EasyChair Preprint

№ 12453

A Spatio-Temporal Graph Neural Network for EEG Emotion Recognition Based on Regional and Global Brain

Xiaoliang Wang, Chuncao Li, Yuzhen Liu, Kuan-Ching Li,
Wei Liang and Shiqi Zheng

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 12, 2024

A spatio-temporal Graph Neural Network for EEG Emotion Recognition Based on Regional and Global Brain

1st Xiaoliang Wang
School of Computer Science and Engineering
Hunan University of Science and Technology
Xiangtan, China
fengwxl@hnust.edu.cn

2nd Chuncao Li
School of Computer Science and Engineering
Hunan University of Science and Technology
Xiangtan, China
2233125215@qq.com

3rd Yuzhen Liu
School of Computer Science and Engineering
Hunan University of Science and Technology
Xiangtan, China
yzhenliu@126.com

4th Kuan-Ching Li*
Dept. of Computer Science and Information Engineering
Providence University
Taiwan, China
kuancli@pu.edu.tw

5th Wei Liang
School of Computer Science and Engineering
Hunan University of Science and Technology
Xiangtan, China
wliang@hnust.edu.cn

6th Shiqi Zheng
School of Computer Science and Engineering
Hunan University of Science and Technology
Xiangtan, China
772938731@qq.com

***Abstract*—Effective emotion recognition based on electroencephalography (EEG) is of relevant importance for the investigation of intelligence of the Brain-Computer Interface (BCI). Neuroscientific studies suggest that investigating localized brain activities contributes to a deeper understanding of the functionality of specific brain regions and the activity patterns under different emotional states. Many deep learning-based methods have been employed for EEG emotion recognition in recent years; however, most of these methods fail to extract the spatio-temporal features of EEG signals adequately. To further improve the efficiency of EEG emotion recognition, we propose in this work a novel spatio-temporal graph neural network, namely MSL-TGNN, by integrating local and global brain information. That is, the multi-scale temporal learner is employed to extract temporal features of EEG data. To explore the spatial features of EEG signals, considering the varying roles of different brain regions in EEG emotion classification, we propose a brain region learning block and an extended global graph attention network. The brain region learning block aggregates local channel information, and the extended global graph attention network can effectively capture nonlinear dependencies among regions and global brain information, thereby enhancing the learning capability for the EEG data. We conducted subject-dependent and subject-independent experiments on the DEAP dataset, and the results obtained indicate that our proposed model outperforms compared to state-of-the-art methods.**

***Index Terms*—Bidirectional gated recurrent unit, EEG emotion recognition, Graph attention network, Deep learning**

I. Introduction

Emotions reflect an individual's current psychological and physiological states, influencing various aspects of our daily lives [1]. Accurate and efficient emotion recognition is crucial for advancing the development of BCI. Both physiological and non-physiological signals can convey a person's emotional state. Non-physiological signals include facial expressions [2], language [3], body posture [4], among others. Physiological signals encompass EEG, EOG, ECG, and the like. Compared to non-physiological signals' more easily disguised nature, physiological signals authentically represent a person's emotions. Additionally, due to its non-invasive, convenient, and cost-effective nature, EEG is widely employed in emotion recognition [5].

One of the challenges in EEG emotion recognition is designing a more efficient method with good adaptability and generalization for automatically extracting relevant information from EEG signals. Traditional EEG emotion recognition often relies heavily on manual feature extraction. The most commonly used features are frequency domain features. The brainwave signals are initially decomposed into five frequency bands through Fourier Transform [6], [7]. Features are then extracted from each frequency band. The Power Spectral

This work was supported by the Scientific Research Fund of the National Natural Science Foundation of China (Grant No. 62372168), Hunan Provincial Natural Science Foundation of China (Grant No. 2023JJ30266), Research Project on teaching reform in Hunan province (No. HNJG-2022-0791) and Hunan University of Science and Technology (No. 2022-44-8), and National Social Science Funds of China (19BZX044).

Density (PSD) [8], the Differential Caudality (DCAU) [9], the Differential Entropy (DE) [10], [11], the Rational Asymmetry (RASM) [12], and the Differential Asymmetry (DASM) [13] are examples of frequently used EEG features. Shi *et al.* [14] first proposed the DE features and demonstrated their effectiveness in EEG signal characterization. Zheng *et al.* [15] extended traditional DE features to dynamic DE features, achieving higher accuracy in EEG emotion classification. Gao *et al.* [16] fused both frequency domain and time domain features for EEG emotion recognition, showing that feature fusion can effectively enhance recognition accuracy. However, these traditionally manually extracted features often fail to extract the temporal and spatial information from EEG signals fully. Additionally, frequency domain features are typically based on static analysis of the entire EEG signal, failing to capture the dynamic changes in the signal over time.

The growth of deep learning has recently been boosted fast, and its application to EEG emotion recognition has expanded dramatically. Fourati *et al.* [17] proposed a model based on Echo State Network (ESN) and utilized filtered signals as network inputs without employing any feature extraction methods. Li *et al.* [18] utilized the Bidirectional Long Short-Term Memory (BiLSTM) to extract spatio-temporal features, introducing a collaborative working mechanism between a classifier and discriminator to help reduce the disparities in emotion recognition across domains. Tao *et al.* [19] employed a Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) combined with attention mechanisms to address EEG emotion recognition problems. Ding *et al.* [20] used multiple one-dimensional convolutions and Graph Neural Network (GNN) to explore the spatio-temporal features of EEG signals. Gu *et al.* [21] utilized Generative Adversarial Networks (GAN) in conjunction with GNN and long short-term memory (LSTM) to explore EEG emotion classification. Zhang *et al.* [22] explored deep-level information of graph-structured data by stacking multiple graph convolution layers. Nevertheless, several problems still require research to increase the accuracy of EEG emotion classification, since EEG signals have excellent temporal resolution but poor spatial resolution [1], as many methods struggle to extract spatial information from different brain regions adequately. The brain is a highly organized and differentiated organ, composed of various regions responsible for different functions. Complex connections and interactions exist between these regions, and studies indicate that the strength of interaction between brain regions attenuates with increasing physical distance [23]. Understanding the local activities of the brain is crucial in neuroscience and clinical fields for exploring cognitive functions, neurological disorders, and the impact of brain injuries. In EEG emotion recognition, another challenge is how to simultaneously focus on the brain's overall structure and information from individual local regions.

