FG-MSTGNN: Cross-subject EEG Emotion Recognition via Frequency-guided Multi-period Spatial-temporal Graph Neural Network

Chenchen Zhang 2022511314@link.tyut.edu.cn
Yanrong Hao* haoyanrong@tyut.edu.cn
Xin Wen xwen@tyut.edu.cn
Mengni Zhou zhoumengni@tyut.edu.cn
Fei Yuan 2023511251@link.tyut.edu.cn
Rui Cao* caorui@tyut.edu.cn
School of Software, Taiyuan University of Technology, Taiyuan, China

Editors: Hung-yi Lee and Tongliang Liu

Abstract

Accurate decoding of emotional EEG signals constitutes a critical challenge for developing affective brain-computer interfaces. Contemporary methods for cross-subject EEG-based emotion recognition confront two critical challenges: 1) inadequate investigation of the distinct affective features of the EEG rhythm; 2) insufficient capability to extract the various neurophysiological connectivity patterns across subjects in the same experimental setting. To address these limitations, we propose FG-MSTGNN, a dual-stage adaptive learning framework comprising the Frequency-guided Multi-period Spatial-temporal Graph Neural Network. The Feature Learning Stage utilizes a Multi-period Time-Frequency Cooperative Encoder Module to hierarchically extract cross-frequency rhythmic dynamics. The Topology Optimization Stage utilizes a Dual-Phase Graph Pooling Module to dynamically generate personalized sparse neurophysiological connectivity patterns. Systematic evaluation under cross-subject experiments demonstrates the framework achieves average classification accuracies of 94.67% and 85.28% on SEED and SEED-IV respectively, showing statistically distinctive improvements over state-of-the-art EEG emotion recognition methods. The proposed framework reveals that both functional brain network topology and EEG spectral dynamics varies from different emotional states.

Keywords: EEG; Emotion recognition; Graph Neural Network; Multi-period rhythmic dynamics

1. Introduction

Emotion recognition, a pivotal research domain in neuroscience, psychology, and brain-computer interfaces (BCIs), holds substantial clinical and technological value across mental health diagnostics, affective computing systems, and personalized human-computer interaction paradigms. Current methods in emotion recognition are broadly classified into two categories: non-physiological signal based techniques and physiological signal based methods. Among all physiological modalities, electroencephalography (EEG) has emerged as a principal investigative tool for emotion computation and neurologic disease research, owing to its millisecond-level temporal resolution and low susceptibility to artifacts during emotional state characterization Geng et al. (2024).

© 2025 C. Zhang, Y. Hao*, X. Wen, M. Zhou, F. Yuan & R. Cao*.

Electroencephalogram (EEG)-based emotion recognition, which captures direct electrophysiological responses from the central nervous system, has become a vital approach for decoding emotional and cognitive processes. Early studies largely relied on handcrafted features, yet these often proved insufficient in capturing the non-stationary nature and cross-subject variability of EEG signals. Recent research has focused on two key dimensions: emotion-relevant feature extraction and optimized deep learning architectures. The former involves time-frequency analysis, nonlinear dynamics, or brain network topology to identify emotion-related patterns, while the latter leverages deep neural networks to automatically learn temporal, spectral, and spatial representations from EEG. A growing trend is toward feature embedding, model optimization, and multi-domain fusion of temporal, frequency, and spatial information, aiming to enhance the practicality and robustness of EEG-based emotion recognition in real-world applications.

To capture neural oscillatory patterns with temporal stability for EEG emotion recognition, researchers have proposed a variety of deep learning methods. Temporal Convolutional Network (TCN) Yang et al. (2023) captures long-range dependencies through expansive convolution. Recurrent Neural Network (RNN) and its improvement methods Tao et al. (2023), use gating mechanism to model sequence dynamics, and TSception Ding et al. (2023) combines multi-scale convolution to improve temporal-frequency feature extraction. Although the methods above can effectively model local temporal relations, they still lack the ability to extract individual cross-period rhythmic features by combining frequency domain information, which is slightly insufficient in exploring the physiological significance.

Although traditional time-series models can extract high temporal resolution features from EEG signals, it is difficult to effectively characterize the non-Euclidean topology. In this regard, Graph Neural Network (GNN) shows unique advantages.

From the perspective of cognitive neuroscience, cognitive processing in the brain involves dynamic information interaction and coordination across different brain regions, with functional connectivity between these regions changing in a task-dependent manner Li et al. (2024). Graph Neural Networks (GNNs) automatically learn node features through neighborhood aggregation mechanisms, yet their classification performance is highly dependent on the quality of the graph topology. While fully-connected graphs may introduce redundant or noisy edges—increasing the risk of overfitting—random pruning can disrupt the physiological meaningfulness of functional connections. To address this, graph pooling methods have been incorporated into GNNs with the goal of providing optimized topologies that align with physiological constraints, thereby improving both model performance and computational efficiency in handling complex network data. However, conventional graph pooling operations often fail to adequately adapt to emotion-related brain networks derived from EEG, which limits their applicability and effectiveness in emotion recognition tasks.

In the current study, we propose a Frequency-guided Multi-period Spatial-Temporal Graph Neural Network (FG-MSTGNN) framework for cross-subject analysis, comprising two stages: 1) Feature Learning Stage. This stage is implemented through the Multi-period Time-Frequency Cooperative Encoder (MTFCE), which transforms 1D EEG signals into 2D temporal-frequency representations, while simultaneously capturing both intra-rhythmic fine-grained patterns reflecting localized spectral dynamics and inter-rhythmic interactions that characterize cross-period coupling relationships. 2) Structure optimization phase. This phase is implemented by the Dual-Phase Graph Pooling Module (DPGPM), which optimizes

the initial graph using a dual graph pooling operation to generate a sparse EEG topology, thereby reducing the noisy computational complexity while preserving neurophysiological plausibility. Finally, the optimized sparse graph is fed into GIN encoder model for non-linear feature aggregation and emotion classification.

