by various sources of interference, such as inter-subject variability and environmental noise, which weaken the effective representation of emotion-related information (Dong et al., 2024). As these interference factors are difficult to eliminate directly during feature extraction, the stability and generalizability of cross-subject emotion recognition are adversely affected. Therefore, it is necessary to disentangle emotion-related factors from interference factors.

It is worth noting that in GNN-based modeling, many prior methods integrate node features with graph topology and then focus on learning discriminative decision boundaries that separate emotion classes in the embedding space (Feng et al., 2026; T. Wang et al., 2024). However, they often lack explicit constraints that promote intra-class compactness, leading to dispersed within-class clusters and degraded cross-subject

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

generalization.

Motivated by these challenges, we propose PEPBNet. Our main contributions are summarized as follows:

1. PEPBNet enhances the stability and discriminability of emotion-related representations while also providing interpretability, and extensive experiments on multiple benchmark datasets demonstrate its competitive performance.
2. We develop a multi-granularity periodic evolution representation module to explicitly model EEG rhythmicity and capture more stable emotion-evoked oscillatory patterns. We further design an emotion-interference factor decomposition module to disentangle emotion factors from interference factors, yielding cleaner features for cross-subject recognition.
3. A prototype-centralized representation module is proposed to couple prototype learning with decision-boundary optimization. Through discriminative-prototype joint optimization, the model enhances inter-class separability while enforcing intra-class compactness, leading to stronger cross-subject generalization.

Methodology

As shown in Figure 1, we first employ a Multi-Granularity Periodic Evolution Representation (MGPER) module to capture the periodic evolution of EEG signals and uncover latent rhythmic patterns. Next, the Emotion-Interference Factor Decomposition Module (EIFDM) disentangles the representation into emotion-related factors and interference factors. Based on the extracted features, we construct a Prior-Dynamic Functional Connectivity (PDFC) graph that integrates structural priors with state-adaptive connections. We then apply the Minimum-Redundancy Topology Backbone (MRTB) and Tri-View Salience Node Pooling (TVSNP) to sparsify the graph structure and preserve salient nodes. Finally, the Prototype-Centralized Representation Module (PCRM) is trained with a Discriminative-Prototype Joint Optimization (DPJO) objective to jointly enforce intra-class compactness and inter-class separability.

Multi-Granularity Periodic Evolution Representation

This module adaptively constructs candidate periods via frequency-domain analysis, captures intra-period patterns and dynamic variations between adjacent periods, and extracts multi-granularity features with multi-scale convolutions.

Multi-Period 2D Reconstruction Given an input signal $\mathbf{X} \in \mathbb{R}^{N \times B \times L}$, where N denotes the number of channels, B denotes the number of frequency bands, and L is the feature length within each band, we first apply the Fast Fourier Transform (FFT) to map $\mathbf{X}$ into the frequency domain. Then, the amplitude is computed via $\mathrm{Amp}(\cdot)$ and averaged using $\mathrm{Avg}(\cdot)$ to obtain the global intensity of each frequency:

$$A_f = \mathrm{Avg}\big(\mathrm{Amp}(\mathrm{FFT}(\mathbf{X}))\big) \tag{1}$$

The DC component is set to 0 Hz to avoid biased period selection. Meanwhile, due to the conjugate symmetry of the frequency domain, we consider only frequencies in the range $\left[1, \frac{B}{2}\right]$. Then, we select the top-k amplitudes to get the corresponding frequencies as $(f_1, \ldots, f_k)$ and map each f_i to its corresponding period p_i $(i \in 1, \ldots, k)$.

$$f_i = \arg\mathrm{Topk}(A_f) \tag{2}$$

$$p_i = \left\lceil \frac{B}{f_i} \right\rceil \tag{3}$$

After zero-padding and reshaping, $\mathbf{X}$ is converted into a set of 2D tensors, one for each period, providing a 2D representation of periodic features. Padding is applied along the temporal dimension to make the sequence length divisible by the period. Here, p_i and f_i denote the numbers of rows and columns of the 2D tensor, respectively:

$$\mathbf{X}^i_{\mathrm{2D}} = \mathrm{Reshape}_{p_i, f_i}\big(\mathrm{Padding}(\mathbf{X})\big), \quad i \in \{1, \ldots, k\} \tag{4}$$

Dynamic Multi-Scale Periodic Feature Encoding Module Within each periodic branch, we use multi-scale convolutions to capture local patterns with varying receptive fields. The m-th convolution uses kernel size s_m and stride 1. Z_m denotes the output of the convolution on the m-th scale.