To address the challenges mentioned above in EEG emotion recognition, we propose a novel neural network model, namely MSL-TGNN, to automatically extract spatio-temporal information from EEG signals through deep learning whilst considering the distinctiveness of different brain regions in emotional expression. Inspired by Ding *et al.* [24], we introduce a multi-scale temporal learner, employing three parallel Bidirectional Gated Recurrent Units (BiGRU) to capture different frequency representations in EEG signals. EEG data typically contain signals of multiple frequencies, and these frequencies may play crucial roles at different time points. The superposition of different hidden layer states in BiGRU can comprehensively handle information from different frequency dimensions, aiding the model in achieving a multi-scale representation of the signals. Additionally, by introducing a brain region learning block to aggregate local channel information, the model better understands the roles played by various brain regions in EEG emotion classification. Simultaneously, we integrate the multi-dimensional "t2t" self-attention mechanism [25] into the Graph Attention Network (GAT) to capture intra-node dependencies and inter-node connectivity. GAT is effective in learning interactions between nodes, and the multi-dimensional "t2t" self-attention mechanism can learn relationships between multi-dimensional features within nodes. The extended global graph attention network comprehends relationships between nodes at a higher level and captures global patterns more effectively. The main contributions of this work include the following aspects:

- To propose a novel end-to-end deep learning framework to overcome the limitations of manually extracting features. The proposed model can automatically learn spatio-temporal features from EEG signals, demonstrating better adaptability and generalization across different subjects. It can comprehensively capture the complexity of EEG signals.
- By leveraging learned local weights, we perform a weighted fusion of information from each brain region. This enables the model to understand better each brain region's unique roles in EEG emotion recognition. Furthermore, introducing the extended global graph attention network strengthens the model's capacity to capture non-linear dependencies between nodes and global information, as integrating local and global modules ensures that the model adequately acquires spatial information from the brain.
- Extensive subject-dependent and subject-independent experiments conducted on the DEAP dataset have demonstrated that the proposed MSL-TGNN method can significantly improve performance. Additionally, ablation studies illustrate the effectiveness of each module in the proposed method.

The remainder of this work is organized as follows. The background of the proposed model is provided briefly in Section 2, a comprehensive description of the proposed MSL-TGNN and its application in EEG emotion recognition is provided in Section 3, the experimental details and the discussions of results are presented in Section 4, and finally, the concluding remarks and future directions are depicted in Section 5.

II. RELATED WORK

This section briefly introduces methods that serve as the foundation for the proposed model.

A. BiGRU

EEG data comprise multi-channel information that varies over time, and RNNs are good at learning long-term dependencies [26]. In recent years, RNN has found widespread application in EEG data. However, RNN suffers from issues such as exploding and vanishing gradients. To overcome the shortcomings of traditional RNN, Hochreiter *et al.* [27] proposed LSTM. To simplify the structure of LSTM for improved training efficiency, Chung *et al.* [28] introduced modifications based on LSTM and presented GRU. Unidirectional GRU performs a state transition from past to future. In specific tasks, particularly those requiring simultaneous consideration of past and future information, unidirectional models may fail to utilize all available contextual information fully. However, EEG data typically contain intricate spatio-temporal information, and the signal feature may depend on both past and future time points. BiGRU can capture past and future information in sequence data by considering forward and backward hidden states simultaneously. So BiGRU assists in handling the time dependencies in EEG data, facilitating a more effective capture of dynamic information at different time points. Abgeena *et al.* [29] proposed a CNN-BiGRU model, demonstrating its effectiveness in emotion classification based on EEG signals. However, it still fails to extract spatial information from the brain entirely.

B. GAT

GNNs are widely used in various fields [30]–[33], and GAT is a special type of GNN. Compared to Graph Convolutional Networks (GCN), GAT possesses unique advantages. Different from GCN that employs uniform neighbor aggregation, thus disregarding the heterogeneity among nodes, GAT introduces an attention mechanism that assigns different weights to neighboring nodes for each node, allowing for a more flexible capture of the graph structure [34]. Because of this, GAT handles complicated graph structures very well and provides a more accurate representation of the relationships between nodes. Zhao *et al.* [35] offered an epilepsy detection method based on GAT and highlighted the potential advantages of GAT in handling multi-channel biological signals. However, little study has been done on applying GAT to EEG emotion recognition.

C. Self-Attention and Multi-dimensional Attention

The self-attention mechanism has been widely used in various natural language processing (NLP) tasks [36]. It allows models to establish weight relationships between positions in a sequence, capturing global contextual information, so the models can be suited for sequences of different lengths. The multi-dimensional attention mechanism is an extension of the attention mechanism at the feature level [25], aimed at more comprehensively capturing relationships across multiple dimensions in input data. Compared to traditional self-attention, multi-dimensional attention introduces independent attention to different feature dimensions. Each feature dimension has its weight allocation in multi-dimensional attention, allowing the model to attend to critical information in different dimensions flexibly. This mechanism is well-suited for handling multi-modal data or data with multiple dimensions, such as multi-channel EEG data. The introduction of multi-dimensional attention enhances the proposed model's perception of the multi-dimensional relationships, thereby exhibiting excellent performance in handling multi-channel data.

III. METHOD

MSL-TGNN consists of two major modules: the multi-scale temporal learner and the spatial feature learner. The multi-scale temporal learner automatically extracts information from different frequency dimensions more comprehensively by overlaying the bidirectional hidden states of multiple dimensions, replacing manual feature extraction. The spatial feature learner includes the brain region learning block and the extended global graph attention network. The brain region learning block captures the neural activity of brain regions, and the aggregated local channel information serves as input to the extended global graph attention network to learn complex relationships among different brain regions. Fig. 1 shows the structure of the proposed MSL-TGNN.

A. Multi-scale Temporal Learner

The multi-scale temporal learner learns information from different frequency dimensions of EEG data by configuring parallel hidden state sizes. To comprehensively capture the temporal dynamics in the input sequence, we set the size of the hidden state in different proportions according to the sampling rate f. The ratio coefficient is denoted as $\lambda_i \in R$, where i represents the layer number of BiGRU, $i = [1, 2, 3]$. The hidden state size $h^{(i)}$ for the i-th layer can be defined as

$$h^{(i)} = \lambda_i \times f, \lambda_i \in [0.5, 1, 2] \tag{1}$$

Given the baseline-corrected EEG data $X_t \in R^{c \times l}$, where t is the time step of the input sequence, c is the number of channels, and l is the sample length along the time dimension. We apply three parallel multi-scale BiGRUs to learn dynamic frequency representations, and the update of the hidden state in a unidirectional GRU can be represented as

$$h_t^{(i)} = z_t^{(i)} \odot \widetilde{h_t^{(i)}} + \left(1 - z_t^{(i)}\right) \odot h_{t-1}^{(i)} \tag{2}$$

where $z_t^{(i)}$ is the output of the update gate, $\widetilde{h_t^{(i)}}$ is the output of the memory cell after activation function and $\odot$ represents the dot product operation.