Contributions of this work include:

- Proposing FG-MSTGNN, a dual-stage adaptive learning framework that integrates graph-theory principles with multi-period rhythmic dynamics to achieve robust cross-subject EEG emotion recognition.
- Proposing two core components. The MTFCE extracts both intra-rhythmic fine-grained patterns and inter-rhythmic interactions, while the DPGPM eliminates redundant neuro-physiological connections.
- Cross-validation experiments on SEED and SEED-IV demonstrate that the proposed framework achieves average accuracies of 94.67% and 85.28% respectively in cross-subject emotion recognition, significantly outperforming existing state-of-the-art (SOTA) methods.

2. Related work

2.1. Temporal Feature Extraction

In scientific practice, many researchers focus on the trend of EEG signals over time, mainly due to the fact that dynamic time series can not only reveal the characteristics of signal cycle fluctuation and continuous change, but also effectively represent the potential pattern of change, which provides multi-dimensional data features for the construction of accurate emotion recognition models.

Deep learning solves the bottleneck of traditional methods that rely on handcrafted features by providing an end-to-end feature learning capability that extracts high-dimensional, abstract feature representations from raw data through multilayered nonlinear transformations. Liang et al. (2021) proposed EEGFuseNet, which uses unsupervised training of a convolutional recurrent hybrid generative adversarial network with an encoder-decoder structure to automatically extract spatial-temporal dynamic features from EEG signals, improving classification accuracy by 12% over traditional methods in emotion recognition tasks. In order to better capture the temporal dependency, Zhang et al. (2024b) combined local GCN and global BiGRU to design the time-aware TAS-Net, which effectively solved the feature redundancy problem.

2.2. Spatial Features Extraction

For the construction of brain topology, previous studies have combined it with convolution for the extraction of correlation information within channels and between short-range electrodes. Although 1D convolution can capture some shorter time series features through dimensional transformations, its fixed receptive fields are difficult to capture the long-range brain connectivity, which may ignore global spatial relationships, and destroy the spatial correlation of EEG signals.

In contrast, GNN, due to its superior graph embedding ability to explicitly define electrode nodes and functionally connected edges, better preserve the deep topology Klepl

et al. (2024). The static graph convolution proposed by Wang et al. (2019) and the dynamic graph convolution (DGCNN) proposed by Song et al. (2020) achieve classification accuracies of 78.23% and 81.91% on SEED, confirming the advantage of dynamic adjacency matrices in characterizing the time-varying properties of emotion-related brain networks. Qiu et al. (2023) apply the idea of residual connectivity to the underlying GCN structure using a two-layer multi-head residual graph convolution network (MRGCN). Nevertheless, fully connected graphs lead to the problem of excessive smoothing of node features, and real brain networks follow small-worldness, while fully connected graphs are inconsistent with the spatial distribution pattern of cortical white matter fibre tracts Bullmore and Sporns (2009). To address this problem, GLFANet, a global-to-local feature aggregation network developed by Liu et al. (2023), nnovatively introduces global topological constraints so that the fully connected graph avoids the problem of excessive smoothing while preserving interactions between different brain regions.

In order to better explore the connection between emotional cognitive mechanisms and EEG signals, researchers have begun to integrate GNN with brain science theories and developed a series of graph models based on physiological prior knowledge Guo et al. (2024). To explore the long-range dependence of the cerebral cortex. Jin et al. (2024) proposed PGCN, a model that implements feature aggregation at three levels: local, mesoscope, and global, with good robustness in EEG-based emotion recognition. Li et al. (2025) used a cognitively-inspired graph-learning neural network model, BF-GCN, that contains three graph branches, which combines data-driven and cognition- inspired strategies in order to automatically learn emotional cognitive graph patterns from emotional EEG signals, achieving 92.72% on SEED and 82.03% on SEED-IV results in subject-independent experiments.

3. Methodology

3.1. Bandpass filter bank data processing layer

In order to optimize the feature extraction process and more accurately capture the EEG activity features associated with changes in emotional states, we extracts the α, β, γ, δ, and θ bands of the original EEG signal x using a band-pass filter bank based on the characteristics of the EEG signals and performs a sliding window on each band in order to extract the DE features, which defined as:

$$DE(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \log_2\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-t)^2}{2\sigma^2}}\right) = \frac{1}{2}\log_2(2\pi e) + \log_2(\sigma) \quad (1)$$

where $x \sim \mathcal{N}(\mu, \sigma^2)$ and e and π are constants. Assume that the DE feature of the computed EEG signals are $X_D \in \mathbb{R}^{C\times N\times L}$. Feature extraction and processing are performed separately for each frequency band of the data $X_{1D} = \{x_{1D}^1, x_{1D}^2, \ldots, x_{1D}^n\}$ (e.g. the SEED dataset represents having the original data divided into 5 frequency bands), where $x_{1D}^n \in \mathbb{R}^{C\times L}$ denotes the feature in the nth band, C denotes the number of EEG electrode channels, N denotes the number of bands, and L denotes the dimension of the feature. Finally, the DE features are fed into the FG-MSTGNN model, band by band.

