Cognitive Computation (2025) 17:3
https://doi.org/10.1007/s12559-024-10358-1

A Graph Neural Network with Spatial Attention for Emotion Analysis

Tian Chen[1] · Lubao Li[1] · Xiaohui Yuan[2]

Received: 11 July 2023 / Accepted: 3 November 2024

Abstract
Emotion recognition plays a crucial role in the diagnosis and treatment of various mental disorders. Research studies revealed the close relationship between brain regions and their functional roles in emotions. Propose a learning method that extends graph neural networks and takes into account the spatial relationship between EEG channels and their contributions of different regions of the brain to human emotions. Our method uses the adjacency matrix to model the spatial topological relationships in multi-channel EEG signals and learns weights to adjust their contributions to the classification. Extensive evaluation is conducted using public data sets, including comparison studies with state-of-the-art methods and performance analysis. In our comparison studies, our method demonstrates superior performance in terms of average accuracy. It is demonstrated that the proposed method improves the accuracy of emotion recognition and analyzes the brain at a fine granularity to decide the part that is most related to the triggering of the emotion.

Keywords EEG · Emotion recognition · GNN · Attention

Introduction

Emotion recognition plays a crucial role in the diagnosis and treatment of various mental disorders such as depression and autism [2, 5]. Extensive research has been conducted on emotion recognition using both physiological and non-physiological signals, e.g., images, speech, and text. However, physiological signals, such as the electrocardiogram [1], electromyogram [10], eye movement [26], galvanic skin response [25], and electroencephalogram (EEG) [9], have gained significant attention due to their ability to provide valuable information for emotion recognition. Among these signals, EEG stands out as it directly reflects the electrical activity of the brain, which is closely linked to the generation of emotions [23]. This makes EEG signals highly influential in emotion recognition research.

Neuroscientific research revealed the close relationship between brain regions and their functional roles in emotions [13, 14]. The emotional categories are specifically associated with the activity of neural systems distributed in the cerebral cortex and subcortex [16]. Studies have been conducted to convert EEG signals into grid data. For instance, Yang et al. [29] converted 1D EEG timing signals into 2D EEG signal frames and utilized a hybrid neural network combining convolutional neural network (CNN) and recurrent neural network (RNN) to capture the spatial and temporal representations of the raw brain current components. However, this approach disregards the spatial information of the EEG channels, which are irregularly arranged in 3D space, thereby neglecting the structural information of the EEG channels.

Graph neural networks (GNNs) were developed for analyzing irregular data and have demonstrated their effectiveness in capturing representations for uncertain and unevenly distributed information. GNNs offer a promising avenue for exploring the relationships between different brain regions and emotions during EEG-based emotion recognition. Song et al. [24] employed a graph convolutional neural network (GCNN) on EEG signals for emotion recognition. This study proposed a dynamic graph convolutional neural network (DGCNN) to investigate the intrinsic connections between EEG channels, achieving an individual-dependent accuracy of 90.4% on the SEED database. While Song's method yielded good results, there is still room for improvement in terms of classification accuracy. Furthermore, it remains challenging to intuitively explain the factors influencing the classification accuracy in emotion recognition tasks.

✉ Xiaohui Yuan
xiaohui.yuan@unt.edu

[1] School of Computer and Information, Hefei University of Technology, Hefei, China

[2] Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA

Published online: 21 November 2024

In this paper, we propose a learning method that extends graph neural networks and takes into account the spatial relationship between EEG channels and their contributions of different regions of the brain to human emotions. Our method uses the adjacency matrix to model the spatial topological relationships in multi-channel EEG signals and learns weights to adjust their contributions to the classification. Extensive evaluation is conducted using public data sets, including comparison studies with state-of-the-art methods and performance analysis. Our main contributions of this paper include the following:

1. The proposed method integrates the hardware properties with the deep learning method by leveraging the spatial topology of electrode channels to construct the graph (represented as an adjacency matrix) and devises a channel attention mechanism for weight assignment to electrodes to encode the information in the EEG signals.
2. A graph neural network enhanced with spatial attention is developed for emotion recognition from EEG signals, which integrates feature visualization and attention mechanisms for localizing prominent features.

The remainder of this paper is organized as follows. Section 2 reviews the related work on emotion recognition using physiological signals. Section 3 introduces the basic scheme of this paper. Section 4 describes the experiments and analysis. Section 5 concludes this paper with a summary.

Related Work

The use of EEG signals for emotion recognition has garnered significant interest among researchers due to the close relationship between human emotions and cortical activity, providing a more realistic reflection of human emotional states. Several studies have achieved notable results in EEG-based emotion recognition by employing different features and classification methods. To address the variation in EEG signals across different individuals and improve the generalization ability to unseen subjects, Su et al. [27] employed the projection dictionary pair learning that uses a synthesis dictionary and an analysis dictionary to enhance the representation of features. Xu et al. [28] proposed the domain adversarial graph attention, which generates a graph using biological topology to model multi-channel EEG signals and uses self-attention pooling to extract salient EEG features from the graph. Zheng [35] developed a group sparse canonical correlation analysis (GSCCA) method for simultaneous channel selection and emotion recognition. Zheng et al. [33] leveraged a deep belief network (DBN), and a hidden Markov model is integrated to capture the emotional stage switching between positive and negative. To learn the discrepancy between the two brain hemispheres for improved emotion recognition, Li et al. [17, 19] proposed a bi-hemispheric discrepancy model (BiHDM) that learns discriminative emotional features for each hemisphere using global and local domain discriminators. Zheng et al. [34] constructed a deep belief network (DBN) and extracted features such as power spectral density (PSD), differential entropy (DE), and differential asymmetry (DASM) from the SEED dataset. Chen et al. [7] utilized features such as Lempel-Ziv complexity, wavelet detail coefficients, covariance degree, and approximate entropy after EMD decomposition and employed the LIBSVM classifier for classification, followed by fuzzy integration of channel results.

Graph neural networks (GCNs) are neural networks designed to process graph-structured data, including traffic networks, social networks, and brain networks. Inspired by the convolutional operation of CNNs in Euclidean domains, researchers have integrated spectral graph theory and neural networks to define convolution in graph domains. For example, Bruna et al. [4] combined spectral graph theory with neural networks and utilized the normalized graph Laplace operator for convolution in graph domains. Defferrard et al. [12] proposed fast local convolution using K-order Chebyshev polynomials to approximate the convolution kernel, aggregating information from K-order neighbors for each node. Kipf et al. [15] further limited K to 1 and introduced standard graph convolution networks with faster local graph convolution operations. The convolution layers in GCNs can be stacked to efficiently process K-order neighborhoods of nodes. Compared to classical CNN approaches, GCNs offer advantages in processing and extracting discriminative features from signals in discrete spatial domains.

