Check for updates

STGATE: Spatial-temporal graph attention network with a transformer encoder for EEG-based emotion recognition

Jingcong Li, Weijian Pan, Haiyun Huang, Jiahui Pan and Fei Wang*

School of Software, South China Normal University, Guangzhou, China

Electroencephalogram (EEG) is a crucial and widely utilized technique in neuroscience research. In this paper, we introduce a novel graph neural network called the spatial-temporal graph attention network with a transformer encoder (STGATE) to learn graph representations of emotion EEG signals and improve emotion recognition performance. In STGATE, a transformer-encoder is applied for capturing time-frequency features which are fed into a spatial-temporal graph attention for emotion classification. Using a dynamic adjacency matrix, the proposed STGATE adaptively learns intrinsic connections between different EEG channels. To evaluate the cross-subject emotion recognition performance, leave-one-subject-out experiments are carried out on three public emotion recognition datasets, i.e., SEED, SEED-IV, and DREAMER. The proposed STGATE model achieved a state-of-the-art EEG-based emotion recognition performance accuracy of 90.37% in SEED, 76.43% in SEED-IV, and 76.35% in DREAMER dataset, respectively. The experiments demonstrated the effectiveness of the proposed STGATE model for cross-subject EEG emotion recognition and its potential for graph-based neuroscience research.

1. Introduction

Emotion is a generalization of subjective human experience and behavior. Emotions affect our perceptions and attitudes dramatically and play an essential role in human-computer interaction (HCI) (Jerritta et al., 2011). Emotions have a significant impact on our evaluation, attitudes, behavior, decisions, cognition, learning, perception, and understanding (Brosch et al., 2013). Moreover, emotions can act as a motivational mechanism that enhances users' attention, interest, and motivation, thereby promoting their learning and cognition (Tyng et al., 2017). In the context of human-computer interaction (HCI), emotions can serve as a valuable feedback mechanism that increases user satisfaction, engagement, and loyalty (Brave and Nass, 2007; Jeon, 2017).

As a crucial and fundamental research area of affective computing and neuroscience, emotion recognition has attracted great attention from the academy and enterprise fields in recent years (Cambria et al., 2017; Torres et al., 2020). Emotion recognition technology can generally be categorized into two major categories (Shu et al., 2018). The first category involves non-physiological signals such as facial expressions, speech, gestures, and posture (Schuller et al., 2003; Anderson and McOwan, 2006; Castellano et al., 2008).

The second category is based on physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), electromyography (EMG), skin temperature (SKT), and others (Egger et al., 2019). Physiological-based emotion recognition is considered more reliable as it is difficult for individuals to deliberately control their physiological signals. Among the physiological signals, EEG signals are widely used in neural engineering and brain-computer interfaces (BCIs) research due to their high temporal resolution, non-invasiveness, and low cost (Craik et al., 2019). Emotional states are closely related to neural activity produced by the central nervous system (Torres et al., 2020). This neural activity can be directly measured using EEG devices, making EEG-based emotion recognition increasingly popular in various fields, such as education, health, entertainment (Xu et al., 2018; Suhaimi et al., 2020; Abdel-Hamid, 2023; Moontaha et al., 2023).

A major problem with recognizing emotions is that emotions should be defined and accessed quantitatively. There are two different models used to define emotions: the discrete model and the dimensional model (Shu et al., 2018). According to the discrete model, emotions are divided into several basic categories, such as sadness, fear, disgust, surprise, happiness, and anger. These emotions can form more complex emotion categories through a certain combination of patterns (Peter and Herbon, 2006; Van den Broek, 2013). The dimensional emotion model maps emotional states into the points on a certain coordinate system. Different emotional states are distributed in different positions in the coordinate system, and the distance between positions reflects the difference between different emotional states (Wioleta, 2013; Poria et al., 2017; He et al., 2020). Different from discrete emotion models, the dimensional emotion model is continuous and has the advantages of a wide range of emotions and the ability to describe the evolution of emotions.

In recent decades, EEG-based emotion recognition has attracted much attention from researchers (Jerritta et al., 2011). A typical recognition process of emotional EEG usually consists of two parts: EEG feature extraction and emotion classification (Alarcao and Fonseca, 2017). EEG is a highly dynamic and nonlinear signal with a large amount and redundancy of data. Thus, feature extraction is an important step in emotion evaluation because high-resolution features are essential for effective pattern recognition (He et al., 2020). EEG features can be mainly divided into time-domain features, frequency-domain features, and time-frequency features (Jenke et al., 2014; Stancin et al., 2021; Huang et al., 2022). One of the widely used methods of frequency domain feature analysis of EEG signals is to decompose EEG signals into several frequency bands, including delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (>31 Hz) (Aftanas et al., 2004; Davidson, 2004; Li and Lu, 2009). EEG features can be extracted from each band. The common time domain features include statistical features (Liu and Sourina, 2014) and Hjorth features (Hjorth, 1970). Commonly used frequency domain features include power spectral density (PSD) (Thammasan et al., 2016), differential entropy (DE) (Shi et al., 2013), and rational asymmetry (RASM) (Zheng et al., 2017). The common time-frequency domain features include wavelet features (Akin, 2002), short-time Fourier transform (Kıymık et al., 2005), and Hilbert-Huang transform (Hadjidimitriou and Hadjileontiadis, 2012).

One of the most successful methods for recognizing emotions based on EEG signals is Deep Neural Networks (DNN) (Zhang et al., 2020; Ozdemir et al., 2021). The Convolutional Neural Networks (CNN) method was proven to be a powerful tool to model structured data in many applications, ranging from image classification and video processing to speech recognition and natural language understanding (Gu et al., 2018). However, EEG signals can be considered non-Euclidean data in order to extract the relationship of different brain regions, and Convolutional Neural Networks (CNN) may not be effective in capturing the hidden patterns of non-Euclidean data (Micheloyannis et al., 2006). In recent years, Graph Neural Networks (GNNs) have been developed rapidly and offer a potential solution to extract correlation features among EEG channels in emotion recognition tasks (Wu et al., 2020). In the graph representation of emotional EEG signals, each EEG channel corresponds to a vertex node, and the connections between vertex nodes correspond to edges in the graph, making it suitable for encoding the correlation among the brain regions in the multichannel EEG signal (Jia et al., 2020). However, constructing a better graph representation of EEG signals for emotion recognition problems remains challenging as the spatial position, which must be predetermined before building the EEG emotion recognition model, is different from the functional connections among EEG channels (Song et al., 2018).

To leverage both spatial relationships and time-frequency information, many researchers have extended graph neural networks by spatial-temporal attention. Spatial temporal attention is a mechanism that captures the dynamic relationship between spatial and temporal dimensions in data. It consists of two kinds of attention: spatial attention, which focuses on the relevant regions or nodes in space; and temporal attention, which focuses on the time steps in time dimension. Sartipi et al. proposed the novel spatial-temporal attention neural network (STANN) to extract discriminative spatial and temporal features of EEG signals by a parallel structure of the multi-column convolutional neural network and attention-based bidirectional long-short term memory (Sartipi et al., 2021). Li X. et al. (2021) proposed a model called attention-based spatial=temporal graphic long short-term memory (ASTG-LSTM), in which a specific spatial-temporal attention embedded into the model to improve the invariance ability against the emotional intensity fluctuation. Liu et al. (2022) proposed a spatial-temporal attention to explore the relationship between emotion and spatial-temporal EEG features. Therefore, it is reasonable to consider incorporating spatial-temporal attention to improve classification accuracy.