The output sequence of the forward GRU in the i-th layer can be expressed as

$$\overrightarrow{h_t^{(i)}} = \overrightarrow{GRU}\left(\overrightarrow{h_{t-1}^{(i)}}, X_t\right) \tag{3}$$

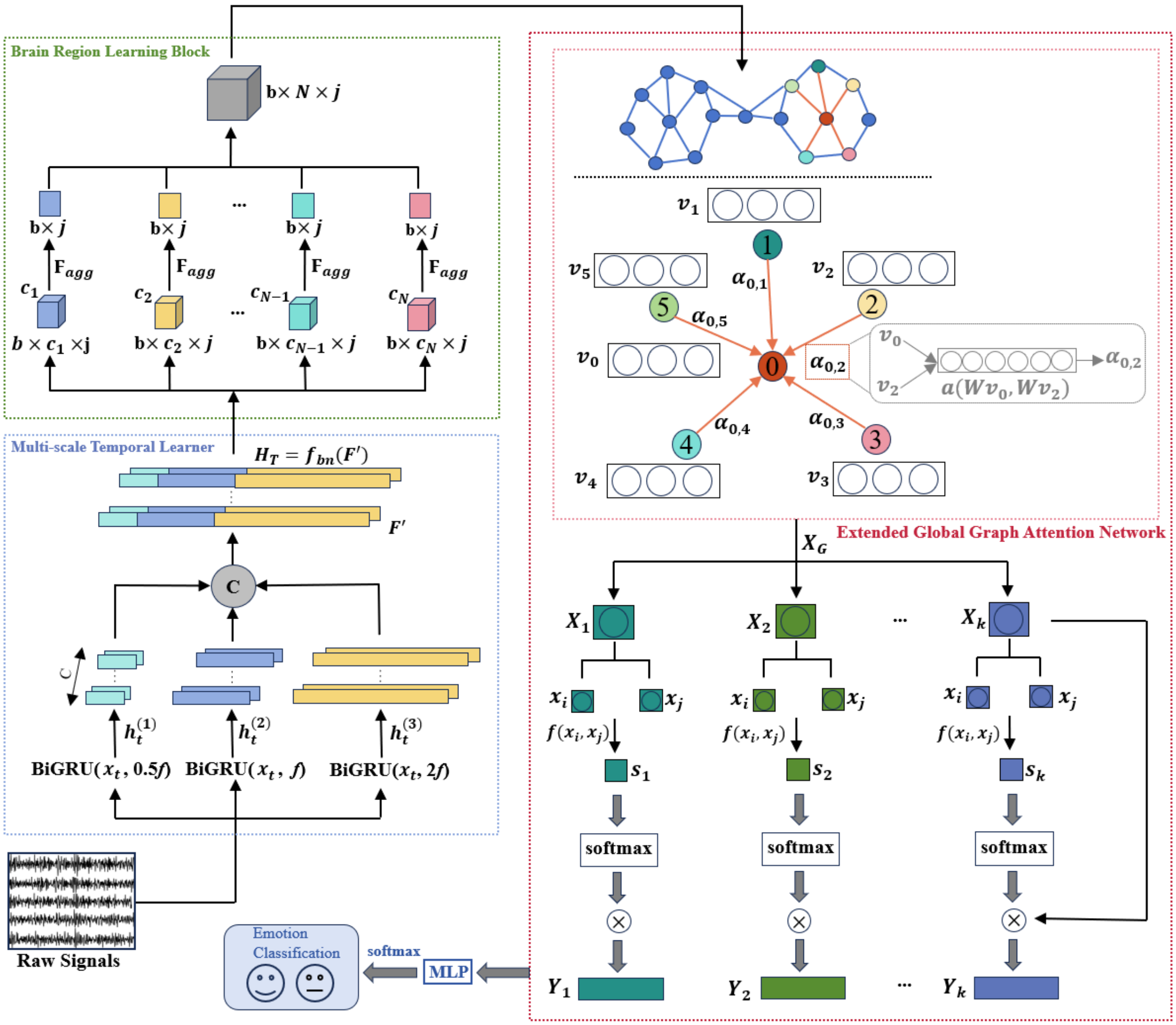


Fig. 1: The structure of MSL-TGNN. MSL-TGNN consists of the multi-scale temporal learner and the spatial feature learner, which further comprises the brain region learning block and the extended graph attention network. The multi-scale temporal learner illustrates three parallel BiGRUs learning information from different frequency dimensions of multi-channel EEG data. The brain region learning block demonstrates the information aggregation process of four brain regions. The extended global graph attention network displays the weight distribution between nodes and the weight distribution of internal multiple feature dimensions at the node level.

And the output sequence of the backward GRU in the i-th layer can be expressed as

$$\overleftarrow{h_t^{(i)}} = \overleftarrow{GRU}\left(\overleftarrow{h_{t+1}^{(i)}}, X_t\right) \tag{4}$$

We concatenate the outputs of all parallel BiGRUs along the feature dimension. Therefore, the final output of the multi-scale temporal learner is represented as

$$H_T = f_{bn}\left(\Gamma\left(\overrightarrow{h_t^{(1)}}, \overleftarrow{h_t^{(1)}}, \ldots, \overrightarrow{h_t^{(i)}}, \overleftarrow{h_t^{(i)}}\right)\right) \tag{5}$$

where $\Gamma(\cdot)$ represents the concatenation operation along the feature dimension and f_{bn} denotes batch normalization operation.