Given that the brain consists of multiple functional regions that work together, graph neural networks can effectively represent the relationships between these topologies and better model brain mechanisms. Consequently, researchers have increasingly explored the application of graph neural networks in EEG-based emotion recognition. For instance, Yin et al. [30] proposed the ERDL model, which combines graph convolutional neural networks (GCNN) and long short-term memory neural networks (LSTM), utilizing GCNN for extracting graph domain features and LSTM for capturing temporal features. Zhang et al. [32] designed a graph convolutional broad network (GCB-net) to explore deeper information about graph structure data, utilizing graph convolutional layers for extracting graph structure input features and stacking multiple regular convolutional layers for extracting abstract features. While these methods achieve improved recognition results, they lack transparency and fail to provide explanations for the decision-making process of deep learning models.

The attention mechanism in deep learning is akin to human visual attention, where humans scan the global image to focus on specific areas of interest and gather more detailed information about the target. Similarly, attention mechanisms in deep learning select relevant information from a pool of data. Channel-wise attention, for instance, compresses global information and generates statistical information for each channel [6]. As multi-channel EEG signals often contain spatial information, attention can be applied to the GCN to explore the importance of EEG signal channels and extract spatial information based on their discriminative power. Li et al. [20] proposed a transferable attention neural network (TANN) that highlights the transferable EEG brain region data and samples through the attention mechanism. Zhong et al. [37] proposed a regularized graph neural network (RGNN) that captures local and global relations among different EEG channels. Cui et al. [11] developed a Gated Recurrent Unit-Minimum Class Confusion (GRU-MCC) model. A gated recurrent unit is applied to model the spatial dependence of electrodes and extract features. Qian et al. [21] proposed an Adaptive Graph Convolutional Network with Spatial Attention and Transformer (AGCN-SAT) for emotion recognition from EEG signals. Transformer is applied to extract global spatial features, which are concatenated with local spatial features. Spatial attention is used to obtain discriminative features. Zhang et al. [31] proposed a spatial-temporal recurrent neural network (STRNN) to integrate spatial and temporal features. A multidirectional recurrent neural network (RNN) layer is employed to capture long-range contextual cues, and discriminative features are derived with a bi-directional temporal RNN layer.

Techniques such as Class Activation Mapping (CAM) [38] and Gradient-weighted Class Activation Mapping (Grad-CAM) [22] are commonly used for visualizing neural networks and explaining the importance of input data about the target category, facilitating better decision-making. Neural networks typically consist of a feature extractor and a classifier, where the feature extractor captures image features, and the classifier assigns class labels to the extracted features. CAM replaces the fully connected layer of the neural network with a global average pooling (GAP) layer and retrains the model. Given a feature map A_1,A_2,...,A_n in the last layer, where each neuron in the classification layer corresponds to a class with a weight w_1,w_2,...,w_n, the generation of class C is performed accordingly.

$$L_{CAM}^{C} = \sum w_i^C A_i. \tag{1}$$

A feature map pixel corresponds to a specific region in the original input image, where its value represents the extracted features from that region. The size of the receptive field in the feature map, denoted as S_c, is determined by the pixel values and weights within the feature map. The global average pooling (GAP) layer is introduced to enable the network to learn which regions in the original input image contain category-related features during training. By removing the fully connected layer and not introducing any new parameters, GAP helps reduce the risk of overfitting. However, to derive the weights within GAP, the model needs to be retrained after replacing the last classifier. To address these limitations, Grad-CAM improves upon Class Activation Mapping (CAM) by utilizing the gradient information flowing into the last convolutional layer of the CNN. This allows Grad-CAM to assign weight values to each neuron based on the gradient information, thereby enabling a more focused analysis of specific decision-making processes. Given a classification score S_c for class C, a feature map size $Z = C_1 * C_2$, and the weight of the A_{kj}^i neuron in the ith feature map represented by the partial derivative ∂_k^c, Grad-CAM provides a refined approach for assigning weights to individual neurons.

$$\alpha_k^c = \frac{1}{Z} \sum_{i \in w} \sum_{j \in h} \frac{\partial y^c}{\partial A_{ij}^k}. \tag{2}$$

By applying ReLu to the weighted summation of linear activation maps, the attention is computed as follows:

$$L_{Grad-CAM}^{c} = ReLu(\sum_k \alpha_k^c A^k). \tag{3}$$

The output provides insights into the regions of interest that the neural network focuses on during emotion recognition. By calculating the heat map for each emotion category across all samples, we can examine the variations in how different emotions are classified and gain an understanding of the importance of different channels in the emotion recognition process. These heat maps offer valuable visualizations that aid in interpreting and analyzing the network's decision-making mechanisms.

Spatial-Attention Graph Neural Network

Network Architecture

The architecture of our proposed Spatial-Attention Graph Neural Network is depicted in Fig. 1. The network comprises five stages: an input layer, a front-end graph convolution layer, a Grad-CAM filtering layer, a back-end graph convolution layer, and an output layer. In the input layer, we construct an adjacency matrix that captures the spatial relationships among EEG channels. Channel-wise attention is employed to adjust the weight parameters of each graph node, adapting them to the input EEG data. Note that the channels are spatially placed on the human skull and it is hence representing the spatial attention in terms of human brain activities. Subsequently, the data is fed into the graph convolution layer,

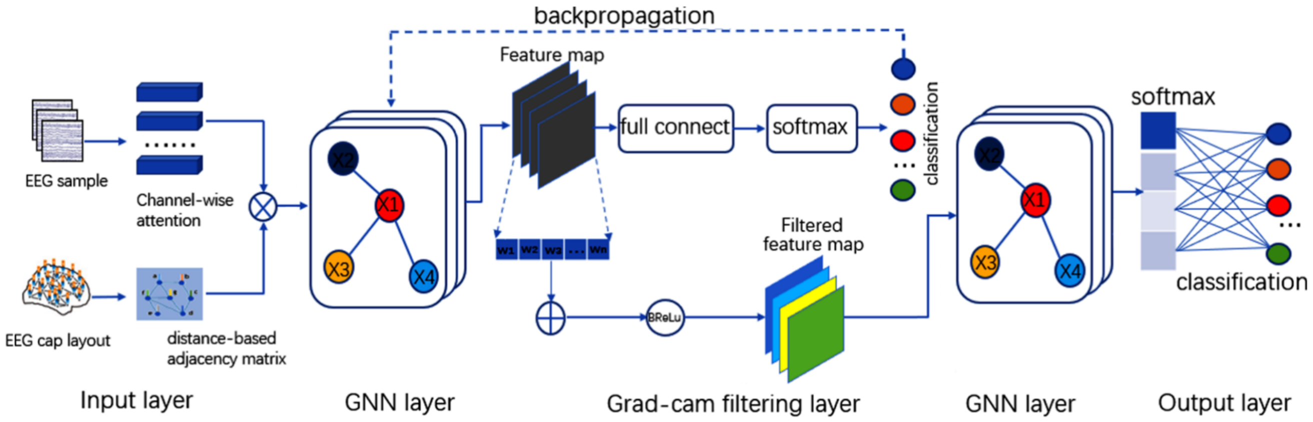


Fig. 1 The network architecture of the proposed method

where the EEG data is processed to extract its feature map. The resulting output consists of feature maps corresponding to the electrode channels.