In this paper, we propose a novel model, STGATE, which combines a transformer learning block (TLB) and a Spatial-temporal Graph Attention (STGAT) mechanism. TLB utilizes 2D convolutional layers and a transformer encoder to extract time-frequency information, while the STGAT utilizes both spatial and temporal attention mechanisms to learn connections between brain regions and temporal information, respectively. Our approach treats EEG signals as graph data and incorporates them into graph neural networks to capture correlations between EEG channels. Unlike the GNN methods, the adjacency matrix learned

by STGATE can provide a better graph representation because it is adaptively updated by spatial attention during the training process. The main contributions of this paper can be summarized as follows:

- This paper proposes a novel spatial-temporal graph attention network with a transformer encoder (termed STGATE) for EEG-based emotion recognition.
- STGATE utilizes a transformer learning block and spatial-temporal graph attention. This allows it to capture electrode-level time-frequency representations. It also helps STGATE learn the emotional brain activities within and among different brain functional areas.
- STGATE achieved state-of-the-art performance with a cross-subject accuracy of 90.37% in SEED, 76.43% in SEED-IV, and 77.44% and 75.26% in the valence and arousal dimensions of the DREAMER dataset, respectively. Extensive ablation studies and analysis experiments were conducted to validate the efficiency of the proposed STGATE.

Remainder of this paper is organized as follows. The proposed STGATE method is presented in Section 3. The datasets and experiment settings are presented in Section 4. In Section 5, numerical emotion recognition experiments on the SEED, SEED-IV, and DREAMER datasets are carried out. In addition, the performance of the current methods and the proposed methods are presented and compared. Some discussions and analyzes of the proposed model are presented in Section 5. The conclusions of this paper are given in Section 6.

2. Related work

2.1. Emotion recognition

Emotion recognition is crucial for research in affective computing and neuroscience. Many studies on EEG-based emotion recognition focus on feature engineering or deep learning. Long short-term memory (LSTM) has been utilized to learn features from raw EEG signals and has achieved higher average accuracy than traditional techniques (Alhagry et al., 2017). A deep adaptation network has also been used to eliminate individual differences in EEG signals for effective model implementation (Li et al., 2018).

However, the inter-channel correlation of EEG signals for emotion recognition is critical. Song et al. (2018) proposed a novel dynamic graph convolutional neural network to dynamically learn the intrinsic relationship between different channels. To capture both local and global relations among different EEG channels, Zhong et al. proposed a regularized graph neural network (RGNN) for EEG-based emotion recognition, which models the inter-channel relations in EEG signals *via* an adjacency matrix (Zhong et al., 2020). A graph convolutional broad network was designed to explore the deeper-level information of graph-structured data and achieved high performance in EEG-based emotion recognition (Zhang et al., 2019). Li et al. proposed a Multi-Domain Adaptive Graph Convolutional Network (MD-AGCN), fusing the knowledge of both the frequency domain and the temporal domain to

fully utilize the complementary information of EEG signals (Li R. et al., 2021) designed a model called ST-GCLSTM, which utilizes spatial attention to modify adjacency matrices to adaptively learn the intrinsic connection among different EEG channels (Feng et al., 2022).

Various methods and classifiers have been proposed and applied to the problem of EEG-based emotion recognition. To improve the accuracy of emotion recognition, this paper proposes STGATE, a model that extracts time-frequency and spatial features from EEG signals.

2.2. Graph attention network

According to previous studies, graph convolutional neural networks are divided into spectral and spatial methods (Chen et al., 2020). The spectral method uses the convolution theorem to map the signal to the spectral space, which overcomes the non-Euclidean data missing translation invariance feature. The spatial method operates directly on the graph data and achieves the convolution effect by aggregating the information of neighboring nodes.

Graph attention networks (GATs) are a kind of network based on an attention mechanism to classify graph-structured data, which belongs to the spatial method of graph convolutional neural network (Veličković et al., 2017). The basic idea is to calculate the hidden representation of each graph node in the graph data by aggregating the information of neighboring points using the self-attention strategy and to define the information fusion using the attention mechanism function. Unlike other graph networks, GAT calculates the association weights by the feature representations of the nodes instead of calculating the weights based on the information of the edges. The input to a graph attention network is a series of feature vectors of nodes, which can be expressed as $H = \left\{ \vec{h}_1, \vec{h}_2, \ldots, \vec{h}_N \right\}$, $\vec{h}_i \in \mathbb{R}^{N \times F}$, where N is the number of vertices, and F represents feature dimensions. The graph attention network uses a self-attentive mechanism to compute the attention coefficients of the input feature vectors and normalize them as follows:
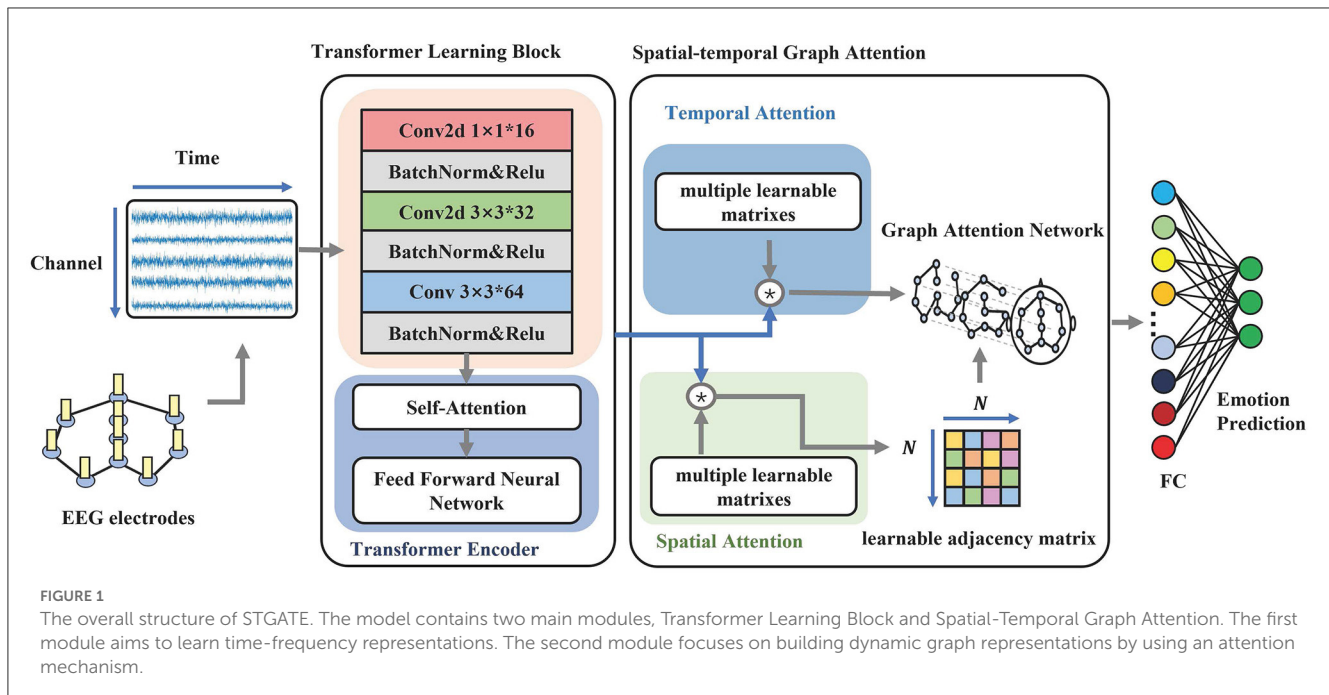
$$e_{ij} = a \left(W \vec{h}_i, W \vec{h}_j \right) \quad (1)$$