$$Z_m = \mathrm{AvgPool}\Big(\mathrm{LeakyReLU}\big(\mathrm{Conv1D}(\mathbf{X}^i_{\mathrm{2D}}, s_m)\big)\Big) \tag{5}$$

Here, $m \in \{1, \ldots, M\}$ indexes the M scales. We apply $\mathrm{LeakyReLU}(\cdot)$ to increase nonlinearity and mitigate vanishing gradients, followed by $\mathrm{AvgPool}(\cdot)$ for downsampling and smoothing. Multi-scale outputs are concatenated along the feature dimension and batch-normalized via f_{bn} to stabilize the distribution, yielding the intra-period features:

$$\mathbf{Z}_q = f_{bn}\big([Z_1, \ldots, Z_M]\big) \tag{6}$$

$[\cdot]$ denotes concatenation along the feature dimension. We then apply 1D convolutions along the periodic-sequence dimension with $s_t \in \left\{(B,1), \left(\frac{B}{2}, 1\right)\right\}$ to model evolutionary trends across adjacent periods, as in Eq. (7):

$$Z_p = \mathrm{AvgPool}\Big(\mathrm{LeakyReLU}\big(\mathrm{Conv1D}(Z_q, s_t)\big)\Big) \tag{7}$$

The final output of periodic feature extraction is:

$$Z_f = f_{bn}\big([Z_1, \ldots, Z_p]\big) \tag{8}$$

We map the resulting 2D representation back to the 1D temporal space and apply $\mathrm{Trunc}(\cdot)$ to restore the original sequence length, producing a 1D feature sequence for the subsequent fusion layer:

$$Z^i_{1D} = \mathrm{Trunc}\Big(\mathrm{Reshape}_{1,(p_i \times f_i)}(Z_f)\Big) \tag{9}$$

Adaptive Aggregation Layer Based on Period Weights Since A_f reflects the contribution of each periodic component, we apply $\mathrm{softmax}(\cdot)$ to obtain normalized weights and use them to adaptively aggregate the Z^i_{1D} features, dynamically integrating key information across different periods:

$$\tilde{Z}_{1D} = \sum_{i=1}^{k} \mathrm{softmax}\big(A_{f_i}\big) \times Z^i_{1D}. \tag{10}$$

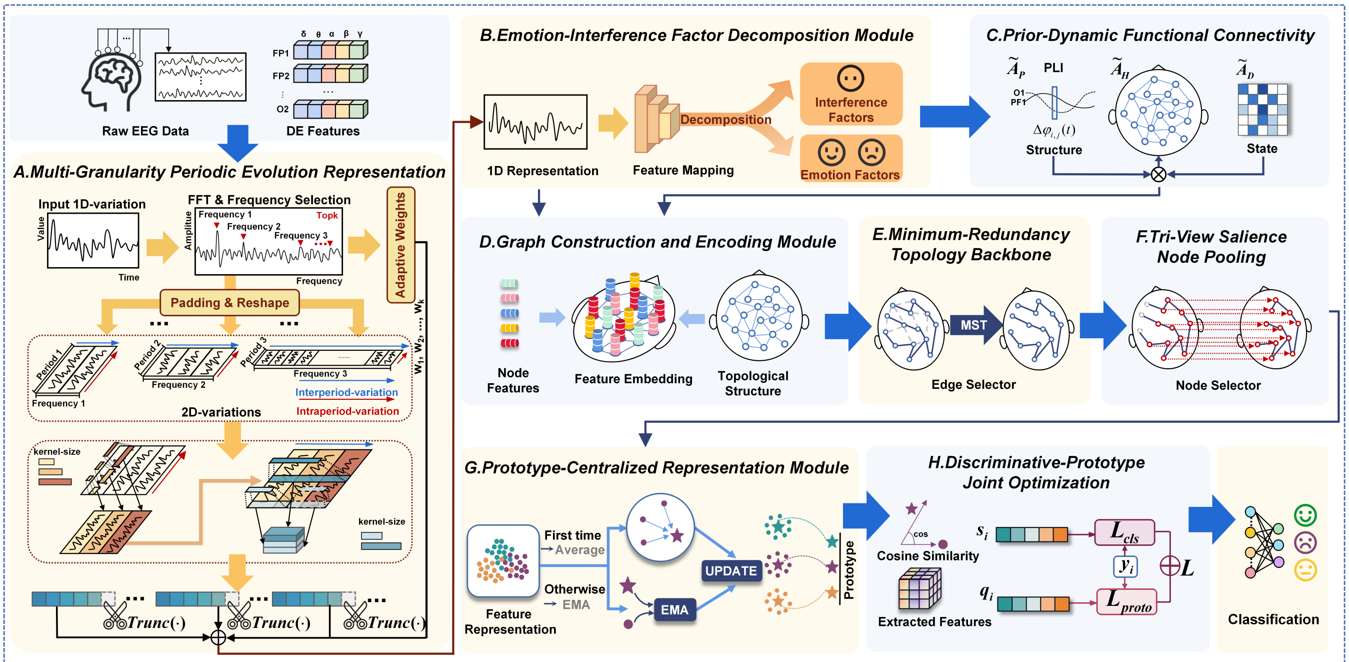


Figure 1: Architecture of PEPBNet.