In the Grad-CAM filtering layer, we utilize backpropagation to obtain the gradient information of the feature map, which is converted into weight information for each channel, denoted as $W = [w_1, w_2, ..., w_n]$. Considering the high-dimensional nature and redundant information present in EEG signals collected through multi-channel electrodes, we incorporate attention coefficients $\rho = [0.1, 0.2, ..., 0.9]$ in the Grad-CAM filtering layer to selectively filter out features with minimal impact on the classification outcome. The graph convolution layer then trains the filtered data, followed by classification through a softmax activation function.

Emotion Recognition

To represent EEG signals as a graph, we assign each EEG channel as a node, and the spatial proximity between two channels is captured by an edge connecting them. The graph is represented using an adjacency matrix. The input layer of CA-GNN initializes the adjacency matrix based on the spatial topology of the EEG channels, as illustrated in Fig. 2. Additionally, initial weights are assigned to each edge based on channel attention mechanisms.

We introduce CA-GNN, a novel network architecture for emotion recognition, as illustrated in Fig. 1. The network consists of five main components: an input layer, a graph convolution layer, a Grad-CAM filtering layer, a second graph convolution layer, and an output layer. In the input layer, we initialize the adjacency matrix based on the standard 10-20 electrode distribution in the cerebral scalp, assigning edge weights as follows:

$$A_{ij} = \frac{\alpha}{d_{ij}^2}, \tag{4}$$

where A_{ij} represents the weight of the edge connecting channels i and j, and d_{ij} represents the spatial distance between the two channels. The adjacency matrix captures the spatial relationships among EEG channels. The strength of connections between brain regions is inversely correlated with the square distances [36].

To address the issue of redundant information in EEG signals collected through multi-channel electrodes, we introduce channel attention to identify emotionally relevant channels. Unlike manual screening, our adaptive approach assigns weights to different channels based on their importance.

$$X = [x_1, x_2, ..., x_c], u \in \mathbb{R}^{W\times H}, \tag{5}$$

where $[x_1, x_2, ..., x_c]$ are the EEG data of different channels. H denotes different frequency bands, and W denotes EEG sample feature points.

$$U = [u_1, u_2, ..., u_c], u \in \mathbb{R}^{W\times H}. \tag{6}$$

Channel attention probability U is obtained through a softmax function applied to the output of a neural network. The weights in U are used to calculate the channel attention features, C_j, by multiplying them with the corresponding spectral features, S_j:

$$C_j = U_j \cdot S_j. \tag{7}$$

The convolution part of the graph utilizes the ChebNet convolution kernel, which processes the graph input data X and outputs $Z \in \mathbb{R}^{n\times d}$, where n is the number of electrode channels, and d is the dimensionality of the output features. The transformation between adjacent layers of the graph neural network (GNN) is computed as follows:

$$X^{i+1} = f(X^i, A), \tag{8}$$

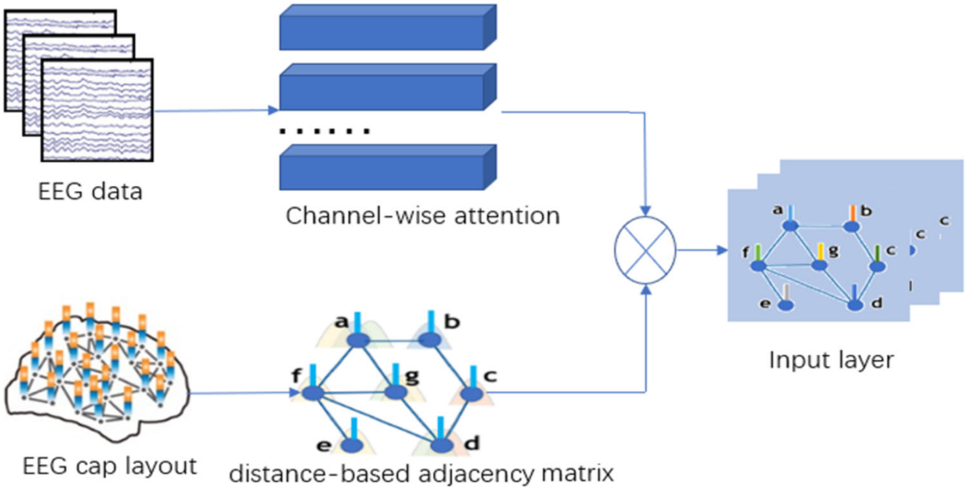


Fig. 2 The input layer of the network

where X^{i+1} is the features at the $(i+1)$-th layer, X^i denotes the features at the i-th layer, A is the adjacency matrix, and f denotes the mapping functions. For a graph convolution network, we have

$$X^{i+1} = \sigma(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X^i W^i), \tag{9}$$

where σ is the sigmoid function, and D is the diagonal matrix of A.

To understand the influence of different brain regions on classification results, we employ Grad-CAM and attention mechanisms in EEG emotion recognition. Class activation maps (CAMs) are generated to visualize the contribution of each region to the classification process. By applying backpropagation, we obtain the gradient information of the target layer and normalize the gradient vector. Attention coefficients are used to filter out channels below a threshold, generating a new vector for the subsequent layer. The CAMs are created by superimposing the original scalp layer with heat maps representing each channel's gradient information, indicating the influence of the channel in the classification process.

In our network, the input data X_i is a collection of multiple frequency band features extracted from EEG signals. It is represented as $X_i \in \mathbb{R}^{N \times n \times d}$, where N is the number of training samples, n is the number of electrode channels, and d is the number of features. The corresponding labels are denoted as Y_i, $Y_i \in 0, 1, ..., C-1$, where C is the number of categories. The probability of Y_i given X_i and the parameters θ is calculated as follows:

$$P(Y_i \mid X_i, \theta) = softmax(H\sigma(HX_iW^i)W^{i+1}), \tag{10}$$

$$H = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}, \tag{11}$$

where H represents the normalized adjacency matrix, and W represents the weight matrices.