3.2. Frequency-guided Multi-period Spatial-Temporal Graph Neural Network

3.2.1. Multi-period Time-Frequency Cooperative Encoder Module

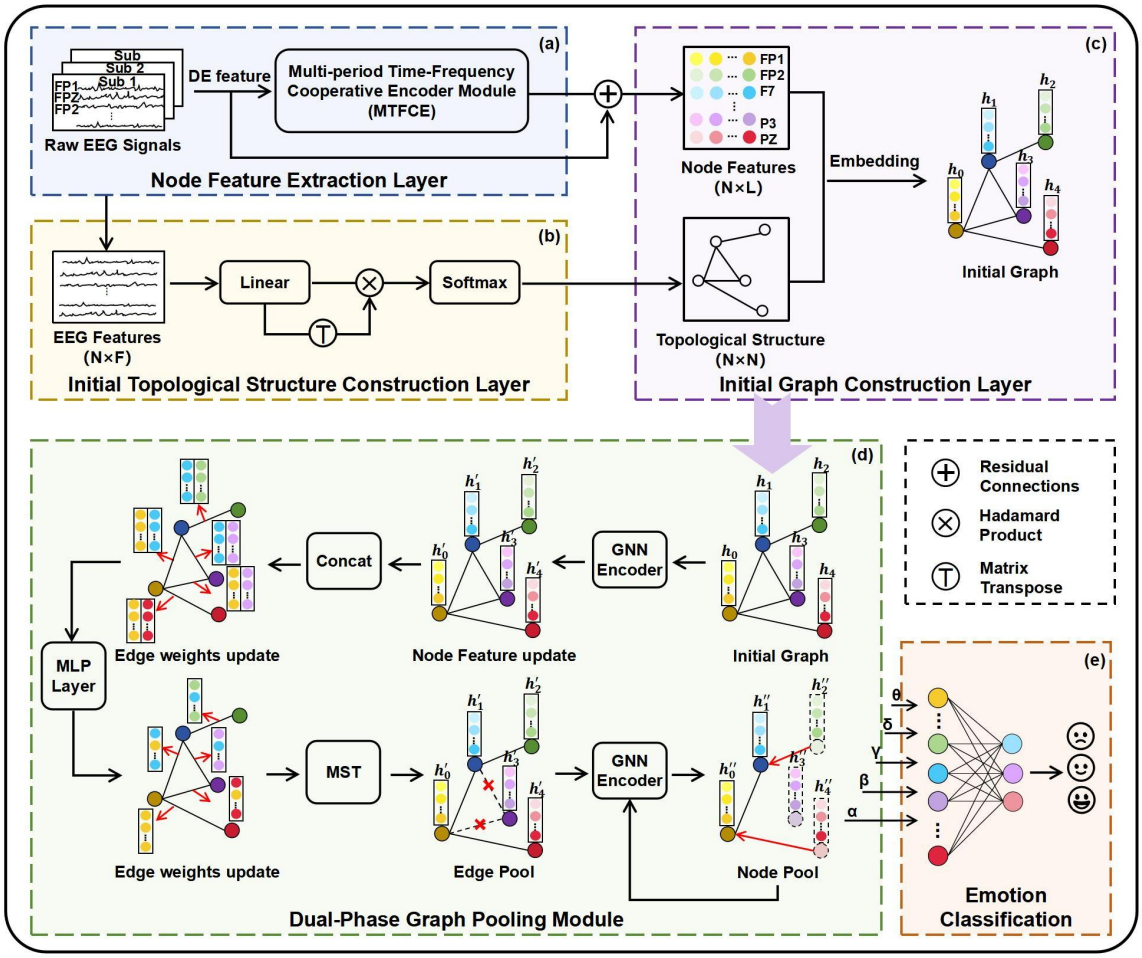

Figure 1: The proposed Frequency-guided Multi-period Spatial-Temporal Graph Neural Network.

1) Data Transformation layer

EEG frequency features Response rhythm dynamic characteristics. This layer uses the Fast Fourier Transform (FFT) to transform the input signal X_{1D} in the frequency domain and to find the variation between periods, revealing the strength of its frequency components. Specifically, a 1D FFT operation is performed for each time step:

$$X_f = FFT(X_{1D}) \tag{2}$$

Next, the mean value of the signal amplitude at each frequency in the frequency domain is calculated channel by channel:

$$A_f = Avg(Amp(X_f)) \tag{3}$$

In Eq.(3), $Amp(\bullet)$ is used to calculate the amplitude values for each frequency X_f , and is averaged using $Avg(\bullet)$ to obtain an amplitude average A_f for each frequency. In particular, the DC component is set to zero.

Based on the calculated mean frequency values, the frequency indexes of the top-k most significant values were selected.

$$f_i = \arg Topk(A_f) \tag{4}$$

where f_i is the set of frequency corresponding to the first k largest values. Assuming that each selected frequency is $\{f_1, f_2, \ldots, f_k\}$ where $i \in \{1, \ldots, k\}$, the period p_i ($i \in \{1, \ldots, k\}$) of each frequency can be calculated by the following equation:

$$p_i = \left\lceil \frac{L}{f_i} \right\rceil \tag{5}$$

where p_i is the period corresponding to the frequency f_i, indicates the repetition pattern of the signal in the time domain, and L is the length of the time series.

According to the conjugate nature of the frequency domain, only the frequencies within $\left\{1, \ldots, \frac{L}{2}\right\}$ are considered in the calculation process, so the X_{1D} time series can be reconstructed into multiple X_{2D} tensors as shown in Eq.(6):

$$X_{2D}^i = Reshape_{p_i, f_i}(Padding(X_{1D})), \quad i \in \{1, \ldots, k\} \tag{6}$$

For the period corresponding to each of the most important frequency components, zero-padding is performed along the time dimension by the $Padding(\bullet)$ operation so that its length is a multiple of the period, facilitating transformation in 2D space, and reshape the filled data into 2D form using the $Reshape(\bullet)$ operation to obtain X_{2D}^i. In which, p_i, f_i is the period of each frequency component.

2) Dynamic multi-scale feature extraction layer

The dynamic multi-scale feature extraction layer uses the multi-scale 1D convolution to extract features from the 2D tensor X_{2D}^i and learn its rich temporal information.

$$\tilde{X}_{2D}^i = AvgPool(LeakyReLU(Conv1D(X_{2D}^i, (1, t)))), \quad i \in \{1, \ldots, k\} \tag{7}$$

As in Eq.(7), dynamic features are obtained by 1D convolution kernels of different sizes to the samples one by one. Where X_{2D}^i is the input sample, and $Conv1D(\bullet)$ is the 1D convolution operation with a convolution kernel of $(1, t)$ and the step size of $(1, 1)$ applied to the input sample. Secondly, an activation function $LeakyReLU(\bullet)$ is used in the convolution operation and the feature map is downsampled by an average pooling function $AvgPool(\bullet)$.

Finally, the learned 2D tensor $\tilde{X}_{2D}^i$ is transformed back into the 1D space $\tilde{X}_{1D}^i$ for dimensional reduction and reshaping, which is defined as:

$$\tilde{X}_{1D}^i = Trunc\left(Reshape_{1,(p_i \times f_i)}\left(\tilde{X}_{2D}^i\right)\right), \quad i \in \{1, \ldots, k\} \tag{8}$$

In this case, the length is truncated to the original length using the $Trunc(\bullet)$ operation. During the process of data conversion, the shape of the 2D tensor obtained will be different due to the different periods.

3) Adaptive feature aggregation layer

This layer takes the output of the feature extraction layer, fuses k 1D representations $\left\{\tilde{X}_{1D}^{1}, \ldots, \tilde{X}_{1D}^{k}\right\}$ .The amplitudes A_f computed in the data transformation layer reflect the selected significance frequencies and the relative importance of the corresponding cycles.