Emotion-Interference Factor Decomposition Module

Because EEG features often entangle emotion-related and task-irrelevant factors, we propose EIFDM to decompose them into two complementary components. Concretely, $\tilde{Z}_{1D}$ is passed through a Multilayer Perceptron (MLP) for nonlinear mapping to obtain an emotion representation space:

$$U = \mathrm{ReLU}\big(\mathrm{MLP}(W_r \tilde{Z}_{1D} + b_r)\big) \tag{11}$$

where W_r and b_r are learnable parameters. We then construct a learnable projection matrix P to model the affective subspace, where W_p is a trainable projection basis matrix used to characterize the subspace structure, and λ is a regularization coefficient that stabilizes the projection computation:

$$P = W_p \left(W_p^{\top} W_p + \lambda I\right)^{-1} W_p^{\top} \tag{12}$$

By projecting into a learnable emotion subspace, EIFDM decomposes the representation into two orthogonal components: the emotion factor F_r and the interference factor F_i.

$$F_r = UP \tag{13}$$

$$F_i = U - F_r \tag{14}$$

Prior-Dynamic Functional Connectivity

We model the adjacency matrix as a hybrid connectivity pattern combining Structure Prior Functional Connectivity (SPFC) with State Adaptive Functional Connectivity (SAFC). The structural component captures stable connectivity patterns, while the state component characterizes dynamic structures that vary over time.

Structure Prior Functional Connectivity The Phase-Lag Index (PLI) is used to quantify phase synchronization between EEG channels and to construct neurophysiologically interpretable, stable connectivity. For channels i and j, let $\Delta\varphi_{i,j}(t)$ denote the phase difference at time t. PLI measures the sign consistency of $\Delta\varphi_{i,j}(t)$ to quantify a stable phase lead or lag relationship.

$$\mathrm{PLI}_{i,j} = \left| \frac{1}{T} \sum_{t=1}^{T} \Big(\mathrm{sgn}\langle \sin(\Delta\varphi_{i,j}(t)) \rangle - \mathrm{sgn}\langle \cos(\Delta\varphi_{i,j}(t)) \rangle \Big) \right| \tag{15}$$

where $\mathrm{sgn}\langle\cdot\rangle$ is the sign function and T is the number of time points. We compute $\mathrm{PLI}_{i,j}$ for all channel pairs and map them to the functional adjacency matrix via $f(\cdot)$:

$$\tilde{A}_P = f(\mathrm{PLI}_{i,j}) \tag{16}$$

State Adaptive Functional Connectivity To capture the inter-channel connectivity under the current temporal state, SAFC randomly initializes a learnable adjacency matrix $A_D \in \mathbb{R}^{N\times N}$ and applies a nonlinear mapping with parameters W_1 and W_2, where $\delta(\cdot)$ and $\sigma(\cdot)$ denote tanh and ELU, respectively. During training, all connection weights in A_D are iteratively updated via backpropagation, enabling adaptive modeling of dynamic inter-channel relationships. The update is formulated as follows:

$$\tilde{A}_D = \mathrm{ReLU}\Big(\sigma\big(W_2\, \delta(W_1 A_D)\big)\Big) \tag{17}$$

Finally, we fuse $\tilde{A}_D$ and $\tilde{A}_P$ to form a hybrid adjacency matrix $\tilde{A}_H$, balancing adaptive modulation with the stability and interpretability of prior connectivity:

$$\tilde{A}_H = \tilde{A}_P \otimes \tilde{A}_D \tag{18}$$

Graph Construction and Encoding Module

Given the node features and the constructed brain topology, we represent each electrode as a node and employ Graph Isomorphism Network (GIN) to aggregate spatial information. Let $G = \{V, A, X, E\}$, where $V = \{v_i\}_{i=1}^{N}$ is the node set, A is the adjacency matrix, and $E = \{e_{uv}\}_{u,v \in V}$ denotes edge weights encoding connection strengths. The node feature update at the k-th layer is defined as follows, where $u \in N_v$ and N_v denotes the set of neighboring nodes of node v.