B. Spatial Feature Learner

1) Brain region learning block: In neuroscience studies, researchers gain insights into the functions of specific brain regions by focusing on the local activities of the brain. To understand the neural activities of brain regions, we input the output of the multi-scale temporal learner into the brain region learning block to aggregate the local information within brain regions. The placement of EEG electrodes on the scalp follows the 10-20 system [37]. Following the definition in [38], the 62 electrodes were split into 17 regions, as shown in Fig. 2,

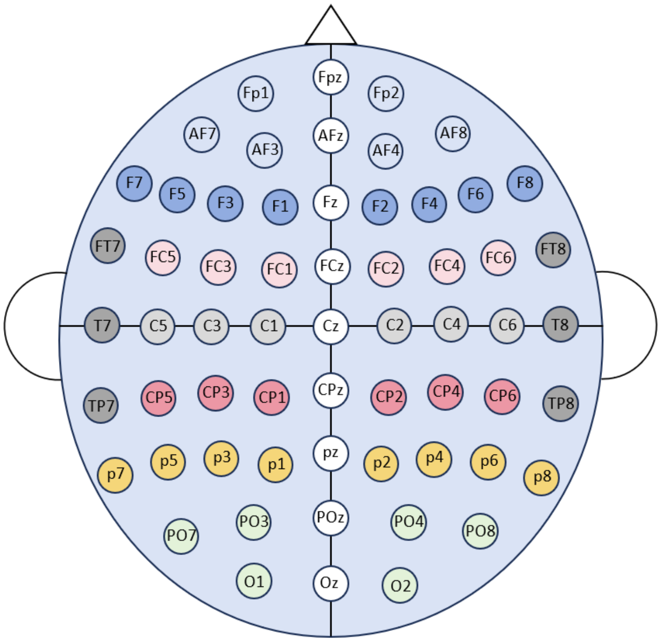


Fig. 2: Schematic diagram of EEG electrode positions. Electrodes with the same color in the same hemisphere represent a brain region. The 62 electrodes are divided into 17 regions.

with %ERS/ERD aggregated over eight areas per hemisphere. Taking the right hemisphere as an example: anteriofrontal (AF: Fp2, AF4, AF8), frontal (F: F2, F4, F6, F8), temporal (T: FT8, T8, TP8), frontocentral (FC: FC2, FC4, FC6), central (C: C2, C4, C6), centroparietal (CP: CP2, CP4, CP6), parietal (P: P2, P4, P6, P8), and parietooccipital (PO: PO4, PO8, O2).

In the process of channel information aggregation, we first introduce the local weight matrix $W_{c \times j}$, where c denotes the number of channels, and j represents the number of features for every channel. Through initialization and learning processes, this matrix can adaptively adjust the weight for each channel. By applying different weights to each channel, we can more finely capture local features of brain regions. Then, we use the aggregation function $F_{agg}(\cdot)$ to aggregate the channel information output by the multi-scale temporal learner. After extensive experimentation with aggregation operations such as average, sum, and variance, we chose the average operation. Therefore, the output H_{area} of the brain region learning block can be represented as

$$H_{area} = F_{agg}\left(W_{c \times j} \odot H_T\right) \tag{6}$$

2) Extended global graph attention network: As seen in the right half of Fig. 1, the extended global graph attention network aims to learn the correlations between different brain regions. We integrate the multi-dimensional ” t2t ” self-attention mechanism into the GAT to focus on the multi-dimensional features within nodes. Each subject's raw data are represented as a new structured time series and serve as input to the extended global graph attention network after being processed by the multi-scale temporal learner and brain region learning block. Initially, each subject's data is treated as a cyclic-free graph. Then, a correlation matrix is generated based on the neural activity relationships between brain regions, representing the adjacency matrix of the corresponding graph. After obtaining the associations between nodes through GAT, we further enhance the expression capability of nodes' internal features using the multi-dimensional ” t2t ” self-attention mechanism.

Inspired by Zhao *et al.* [35] and Wang *et al.* [39], we consider the information of each subject's brain as a graph. The information aggregated from each region is regarded as a node in the graph, and the associations between each pair of regions are considered edges. Pearson correlation matrix is employed to calculate spatial correlations.

The input of GAT consists of a set of node features $V = \{v_1, v_2, \ldots, v_N\}, v_N \in R^F$, and the corresponding adjacency matrix. Here, N denotes the number of brain regions or nodes, and F denotes the feature dimensions of each node. GAT initially performs a linear transformation on each node by multiplying it with a weight matrix. Subsequently, the attention coefficients between each pair of nodes are calculated as

$$e_{ij} = a\left(Wv_i, Wv_j\right) \tag{7}$$

where $W \in R^{F \times F'}$, with F' representing the feature dimensions of the output nodes, i and j denote any two nodes and $a(\cdot)$ represents a feedforward neural network that concatenates the resulting vectors to accomplish feature mapping. Next, we compute the attention coefficients of node i to all other nodes and employ softmax to normalize the attention weights, obtaining the ultimate attention coefficients α_{ij}. The calculation formula is defined as

$$\alpha_{ij} = \frac{\exp\left(LeakyRelu\left(\vec{a}^T\left[Wv_i \,||\, Wv_j\right]\right)\right)}{\sum_{k \in N_i} \exp\left(LeakyRelu\left(\vec{a}^T\left[Wv_i \,||\, Wv_k\right]\right)\right)} \tag{8}$$

where k represents any node, $||$ is the concatenation operator. Finally, during the convolution process, we employ the multi-head attention mechanism. After being processed by the GAT layer, the features of node i can be represented as

$$v_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} \alpha_{ij}^k W^k v_j\right) \tag{9}$$

where k denotes any head, and i denotes any node.

The attention mechanism is applied in a shared manner to all edges in the graph. Unlike GCN, where each neighboring node equally influences the representation of the target node, the proposed model assigns different attention weights to adjacent nodes, permitting the model to flexibly consider different relationships between nodes when updating node representations, aiding in capturing local information within the graph. The final output of the GAT layer is denoted as X_G.

To simultaneously focus on the multiple dimensions of features within nodes, we introduce a multi-dimensional ” t2t ” self-attention mechanism to assign weights to different features of nodes. In contrast to traditional self-attention mechanisms that primarily focus on relationships between different nodes within a sequence, the multi-dimensional "t2t" self-attention mechanism can comprehensively capture information in the input sequence by considering the multi-dimensional features of each node and calculating attention scores across multiple feature dimensions. Let X_k represent the k-th sample in X_G, and s_k denote the inherent correlation between different feature dimensions x_i and x_j of X_k. The multi-dimensional "t2t" self-attention mechanism adds biases both inside and outside the activation function. Let W and b represent the weight and bias of the σ function, respectively. Thus, s_k can be expressed as

$$s_k = f\left(x_i, x_j\right) = W^T \sigma\left(W_1 x_i + W_2 x_j + b_1\right) + b \tag{10}$$

For each x_i, we calculate a probability matrix $P = \{p_1, p_2, \ldots, p_r\}$. The calculation output for x_i is defined as

$$Y_i = \sum_{j=1}^{n} p_j \odot x_j \tag{11}$$

The output of the multi-dimensional "t2t" self-attention mechanism for all samples X_G is denoted as $Y = [Y_1, Y_2, \ldots, Y_k]$.