The loss function of CA-GNN combines the cross-entropy function with an L1 regularization term:

$$-\sum_{i=1}^{N} \log(P(Y \mid X_i, \epsilon), \hat{Y}_i) + \alpha \|A\|_1. \tag{12}$$

The cross-entropy loss measures the difference between the actual emotion labels and the model's predicted labels, while the L1 regularization term encourages sparsity in the adjacency matrix A.

Results and Discussion

Datasets and Evaluation Metrics

We conducted experiments on three widely used public datasets: SEED [34], SEED-IV [36], and MPED [25]. The SEED dataset consists of 62 channels of EEG signals recorded from 15 subjects (8 females and 7 males). The subjects watched 15 movie clips and generated three emotions: neutral, positive, and negative. Each participant underwent three sets of experiments at different times, with each set comprising 15 trials. Figure 3 illustrates four example signals of the SEED dataset.

The SEED-IV dataset includes 62 channels of EEG signals collected from 15 subjects (8 females and 7 males). The subjects watched 72 movie clips designed to elicit happy, sad, fearful, and neutral emotions. Similar to the SEED dataset, each participant completed three sets of experiments at different times, with each set consisting of 24 trials.

The MPED dataset is a multimodal physiological signal emotion dataset comprising EEG signals recorded from 23 subjects using 62 EEG electrodes. The subjects watched 28 movie clips representing different emotions, including joy, fun, anger, fear, sadness, disgust, and neutrality.

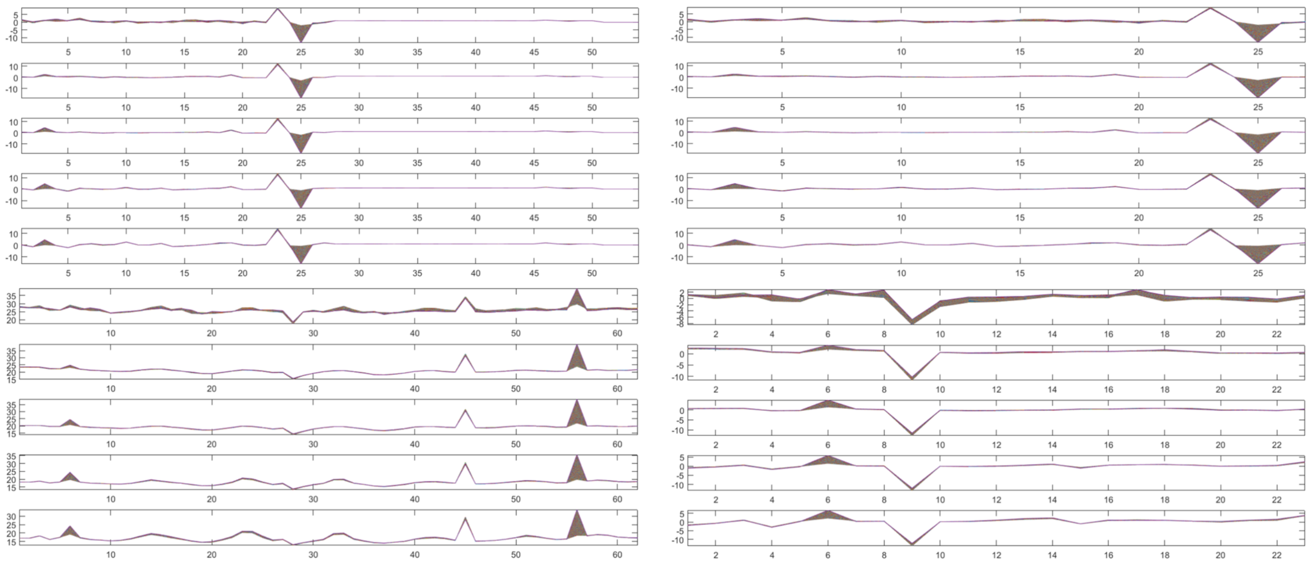


Fig. 3 Example signals of the SEED dataset. Multiple channels are shown in each plot

In our study, average accuracy and confusion matrices are used to evaluate the performance of EEG emotion recognition methods. Accuracy is the proportion of correctly classified emotions out of all emotions. This metric gives an overall view of the method performance without differentiating the performance for the individual emotion. The confusion matrix refers to a table that is used to evaluate the performance of a classification model by comparing its predicted and actual outputs. The matrix consists of four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The confusion matrix provides a more quantitative evaluation of each class and, hence, is useful in understanding the performance with respect to the accuracy of recognizing each emotion and the confusion among classes.

Experimental Settings

Our experiments were conducted on a computer with Intel i7 10th CPU and NVIDIA GTX 1660Ti GPU. The programming packages include Anaconda 3, PyTorch 1.9.0, and Cuda 10.2. The parameter settings of our method include the number of EEG channels, 62; the number of frequency bands, 5; and the number of convolution kernels, 32; and the convolution kernel was a second-order Chebyshev polynomial.

In our experiments, we conducted individual correlation classification experiments. For the SEED dataset, the training data and testing data were obtained from different trials of the same experiment. The training set comprised data from nine trials, while the test set contained data from the remaining six trials of the same experiment. The classification labels included negative, neutral, and positive emotions. Similarly, for the SEED-IV dataset, the training data consisted of data from 16 trials, and the test data included data from the other 8 trials of the same experiment. The classes include happy, sad, fearful, and neutral emotions. For the MPED dataset, the training data consisted of data from 21 trials, and the test data contained data from the remaining 7 trials of the same experiment. The classification labels included happy, funny, angry, fearful, sad, disgusted, and neutral emotions.

In all experiments, the parameters of our method were set as follows: the number of EEG channels was 62, the number of frequency bands was 5, the number of convolution kernels was set to 32, and the convolution kernel was a second-order Chebyshev polynomial. During the training process, the activation function used was BReLU. We employed the Adam optimizer with a learning rate of 0.001 and a dropout rate of 0.3. The graphical convolutional network consisted of 2 or 3 layers, and the entire model was implemented using PyTorch.

Comparison Study

We conduct a comparison study with fourteen state-of-the-art methods, including PDPL [27], GSCCA [35], TANN [20], RGNN [37], DBN [33], STRNN [31], DGCNN [24], BiDANN [17], AGCN-SAT [21], GCB-net [32], Emotion-Meter [36], GRU_MCC [11], BiHDM [19], and DAGAM [28]. In addition, we include SVM as a baseline, which has been applied to all three datasets. The average accuracy and standard deviation of these methods together with the performance of our method are reported in Table 1.