$$H_v^k = \mathrm{MLP}^k\left(H_v^{k-1} + \sum_{u \in N_v} H_u^{k-1}\, e_{uv}\right) \tag{19}$$

Minimum-Redundancy Topology Backbone

To sparsify the topology while preserving key connections, we adopt a minimum spanning tree (MST), which retains all nodes while enforcing connectivity, acyclicity, and minimal total weight. We therefore assign each edge a weight score for selective retention. Specifically, we concatenate the source and target embeddings (H_s^k, H_t^k) to form an edge representation, which is then fed into a two-layer MLP parameterized by (W_3, b_3) and (W_4, b_4), with a ReLU nonlinearity between the two layers. Finally, a Sigmoid function maps the output to $(0, 1)$ to produce a normalized edge score e_{st}.

$$e_{st} = \mathrm{Sigmoid}\Big(\mathrm{ReLU}\big([H_s^k; H_t^k]^\top W_3 + b_3\big) W_4 + b_4\Big) \tag{20}$$

Tri-View Salience Node Pooling

In TVSNP, MVPool is introduced to model node importance from three views and to select a discriminative node subset that aims to preserve salient topology (Ma et al., 2024). Specifically, Eqs. (21)-(23) compute degree centrality score s_i^1, feature-based importance score s_i^2, and PageRank score s_i^3, capturing node criticality from local to global perspectives.

$$s_i^1 = \sigma_{\mathrm{sigmoid}}\big(\alpha \log\big(\deg(v_i) + \varepsilon\big) + \beta\big) \tag{21}$$
$$s_i^2 = \sigma_{\mathrm{sigmoid}}\big(\mathrm{MLP}(H_v^k)\big) \tag{22}$$
$$s_i^3 = \sigma_{\mathrm{sigmoid}}\big(\mathrm{PageRank}(v_i)\big) \tag{23}$$

Here, $\deg(\cdot)$ denotes node degree, α and β are learnable scale and shift parameters, and ε is a constant. The final node importance score is obtained by aggregating the three scores.

$$s = \sigma_{\mathrm{sigmoid}}\left(w^1 s_i^1 + w^2 s_i^2 + w^3 s_i^3\right) \tag{24}$$

where w^1, w^2, and w^3 are learnable weights. We rank nodes by score s and retain the top $\lceil r \cdot n_i^k \rceil$ nodes under pooling ratio r, where n_i^k is the number of nodes in layer k of graph i:

$$\mathrm{idx} = \mathrm{top-rank}(s, \lceil r \cdot n_i^k \rceil) \tag{25}$$
$$\tilde{H}^{k+1} = H^k(\mathrm{idx}, :) \tag{26}$$
$$\tilde{A}^{k+1} = A^k(\mathrm{idx}, \mathrm{idx}) \tag{27}$$

top−rank$(\cdot)$ returns the indices of selected nodes. Here, H^k is the node-feature matrix at layer k, constructed by stacking node feature $\{H_v^k\}$, and A^k denote the adjacency matrix, respectively. Eqs. (25)-(27) extract the subgraph induced by important nodes while preserving their connection strengths.

Prototype-Centralized Representation Module

To improve inter-class separability and stability in cross-subject emotion recognition, PCRM maintains a learnable prototype η_c for each class $c \in \{1, \ldots, C\}$ in the prototype subspace. During training, we update class prototypes with an Exponential Moving Average (EMA). Let $B_c = \{\, i \mid y_i = c \,\}$ denote the index set of class-c samples in the current mini-batch. If class c is observed for the first time, we initialize its prototype with the class-c mini-batch mean feature:

$$\mu_c = \frac{1}{|B_c|} \sum \tilde{H}_v^{k+1} \tag{28}$$

Otherwise, we update the prototype via EMA:

$$\eta_c = \alpha \eta_c + (1 - \alpha)\mu_c \tag{29}$$

where $\alpha \in [0, 1)$ is the EMA momentum factor, used to balance the historical prototype and the current batch statistics.