IV. EXPERIMENTAL RESULTS

In this section, we first briefly introduce the dataset and the pre-processing steps. Then, we describe the experimental settings and model parameters. Subsequently, we present the results of subject-dependent and subject-independent experiments of MSL-TGNN and engage in relevant discussions.

A. DEAP Dataset

The DEAP dataset [40] was collected from 32 volunteers (16 males, 16 females) with ages ranging from 19 to 37 years, and 26.9 as average age. Each participant underwent 40 trials, where they watched emotionally evocative music videos lasting one minute each to induce corresponding emotional states. Simultaneously, EEG data from 32 channels and peripheral physiological signals from 8 channels of each participant were collected. After each trial, participants rated their arousal, valence, dominance, and liking for each video using a continuous 9-point scale. In this investigation, we utilized EEG data from 32 channels.

B. Data Preprocessing

Due to our model being an end-to-end framework, for the DEAP dataset, we only performed baseline correction on the pre-processed data provided by the authors. Baseline correction is applied to reduce errors caused by scalp potential variations, equipment drift, or other environmental interferences, aiming to obtain EEG data with a high signal-to-noise ratio [41]. The original EEG data were downsampled to 128 Hz. We selected the stable-state EEG data before the stimulus as the baseline, corresponding to the first 3 seconds of each trial in the DEAP dataset. The average of this baseline was calculated and considered as the baseline level, representing the resting state of the measured brain region. Subsequently, the value at each time point in the entire EEG signal was subtracted by the corresponding value of the baseline at that time point. This process shifts the signal as a whole to a zero baseline level. Finally, the 3-second baseline data were removed. For the label processing, we projected the continuous 9-point scale onto high and low classes for each dimension by thresholding the valence and arousal at 5. Following the approach outlined in [42], we segmented the data from each trial into non-overlapping segments of 3 seconds, further splitting each segment into three 1-second data segments.

C. Experiment Settings

We conducted subject-dependent and subject-independent experiments on the DEAP dataset to evaluate MSL-TGNN. In the subject-dependent experiments, after the pre-processing stage, each data sample from a subject is represented as $X_i \in R^{3\times32\times128}$, where $i = [1, 2, \ldots, 800]$. All samples from different trials were shuffled for every subject. In the subject-independent experiments, we combined all subjects' samples and shuffled them. The data samples can be represented as $X_j \in R^{3\times32\times128}$, where $j = [1, 2, \ldots, 25600]$. The experiments employed 10-fold cross-validation to evaluate the model's performance, and the average performance was taken as the final experimental results. Our model was implemented with the PyTorch framework and trained on an NVIDIA GeForce RTX 3080 Ti GPU. The GAT layer was set to 1, and the number of nodes was set to 17. We used the Adam optimizer with a learning rate 10^{-4} to update the model parameters, minimizing the cross-entropy loss function. During the training process, dropout operations randomly discarded input neurons with probability 0.5, and batch normalization was applied for each mini-batch, addressing the vanishing gradient problem, accelerating the training process, and improving model generalization.

D. Comparative Studies

In this subsection, we present the results of subject-dependent and subject-independent experiments to validate the effectiveness of the proposed method, followed by a brief analysis.

1) Subject-dependent experiments: We compared our method with five recent deep learning methods and one traditional machine learning method, including GAT [35], STFFNN [43], GCNN [6], DCNN+ConvLSTM [44], STS-Transformer [45], and Decision Tree (DT) [46]. In [35], GAT was used for epilepsy detection based on EEG data. STFFNN captures electrode dependencies using power topography maps, employs CNN for spatial feature learning, utilizes feedforward networks for temporal feature learning, and integrates spatial-temporal features using BiLSTM. GCNN is a traditional graph convolutional neural network, and we use DE features as the input to GCNN. In [44], the authors used Deep Convolutional Neural Network (DCNN) and ConvLSTM to extract features separately and then concatenated the features with attention mechanism-weighted fusion. STS-Transformer relies on the transformer and attention mechanisms for feature extraction and weight allocation. We either directly cite their results from the literature or reproduce them based on the code they have released to guarantee an effective comparison with our method. In Table 1, we list all the features that each method used in detail.

Table 1 shows that MSL-TGNN achieves the highest accuracy of 93.09% (valence) and 93.74% (arousal). Additionally, DT outperforms GAT significantly on both valence and arousal classification tasks. We speculate that this may stem from DT utilizing DE features, which providing more favorable abstract representations for emotion recognition. In contrast, GAT directly employs raw data, even though it offers more flexibility in handling spatial relationships, it might result in relatively lower performance due to the multiple channels and complexity of the data. This also emphasizes the crucial role of feature selection in emotion recognition tasks. Our method employs raw data as the model input. Compared with GAT and DT, MSL-TGNN achieves an average accuracy improvement of approximately 20.38% and 16.35%, respectively. Although GCNN takes manually extracted DE features as input, our approach still outperforms GCNN by approximately 5.5%, indicating the effectiveness of our improvements to the GCNN. For STFFNN, DCNN+ConvLSTM, and STS-Transformer, we used the same features mentioned in the original papers as inputs. Our method outperforms these approaches by approximately 7.6%, 5.65% and 5%, respectively. This significant performance improvement indicates that our method has a clear advantage in feature extraction.

2) Subject-independent experiments: In the subject-independent experiments, we compared our approach with four advanced deep learning methods: GAT, CapsNet [47], STFFNN, and STS-Transformer. In [47], the frequency domain, frequency band characteristics, and the spatial characteristics of the EEG signals are fused and input into CapsNet for emotion classification. Table 2 displays the results of the comparison. Even when using raw data as the model input, our method performs well compared to other models. Notably, our approach's accuracy (84.14%) is comparable to STS-Transformer (84.75%), significantly outperforming GAT and CapsNet in valence classification tasks. Additionally, our method surpasses STS-Transformer and the other three methods, achieving an accuracy of 83.99% in arousal classification tasks, demonstrating the effectiveness of our method in capturing emotional states. Furthermore, our method achieves better classification accuracy despite STFFNN taking pre-processed EEG features as model input. Our method achieved better F1 scores in both valence and arousal tasks, which further confirms the robustness of our method. The experiment results indicate that by effectively capturing the spatio-temporal features of EEG data, the model can better understand the similarities and differences of EEG signals among different subjects.