In general, both datasets SEED-IV and MPED are much more challenging than SEED as the performance of all methods is relatively lower. The average accuracy of various methods for the SEED dataset is in the range of 80% and mid 90%, whereas the average accuracy of the SEED-IV

Table 1 Average accuracy (%) and standard deviation

Method	Year	Dataset					
		SEED		SEED-IV		MPED	
SVM	–	83.99	9.92	56.62	20.05	31.14	8.06
GA-PDPL [27]	2023	69.89	14.39	–	–	24.87	5.83
GSCCA [35]	2016	82.96	9.95	69.08	16.66	–	–
TANN [20]	2021	84.41	8.75	68.00	8.35	28.32	5.11
RGNN [37]	2020	85.30	6.72	73.84	8.02	–	–
BiHDM [19]	2020	85.40	7.53	69.03	8.66	28.27	4.99
DBN [33]	2014	86.08	8.34	66.77	7.38	29.26	9.19
GRU_MCC [11]	2022	88.07	5.27	–	–	31.22	4.48
STRNN [31]	2018	89.50	7.63	–	–	–	–
DGCNN [24]	2018	90.40	8.49	69.88	16.29	36.92	12.78
BiDANN [17]	2018	92.38	7.04	70.29	12.63	–	–
EmotionMeter [36]	2018	–	–	70.58	17.01	–	–
DAGAM [28]	2023	92.59	3.21	**80.74**	4.14	–	–
AGCN-SAT [21]	2023	92.76	6.16	–	–	–	–
GCB-net [32]	2019	94.24	6.70	–	–	–	–
Our	–	**94.73**	5.63	74.86	10.81	**39.04**	4.46

dataset is about mid 50% to mid 70%. The performance of the MPED dataset is much lower about 30%. This is partly due to the large number of emotions in the MPED dataset. MPED has seven emotion classes, which is more than the number of classes in SEED or SEED-IV.

In this table, the best performance is highlighted in bold font face and the second best is marked with underline. For the SEED dataset, our method achieves an average accuracy of 94.73% with a standard deviation of 5.63%. Compared to GCB-net [37], our method improves the accuracy by 0.52%. In addition, the standard deviation of our method is among the smallest, which implies a more consistent emotion recognition. A similar trend is observed in the results using SEED-IV and MPED datasets. It is important to note that the improvement of our method with respect to the best among the state-of-the-art methods is much more significant for the results of using the MPED datasets at 5.74%. Although the average accuracy of DAGAM for the SEED-IV dataset is at 80.74%, our method exhibited a much-improved accuracy for the SEED dataset at a rate of 2.3%.

Part of the performance gain of our method can be attributed to the spatial-functional relationships encoded with our attention technique. The spatial topology-based adjacency matrix and the attention-based graph neural network provide a close synthesis of EEG signals of the human brain. In addition, the attention pattern-based Grad-CAM effectively filters out irrelevant regions for emotion recognition highlights the regions that contribute more, and enhances classification performance.

Asadzadeh et al. [3] proposed a GNN node (ESB-G3N) method and performed binary classification experiments on self-built data sets, which are obtained with sLORETA. On the other hand, public data sets, e.g., SEED, SEED-IV, and MPED, obtain emotion labels based on the forms completed by the subjects after the test. Due to the delay of feedback and factors other than the videos presented to the subjects, the self-reports could be different from the subjects' true emotions at the time of the visual excitement.

In comparison to the method by Chen et al. [8], our experiments utilized EEG data collected using Emotiv EPOC+. The classes include arousal and valence. Figure 4 illustrates the accuracy of our proposed graph neural network with attention over different epochs. The left figure shows the accuracy of arousal and the right one shows the accuracy of valence. It is clear that the arousal dimension stabilizes after 30 epochs, and the valence dimension stabilizes after 50 epochs. The model achieves an accuracy of 87.89% in the arousal dimension and 89.45% in the valence dimension at about 100 epochs. In contrast, the classification accuracy of the method presented in [8] for arousal and valence dimensions is 74.88% and 82.63%, respectively. These results highlight the superior performance of our proposed method over the approach in [8].

Performance Analysis

Figure 5 depicts the confusion matrix of our method applied to the three datasets. For both SEED and SEED-IV datasets, our method exhibits a fairly close performance to all emotion classes. The numbers on the diagonal line give the correct recognition rates. For the SEED dataset, the correct rate is above 90%. For the SEED-IV dataset, the rate is about

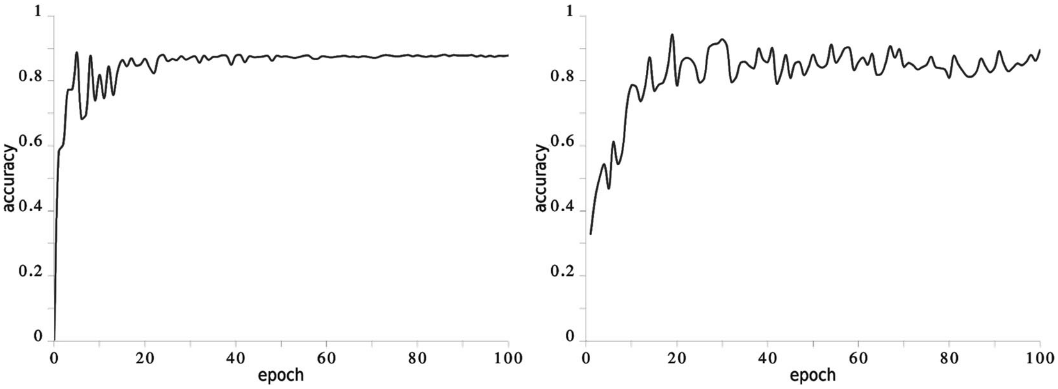


Fig. 4 Accuracy of graph neural network and attention-based models for EEG signals of arousal dimensions (left) and valence dimensions (right)

80%. However, for the MPED dataset, there exists a disparity among some emotions. Specifically, Digust appears to be an emotion that is often misclassified into Angry, Funny, and Sad. Also, the confusion between Joy and Funny is quite significant. It is arguable that there is a clear distinction between the two emotions.

The proposed network achieves a recognition rate of 96% for neutral and positive emotions and 90% for negative emotions in the SEED dataset. This indicates that the proposed method performs well in identifying positive and neutral emotions. However, the model shows relatively lower performance in detecting negative emotions, possibly due to the presence of various patterns corresponding to different negative emotions.