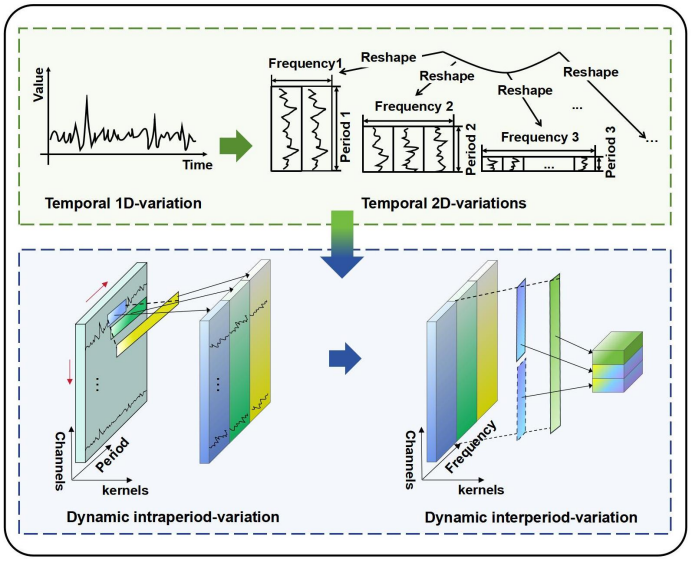

Figure 2: The proposed Multi-period Time-Frequency Cooperative Encoder Module.

Therefore, the $Softmax(\bullet)$ function is applied to each frequency, followed by multiplying it by the corresponding $\overline{x}_{1D}^{i}$ and summing it. This is defined as:

$$\tilde{X}_{1D} = \sum_{i=1}^{k} Softmax(A_f) \times \tilde{X}_{1D}^{i} \tag{9}$$

In the equation above, period weights are applied to the output of each period. The outputs of all periods are then adaptively aggregated.

3.2.2. Dual-Phase Graph Pooling Module

1) Initial graph topology construction layer

Each electrode of the input signal is considered as a single node in the constructed brain network. The dynamic time-frequency representation of the learned individual electrodes is considered as a node attribute. The global connectivity of the initial graph is defined by the relationships between electrode nodes. Neuroscience studies have shown that activation of a specific brain region tends to activate bundles of neurons within that region to carry out higher cognitive processes.

We reflect the relationship between electrode nodes by calculating the dot product between the features of electrode nodes within each frequency band. It is worth noting that the similarity adjacency matrix is dynamic and instance-specific. Assume that the adjacency

matrix $A_{\text{initial}} \in \mathbb{R}^{C\times C}$ of the Initial-graph as:

$$A_{\text{initial}} = \begin{bmatrix} h_1 \cdot h_1 & \cdots & h_1 \cdot h_C \\ \vdots & \ddots & \vdots \\ h_C \cdot h_1 & \cdots & h_C \cdot h_C \end{bmatrix} \tag{10}$$

where $(\bullet)$ is the dot product, and h_i, where $i \in \{1, 2, \ldots, C\}$ is the vector generated for each node after the MTFCE.

2) Initial graph edge embedding construction layer

The node embeddings are first updated using the message-passing process of Graph Isomorphism Network (GIN), refer to Eq.(13).

Each node feature h_i in the graph is updated based on the information of its neighbour nodes N_v and itself. Assuming that the embeddings of the source and target nodes are obtained as h_s and h_i respectively, the embedding e_{st} is obtained by cascading the embeddings of the source and target nodes:

$$e_{st} = g(Concat(h_s, h_t)) \tag{11}$$

i.e., the learned node features $\tilde{x}_v = h_v$ are concatenated to create embedded edge features $\tilde{x}_{st} = [\tilde{x}_s; \tilde{x}_t] \in \mathbb{R}^{2d}$. Here the aggregation function $g(\cdot)$ uses a multi-layer perceptron MLP, which consists of a linear layer with trainable weights $W_3 \in \mathbb{R}^{2d\times 2d}$, $b_3 \in \mathbb{R}^{1\times 2d}$, a ReLU activation function σ_{ReLU}, and a weight $W_4 \in \mathbb{R}^{2d\times 1}, b_4 \in \mathbb{R}$, and converts the result to the range (0,1) using a Sigmoid function:

$$e_{st} = \sigma_{\text{Sigmoid}}\left(\sigma_{\text{ReLU}}\left(\tilde{x}_{st}^{\top} W_3 + b_3\right) W_4 + b_4\right) \tag{12}$$

3.2.3. GIN Encoder Module

For advanced graphs $T = \{V, A_T, \tilde{X}_{1D}, E_T\}$, an effective graph representation is obtained by learning the complex relationships between individual electrode nodes and aggregating their spatial features by means of GIN. The message passing process can be referred to the following equation:

$$h_v^{(k)} = MLP^{(k)}\left(h_v^{(k-1)} + \sum_{u\in\mathcal{N}_v} h_u^{(k-1)} e_{uv}\right) \tag{13}$$

Thus, for the k-th layer in GIN, $h_v^{(k)}$ corresponding to a node v in the k-th layer of the graph is updated based on the information of its neighbouring nodes $u \in \mathcal{N}_v$ and itself, resulting in a more globally informative representation of the node, $h_v^{(0)} = \tilde{x}_v$.

3.2.4. Dual-Phase Graph Pooling Module

We proposed a dual graph pooling method aimed at optimizing the representation and processing of EEG brain network graphs. The method consists of two steps, edge pooling and node pooling, and improves the performance of emotion recognition tasks by effectively simplifying the network structure and retaining key information.

1) Edge pruning graph pooling layer

Let $G = (V, A_{\text{initial}}, \tilde{X}_{1D}, E)$ denotes the graph constructed based on EEG data, where $V = \{v_i : v = 1, 2, \ldots, C\}$ denotes the set of nodes. $A_{\text{initial}} = [a_{st} : s, t \in V] \in \{0, 1\}^{C \times C}$ is the adjacency matrix describing their connectivity information, and $a_{st} = 1$ indicates the existence of edges between the nodes, otherwise there is no connection. $\tilde{X}_{1D} = \{\tilde{x}_v : v \in V\}$ is the set of node attributes denoting the attributes corresponding to each node v_i. $E = \{e_{st} : s, t \in V\} \in \mathbb{R}^{C \times C}$ is the set of edge weights denoting the strength of node-to-node connectivity, and C denotes the number of EEG electrode channels.