We perform ℓ_2 normalization for the sample representation and the class prototype, respectively:

$$\hat{h}_i = \frac{\tilde{H}_v^{k+1}}{\big\|\tilde{H}_v^{k+1}\big\|_2}, \qquad \hat{\eta}_c = \frac{\eta_c}{\|\eta_c\|_2} \tag{30}$$

To characterize the geometric relationship between samples and prototypes, we compute the cosine-similarity vector between sample i and all class prototypes:

$$s_i = \big[\hat{h}_i^\top \hat{\eta}_1, \hat{h}_i^\top \hat{\eta}_2, \ldots, \hat{h}_i^\top \hat{\eta}_C\big]^\top \tag{31}$$

Discriminative-Prototype Joint Optimization

We jointly optimize the classification and prototype objectives. Let q_i be the classifier prediction, s_i the prototype-head prediction, and y_i the ground-truth label. The classification loss and prototype constraint loss are defined as follows.

$$L_{\mathrm{cls}} = \mathrm{CE}(q_i, y_i), \qquad L_{\mathrm{proto}} = \mathrm{CE}(s_i, y_i) \tag{32}$$

where $\mathrm{CE}(\cdot)$ denotes the cross-entropy loss. Their weighted sum yields the final objective:

$$L = L_{\mathrm{cls}} + \lambda_{\mathrm{proto}} L_{\mathrm{proto}} \tag{33}$$

The hyperparameter $\lambda_{\mathrm{proto}} > 0$ controls the strength of the prototype-based supervision.

Results and Discussion

We evaluate PEPBNet on three public datasets: SEED (Zheng & Lu, 2015), SEED-IV (Zheng et al., 2018), and DEAP (Koelstra et al., 2011). SEED includes EEG recordings from 15 subjects watching 15 film clips over 3 sessions with 3 emotion labels (positive, neutral, negative). SEED-IV expands to four classes (happy, sad, neutral, fear) with 15 subjects viewing 24 clips across 3 sessions. DEAP contains EEG recordings from 32 subjects watching 40 one-minute music videos, with valence and arousal ratings on a 1-9 scale; we binarize both dimensions using 5 as the threshold. For all datasets, we use leave-one-subject-out (LOSO) cross-validation, testing on one held-out subject and training on the remaining subjects, repeated for each subject.

Table 1: Accuracy (%) comparison on SEED and SEED-IV (mean ± std).

Model	SEED	SEED-IV
PGCN	84.59±8.68	73.69±7.16
BGAGCN-MT	88.50±4.96	75.78±8.17
EmotionCLIP-32	88.69±4.82	73.50±9.73
STCBI-Nets	90.21±6.32	75.17±8.82
HEDN	94.00±7.05	79.36±9.62
Ours	**96.30±4.11**	**87.04±4.61**

Table 2: Accuracy (%) comparison on DEAP (mean ± std).

Model	Valence	Arousal
CAMLAF-DANN	52.28±12.64	58.21±13.67
DSSN	59.84±13.42	59.68±7.23
UDA-DDA	61.44±7.15	62.86±10.58
MACTN	66.14±6.11	67.83±8.07
MSGDAN	68.25±4.37	69.54±4.67
Ours	**71.88±5.96**	**71.80±5.55**

Comparative Experiments

As shown in Table 1, we compare PEPBNet with several representative methods (PGCN (Jin et al., 2024), BGAGCN-MT (H. Yan et al., 2024), EmotionCLIP-32 (R. Yan et al., 2025), STCBI-Nets (An et al., 2025), and HEDN (Q. Wang & Yang, 2025)) on the SEED and SEED-IV datasets. PEPBNet achieves the best results on both datasets, reaching 96.30% ± 4.11% on SEED and 87.04%±4.61% on the more challenging four-class SEED-IV task, indicating robust performance under finer-grained categories and more complex distributions.

For the DEAP dataset (Table 2), we compare PEPBNet with CAMLAF-DANN (Sun et al., 2025), DSSN (Li et al., 2024), UDA-DDA (J. Tang et al., 2025), MACTN (Si et al., 2024), and MSGDAN (J. Wang et al., 2024) on both the Valence and Arousal dimensions. PEPBNet consistently performs best, demonstrating strong emotion discriminative modeling and stable cross-subject generalization across emotion dimensions. Overall, PEPBNet delivers consistently strong performance and stability compared with existing methods.

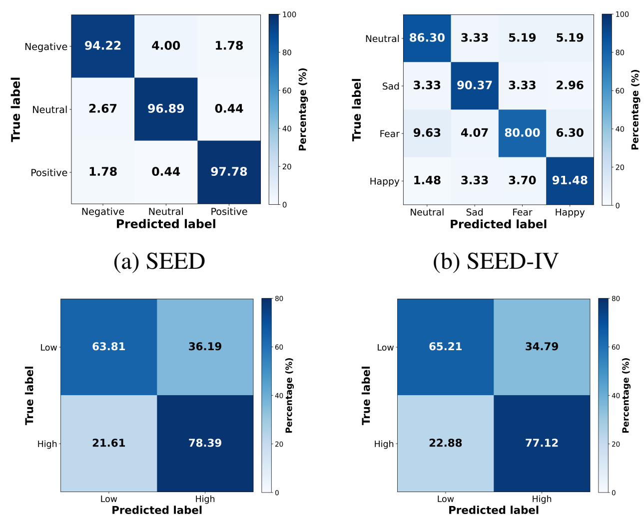
Ablation Study

To assess the contribution of each component, we conduct ablation studies (Table 3). Removing MGPER consistently degrades performance, confirming the importance of modeling periodic EEG evolution. Excluding EIFDM also hurts accuracy, indicating that disentangling emotion-related and interference factors reduces the impact of task-irrelevant components.