When it is applied to the SEED-IV dataset (as shown in the middle of Fig. 5), our method achieves recognition rates of 80%, 89%, and 75% for different negative emotions, indicating its ability to differentiate between different negative emotions. The confusion matrix of our method on the MPED dataset is shown on the right of Fig. 5. Our method achieves recognition rates of 46% and 49% for joy and fun, respectively. Positive emotions exhibit better performance than negative emotions, and joy and fun tend to be more easily confused within the same emotion category.

Ablation Study

Adjacency Matrix

To understand the contribution of the adjacency matrix to the learning topology, we conducted experiments using the SEED dataset and analyzed the impact of model structure on emotion recognition performance. Our experiments consider four adjacency matrix designs:

1. Fully connected adjacency matrix (FC): All elements of the adjacency matrix are set to 1, indicating that all channels are connected to each other with equal weight.
2. Locally connected adjacency matrix (LC) [18]: The entire brain is divided into 16 regions, and connections between channels within the same region have a weight of 1.
3. Randomly connected adjacency matrix (RC): All elements of the adjacency matrix are randomly initialized with values ranging from 0 to 1. Non-zero values indicate connections, while 0 represents no connection.
4. Adjacency matrix based on spatial topological relations (STR): The connection matrix between channels is determined based on their spatial proximity and distance, with closer channels having greater connection weights.

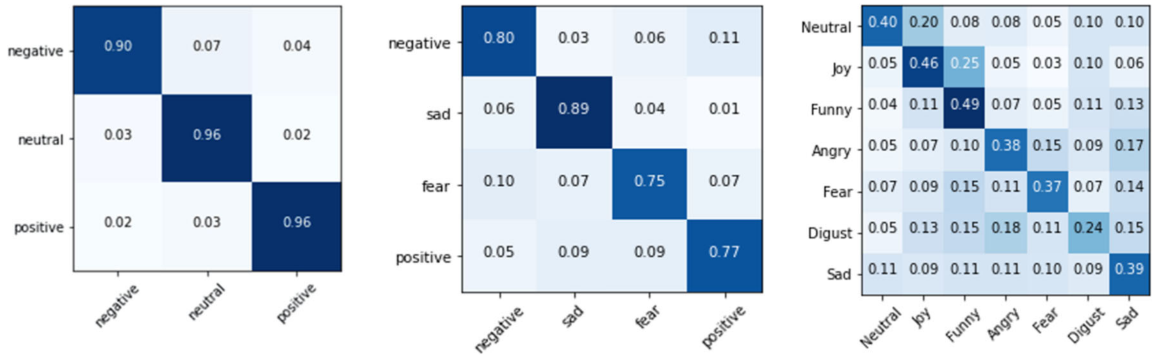


Fig. 5 Confusion matrix of our method on the SEED dataset (left), SEED-IV dataset (middle), and MPED dataset (right)

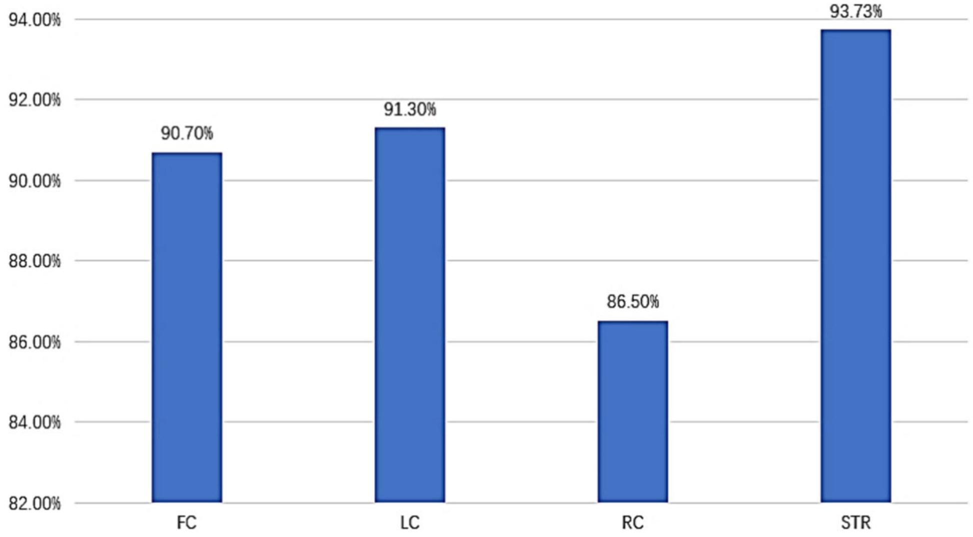


Fig. 6 Average accuracy of our method using different adjacency matrices

Figure 6 illustrates the average accuracy of our method using the four different adjacency matrices. Among the four adjacency matrics, the one based on spatial topological relations, i.e., STR, achieves the best accuracy, which aligns well with findings in neuroscience research [36]. On the other hand, the randomly connected adjacency matrix (RC) yielded the worst accuracy at 86.5%. The relatively poor performance of RC and the large difference between STR and RC is likely due to the fact that randomly initialized adjacency matrices introduce errors in the propagation of features on the graph network, resulting in lower accuracy. The accuracies of using the locally connected adjacency matrix and fully connected adjacent matrix are both quite competitive at 91.3% and 90.7%, respectively. This implies that the key connections may exist within local areas.

Attention Rate

To evaluate the effectiveness of the attention mechanism in improving emotion recognition, we conducted experiments on the SEED dataset to analyze the impact of different attention coefficients on the results. Different attention coefficients were set in the Grad-CAM filtering layer to filter out features with weights below the attention threshold. The filtered data was then fed into the second layer of the graph convolutional neural network for model training. The results of the experiments are illustrated in Fig. 7.

Figure 7 presents the recognition accuracy of our method using different attention coefficient rates. The results from the three datasets are depicted with different colors. Despite the subtle difference among the three accuracy curves, the overall trend is consistent with a peak near the attention rate at 0.5, which gives the highest recognition accuracy. When the attention coefficient is too small, numerous invalid features may not be filtered out, leading to a lower emotion recognition accuracy. As the attention coefficient increases, more important features for result classification are retained, resulting in improved emotion recognition rates. However, when the attention coefficient becomes too high, the majority of features are filtered out, resulting in an insufficient number of features for model classification and a lower emotion recognition accuracy.

Spatial Activity Analysis

Figure 8 illustrates the brain cortex activation distribution for two-layer map convolutional network classification. Color is used to encode the activity level of human brain regions, with darker colors indicating stronger brain activities. It is clear that significant brain activities are observed in the prefrontal, parietal, and occipital regions across all frequency bands, suggesting their involvement in emotion processing within the brain. Additionally, the asymmetrical activity in the left and right hemispheric sites indicates their critical role in emotion recognition, as evident from channel activation.