E. Ablation Study

To validate the performance of each module in our proposed method, we designed three models, namely L-TGNN, MS-TGNN, and MSL-GAT. In the first ablation study, aiming to emphasize the contribution of the multi-scale temporal learner, we replaced it with a single BiGRU, resulting in L-TGNN. In the second ablation experiment, we removed it from the proposed model to verify the importance of the brain region learning block, resulting in MS-TGNN. Finally, in the third ablation experiment, we replaced the original model's extended global graph attention network with a regular GAT layer to verify its effect, resulting in MSL-GAT. To control variables, we maintained consistency with the original model in data pre-processing methods. Also, we utilized the average performance from 10-fold cross-validation as the final results for all ablation experiments.

From Table 3, we can observe that MSL-GAT and L-TGNN, in the subject-dependent experiments, the accuracy decreased by approximately 3.55% and 3.44%, respectively, compared to MSL-TGNN. However, in the subject-independent experiments, the accuracy decreased by 7.38% and 5.09%. This indicates that these models are more likely to adapt to individual-specific patterns in subject-dependent experiments, resulting in relatively minor performance differences between models. This also suggests that the extended global graph attention network contributes more to the model than the multi-scale temporal learner. MS-TGNN exhibited a minor decrease in accuracy compared to MSL-GAT and L-TGNN, indicating a relatively more minor contribution from the brain region learning block than the significant contributions of the multi-scale temporal learner and extended global graph attention network. In the ablation study of subject-dependent, each of our methods was experimented on all subjects to validate the contributions of various modules in our model. Fig. 3 and Fig. 4 display the classification accuracy and standard deviation for each subject on each label task. The figures show that MSL-TGNN achieves higher accuracy on both labels than L-TGNN, MS-TGNN, and MSL-GAT, with more minor standard deviations. This indicates that each module in our model plays a unique role, resulting in better generalization and adaptability of the final model. We can observe that there is variability in the accuracy of different individuals. Since our model takes raw EEG data as input without additional feature extraction, differences in age, gender, and physical conditions are prominently reflected in our experimental results. In addition, due to individual differences, the classification accuracy of the subject-dependent experiments is higher than that of the subject-independent experiments.

F. Discussion

As shown in Fig. 5, MSL-TGNN achieves recognition accuracies of 94.35% (positive) and 91.52% (negative) on valence, and 95.45% (positive) and 91.41% (negative) on arousal in the subject-dependent experiments. This indicates that the model performs better in identifying high valence and high arousal emotions, suggesting a better ability to recognize positive emotions. The confusion matrices of the subject-independent experiments also show similar results. This observation is consistent with previous research [48].

TABLE I: Comparison of Input Features and Performance of Different Methods on the DEAP Dataset in the Subject-dependent Experiments

Methods	Features	Valence		Arousal	
		Acc(%)	F1(%)	Acc(%)	F1(%)
GAT(2021)	Raw signals	72.05	73.29	74.03	73.2
DT(2018)	Differential entropy	75.95	-	78.18	-
STFFNN(2022)	PSD+temporal statistics	85.42	84.33	86.16	85.5
DCNN+ConvLSTM(2021)	Raw signals	87.84	-	87.69	-
GCNN(2018)	Differential entropy	88.24	-	87.72	-
STS-Transformer(2023)	Raw signals	89.86	-	86.83	-
MSL-TGNN	Raw signals	**93.09**	**93.39**	**93.74**	**93.94**

TABLE II: Comparison of Input Features and Performance of Different Methods on the DEAP Dataset in the Subject-independent Experiments

Methods	Features	Valence		Arousal	
		Acc(%)	F1(%)	Acc(%)	F1(%)
GAT(2021)	Raw signals	62.88	72.19	64.62	74.27
CapsNet(2019)	Band power feature matrix	66.73	-	68.28	-
STFFNN(2022)	PSD+temporal statistics	80.17	79.97	81.28	81.09
STS-Transformer(2023)	Raw signals	**84.75**	-	82.16	-
MSL-TGNN	Raw signals	84.14	**85.83**	**83.99**	**86.05**

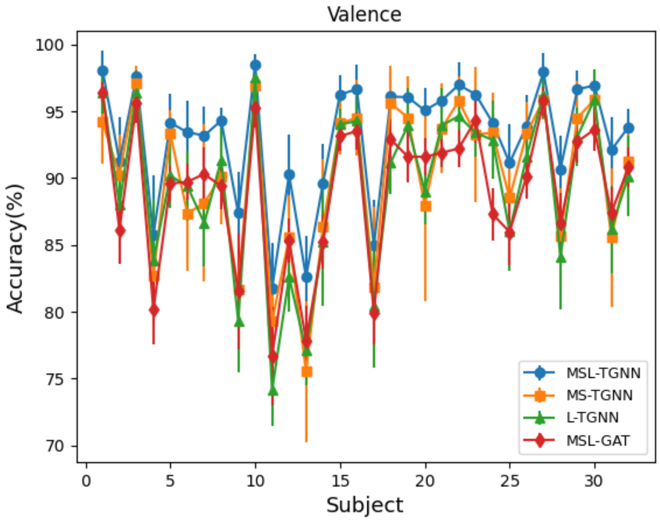


Fig. 3: Average accuracies on each subject of ablation experiments on valence classification tasks in the subject-dependent experiments