For topology learning, removing either SPFC or SAFC lowers performance: SPFC provides stable priors, whereas SAFC enables adaptive, data-driven connectivity updates, and the two are complementary. Dropping MRTB further reduces accuracy, suggesting improved robustness by suppressing redundant connections. Disabling TVSNP similarly degrades results, showing that emphasizing high-contribution nodes benefits representation learning. Removing PCRM also leads to a performance drop, validating its role in promoting intra-class compactness and inter-class separability. Overall, the complete model performs best across all datasets, demonstrating the effectiveness and robustness of the PEPBNet.

Table 3: Ablation study of PEPBNet on accuracy (%) (‘w/o’ denotes ‘without’).

Method	SEED	SEED-IV	DEAP	
			Valence	Arousal
w/o MGPER	93.48	85.18	65.86	66.09
w/o EIFDM	95.26	86.11	68.28	67.81
w/o SPFC	94.96	85.09	67.66	69.06
w/o SAFC	95.11	86.02	69.45	68.59
w/o MRTB	94.37	85.83	69.61	69.38
w/o TVSNP	95.70	86.67	69.53	68.83
w/o PCRM	95.71	86.67	70.47	70.31
Ours	**96.30**	**87.04**	**71.88**	**71.80**



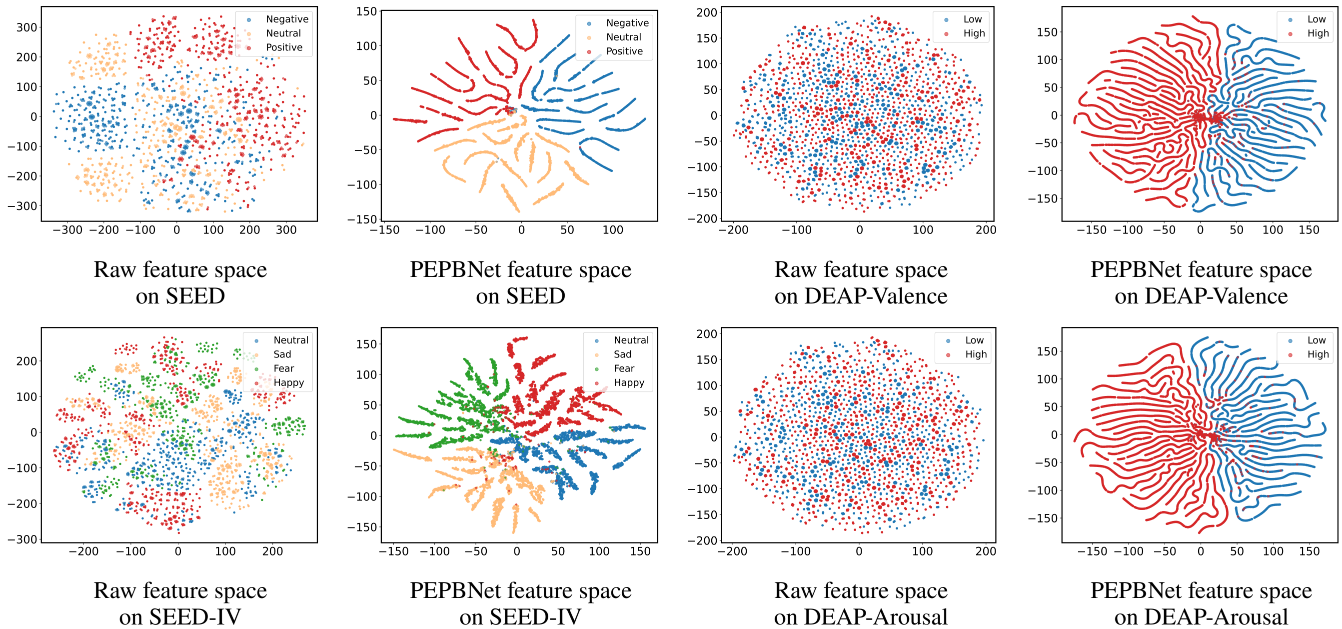
(a) SEED (b) SEED-IV

(c) DEAP-Valence (d) DEAP-Arousal

Figure 2: Confusion matrices of PEPBNet.