Figures 9 and 10 display the activation distribution maps generated by both the two-layer and three-layer graph neural networks, revealing that the multilayer graph neural network has minimal impact on the classification results. However, the activation distribution maps produced by the three-layer graph neural network exhibit clearer activity points and improved localization, indicating its ability to capture finer details.

Conclusion

In this study, we present a novel approach for EEG emotion recognition by leveraging graph neural networks (GNNs) and attention mechanisms. Our proposed method incorporates

Fig. 7 Accuracy of different attention coefficients

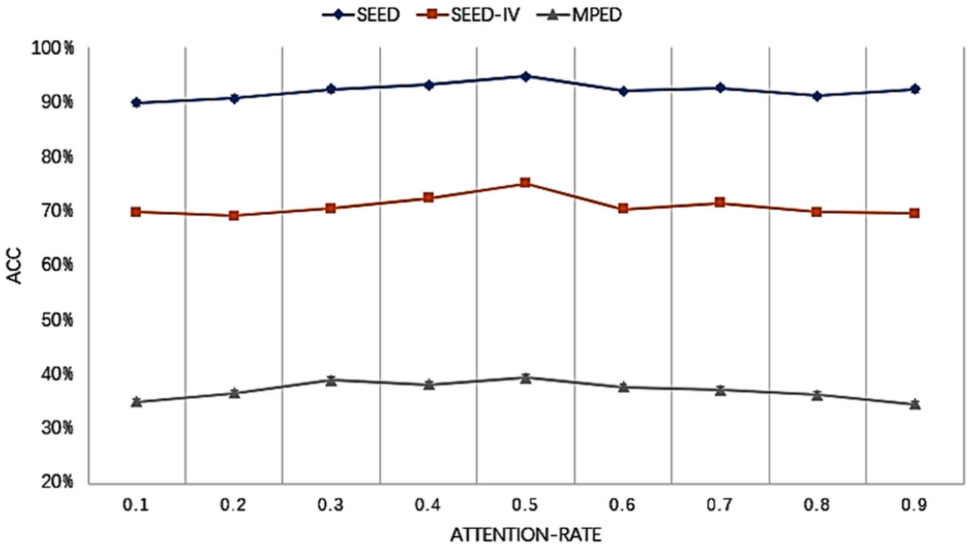


Fig. 8 Brain cortex activation distribution for two-layer map convolutional network classification

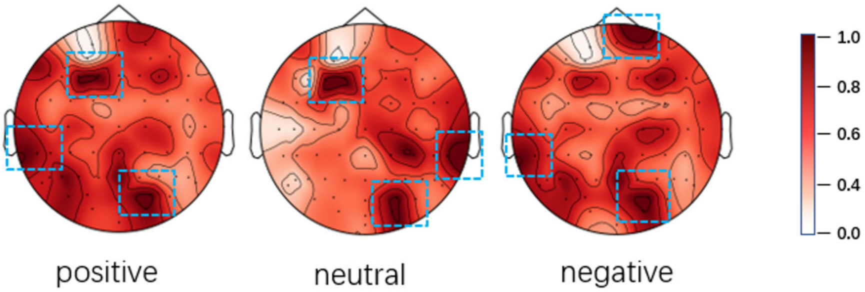


Fig. 9 Distribution of first-level attentional activation of three-layer graph neural network

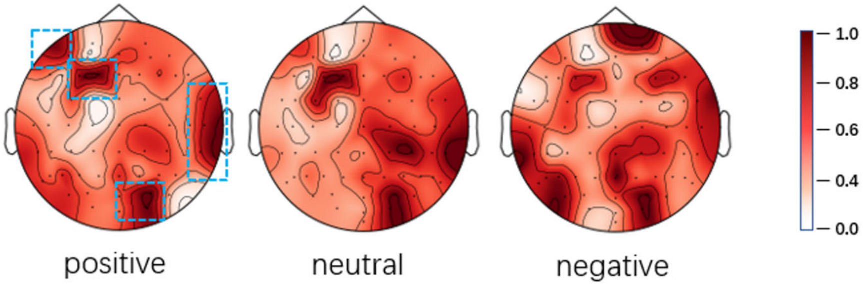
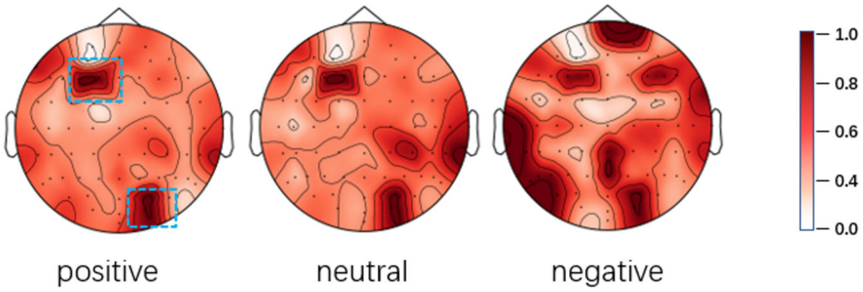


Fig. 10 Distribution of secondary attentional activation of three-layer graph neural network



spatial and attentional information from EEG signals to enhance the accuracy of emotion recognition while providing insights into the specific regions of the cerebral cortex that contribute significantly to the process. Drawing inspiration from neuroscience and the attention mechanism, we construct a graph using an adjacency matrix based on the topological relations between EEG channels. Additionally, we employ channel attention to assign weights to individual electrode channels, effectively capturing the topological relations inherent in the EEG signals. By utilizing the Grad-CAM and adjusting the attention parameters, we prioritize the features that are most critical for EEG emotion recognition.

Experiments were conducted with three public datasets and the results demonstrate the superior performance of our approach, achieving classification accuracy of 94.73%, 74.86%, and 39.04% on the SEED, SEED-IV, and MPED datasets, respectively. The analysis of activation distribution maps reveals pronounced activity in the prefrontal, parietal, and occipital regions across all frequency bands, suggesting their close association with emotional processing in the brain.

Acknowledgements The authors want to thank Siyuan Liu at the University of North Texas for his contribution to the revision of this manuscript.

Funding This work was supported in part by the National Natural Science Foundation of China (Grant numbers: 62174048 and 62027815).

Data Availability The data used in this study is available upon request.