The edges in the graph are pruned using either the Pulliam or Kruskal algorithms to construct a minimum spanning tree(MST) and to implement edge graph pooling operations. The process reduces the complexity of the graph by optimising the choice of edges while preserving important topologies.

To compute the MST for the graph G, the goal is to find a subgraph that contains all nodes in such a way that the subgraph is connected and the sum of the weights of the edges is minimum. Thus, to find a subgraph $T = \{V, A_T, \tilde{X}_{1D}, E_T\}$ that is compatible with the following conditions: 1) connectivity, where all nodes in the graph T are connected, 2) acyclicity, where the graph T does not contain any loops, and 3) minimum weight, the sum of weights of edges in the graph T is minimum. In this paper, the minimum spanning tree constructed contains 62 nodes and 61 edges.

2) Node clustering graph pooling layer

Specifically, an importance score is calculated for each node, for node v_i, the importance score s_i can be expressed by the following equation:

$$s_i^1 = \sigma_{\text{Sigmoid}} \left(\alpha \log(\deg(v_i) + \varepsilon) + \beta\right) \tag{14}$$

$$s_i^2 = \sigma_{\text{Sigmoid}} \left(\text{MLP}(H^{(k)})\right) \tag{15}$$

$$s_i^3 = \sigma_{\text{Sigmoid}} \left(\text{Pagerank}(v_i)\right) \tag{16}$$

Eq.(14) calculates the node degree centrality score, Eq.(15) calculates the node feature importance score, and Eq.(16) calculates the node's PageRank score.

Particularly, $\deg(v_i)$ denotes the degree centrality of the node v_i. $\alpha, \beta, \varepsilon$ is the constant, and $H^{(k)}$ is the node representation matrix of the k-th layer.

These scores are combined into a vector s to get the scores of all nodes, w^1, w^2, w^3 are trainable weights:

$$s = \sigma_{\text{Sigmoid}} \left(w^1 s_i^1 + w^2 s_i^2 + w^3 s_i^3\right) \tag{17}$$

Based on the score s, the nodes in the graph are first reordered, and then the top-ranked nodes is selected. The ratio of nodes selected by pooling is r, i.e., $r \times C$ nodes are retained:

$$\text{idx} = \text{top-rank}\left(p, \left\lceil r * n_i^k \right\rceil\right) \tag{18}$$

$$\widetilde{H}_i^{k+1} = H_i^k(\text{idx}, :) \tag{19}$$

$$A_i^{k+1} = A_i^k(\text{idx}, \text{idx}) \tag{20}$$

top-rank$(\cdot)$ denotes the function that returns the index of the top $n_i^{k+1} = \left\lceil r * n_i^k \right\rceil$, and $H_i^k(\text{idx}, :)$ and $A_i^k(\text{idx}, \text{idx})$ denote the rows or columns extracted to form the node representation matrix and the adjacency matrix of the subgraph. Finally, and $\widetilde{H}_i^{k+1}$ and A_i^{k+1} denote the node features and graph structure information of the next layer.

3.2.5. Classification Module

The output corresponding to each frequency band is aggregated and fed to the fully connected layer to obtain the final classification result:

$$Output = Linear(Dropout(\sigma_{\mathrm{ReLU}}(\tilde{H}_i^{k+1}))) \tag{21}$$

Where $Dropout(\bullet)$ is to randomly set the outputs of some neurons to zero during the training process. $Linear(\bullet)$ is the fully-connected layer, which obtains the final prediction.

4. Experiments

4.1. Datasets

The SEED dataset included 15 healthy subjects exposed to three discrete emotion categories (positive, neutral, negative) induced through standardized film clips. The stimulus set included 15 emotionally film excerpts, with five clips systematically allocated to each affective category. Subjects participated in three experimental sessions spaced one week apart, with each session comprising 15 trials. EEG signals were acquired using an ESNeuro 62-electrode system with initial sampling at 1000Hz, subsequently downsampled to 200Hz. Data preprocessing involved band-pass filtering (0.5-70 Hz), baseline correction, and ocular artifact removal via Independent Component Analysis (ICA).

The SEED-IV dataset, utilizing identical acquisition equipment to SEED, incorporates four discrete emotion categories: happiness, sadness, fear, and neutrality. Fifteen participants viewed 24 emotionally film stimuli, systematically grouped into six excerpts per affective category. Each subject participated in three experimental sessions on separate days, comprising 24 trials per session.

4.2. Results

4.2.1. Performance on SEED and SEED-IV

To evaluate the overall performance of the proposed FG-MSTGNN, we conducted Leave-One-Subject-Out (LOSO) cross-validation experiments on the public datasets SEED and SEED-IV. The results are shown in Table 1, which contains the classification performance metrics for each subject, including the accuracy (ACC), F1-score and AUC values. The experiment involved all 15 subjects, and the results for each subject are listed with the final mean and standard deviation (Std) calculated.

As can be seen from Table 1, the SEED shows distinctive advantages in all three core metrics, with a mean value of accuracy of 94.67%. More than 93% of the subjects (14 subjects) had an ACC of more than 90%, indicating that the model performs excellently in classification tasks. The average of the F1-score values can be up to 93.74%, from which it can be concluded that the overall model has a better performance in dealing with the unbalanced category. The feature discriminative index AUC is particularly outstanding with a mean value of 98.76%±1.39%, and the AUC of all subjects is higher than 94.52%, which confirms the high separability of the EEG affective features in the three-classification task.

Table 1: Accuracy of subject-independent experiments of the FG-MSTGNN model on the SEED and SEED-IV (%).