$$\alpha_{ij} = \text{Softmax}_j \left(e_{ij} \right) = \frac{\exp \left(e_{ij} \right)}{\sum_{k \in \mathcal{N}_i} \exp \left(e_{ik} \right)} \quad (2)$$

where e_{ij} represents attention weights between node i and node j, and a_{ij} is the normalized attention weight, indicating the importance of node i to node j, $\vec{h}$ is the eigenvector; W is the weight matrix in (1, 2). The attention weights and expressions can be represented as follows:

$$\alpha_{ij} = \frac{\exp \left(\text{LeakyReLU} \left(\vec{\mathbf{a}}^T \left[W\vec{h}_i \| W\vec{h}_j \right] \right) \right)}{\sum_{k \in \mathcal{N}_i} \exp \left(\text{LeakyReLU} \left(\vec{\mathbf{a}}^T \left[W\vec{h}_i \| W\vec{h}_k \right] \right) \right)} \quad (3)$$

where $\|$ denotes the concatenation operation, N_i denotes the set of neighboring nodes of the i th node, $\vec{\mathbf{a}}^T$ represents the

FIGURE 1
The overall structure of STGATE. The model contains two main modules, Transformer Learning Block and Spatial-Temporal Graph Attention. The first module aims to learn time-frequency representations. The second module focuses on building dynamic graph representations by using an attention mechanism.

transpose of the attention weight vector, and LeakyReLU denotes the nonlinear activation function in Equation (3). To make the network more informative, the graph attention network uses a multi-head mechanism that makes each head capture different information. The information from multiple heads is fused through a linear layer, and the attention coefficients are combined with the corresponding feature vectors to compute the final output features of each node.

$$\vec{h}_i' = \sigma \left(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right) \tag{4}$$

where W is the weight matrix of the linear layer, σ is the nonlinear activation function, and $\vec{h}_i'$ is the final output vector of the graph attention network in Equation (4).

The graph attention network assigns different weights to the nodes (EEG channels) through the attention mechanism, which effectively improves the representational capability of the network. At the same time, the graph attention network operates very efficiently with a computational complexity of $O\left(|V|FF' + |E|F'\right)$, where F is the dimension of the input vector, $|V|$ is the number of nodes, and $|E|$ is the number of edges.

3. Methods

As shown in Figure 1, the architecture for our proposed model consists of two modules. The upper module is an electrode-level learning block for extracting time-frequency information. The bottom module is a dynamic graph convolution for the correlation of EEG channels by constructing the adjacency matrix during the training and testing processes.

3.1. Transformer learning block

Figure 1 illustrates the transformer learning block (TLB), which aims to learn electrode-level time-frequency representations from EEG signals. TLB comprises two main components. The first component is a stack of multi-kernel convolutions that downsample five frequency bands and extract multi-scale features. Previous studies have reported that network connections in the high gamma band are denser among different emotional states, such as happiness, neutrality, and sadness, compared to other frequency bands (Yang et al., 2020). Similarly, Newson and Thiagarajan (2019) found that emotional disorders are more related to higher frequencies, including the alpha, beta, and gamma bands. Therefore, emotional states are more relevant to the alpha (8–13 Hz), beta (14–30 Hz), and gamma (>31 Hz) frequency bands (Ding et al., 2020). To better make use of all informative frequency bands, the kernels of the convolution layers are set to 1, 3, and 3. The 3*3 convolution layers are adopted followed by the 1*1 convolution layer, which aims to add network nonlinear features. Transformer-based methods have achieved great success in many areas (Raganato and Tiedemann, 2018; Liu et al., 2021). The multi-head self-attention and parallel inputting have superior abilities to capture long-range dependencies. The positional embedding learns the positional information of the sequence. To enhance the long-range dependencies captured in EEG, the second component uses a transformer encoder to map the EEG sequence to a new encoded sequence that contains more temporal information to enhance the long-range dependency capturing ability in EEG.

3.2. Graph representation

Usually, EEG signals are measured by placing the electrodes on the corresponding locations in the human brain scalp, and

the brain electrodes measure voltage changes generated by neural activity in the cerebral cortex (Subha et al., 2010). The distribution position of the brain electrodes is defined by some standards, such as the international 10/20 system. The distribution position of the brain electrodes is fixed and regular, so the EEG signal channels can be considered classical non-Euclidean structured data (Micheloyannis et al., 2006), which are well suited to be represented by graphical data.

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \tag{5}$$

$$\mathcal{V} = \{v_i \mid i = 1, \dots, N\} \tag{6}$$

$$\mathcal{E} = \{e_{ij} \mid v_i, v_j \in \mathcal{V}\} \tag{7}$$

$$A = \{a_{ij}\} \tag{8}$$

A segment of EEG signals collected by a brain electrode can be considered as a node of the graph. Therefore, we regard multi-channel EEG signals as a graph. $\mathcal{G}$ denotes a graph, $\mathcal{V}$ denotes the set of vertices in graph $\mathcal{G}$, and $\mathcal{E}$ represents the set of edges in Equations (5–7). N is the number of brain electrodes in Equation (6). In the graph representation of EEG signals, a node v_i is usually used to represent an EEG electrode, while an edge e_{ij} represents the correlation between nodes v_i and v_j. A is the adjacency matrix of graph $\mathcal{G}$. a_{ij} represents the strength of the correlation of nodes v_i and v_j in Equation (8). The adjacency matrix A is a learnable matrix and can be dynamically modified during the training process. Generally, we model the EEG signal as an undirected graph and use this undirected graph as the input to the adaptive graph module. The initial set of edges of the undirected graph obtained from the above modeling is determined by the kNN algorithm, which computes graph edges to the nearest k points.

3.3. Spatial–temporal graph attention

Neural activity in different brain regions has an intrinsic correlation during the emotional experience. EEG signals recorded by brain electrodes can also reflect some intrinsic correlation in different brain regions. Therefore, we proposed spatial-temporal graph attention (STGAT) to capture correlations between EEG electrodes in the spatial domain and temporal EEG information in the temporal domain. Specifically, STGAT dynamically learns the adjacency matrix A through a spatial attention mechanism during the training process and uses temporal attention to further learn the temporal information in EEG.