Declarations

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

References

1. Foteini A, Dimitris H, Adam KA. ECG pattern analysis for emotion detection. IEEE Trans Affect Comput. 2011;3(1):102–15.
2. Al-Kaysi AM, Al-Ani A, Colleen KL, Tamara YP, Donel MM, Michael B, Tjeerd WB. Predicting TDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. J Affect Dis. 2017;208:597–603.
3. Shiva A, Tohid Yousefi R, Soosan B, Saeed M. Accurate emotion recognition utilizing extracted EEG sources as graph neural network nodes. Cogn Comput. 2023;15(1):176–89.
4. Joan B, Wojciech Z, Arthur S, Yann L. Spectral networks and locally connected networks on graphs. 2013. arXiv:1312.6203,
5. Andrey VB, Gennady GK, Alexander NS. Depression and implicit emotion processing: an EEG study. Neurophys Clinique/Clinical Neurophys. 2017;47(3):225–30.
6. Long C, Hanwang Z, Jun X, Liqiang N, Jian S, Wei L, Tat-Seng C. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017;5659–5667.
7. Tian C, Sihang J, Fuji R, Mingyan F, Yu G. EEG emotion recognition model based on the LibSVM classifier. Measurement. 2020;164:108047
8. Tian C, Hongfang Y, Xiaohui Y, Yu G, Fuji R, Xiao S. Emotion recognition based on fusion of long short-term memory networks and SVMs. Digital Signal Process. 2021;117:103153
9. Qiang C, Yingying L, Xiaohui Y. A hybrid method for muscle artifact removal from EEG signals. J Neurosci Methods. 2021;353:109104–1.
10. Bo C, Guangyuan L. Emotion recognition from surface EMG signal using wavelet transform and neural network. In: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, 2008;1363–1366.
11. Heng C, Aiping L, Xu Z, Xiang C, Jun L, Xun C. EEG-based subject-independent emotion recognition using gated recurrent unit and minimum class confusion. IEEE Trans Affect Comput, 2022.
12. Michaël D, Xavier B, Pierre V. Convolutional neural networks on graphs with fast localized spectral filtering. Adv Neural Inf Process Syst, 2016;29.
13. Hedy K, Lisa Feldman B, Josh J, Eliza Bliss-M, Kristen L, Tor D W. Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. Neuroimage, 2008;42(2):998–1031.
14. M Justin K, Rebecca A L, Amy L P, Annemarie C B, Kimberly M S, Ashley N M, Paul J W. The structural and functional connectivity of the amygdala: from normal emotion to pathological anxiety. Behav Brain Res, 2011;223(2):403–410.
15. Thomas N K, Max W. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2016.
16. Philip AK, Kevin SL. Decoding the nature of emotion in the brain. Trends in Cogn Sci. 2016;20(6):444–55.
17. Yang L, Wenming Z, Yuan Z, Zhen C, Tong Z, Xiaoyan Z. A bi-hemisphere domain adversarial neural network model for EEG emotion recognition. IEEE Trans Affect Comput. 2018;12(2):494–504.
18. Peiyang L, Huan L, Yajing S, Cunbo L, Fali L, Xuyang Z, Xiaoye H, Ying Z, Dezhong Y, Yangsong Z, et al. EEG based emotion recognition by combining functional connectivity network and local activations. IEEE Trans Biomed Eng. 2019;66(10):2869–81.
19. Yang L, Lei W, Wenming Z, Yuan Z, Lei Q, Zhen C, Tong Z, Tengfei S. A novel bi-hemispheric discrepancy model for EEG emotion recognition. IEEE Trans Cogn Development Syst. 2020;13(2):354–67.
20. Yang L, Boxun F, Fu L, Guangming S, Wenming Z. A novel transferability attention neural network model for EEG emotion recognition. Neurocomput. 2021;447:92–101.
21. Bincheng Q, Zhiqin Q, Huishan L, Qi L, Lei X. Agcn-sat: Adaptive graph convolutional network with spatial attention and transformer for EEG emotion recognition. In: 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information, 2023;418–423. IEEE,
22. Ramprasaath R S, Michael C, Abhishek D, Ramakrishna V, Devi P, Dhruv B. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, 2017;618–626.
23. Mohammad S, Maja P, Thierry P. Multimodal emotion recognition in response to videos. IEEE Trans Affect Comput. 2011;3(2):211–23.
24. Tengfei S, Wenming Z, Peng S, Zhen C. EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Trans Affect Comput. 2018;11(3):532–41.
25. Tengfei S, Wenming Z, Cheng L, Yuan Z, Xilei Z, Zhen C. MPED: a multi-modal physiological emotion database for discrete emotion recognition. IEEE Access. 2019;7:12177–91.

26. Soundariya RS, Renuga R. Eye movement based emotion recognition using electrooculography. In: 2017 Innovations in Power and Advanced Computing Technologies, 2017;1–5. IEEE,
27. Jipu S, Jie Z, Tiecheng S, Hongli C. Subject-independent EEG emotion recognition based on genetically optimized projection dictionary pair learning. Brain Sci. 2023;13(7):977.
28. Tao X, Wang D, Jiabao W, Yun Z. DAGAM: a domain adversarial graph attention model for subject-independent EEG-based emotion recognition. J Neural Eng. 2023;20(1): 016022.
29. Yilong Y, Qingfeng W, Ming Q, Yingdong W, Xiaowei C. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In: 2018 International Joint Conference on Neural Networks, 2018;1–7. IEEE,
30. Yongqiang Y, Xiangwei Z, Bin H, Yuang Z, Xinchun C. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. Applied Soft Comput. 2021;100: 106954.
31. Tong Z, Wenming Z, Zhen C, Yuan Z, Yang L. Spatial-temporal recurrent neural network for emotion recognition. IEEE Trans Cybernetics. 2018;49(3):839–47.
32. Tong Z, Xuehan W, Xiangmin X, CL Philip C. GCB-Net: graph convolutional broad network and its application in emotion recognition. IEEE Trans Affect Comput, 2019;13(1):379–388.
33. Wei-Long Z, Jia-Yi Z, Yong P, Bao-Liang L. EEG-based emotion classification using deep belief networks. In: 2014 IEEE International Conference on Multimedia and Expo, 2014;1–6. IEEE.
34. Wei-Long Z, Bao-Liang L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans Autonom Mental Development. 2015;7(3):162–75.
35. Wenming Z. Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. IEEE Trans Cogn Development Syst. 2016;9(3):281–90.
36. Wei-Long Z, Wei L, Yifei L, Bao-Liang L, Andrzej C. Emotion-Meter: a multimodal framework for recognizing human emotions. IEEE Trans Cybernetics. 2018;49(3):1110–22.
37. Peixiang Z, Di W, Chunyan M. EEG-based emotion recognition using regularized graph neural networks. IEEE Trans Affect Comput. 2020;13(3):1290–301.
38. Bolei Z, Aditya K, Agata L, Aude O, Antonio T. Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016;2921–2929.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.