Confusion Matrices

As shown in Figure 2, all confusion matrices exhibit strong diagonal dominance, indicating that PEPBNet learns stable EEG representations across emotion classes. On SEED, the accuracies for negative, neutral, and positive reach 94.22%, 96.89%, and 97.78%, respectively, with most errors arising from confusion between negative and neutral. On the finer-grained SEED-IV, sad and happy are best recognized (90.37%



Raw feature space on SEED
PEPBNet feature space on SEED
Raw feature space on DEAP-Valence
PEPBNet feature space on DEAP-Valence
Raw feature space on SEED-IV
PEPBNet feature space on SEED-IV
Raw feature space on DEAP-Arousal
PEPBNet feature space on DEAP-Arousal

Figure 3: t-SNE visualization of the raw and PEPBNet feature spaces.

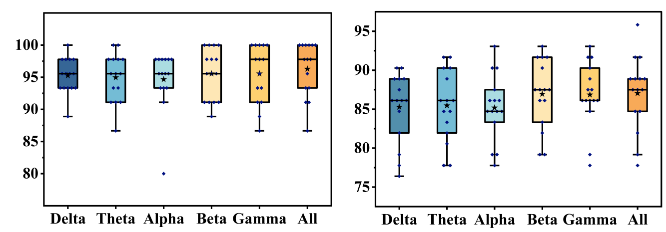
and 91.48%), whereas fear and neutral are more frequently misclassified. Given the consistent patterns on SEED and SEED-IV, we infer higher representational similarity between neutral and negative emotions. On DEAP, for both arousal and valence, high-level is recognized more reliably than low-level, implying stronger neurophysiological separability for high-level affective states.

Visualization

We assess feature discriminability via t-SNE visualization (Figure 3). Compared with raw features that exhibit substantial overlap and diffuse structure, PEPBNet produces compact, well-separated clusters with clearer class boundaries. This behavior aligns with our prototype-centered objective, which pulls samples toward their class prototypes and pushes them away from others, improving intra-class compactness and inter-class separability. Overall, Figure 3 corroborates that PEPBNet consistently enhances emotion feature separability across settings and class granularities, producing more stable and structured clustering.

Frequency-Band Analysis

As shown in Figure 4, the single-band results reveal a clear spectral disparity: Gamma and Beta achieve higher accuracy, consistent with evidence that high-frequency oscillations are closely linked to emotion-related cognition. Beta likely reflects enhanced emotional appraisal and cognitive processing during affective video viewing, providing richer discriminative cues. Gamma performs best, suggesting it carries the most discriminative emotion-related information. In contrast, Alpha is more susceptible to subject-specific factors (e.g., attention and fatigue), which may reduce its discriminability under continuous stimulation when used alone. Notably, using the full-band signal yields the best overall performance, indicating that integrating cross-band complementary information is beneficial.



(a) SEED (b) SEED-IV

Figure 4: Accuracy (%) of PEPBNet across different frequency bands on the SEED and SEED-IV datasets (the star marker indicates the mean accuracy).

Conclusion

In this paper, we propose PEPBNet, a model that characterizes the neural encoding of affective states to enable more reliable cross-subject EEG emotion recognition. PEPBNet extracts periodic evolution features, explicitly disentangles interference factors from emotion representations, and enhances intra-class compactness and inter-class separability through PCRM and DPJO. Extensive experiments demonstrate that the proposed model achieves strong generalization and robustness. Overall, PEPBNet provides an interpretable and generalizable framework for cross-subject EEG emotion recognition, supporting more reliable affective computing and human-computer interaction.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants [62206196] and [62303445], in part by the Natural Science Foundation of Shanxi Province under Grant [202303021221001], in part by the Natural Science Foundation for Young Scientists of Shanxi Province under Grants [202403021222057] and [202203021212216].

References

An, Y., Hu, S., Liu, S., Gu, Z., Zhang, Y., & Zhang, Y. (2025). Spatiotemporal isomorphic cross-brain region interaction network for cross-subject eeg emotion recognition. *Knowledge-Based Systems*, 114115.

Cao, L., Zhao, W., & Sun, B. (2025). Emotion recognition using multi-scale eeg features through graph convolutional attention network. *Neural Networks*, *184*, 107060.

Ding, S., Wang, K., Jiang, W., Xu, C., Bo, H., Ma, L., & Li, H. (2025). Dgat: A dynamic graph attention neural network framework for eeg emotion recognition. *Frontiers in Psychiatry*, *16*, 1633860.