Subjects	**SEED**			**SEED-IV**		
	ACC	**F1**	**AUC**	**ACC**	**F1**	**AUC**
Sub01	91.11	91.32	97.56	86.11	86.15	95.58
Sub02	95.56	95.62	99.19	79.17	79.48	95.01
Sub03	97.78	97.78	99.85	81.94	81.95	95.99
Sub04	80.00	79.37	94.52	88.89	88.70	96.86
Sub05	91.11	91.07	97.93	87.50	87.58	98.25
Sub06	97.78	97.78	99.48	87.50	87.40	97.81
Sub07	100.00	97.78	100.00	83.33	83.47	96.76
Sub08	95.56	95.55	99.26	91.67	91.45	98.66
Sub09	95.56	93.32	97.63	84.72	84.66	96.14
Sub10	93.33	93.14	98.30	88.89	88.94	97.84
Sub11	95.56	93.25	99.41	80.56	80.55	95.42
Sub12	93.33	91.17	98.89	80.56	80.55	94.06
Sub13	95.56	95.54	99.33	86.11	86.16	98.20
Sub14	97.78	95.62	100.00	84.72	84.66	97.22
Sub15	100.00	97.78	100.00	87.50	87.45	97.87
Avg	**94.67**	**93.74**	**98.76**	**85.28**	**85.28**	**96.78**
Std	**4.72**	**4.51**	**1.39**	**3.38**	**3.38**	**1.27**

For SEED-IV, the mean AUC across all subjects was 96.78%, indicating that the model was effective in discriminating between different emotional categories for most subjects.The mean value of accuracy reached 85.28%, with a difference of 12.5% between Sub08 (ACC=91.67%, F1=91.45%), which had the best classification effectiveness, and Sub02 (ACC=79.17%, F1=79.48%), which had the lowest classification accuracy.

Overall, the results of the LOSO cross-validation experiments show the excellent performance of the model on both SEED and SEED-IV, indicating that the model is highly generalisable and robust.

4.2.2. Ablation Experiments

To further explore the effectiveness of each module of the proposed framework, we conducted an ablation study using LOSO experiments on SEED and SEED-IV. Specifically, as shown in Table 2, we explored the impact of the Feature Learning Stage and the Structure Optimization Stage, with results corresponding to w/o MTFCE and w/o DPGPM, respectively.

From Table 2, we can find that the proposed framework achieved the optimal composite performance in subject-independent experiments. The AUC of the proposed framework in the SEED dataset was improved by 0.66% and 1.18% compared to the w/o DPGPM and w/o MTFCE, respectively (w/o MTFCE: AUC=97.58%; w/o DPGPM: AUC=98.10%; Ours: AUC=98.76%). This advantage is even more distinctive in SEED-IV. The AUC of the proposed framework on SEED-IV is improved by 4.26% and 3.61% compared to w/o DPGPM and w/o MTFCE, respectively (w/o MTFCE: AUC=95.03%; w/o DPGPM: AUC=93.43%;

Table 2: Ablation study for subject-independent classification accuracy on SEED and SEED-IV (%).

Model	SEED						SEED-IV					
	ACC	STD	F1	STD	AUC	STD	ACC	STD	F1	STD	AUC	STD
Ours	**94.67**	**4.72**	**93.73**	4.51	**98.76**	**1.39**	**85.28**	**3.38**	**85.28**	**3.38**	**96.78**	1.27
w/o DPGPM	93.77	4.82	89.93	**4.32**	98.10	1.82	81.02	4.01	80.94	4.14	93.43	2.56
w/o MTFCE	92.59	5.48	90.20	6.37	97.58	2.53	81.67	3.83	79.73	4.96	95.03	**1.19**

Ours: AUC=96.78%). The Std of the proposed complete framework remains lowest in both datasets (SEED: ACC_Std=4.72%, F1_Std=4.51%; SEED-IV: ACC_Std=3.38%, F1_Std= 3.38%), which suggests that it fuses the cross-band temporal-frequency features through the MTFCE.

To quantitatively assess model efficacy, we employed violin plots (Figure 3) to visualize the classification performance of MTFCE and TAGCM, demonstrating their statistically significant differences.

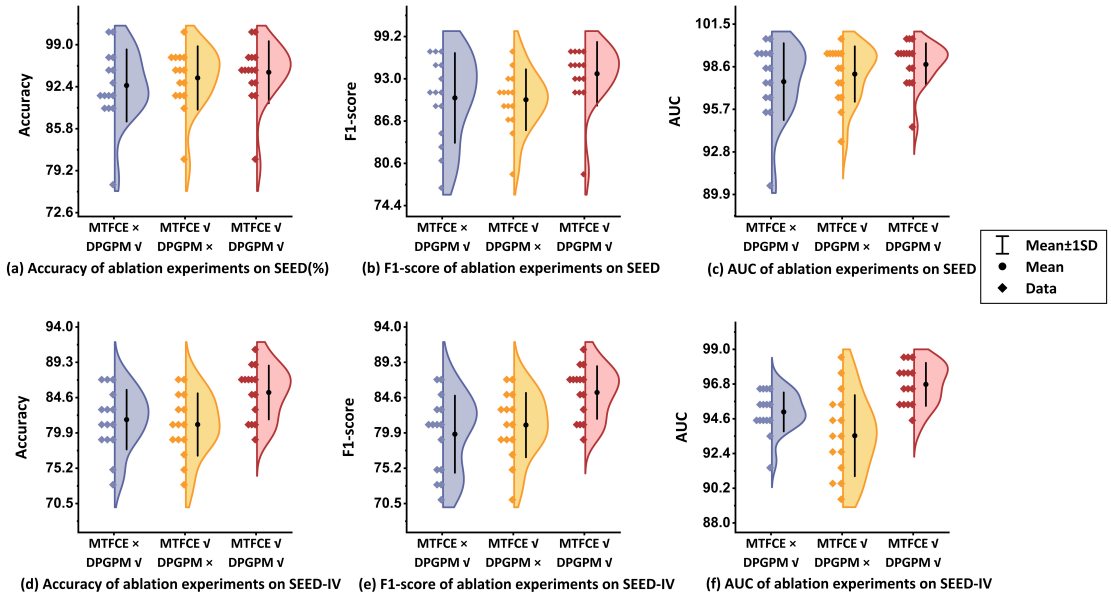

Figure 3: Ablation experiments on SEED and SEED-IV. Three configurations: MTFCE✓ DPGPM×, MTFCE× DPGPM✓, and MTFCE✓ DPGPM✓.