Spatial attention can be implemented with the following formula:

$$S = V \cdot \sigma \left(W_1 X_h W_2 + b_s\right) \tag{9}$$

$$A = \frac{S - E[S]}{\sqrt{\mathrm{Var}[S]}} \tag{10}$$

where $S \in \mathbb{R}^{B \times N \times N}$ is a weight matrix, which represents the importance of edges. A represents the dynamic adjacency matrix. $X_h \in \mathbb{R}^{B \times N \times C \times T_r}$ is the input of the block. B is the batch size. N is the number of vertices of the input data. C represents the 2D convolution channels. T_r denotes the length of the temporal dimension. $V, b_s \in \mathbb{R}^{N \times N}$, $W_1 \in \mathbb{R}^{T_r}$, and $W_2 \in \mathbb{R}^{C \times N}$ are learnable parameters, and σ denotes the Tanh activation function. We adopt batch normalization to reduce internal covariate shifts and accelerate training (Santurkar et al., 2018). $E[\cdot]$ and $Var[\cdot]$ denote the mini-batch mean and mini-batch variance of S. The value of an element S_{ij} indicates the strength of the connection between node i and node j. We use the spatial attention matrix A as the adjacency matrix so that the adjacency matrix can be dynamically constructed by the corresponding input features. To obtain better representations of EEG signals, we adopt a Top-K algorithm to maintain the 10 edges with the highest weight and discard the others. The Top-K operation is applied as follows:

$$\begin{cases} \text{for } i = 1, 2, \dots, N \\ \text{index} = \mathrm{argtopk}(A[i, :]) \\ A[i, \overline{\text{index}}] = 0 \end{cases} \tag{11}$$

where the $argtopk(\cdot)$ is a function to obtain the index of the top-k largest values of each vector in Equation (11). The use of a dynamic adjacency matrix in EEG emotion recognition has contributed to the ability to dynamically learn the intrinsic relationship between different EEG channels, which can reflect the brain connectivity patterns associated with different emotional states. Moreover, the dynamic adjacency matrix can adapt to different subjects, thereby improving the cross-subject generalization ability of EEG emotion recognition models. By applying graph convolution on multichannel EEG features using the dynamic adjacency matrix, more discriminative features can be extracted for emotion classification.

Temporal attention is designed to dynamically capture the correlation between emotional EEG signals in the time domain. The temporal attention mechanism is defined as follows:

$$T = V \cdot \sigma \left(W_3 X_h W_4 + b_t\right) \tag{12}$$

$$\widehat{T} = \frac{T - E[T]}{\sqrt{\mathrm{Var}[T]}} \tag{13}$$

$W_3 \in \mathbb{R}^{T_r}$ and $W_4 \in \mathbb{R}^{C \times N}$ are learnable parameters, and σ denotes the tanh activation function. Having the temporal attention weight matrix, we tuned the input X_h by the temporal attention:

$$\widehat{X_h} = \widehat{T} X_h \tag{14}$$

We utilize temporal attention to focus on valuable temporal information in EEG-based emotion recognition. The purpose of time-domain attention is to uncover the temporal patterns in EEG signals and assign importance weights based on their intrinsic similarities. By combining spatial attention with temporal attention, the model can extract more discriminative features from EEG signals and enhance the accuracy of emotion recognition. We use $\widehat{X_h}$ as the input to the graph attention network and A as the adjacency matrix of the graph data.

4. Experiment

4.1. Datasets

The SEED dataset is an EEG-based dataset collected in the BCMI lab of Shanghai Jiao Tong University, known as the SJTU Emotion EEG Dataset (Zheng and Lu, 2015). The dataset contains a total of 62 channels of EEG signals from 15 subjects for 15 experiments. The researchers prepared 15 movie clips of approximately 4 min, which were divided into 3 categories: negative, neutral, and positive. Positive movies are comedies that stimulate positive emotions such as happiness; negative movies are tragic movies that stimulate negative emotions such as sadness, and neutral movies are world heritage documentaries that do not stimulate positive or negative emotions. The subjects were asked to watch these movie clips and were given 45 s to self-evaluate and calm down after each clip was shown.

The SEED-IV dataset is also from the BCMI lab (Zheng et al., 2018). This dataset features 168 movie clips that serve as a repository for four emotions (happy, sad, fearful, and neutral). Forty-four participants (22 females, all college students) were recruited to assess their emotions while watching the movie clips using keywords from four discrete emotions (happy, sad, neutral, and fearful) and rating 10 points (from −5 to 5) on two dimensions: valence and arousal.

The DREAMER dataset is a commonly used emotion recognition dataset (Katsigiannis and Ramzan, 2017). Researchers had subjects watch edited movie clips to elicit emotions from subjects and recorded EEG data using a 14-channel EEG acquisition device. These film clips consist of selected scenes from various movies that have been demonstrated to elicit a diverse array of emotions (Gabert-Quillen et al., 2015). After each movie clip was played, the researchers classified the emotions based on the subjects' ratings using 3 dimensions: potency, arousal, and dominance. The dataset contained 2 clips each of 9 emotion-evoking movies of happiness, excitement, bliss, calmness, anger, disgust, fear, sadness, and surprise, for a total of 18 movie clips.

4.2. Experiment settings

The STGATE model is implemented by PyTorch 1.10. The hyperparameters were tuned to obtain the best performance on the validation datasets.

The cross-subject experiments are conducted. Since there were multiple subjects, the leave-one-subject-out (LOSO) cross-validation strategy was applied in the experiments. The EEG data of one subject were used as the validation dataset, while the data of the other subjects were used as the training dataset. We repeatedly performed ten rounds of cross-validation experiments and the average accuracy and standard deviation of the test set are adopted as the performance criteria. The experiments in this paper use the Adam optimizer to accelerate the training process of the model with a batch size of 16 and a learning rate of 0.00001 (Kingma and Ba, 2014). Additionally, we use the Dropout algorithm to suppress the overfitting phenomenon of the model. The drop rate is set to 0.3.

During the training process, the training set is stopped when the training loss is lower than 0.15.

5. Results

5.1. Results of experiments

Tables 1, 2 summarize the experimental results in terms of the average EEG emotion recognition accuracies and standard deviations of the STGATE method. To validate the effectiveness of our proposed method for EEG emotion recognition, we compared it with various machine learning and deep learning methods. Conventional classifiers such as supported vector machine (SVM) (Zhong et al., 2020) and transductive SVM (T-SVM) (Collobert et al., 2006) can be applied for cross-subject emotion recognition problems. Domain adaptation methods such as Transfer Component Analysis (TCA) (Pan et al., 2010) can also handle cross-subject emotion recognition problems. Pretrained Convolutional Neural Network (CNN) architectures have also been used in emotion recognition tasks (Cimtay and Ekmekcioglu, 2020). Rahman et al. proposed a method

TABLE 1 Leave-one-subject-out emotion recognition (accuracy/standard deviation) on the SEED and SEED-IV datasets.

Model	SEED	SEED-IV
SVM (Zhong et al., 2020)	56.73/16.29	37.99/12.52
T-SVM (Collobert et al., 2006)	72.53/14.00	(-)/(-)
TCA (Pan et al., 2010)	63.64/14.88	56.56/13.77
CNN (Cimtay and Ekmekcioglu, 2020)	78.34/**6.11**	(-)/(-)
PCA+ANN (Rahman et al., 2020)	84.3/(-)	(-)/(-)
RODAN (Lew et al., 2020)	(-)/(-)	60.75/(-)
DAN (Li et al., 2018)	83.81/8.56	58.87/8.13
BiHDM (Li et al., 2020)	85.40/7.53	69.03/8.66
DGCNN (Song et al., 2018)	79.95/9.02	(-)/(-)
GCN (Kipf and Welling, 2016)	84.95/7.51	72.23/**4.01**
GAT (Veličković et al., 2017)	85.56/6.75	70.33/4.57
STGATE (Ours)	**90.37**/6.18	**76.43**/5.01

The bold values indicate the maximum value of the current column in tables.

TABLE 2 Leave-one-subject-out emotion recognition (accuracy/standard deviation) on the DREAMER dataset.