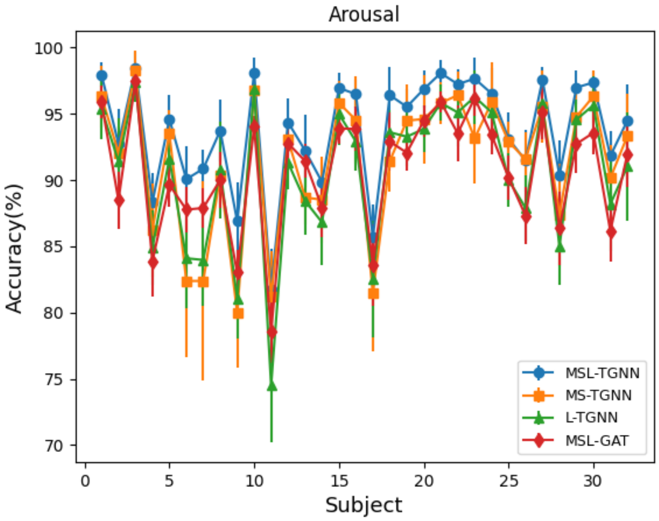


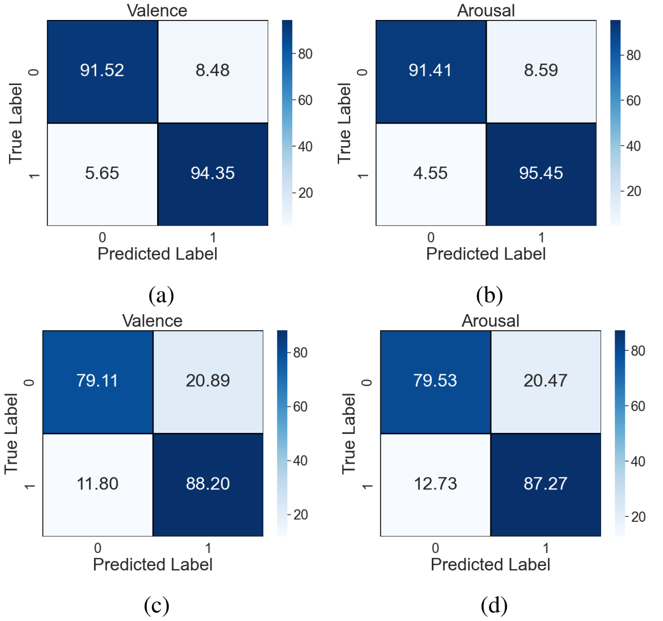
Fig. 4: Average accuracies on each subject of ablation experiments on arousal classification tasks in the subject-dependent experiments

TABLE III: Ablation Study on the DEAP Dataset

Experiment Schemes	Methods	Valence		Arousal	
		Acc(%)	F1(%)	Acc(%)	F1(%)
Subject-dependent	MSL-GAT	89.07	89.58	90.67	91.01
	L-TGNN	89.34	89.95	90.61	90.97
	MS-TGNN	90.1	90.62	91.39	91.53
	MSL-TGNN	**93.09**	**93.39**	**93.74**	**93.94**
Subject-independent	MSL-GAT	76.33	78.69	77.04	80.07
	L-TGNN	78.69	80.76	79.27	81.91
	MS-TGNN	81.70	83.41	81.32	83.36
	MSL-TGNN	**84.14**	**85.83**	**83.99**	**86.05**

The experiment results also show that MSL-TGNN obtains a high F1 score while achieving high accuracy. This implies that the model correctly identifies positive instances and effectively captures negative instances. In other words, for the EEG data of all subjects, the model can robustly capture features of both classes, demonstrating good robustness and generalization of the model.

In this paper, we demonstrate the effectiveness of our proposed model through extensive experiments. In comparative experiments, we contrast MSL-TGNN with six deep learning methods and one traditional machine learning method, encompassing a series of models widely applied in the processing of biosignals. While extracting effective features is crucial for EEG emotion recognition, our model achieves outstanding results even with raw EEG data. Through ablation experiments, we conduct a thorough analysis of the performance of each module. MSL-TGNN exhibits significantly higher accuracy on valence and arousal when compared to other



(a)

(b)

(c)

(d)

Fig. 5: Confusion matrices of MSL-TGNN (a)Subject-dependent experiments of valence (b)Subject-dependent experiments of arousal (c)Subject-independent experiments of valence (d)Subject-independent experiments of arousal

ablation models, confirming the unique contributions of the multi-scale temporal learner, brain region learning block, and extended global graph attention network in the entire model. Overall, our method has achieved a significant performance improvement in emotion classification tasks by investigating multi-scale representations of brain signals in different frequency dimensions, gaining a deeper understanding of the functionality of specific brain regions, patterns of brain activity under different emotional states, and proposing a novel network capable of considering complex graph data structure and multi-dimensional feature representations. This has positive implications for future research in emotion recognition and other fields of bio-signal processing.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel graph neural network model that employs the multi-scale temporal learner to process different frequency dimensions in the raw EEG signals concurrently, the brain region learning block to apply different weights based on the function of each brain region, and the extended global graph attention network to capture the global patterns of brain activity. Our model can effectively capture global and local information, achieving a balanced perspective on global brain activity and detailed attention to specific regions. Moreover, MSL-TGNN is an end-to-end model capable of achieving robust recognition performance on raw EEG data. Extensive subject-dependent and subject-independent experimental results demonstrate the competitiveness of our method compared to state-of-the-art methods. Whereas our method has been proven effective in EEG emotion recognition, there may be significant differences in emotional activity patterns among individuals. Therefore, exploring how to reduce this variability remains to be investigated. As future research directions, we will focus on enhancing the model's generalization capability, especially in cross-dataset scenarios, which may require further exploration of data augmentation, domain adaptation, and transfer learning techniques.

REFERENCES

[1] Soraia M Alarcao and Manuel J Fonseca. Emotions recognition using eeg signals: A survey. *IEEE Transactions on Affective Computing*, 10(3):374–393, 2017.

[2] Wang Kay Ngai, Haoran Xie, Di Zou, and Kee-Lee Chou. Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. *Information Fusion*, 77:107–117, 2022.

[3] Hesam Sagha, Pavel Matejka, Maryna Gavryukova, Filip Povolnỳ, Erik Marchi, and Björn Schuller. Enhancing multilingual recognition of emotion in speech by language identification. 2016.

[4] Nele Dael, Marcello Mortillaro, and Klaus R Scherer. Emotion expression in body action and posture. *Emotion*, 12(5):1085, 2012.

[5] Christopher Niemic. Studies of emotion: a theoretical and empirical review of psychophysiological studies of emotion. 2004.

[6] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2018.

[7] Yun Gu, Xinyue Zhong, Cheng Qu, Chuanjun Liu, and Bin Chen. A domain generative graph network for eeg-based emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[8] Matteo Demuru, Simone Maurizio La Cava, Sara Maria Pani, and Matteo Fraschini. A comparison between power spectral density and network metrics: an eeg study. *Biomedical Signal Processing and Control*, 57:101760, 2020.

[9] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

[10] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu. Eeg-based emotion classification using deep belief networks. In *2014 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2014.

[11] Yifan Jiang, Ning Chen, and Jing Jin. Detecting the locus of auditory attention based on the spectro-spatial-temporal analysis of eeg. *Journal of Neural Engineering*, 19(5):056035, 2022.

[12] Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, 2010.

[13] Aasim Raheel, Muhammad Majid, and Syed Muhammad Anwar. A study on the effects of traditional and olfaction enhanced multimedia on pleasantness classification based on brain activity analysis. *Computers in biology and medicine*, 114:103469, 2019.

[14] Li-Chen Shi, Ying-Ying Jiao, and Bao-Liang Lu. Differential entropy feature for eeg-based vigilance estimation. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6627–6630. IEEE, 2013.