Dong, Y., Jing, C., Mahmud, M., Ng, M. K.-P., & Wang, S. (2024). Enhancing cross-subject emotion recognition precision through unimodal eeg: A novel emotion preceptor model. *Brain Informatics*, *11*(1), 31.

Erat, K., Şahin, E. B., Doğan, F., Merdanoğlu, N., Akcakaya, A., & Durdu, P. O. (2024). Emotion recognition with eeg-based brain-computer interfaces: A systematic literature review. *Multimedia tools and applications*, *83*(33), 79647–79694.

Feng, L., Huo, Y., Kong, Y., & Cheng, C. (2026). A dual-branch multiscale model based on bi-mamba for eeg emotion recognition. *Biomedical Signal Processing and Control*, *117*, 109623.

Garg, D., Verma, G. K., & Singh, A. K. (2024). Eeg-based emotion recognition using mobilenet recurrent neural network with time-frequency features. *Applied Soft Computing*, *154*, 111338.

Jin, M., Du, C., He, H., Cai, T., & Li, J. (2024). Pgcn: Pyramidal graph convolutional network for eeg emotion recognition. *IEEE Transactions on Multimedia*, *26*, 9070–9082.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18–31.

Li, W., Dong, J., Liu, S., Fan, L., & Wang, S. (2024). Dynamic stream selection network for subject-independent eeg-based emotion recognition. *IEEE Sensors Journal*, *24*(12), 19336–19343.

Ma, X., Fu, X., Wang, T., Zhuo, L., & Zou, Q. (2024). Graphadt: Empowering interpretable predictions of acute dermal toxicity with multi-view graph pooling and structure remapping. *Bioinformatics*, *40*(7), btae438.

Meng, W., Hou, F., Zhao, M., Kong, J., Wu, J., Zuo, J., & Liu, Q. (2025). Emotion recognition via affective eeg signals: State of the art. *Neurocomputing*, 130418.

Mishra, A., & Ranjan, P. (2026). Comprehensive analysis of the state of art on emotion recognition using eeg. *Computers and Electrical Engineering*, *131*, 110958.

Mohammadi Foumani, N., Mackellar, G., Ghane, S., Irtza, S., Nguyen, N., & Salehi, M. (2024). Eeg2rep: Enhancing self-supervised eeg representation through informative masked inputs. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5544–5555.

Pan, D., Zheng, H., Xu, F., Ouyang, Y., Jia, Z., Wang, C., & Zeng, H. (2023). Msfr-gcn: A multi-scale feature reconstruction graph convolutional network for eeg emotion and cognition recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 3245–3254.

Si, X., Huang, D., Liang, Z., Sun, Y., Huang, H., Liu, Q., Yang, Z., & Ming, D. (2024). Temporal aware mixed attention-based convolution and transformer network for cross-subject eeg emotion recognition. *Computers in Biology and Medicine*, *181*, 108973.

Sun, W., Su, Y., Zhang, Y., & Wu, X. (2025). Eeg emotion recognition through a domain-adversarial multi-feature fusion network. *Expert Systems with Applications*, 129694.

Tang, J., Li, Y., Su, C.-W., Fan, X., Zheng, Y., Wang, H., Naeem, H., Fang, P., Wang, J., Yao, N., et al. (2025). Udadda: Unsupervised domain adaptation with dynamic distribution alignment network for emotion recognition using eeg signals. *Neurocomputing*, 131715.

Tang, W., Fan, L., Lin, X., & Gu, Y. (2025). Eeg emotion recognition based on efficient-capsule network with convolutional attention. *Biomedical Signal Processing and Control*, *103*, 107473.

Wang, J., Ning, X., Xu, W., Li, Y., Jia, Z., & Lin, Y. (2024). Multi-source selective graph domain adaptation network for cross-subject eeg emotion recognition. *Neural Networks*, *180*, 106742.

Wang, Q., & Yang, L. (2025). Hedn: A hard-easy dual network with task difficulty assessment for eeg emotion recognition. *arXiv e-prints*, arXiv–2511.

Wang, T., Huang, X., Xiao, Z., Cai, W., & Tai, Y. (2024). Eeg emotion recognition based on differential entropy feature matrix through 2d-cnn-lstm network. *EURASIP Journal on Advances in Signal Processing*, *2024*(1), 49.

Yan, H., Guo, K., Xing, X., & Xu, X. (2024). Bridge graph attention based graph convolution network with multi-scale transformer for eeg emotion recognition. *IEEE Transactions on Affective Computing*, *15*(4), 2042–2054.

Yan, R., Li, Y., Ding, H., & Wang, F. (2025). Cross-domain eeg-based emotion recognition with contrastive learning. *arXiv preprint arXiv:2511.05293*.

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

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