Model	Valence	Arousal	Average
SVM (Zheng and Lu, 2015)	56.57/(-)	58.91/(-)	57.74/(-)
DBN (Zheng and Lu, 2015)	56.43/(-)	58.94/(-)	57.685/(-)
DGCNN (Song et al., 2018)	59.64/(-)	62.91/(-)	61.28/(-)
ADDA-TCN (He et al., 2022)	66.56/10.04	63.69/**6.57**	65.13/**8.31**
GCN (Kipf and Welling, 2016)	76.43/10.13	74.01/10.76	75.22/10.45
GAT (Veličković et al., 2017)	75.70/9.64	**75.31**/10.71	75.51/10.18
STGATE (Ours)	**77.44**/8.40	75.26/9.71	**76.35**/8.40

The bold values indicate the maximum value of the current column in tables.

that hybridizes Principal Component Analysis (PCA) and t-statistics for feature extraction, and Artificial Neural Network (ANN) is applied for classification (Rahman et al., 2020). To deal with the domain shift problem between different subjects, a deep domain adaptation network (DAN) was proposed for cross-subject EEG signal recognition (Li et al., 2018). Likewise, to model asymmetric differences between two hemispheres of the EEG signal, a novel bi-hemispheric discrepancy model (BiHDM) was proposed for EEG emotion recognition (Li et al., 2020). He et al. explored the feasibility of combining Temporal Convolutional Networks (TCNs) and Adversarial Discriminative Domain Adaptation (ADDA) algorithms to solve the domain shift problem in EEG-based cross-subject emotion recognition (He et al., 2022). The Dynamical Graph Convolutional Neural Network (DGCNN) is a novel EEG-based emotion recognition model in which graph spectral convolution operation with dynamical adjacent matrix is applied (Song et al., 2018). Lew et al. propose a Regionally-Operated Domain Adversarial Network (RODAN) incorporate the attention mechanism to enable cross-domain learning to capture both spatial-temporal relationships among the EEG electrodes and an adversarial mechanism to reduce the domain shift in EEG signals (Lew et al., 2020).

Machine learning methods such as supported vector machine (SVM) and transfer component analysis (TCA) can be applied to address cross-subject emotion recognition problems. According to the experimental results in the SEED dataset and SEED-IV dataset, both machine learning methods, SVM, T-SVM and TCA, give lower accuracy than deep learning models. The performance of many deep learning methods, such as CNN, PCA+ANN, RODAN, DGCNN, DAN, and BiHDM, are better than that of the traditional machine learning methods (SVM and TCA), indicating that machine learning has difficulty obtaining valid features.

The proposed method achieves the highest accuracy in SEED, SEED-IV and DREAMER dataset. STGATE achieve 90.37% in SEED dataset and 76.43% in SEED-IV dataset. STGATE achieves 77.44% in the valence dimension, 75.26% in the arousal dimension, and 76.35% in the average value in both dimensions in DREAMER dataset, because the proposed STGATE can extract more useful information in the temporal and spatial dimensions. The proposed method treats EEG signals as non-Euclidean data and uses graph representations and attention mechanisms to extract the spatial and temporal characteristics of EEGs. STGATE compensates for the limitations of convolutional neural networks and can handle the feature extraction problem of non-Euclidean data with topological graph structure. The combination of the transformer encoder and STGAT enhances the performance of the network. The modeled graph representations restore the spatial and temporal connectivity of the data and make STGATE extract more discriminative emotional features that can be used to accurately classify and identify the emotional states of subjects. The proposed STGATE extracts electrode-level information through TLB and spatial features based on an adaptive graph structure. Therefore, we can see that our proposed method achieves the best accuracy results on the SEED, SEED-IV, and DREAMER datasets.

5.2. Ablation study

To verify the effectiveness of each module in STGATE, we removed them one at a time or replaced some of the layers and evaluated the performance of the ablated model. As shown in Table 3, we trained several models to verify the impact of the modules in STGATE. For the baseline model, we only deployed a series of multi-kernel 2D convolutions during training and testing. "TGAT" refers to using spatial-temporal graph attention without spatial attention and is similar to "SGAT." The TLB was utilized to attentively fuse the node-level features, and the convolutional layers downsampled the multi-channel EEG input and learned the time-frequency representations. The transformer encoder enhanced the long-range dependency capturing ability. According to the results shown in Table 3, removing the TLB module caused the accuracy to drop from 90.27% to 87.88%, a decrease of 2.39%. When we removed the transformer encoder in the TLB, the accuracy dropped from 86.2% to 83.86%, a decrease of 2.36%. The results show the effectiveness of the TLB module.

The STGAT module aimed to learn the dynamic spatial-temporal representations of the graph. Spatial attention built an adaptive adjacency matrix through several learnable parameters, making the graph structure dynamically change during the training process. The dynamic adjacency matrix had the potential to extract informative correlations among electrodes. Temporal attention was similar to spatial attention and captured temporal information through several learnable parameters. As shown in Table 3, removing the STGAT module caused the accuracy to drop from 90.27% to 86.22%, a decrease of 4.05%. When we removed the spatial and temporal attention in STGAT, the accuracy dropped from 87.88% to 86.84%, decreasing by 1.04%, and 87.88% to 86.67%, decreasing by 1.21%, respectively.

The performance of the STGATE model outperforms that of other models by a significant margin. This is attributed to the ability of the spatial and temporal attention modules to capture potential EEG signal features, while the transformer encoder helps to enhance long-range dependency capturing ability. The models with STGAT and TLB modules significantly outperform those without these modules. The transformer learning block aggregates time-frequency features using convolution and a transformer

TABLE 3 Performance of our proposed modules on the SEED and SEED-IV datasets.

| Model	SEED	SEED-IV
Baseline	83.86/9.88	71.17/4.48
TLB	86.22/8.57	73.98/4.12
SGAT	86.67/6.98	75.39/5.49
SGAT+TLB	88.56/5.91	75.85/5.06
TGAT	86.84/7.40	76.04/4.76
TGAT+TLB	89.30/**5.69**	76.28/**4.06**
SGAT+TGAT (STGAT)	87.88/7.38	75.52/4.30
SGAT+TGAT+TLB (STGATE)	**90.37**/6.18	**76.43**/5.01

The bold values indicate the maximum value of the current column in tables.

encoder, while the spatial-temporal graph attention captures inter-channel connections *via* an adaptive adjacency matrix and temporal information using temporal attention. Therefore, both the transformer learning block and the dynamic graph convolution are essential components of the STGATE model.