[15] Fa Zheng, Bin Hu, Xiangwei Zheng, Cun Ji, Ji Bian, and Xiaomei Yu. Dynamic differential entropy and brain connectivity features based eeg emotion recognition. *International Journal of Intelligent Systems*, 37(12):12511–12533, 2022.

[16] Qiang Gao, Yi Yang, Qiaoju Kang, Zekun Tian, and Yu Song. Eeg-based emotion recognition with feature fusion networks. *International journal of machine learning and cybernetics*, 13(2):421–429, 2022.

[17] Rahma Fourati, Boudour Ammar, Javier Sanchez-Medina, and Adel M Alimi. Unsupervised learning in reservoir computing for eeg-based emotion recognition. *IEEE Transactions on Affective Computing*, 13(2):972–984, 2020.

[18] Yang Li, Wenming Zheng, Lei Wang, Yuan Zong, and Zhen Cui. From regional to global brain: A novel hierarchical spatial-temporal neural network model for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 13(2):568–578, 2019.

[19] Wei Tao, Chang Li, Rencheng Song, Juan Cheng, Yu Liu, Feng Wan, and Xun Chen. Eeg-based emotion recognition via channel-wise attention and self attention. *IEEE Transactions on Affective Computing*, 2020.

[20] Yi Ding, Neethu Robinson, Chengxuan Tong, Qiuhao Zeng, and Cuntai Guan. Lggnet: Learning from local-global-graph representations for brain–computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[21] Yun Gu, Xinyue Zhong, Cheng Qu, Chuanjun Liu, and Bin Chen. A domain generative graph network for eeg-based emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[22] Tong Zhang, Xuehan Wang, Xiangmin Xu, and CL Philip Chen. Gcb-net: Graph convolutional broad network and its application in emotion recognition. *IEEE Transactions on Affective Computing*, 13(1):379–388, 2019.

[23] Raymond Salvador, John Suckling, Martin R Coleman, John D Pickard, David Menon, and ED Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex*, 15(9):1332–1342, 2005.

[24] Yi Ding, Neethu Robinson, Qiuhao Zeng, Duo Chen, Aung Aung Phyo Wai, Tih-Shih Lee, and Cuntai Guan. Tsception: a deep learning framework for emotion detection using eeg. In *2020 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE, 2020.

[25] Tao Shen, Tianyi Zhou, Guodong Long, Jing Jiang, Shirui Pan, and Chengqi Zhang. Disan: Directional self-attention network for rnn/cnn-free language understanding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[26] Na Hu, Dafang Zhang, Kun Xie, Wei Liang, Chunyan Diao, and Kuan-Ching Li. Multi-range bidirectional mask graph convolution based gru networks for traffic prediction. *Journal of Systems Architecture*, 133:102775, 2022.

[27] Long Short-Term Memory. Long short-term memory. *Neural computation*, 9(8):1735–1780, 2010.

[28] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[29] Abgeena Abgeena and Shruti Garg. A novel convolution bi-directional gated recurrent unit neural network for emotion recognition in multichannel electroencephalogram signals. *Technology and Health Care*, 31(4):1215–1234, 2023.

[30] Jiahong Cai, Wei Liang, Xiong Li, Kuanching Li, Zhenwen Gui, and Muhammad Khurram Khan. Gtxchain: A secure iot smart blockchain architecture based on graph neural network. *IEEE Internet of Things Journal*, 10(24):21502–21514, 2023.

[31] Wei Liang, Yuhui Li, Kun Xie, Dafang Zhang, Kuan-Ching Li, Alireza Souri, and Keqin Li. Spatial-temporal aware inductive graph neural network for c-its data recovery. *IEEE Transactions on Intelligent Transportation Systems*, 24(8):8431–8442, 2023.

[32] Chunyan Diao, Dafang Zhang, Wei Liang, Kuan-Ching Li, Yujie Hong, and Jean-Luc Gaudiot. A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. *IEEE Transactions on Intelligent Transportation Systems*, 24(1):904–914, 2023.

[33] Na Hu, Dafang Zhang, Kun Xie, Wei Liang, Kuanching Li, and Albert Zomaya. Multi-graph fusion based graph convolutional networks for traffic prediction. *Computer Communications*, 210:194–204, 2023.

[34] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[35] Yanna Zhao, Gaobo Zhang, Changxu Dong, Qi Yuan, Fangzhou Xu, and Yuanjie Zheng. Graph attention network with focal loss for seizure detection on electroencephalography signals. *International journal of neural systems*, 31(07):2150027, 2021.

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[37] Robert Oostenveld and Peter Praamstra. The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, 112(4):713–719, 2001.

[38] Roland H Grabner and Bert De Smedt. Oscillatory eeg correlates of arithmetic strategies: A training study. *Frontiers in psychology*, 3:428, 2012.

[39] Yao Wang, Yufei Shi, Zhipeng He, Ziyi Chen, and Yi Zhou. Combining temporal and spatial attention for seizure prediction. *Health Information Science and Systems*, 11(1):38, 2023.

[40] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[41] Juan Cheng, Meiyao Chen, Chang Li, Yu Liu, Rencheng Song, Aiping Liu, and Xun Chen. Emotion recognition from multi-channel eeg via deep forest. *IEEE Journal of Biomedical and Health Informatics*, 25(2):453–464, 2020.

[42] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. Emotional state classification from eeg data using machine learning approach. *Neurocomputing*, 129:94–106, 2014.

[43] Zhe Wang, Yongxiong Wang, Jiapeng Zhang, Chuanfei Hu, Zhong Yin, and Yu Song. Spatial-temporal feature fusion neural network for eeg-based emotion recognition. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022.

[44] Yi An, Ning Xu, and Zhen Qu. Leveraging spatial-temporal convolutional features for eeg-based emotion recognition. *Biomedical Signal Processing and Control*, 69:102743, 2021.

[45] Wei Zheng and Bo Pan. A spatiotemporal symmetrical transformer structure for eeg emotion recognition. *Biomedical Signal Processing and Control*, 87:105487, 2024.

[46] Yilong Yang, Qingfeng Wu, Yazhen Fu, and Xiaowei Chen. Continuous convolutional neural network with 3d input for eeg-based emotion recognition. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part VII 25*, pages 433–443. Springer, 2018.

[47] Hao Chao, Liang Dong, Yongli Liu, and Baoyun Lu. Emotion recognition from multiband eeg signals using capsnet. *Sensors*, 19(9):2212, 2019.

[48] Yilong Yang, Qingfeng Wu, Ming Qiu, Yingdong Wang, and Xiaowei Chen. Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network. In *2018 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE, 2018.