The violin diagram for SEED (Figure 3 a-c) show a clear pattern of narrow-waisted, high-density peaks indicating that the model predictions are highly stable in this dataset. while the distribution for SEED-IV (Figure 3 d-f) is more dispersed.And the median line of

our proposed model (red) is significantly higher than all ablation models (blue, yellow) to achieve optimal performance.

In summary, the ablation study analysis reveals that the synergistic integration of MTFCE and TAGCM drives performance improvements, where MTFCE effectively incorporates cross-frequency temporal-spectral features while TAGCM captures emotion-relevant dynamic brain network topology patterns, collectively achieving superior discriminative characterization of emotional EEG signals.

4.2.3. Comparison Experiment Against SOTA Method

As shown in Table 3, our proposed method achieves superior performance compared to other deep learning methods on both the SEED and SEED-IV datasets, with notable improvements in classification accuracy and stability. Entries marked with a dash (-) denote unreported results from the original studies.

Table 3: Comparison between FG-MSTGNN and SOTA methods on SEED and SEED-IV datasets using LOSO cross-validation (%).

Method	**Year**	**SEED**		**SEED-IV**	
		ACC	**Std**	**ACC**	**Std**
GMSS Li et al. (2023)	2023	86.52	6.22	73.48	7.41
V-IAG Song et al. (2023)	2023	88.38	4.80	-	-
GRU-Conv Xu et al. (2023)	2023	87.04	13.35	-	-
PGCN Jin et al. (2024)	2024	84.59	8.68	73.69	7.16
BFE-Net Zhang et al. (2024a)	2024	92.29	4.65	79.81	4.11
BF-GCN Li et al. (2025)	2024	92.72	3.90	82.03	8.42
PR-PL Zhou et al. (2024)	2024	93.06	5.12	81.32	8.53
MSS-JDA Chen et al. (2025)	2025	93.78	**3.39**	78.93	7.39
Ours	**2025**	**94.67**	4.88	**85.28**	**3.38**

For the SEED, the proposed framework achieves the highest performance with an accuracy of 94.67%, which is an improvement of 0.89% over MSS-JDA, the model with the best accuracy among the methods compared (ACC=93.78%). The Std of 4.88% indicates the high stability of the experimental results, which is significantly superior to PGCN (Std=8.68%) and GRU-Conv (Std=13.35%).

For the SEED-IV, the proposed method continues to maintain its dominance with an accuracy of 85.83%, an improvement of 3.8% compared to the next best method, BF-GCN (ACC=82.03%). The performance difference advantage over the domain adaptation method (MSS-JDA: 78.93%±7.39) and graph neural network (BFE-Net: 79.81%±4.11) is significant. Notably, the present method has the lowest Std of all compared methods at 3.38%, reducing volatility by 19.7% compared to the suboptimal stable V-IAG (Std=4.80).

Based on the results in the table, the proposed framework effectively mitigates the problem of 'accuracy-stability' trade-off in deep learning, and achieves simultaneous optimization of both metrics. These results show that the proposed method has better feature characterization ability in the cross-subject EEG emotion recognition task.

5. Conclusion

We propose FG-MSTGNN, a dual-stage adaptive framework for cross-subject EEG-based emotion recognition, which enhances affective decoding through systematic integration of graph-theoretic principles and multi-frequency rhythmic patterns. The proposed framework reveals a change in emotional state brings corresponding changes in both functional brain network topology and EEG rhythmic dynamics. This framework initiates transformative advances in human-machine collaboration by establishing a neurophysiologically based architecture, while providing a fundamental interpretable computational structure for the development of clinically viable personalized affective brain-computer interfaces.

6. Acknowledgements

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 [202403011222002], in part by the Natural Science Foundation for Young Scientists of Shanxi Province under Grant [202403021222057], and in part by the Science and Technology Innovation Program of Shanxi under Grant [RD2300004062].

References

E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–98, 2009. ISSN 1471-0048 (Electronic), 1471-003X (Linking). doi: 10.1038/nrn2575. URL https://www.ncbi.nlm.nih.gov/pubmed/19190637.

Shinan Chen, Weifeng Ma, Yuchen Wang, and Xiaoyong Sun. Mss-jda: Multi-source self-selected joint domain adaptation method based on cross-subject eeg emotion recognition. *Biomedical Signal Processing and Control*, 100:106953, 2025. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2024.106953. URL https://www.sciencedirect.com/science/article/pii/S1746809424010115.

Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. TSception: Capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2023. ISSN 1949-3045. doi: 10.1109/TAFFC.2022.3169001. URL https://doi.org/10.1109/TAFFC.2022.3169001.

Yuxiao Geng, Shuo Shi, and Xiaoke Hao. Deep learning-based EEG emotion recognition: a comprehensive review. *Neural Computing and Applications*, 37(4):1919–1950, 2024. ISSN 0941-0643. doi: 10.1007/s00521-024-10821-y. URL https://doi.org/10.1007/s00521-024-10821-y.

Wenhui Guo, Yaxuan Li, Mengxue Liu, Rui Ma, and Yanjiang Wang. Functional connectivity-enhanced feature-grouped attention network for cross-subject eeg emotion recognition. *Knowledge-Based Systems*, 283:111199, 2024. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2023.111199. URL https://www.sciencedirect.com/science/article/pii/S0950705123009498.

Ming Jin, Changde Du, Huiguang He, Ting Cai, and Jinpeng Li. Pgcn: Pyramidal graph convolutional network for eeg emotion recognition. *IEEE Transactions on Multimedia*, 26:9070–9082, 2024. ISSN 1520-9210, 1941-0077. doi: 10.1109/tmm.2024.3385676. URL https://doi.org/10.1109/tmm.2024.3385676.

D. Klepl, M. Wu, and F. He. Graph neural network-based eeg classification: A survey. *IEEE Trans Neural Syst Rehabil Eng*, 32:493–503, 2024. ISSN 1558-0210 (Electronic), 1534-4320 (Linking). doi: 10.1109/TNSRE.2024.3355750. URL https://www.ncbi.nlm.nih.gov/pubmed/38236670.