5.3. The impact of feature selection

Raw emotional EEG is a non-linear random signal with a large amount of data redundancy and a low signal-to-noise

ratio (Balasubramanian et al., 2018). The EEG signal features, such as power spectral density (PSD) and differential entropy, are more representative of the prominent features of the EEG signal in certain aspects. Therefore, the emotion classification task generally uses the feature of the EEG signal for classification. Therefore, we study the impact of the feature selection of STGATE on classification performance. Figure 2 shows the emotion recognition accuracies and standard deviations of the proposed STGATE model with different features, including DE, PSD, ASM, DASM, and RASM features, in the SEED dataset. The DE feature is the most discriminated feature, while the performance of other features is much lower. The DE feature obtained the highest classification accuracy (90.37%) and lowest standard deviation, followed by PSD (83.17%). RASM (rational asymmetry), DASM (differential asymmetry), and ASM (asymmetry) are calculated from DE features designed to express asymmetry (Shi et al., 2013). The average accuracies of the RASM, DASM, and ASM features are close to each other, 78.07%, 75.76%, and 77.75%, respectively. The result implies that the DE feature is more suitable for EEG emotion recognition than the traditional feature.
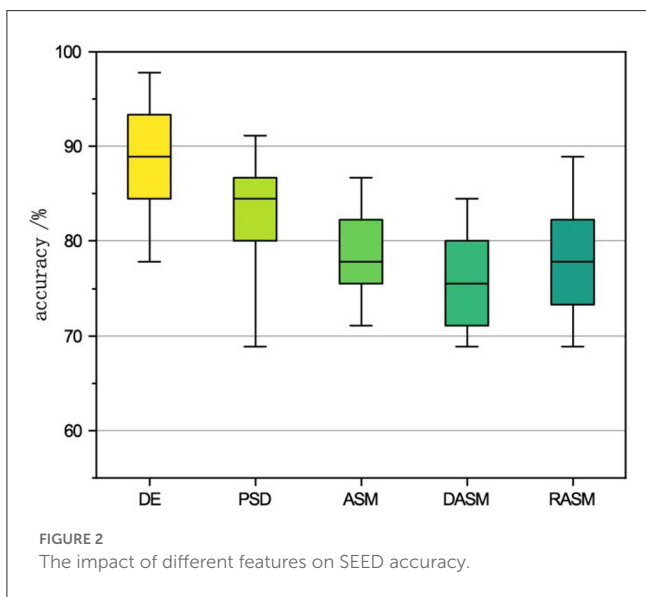
5.4. Visualization

As shown in Figure 3, the topographic map is utilized to analyze the inter-channel connections of the learned graphs for emotion recognition in the SEED dataset. The adjacent matrices are extracted at the end of training and transformed into a topographic map. To better show which part of the connections is more informative, we extracted the adjacency matrices of the EEG samples of all the subjects, averaged all
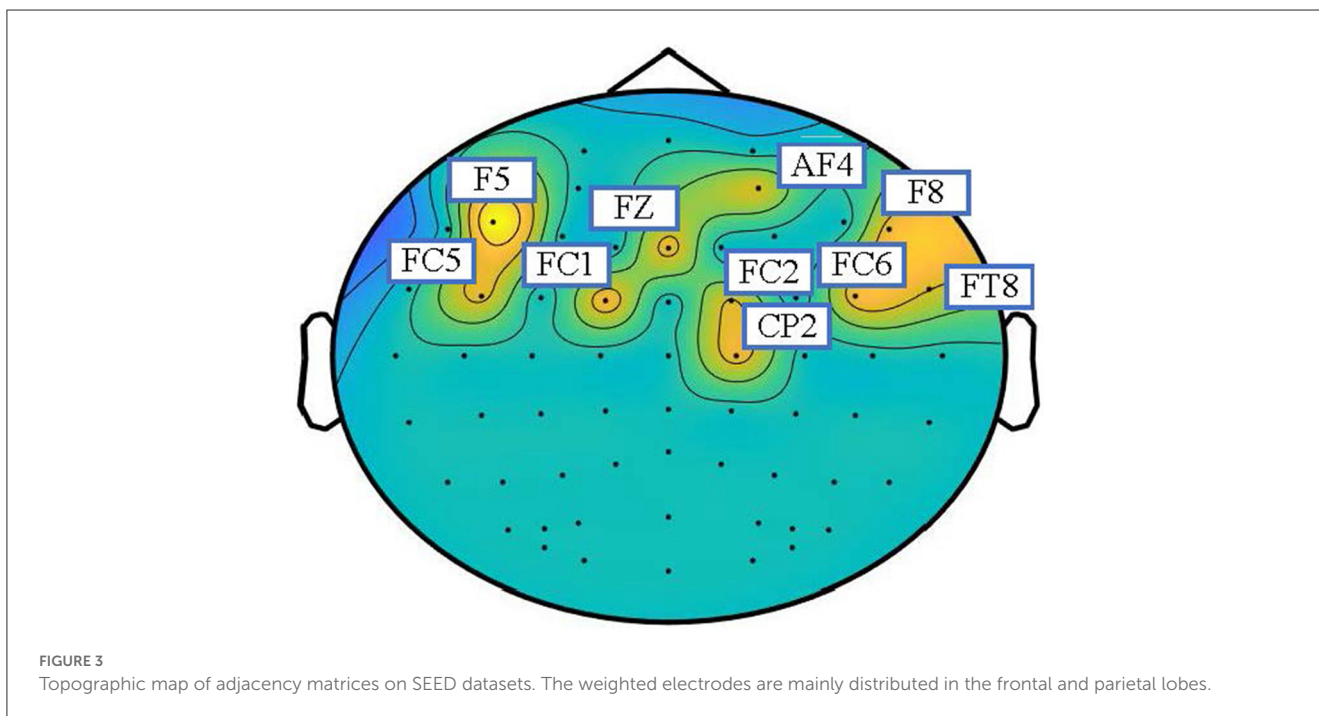
the matrices, took the largest ten values, and set the others to zero. The topographic map of the adjacency matrices is shown in Figure 3. According to the topographic map, the frontal lobe plays an important role in the classification of emotions. The F5, FC5, FC1, FZ, F8, AF8, CP2, FC6, FT8, and C2 channels have more weight than other channels, which means that these channels provide more information during the training process. According to previous studies, the pre-frontal, parietal and occipital channels may be the most associated with emotions (Zheng and Lu, 2015; Zhong et al., 2020; Ding et al., 2021). The visualization results basically coincide with the observations in neuroscience. Therefore, the topographic map indicates that the dynamic adjacency matrix gives more weight to the emotionally relevant EEG channels to enhance the potential ability of STGATE.

6. Conclusion

In this paper, we proposed STGATE, a novel method for EEG-based emotion recognition that can dynamically learn the inter-channel relationships of EEG emotion signals. The STGATE is composed of two modules, TLB and STGAT. The TLB module employs 2D convolutions and a transformer encoder to downsample EEG signals and capture long-range information. The STGAT module dynamically captures correlations between EEG electrodes in the spatial domain and temporal EEG information using a time-spatial attention mechanism. The experimental results demonstrate that STGATE achieves higher classification accuracies compared to existing methods for cross-subject EEG-based emotion recognition. However, a limitation of this study is the small sample size of the publicly available dataset used in the article and the lack of sufficient reliable data within the dataset. Nevertheless, our proposed method has the potential to inspire new methodologies for emotion recognition and affective computing.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JL proposed the idea and wrote the manuscript. WP conducted the experiments. HH and JP provided advice on the research approaches, signal processing, and checked and revised the manuscript. FW offered important help on guided the experiments and analysis methods. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdel-Hamid, L. (2023). An efficient machine learning-based emotional valence recognition approach towards wearable eeg. *Sensors* 23, 1255. doi: 10.3390/s23031255

Aftanas, L. I., Reva, N. V., Varlamov, A. A., Pavlov, S. V., and Makhnev, V. P. (2004). Analysis of evoked eeg synchronization and desynchronization in conditions of emotional activation in humans: temporal and topographic characteristics. *Neurosci. Behav. Physiol.* 34, 859–867. doi: 10.1023/B:NEAB.0000038139.39812.eb

Akin, M. (2002). Comparison of wavelet transform and fft methods in the analysis of eeg signals. *J. Med. Syst.* 26, 241–247. doi: 10.1023/A:1015075101937

Alarcao, S. M., and Fonseca, M. J. (2017). Emotions recognition using eeg signals: a survey. *IEEE Trans. Affect. Comput.* 10, 374–393. doi: 10.1109/TAFFC.2017.2714671

Alhagry, S., Fahmy, A. A., and El-Khoribi, R. A. (2017). Emotion recognition based on eeg using lstm recurrent neural network. *Emotion* 8, 355–358. doi: 10.14569/IJACSA.2017.081046

Anderson, K., and McOwan, P. W. (2006). A real-time automated system for the recognition of human facial expressions. *IEEE Trans. Syst. Man Cybern. B* 36, 96–105. doi: 10.1109/TSMCB.2005.854502

Balasubramanian, G., Kanagasabai, A., Mohan, J., and Seshadri, N. G. (2018). Music induced emotion using wavelet packet decomposition–an EEG study. *Biomed. Signal Process. Control.* 42:115–128. doi: 10.1016/j.bspc.2018.01.015

Brave, S., and Nass, C. (2007). "Emotion in human-computer interaction," in *The Human-Computer Interaction Handbook* (CRC Press), 103–118.

Brosch, T., Scherer, K., Grandjean, D., and Sander, D. (2013). The impact of emotion on perception, attention, memory, and decision-making. *Swiss Med. Wkly.* 143, w13786-w13786. doi: 10.4414/smw.2013.13786

Cambria, E., Das, D., Bandyopadhyay, S., and Feraco, A. (2017). *Affective Computing and Sentiment Analysis*. Springer.

Castellano, G., Kessous, L., and Caridakis, G. (2008). *Emotion Recognition Through Multiple Modalities: Face, Body Gesture, Speech*. Springer.

Chen, Z., Chen, F., Zhang, L., Ji, T., Fu, K., Zhao, L., et al. (2020). Bridging the gap between spatial and spectral domains: A survey on graph neural networks. *arXiv preprint* arXiv:2002.11867. doi: 10.48550/arXiv.2002.11867

Cimtay, Y., and Ekmekcioglu, E. (2020). Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition. *Sensors* 20, 2034. doi: 10.3390/s20072034

Collobert, R., Sinz, F., Weston, J., Bottou, L., and Joachims, T. (2006). Large scale transductive svms. *J. Mach. Learn. Res.* 7, 1687–1712. doi: 10.1016/j.neucom.2017.01.012

Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.* 16, 031001. doi: 10.1088/1741-2552/ab0ab5

Davidson, R. J. (2004). What does the prefrontal cortex "do" in affect: perspectives on frontal eeg asymmetry research. *Biol. Psychol.* 67, 219–234. doi: 10.1016/j.biopsycho.2004.03.008

Ding, Y., Robinson, N., Zeng, Q., Chen, D., Wai, A. A. P., Lee, T.-S., et al. (2020). "Tsception: a deep learning framework for emotion detection using EEG," in *2020 International Joint Conference on Neural Networks (IJCNN)* (Glasgow, UK: IEEE), 1–7.

Ding, Y., Robinson, N., Zhang, S., Zeng, Q., and Guan, C. (2021). Tsception: capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *arXiv preprint* arXiv:2104.02935. doi: 10.1109/TAFFC.2022.3169001

Egger, M., Ley, M., and Hanke, S. (2019). Emotion recognition from physiological signal analysis: a review. *Electron. Notes Theor. Comput. Sci.* 343, 35–55. doi: 10.1016/j.entcs.2019.04.009

Feng, L., Cheng, C., Zhao, M., Deng, H., and Zhang, Y. (2022). Eeg-based emotion recognition using spatial-temporal graph convolutional lstm with attention mechanism. *IEEE J. Biomed. Health Inform.* 26, 5406–5417. doi: 10.1109/JBHI.2022.3198688

Gabert-Quillen, C. A., Bartolini, E. E., Abravanel, B. T., and Sanislow, C. A. (2015). Ratings for emotion film clips. *Behav. Res. Methods* 47, 773–787. doi: 10.3758/s13428-014-0500-0

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognit.* 77, 354–377. doi: 10.1016/j.patcog.2017.10.013

Hadjidimitriou, S. K., and Hadjileontiadis, L. J. (2012). Toward an eeg-based recognition of music liking using time-frequency analysis. *IEEE Trans. Biomed. Eng.* 59, 3498–3510. doi: 10.1109/TBME.2012.2217495

He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., et al. (2020). Advances in multimodal emotion recognition based on brain-computer interfaces. *Brain Sci.* 10, 687. doi: 10.3390/brainsci10100687

He, Z., Zhong, Y., and Pan, J. (2022). "Joint temporal convolutional networks and adversarial discriminative domain adaptation for EEG-based cross-subject emotion recognition," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Singapore: IEEE), 3214–3218.

Hjorth, B. (1970). Eeg analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 306–310. doi: 10.1016/0013-4694(70)90143-4

Huang, E., Zheng, X., Fang, Y., and Zhang, Z. (2022). Classification of motor imagery eeg based on time-domain and frequency-domain dual-stream convolutional neural network. *IRBM* 43, 107–113. doi: 10.1016/j.irbm.2021.04.004

Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/TAFFC.2014.2339834

Jeon, M. (2017). "Emotions and affect in human factors and human-computer interaction: taxonomy, theories, approaches, and methods," in *Emotions and Affect in Human Factors and Human-Computer Interaction*, 3–26.

Jerritta, S., Murugappan, M., Nagarajan, R., and Wan, K. (2011). "Physiological signals based human emotion recognition: a review," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications* (Penang: IEEE), 410–415.

Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., et al. (2020). "Graphsleepnet: adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *IJCAI*, 1324–1330.

Katsigiannis, S., and Ramzan, N. (2017). Dreamer: a database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE J. Biomed. Health Inform.* 22, 98–107. doi: 10.1109/JBHI.2017.2688239

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint* arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980

Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint* arXiv:1609.02907. doi: 10.48550/arXiv.1609.02907

Kıymık, M. K., Güler, N., Dizibüyük, A., and Akın, M. (2005). Comparison of stft and wavelet transform methods in determining epileptic seizure activity in eeg signals for real-time application. *Comput. Biol. Med.* 35, 603–616. doi: 10.1016/j.compbiomed.2004.05.001

Lew, W.-C. L., Wang, D., Shylouskaya, K., Zhang, Z., Lim, J.-H., Ang, K. K., et al. (2020). "EEG-based emotion recognition using spatial-temporal representation via bi-gru," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (Montreal, QC: IEEE), 116–119.

Li, H., Jin, Y.-M., Zheng, W.-L., and Lu, B.-L. (2018). "Cross-subject emotion recognition using deep adaptation networks," in *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap* (Cambodia: Springer), 403–413.

Li, M., and Lu, B.-L. (2009). "Emotion classification based on gamma-band EEG," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Minneapolis, MN: IEEE), 1223–1226.

Li, R., Wang, Y., and Lu, B.-L. (2021). "A multi-domain adaptive graph convolutional network for eeg-based emotion recognition," in *Proceedings of the 29th ACM International Conference on Multimedia*, 5565–5573.

Li, X., Zheng, W., Zong, Y., Chang, H., and Lu, C. (2021). "Attention-based spatio-temporal graphic lstm for eeg emotion recognition," in *2021 International Joint Conference on Neural Networks (IJCNN)* (Shenzhen: IEEE), 1–8.