C. Li, T. Tang, Y. Pan, L. Yang, S. Zhang, Z. Chen, P. Li, D. Gao, H. Chen, F. Li, D. Yao, Z. Cao, and P. Xu. An efficient graph learning system for emotion recognition inspired by the cognitive prior graph of eeg brain network. *IEEE Trans Neural Netw Learn Syst*, 36(4):7130–7144, 2025. ISSN 2162-2388 (Electronic), 2162-237X (Linking). doi: 10.1109/TNNLS.2024.3405663. URL https://www.ncbi.nlm.nih.gov/pubmed/38837920.

P. Li, X. Gao, C. Li, C. Yi, W. Huang, Y. Si, F. Li, Z. Cao, Y. Tian, and P. Xu. Granger causal inference based on dual laplacian distribution and its application to MI-BCI classification. *IEEE Transactions on Neural Networks and Learning Systems*, 35 (11):16181–16195, 2024. ISSN 2162-2388. doi: 10.1109/TNNLS.2023.3292179. URL https://www.ncbi.nlm.nih.gov/pubmed/37463076.

Yang Li, Ji Chen, Fu Li, Boxun Fu, Hao Wu, Youshuo Ji, Yijin Zhou, Yi Niu, Guangming Shi, and Wenming Zheng. Gmss: Graph-based multi-task self-supervised learning for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2512–2525, 2023. ISSN 1949-3045, 2371-9850. doi: 10.1109/taffc.2022.3170428. URL https://doi.org/10.1109/taffc.2022.3170428.

Zhen Liang, Rushuang Zhou, Li Zhang, Linling Li, Gan Huang, Zhiguo Zhang, and Shin Ishii. Eegfusenet: Hybrid unsupervised deep feature characterization and fusion for high-dimensional eeg with an application to emotion recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1913–1925, 2021. ISSN 1534-4320, 1558-0210. doi: 10.1109/tnsre.2021.3111689. URL https://doi.org/10.1109/tnsre.2021.3111689.

Shuaiqi Liu, Yingying Zhao, Yanling An, Jie Zhao, Shui-Hua Wang, and Jingwen Yan. Glfanet: A global to local feature aggregation network for eeg emotion recognition. *Biomedical Signal Processing and Control*, 85, 2023. ISSN 17468094. doi: 10.1016/j.bspc.2023.104799. URL https://doi.org/10.1016/j.bspc.2023.104799.

X. Qiu, S. Wang, R. Wang, Y. Zhang, and L. Huang. A multi-head residual connection gcn for eeg emotion recognition. *Comput Biol Med*, 163:107126, 2023. ISSN 1879-0534 (Electronic), 0010-4825 (Linking). doi: 10.1016/j.compbiomed.2023.107126. URL https://www.ncbi.nlm.nih.gov/pubmed/37327757.

Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2020. ISSN 1949-3045, 2371-9850. doi: 10.1109/taffc.2018.2817622. URL https://doi.org/10.1109/taffc.2018.2817622.

Tengfei Song, Suyuan Liu, Wenming Zheng, Yuan Zong, Zhen Cui, Yang Li, and Xiaoyan Zhou. Variational instance-adaptive graph for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 14(1):343–356, 2023. ISSN 1949-3045, 2371-9850. doi: 10.1109/taffc.2021.3064940. URL https://doi.org/10.1109/taffc.2021.3064940.

Wei Tao, Chang Li, Rencheng Song, Juan Cheng, Yu Liu, Feng Wan, and Xun Chen. EEG-based emotion recognition via channel-wise attention and self attention. *IEEE Transactions on Affective Computing*, 14(1):382–393, 2023. ISSN 1949-3045. doi: 10.1109/TAFFC.2020.3025777. URL https://doi.org/10.1109/TAFFC.2020.3025777.

Min Wang, Heba El-Fiqi, Jiankun Hu, and Hussein A. Abbass. Convolutional neural networks using dynamic functional connectivity for eeg-based person identification in diverse human states. *IEEE Transactions on Information Forensics and Security*, 14(12): 3259–3272, 2019. ISSN 1556-6013, 1556-6021. doi: 10.1109/tifs.2019.2916403. URL https://doi.org/10.1109/tifs.2019.2916403.

Guixun Xu, Wenhui Guo, and Yanjiang Wang. Subject-independent eeg emotion recognition with hybrid spatio-temporal gru-conv architecture. *Medical and Engineering, Biological and Computing*, 61(1):61–73, 2023. ISSN 0140-0118. doi: 10.1007/s11517-022-02686-x. URL https://doi.org/10.1007/s11517-022-02686-x.

Lijun Yang, Yixin Wang, Xiaohui Yang, and Chen Zheng. Stochastic weight averaging enhanced temporal convolution network for eeg-based emotion recognition. *Biomedical Signal Processing and Control*, 83:104661, 2023. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2023.104661. URL https://www.sciencedirect.com/science/article/pii/S1746809423000940.

J. Zhang, Y. Hao, X. Wen, C. Zhang, H. Deng, J. Zhao, and R. Cao. Subject-independent emotion recognition based on eeg frequency band features and self-adaptive graph construction. *Brain Sci*, 14(3), 2024a. ISSN 2076-3425 (Print), 2076-3425 (Electronic), 2076-3425 (Linking). doi: 10.3390/brainsci14030271. URL https://www.ncbi.nlm.nih.gov/pubmed/38539659.

Yongtao Zhang, Yue Pan, Yulin Zhang, Min Zhang, Linling Li, Li Zhang, Gan Huang, Lei Su, Honghai Liu, Zhen Liang, and Zhiguo Zhang. Unsupervised time-aware sampling network with deep reinforcement learning for eeg-based emotion recognition. *IEEE Transactions on Affective Computing*, 15(3):1090–1103, 2024b. ISSN 1949-3045, 2371-9850. doi: 10.1109/taffc.2023.3319397. URL https://doi.org/10.1109/taffc.2023.3319397.

Rushuang Zhou, Zhiguo Zhang, Hong Fu, Li Zhang, Linling Li, Gan Huang, Fali Li, Xin Yang, Yining Dong, Yuan-Ting Zhang, and Zhen Liang. Pr-pl: A novel prototypical representation based pairwise learning framework for emotion recognition using eeg signals. *IEEE Transactions on Affective Computing*, 15(2):657–670, 2024. ISSN 1949-3045, 2371-9850. doi: 10.1109/taffc.2023.3288118. URL https://doi.org/10.1109/taffc.2023.3288118.