Li, Y., Wang, L., Zheng, W., Zong, Y., Qi, L., Cui, Z., et al. (2020). A novel bi-hemispheric discrepancy model for EEG emotion recognition. *IEEE Trans. Cogn. Dev. Syst.* 13, 354–367. doi: 10.1109/TCDS.2020.2999337

Liu, A. T., Li, S.-W., and Lee, H.-y. (2021). Tera: Self-supervised learning of transformer encoder representation for speech. *IEEE/ACM Trans. Audio Speech Lang. Process.* 29:2351–2366. doi: 10.1109/TASLP.2021.3095662

Liu, J., Wu, H., Zhang, L., and Zhao, Y. (2022). "Spatial-temporal transformers for eeg emotion recognition," in *2022 The 6th International Conference on Advances in Artificial Intelligence*, 116–120.

Liu, Y., and Sourina, O. (2014). *Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm*. Springer.

Micheloyannis, S., Pachou, E., Stam, C. J., Vourkas, M., Erimaki, S., and Tsirka, V. (2006). Using graph theoretical analysis of multi channel EEG to evaluate the neural efficiency hypothesis. *Neurosci. Lett.* 402, 273–277. doi: 10.1016/j.neulet.2006.04.006

Moontaha, S., Schumann, F. E. F., and Arnrich, B. (2023). Online learning for wearable eeg-based emotion classification. *Sensors* 23, 2387. doi: 10.20944/preprints202301.0156.v1

Newson, J. J., and Thiagarajan, T. C. (2019). Eeg frequency bands in psychiatric disorders: a review of resting state studies. *Front. Hum. Neurosci.* 12, 521. doi: 10.3389/fnhum.2018.00521

Ozdemir, M. A., Degirmenci, M., Izci, E., and Akan, A. (2021). Eeg-based emotion recognition with deep convolutional neural networks. *Biomed. Eng.* 66, 43–57. doi: 10.1515/bmt-2019-0306

Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q. (2010). Domain adaptation via transfer component analysis. *IEEE Trans. Neural Networks* 22, 199–210. doi: 10.1109/TNN.2010.2091281

Peter, C., and Herbon, A. (2006). Emotion representation and physiology assignments in digital systems. *Interact Comput.* 18, 139–170. doi: 10.1016/j.intcom.2005.10.006

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: from unimodal analysis to multimodal fusion. *Inf. Fusion* 37:98–125. doi: 10.1016/j.inffus.2017.02.003

Raganato, A., and Tiedemann, J. (2018). "An analysis of encoder representations in transformer-based machine translation," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. The Association for Computational Linguistics.

Rahman, M. A., Hossain, M. F., Hossain, M., and Ahmmed, R. (2020). Employing pca and t-statistical approach for feature extraction and classification of emotion from multichannel eeg signal. *Egyptian Inform. J.* 21, 23–35. doi: 10.1016/j.eij.2019.10.002

Santurkar, S., Tsipras, D., Ilyas, A., and Madry, A. (2018). "How does batch normalization help optimization?" in *Advances in Neural Information Processing Systems, Vol. 31*.

Sartipi, S., Torkamani-Azar, M., and Cetin, M. (2021). "EEG emotion recognition via graph-based spatio-temporal attention neural networks," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Mexico: IEEE), 571–574.

Schuller, B., Rigoll, G., and Lang, M. (2003). "Hidden markov model-based speech emotion recognition," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03), volume 2* (Baltimore, MD: IEEE), II-1.

Shi, L.-C., Jiao, Y.-Y., and Lu, B.-L. (2013). "Differential entropy feature for EEG-based vigilance estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka: IEEE), 6627–6630.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., et al. (2018). A review of emotion recognition using physiological signals. *Sensors* 18, 2074. doi: 10.3390/s18072074

Song, T., Zheng, W., Song, P., and Cui, Z. (2018). Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 11, 532–541. doi: 10.1109/TAFFC.2018.2817622

Stancin, I., Cifrek, M., and Jovic, A. (2021). A review of eeg signal features and their application in driver drowsiness detection systems. *Sensors* 21, 3786. doi: 10.3390/s21113786

Subha, D. P., Joseph, P. K., Acharya, R., and Lim, C. M. (2010). EEG signal analysis: a survey. *J. Med. Syst.* 34, 195–212. doi: 10.1007/s10916-008-9231-z

Suhaimi, N. S., Mountstephens, J., Teo, J., et al. (2020). EEG-based emotion recognition: a state-of-the-art review of current trends and opportunities. *Comput. Intell. Neurosci.* 2020, 8875426. doi: 10.1155/2020/8875426

Thammasan, N., Moriyama, K., Fukui, K.-i., and Numao, M. (2016). Continuous music-emotion recognition based on electroencephalogram. *IEICE Trans. Inf. Syst.* 99, 1234–1241. doi: 10.1587/transinf.2015EDP7251

Torres, E. P., Torres, E. A., Hernández-Álvarez, M., and Yoo, S. G. (2020). EEG-based bci emotion recognition: a survey. *Sensors* 20, 5083. doi: 10.3390/s20185083

Tyng, C. M., Amin, H. U., Saad, M. N., and Malik, A. S. (2017). The influences of emotion on learning and memory. *Front. Psychol.* 8, 1454. doi: 10.3389/fpsyg.2017.01454

Van den Broek, E. L. (2013). Ubiquitous emotion-aware computing. *Pers. Ubiquit. Comput.* 17, 53–67. doi: 10.1007/s00779-011-0479-9

Veličkovi,ć, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv preprint* arXiv:1710.10903. doi: 10.48550/arXiv.1710.10903

Wioleta, S. (2013). "Using physiological signals for emotion recognition," in *2013 6th International Conference on Human System Interactions (HSI)* (Sopot: IEEE), 556–561.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 4–24. doi: 10.1109/TNNLS.2020.2978386

Xu, T., Zhou, Y., Wang, Z., and Peng, Y. (2018). Learning emotions eeg-based recognition and brain activity: a survey study on bci for intelligent tutoring system. *Procedia Comput. Sci.* 130, 376–382. doi: 10.1016/j.procs.2018.04.056

Yang, K., Tong, L., Shu, J., Zhuang, N., Yan, B., and Zeng, Y. (2020). High gamma band eeg closely related to emotion: evidence from functional network. *Front. Hum. Neurosci.* 14, 89. doi: 10.3389/fnhum.2020.00089

Zhang, T., Wang, X., Xu, X., and Chen, C. P. (2019). Gcb-net: Graph convolutional broad network and its application in emotion recognition. *IEEE Trans. Affect. Comput.* 13, 379–388. doi: 10.1109/TAFFC.2019.2937768

Zhang, Y., Chen, J., Tan, J. H., Chen, Y., Chen, Y., Li, D., et al. (2020). An investigation of deep learning models for eeg-based emotion recognition. *Front. Neurosci.* 14, 622759. doi: 10.3389/fnins.2020.622759

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., and Cichocki, A. (2018). Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176

Zheng, W.-L., and Lu, B.-L. (2015). Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, W.-L., Zhu, J.-Y., and Lu, B.-L. (2017). Identifying stable patterns over time for emotion recognition from eeg. *IEEE Trans. Affect. Comput.* 10, 417–429. doi: 10.1109/TAFFC.2017.2712143

Zhong, P., Wang, D., and Miao, C. (2020). Eeg-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* 13, 1290–1301. doi: 10.1109/TAFFC.2020.2994159