Exploring Adaptive Graph Topologies and Temporal Graph Networks for EEG-Based Depression Detection

Gang Luo, Hong Rao, Panfeng An, Yunxia Li, Ruiyun Hong, Wenwu Chen, and Shengbo Chen

***Abstract*— In recent years, Graph Neural Networks (GNNs) based on deep learning techniques have achieved promising results in EEG-based depression detection tasks but still have some limitations. Firstly, most existing GNN-based methods use pre-computed graph adjacency matrices, which ignore the differences in brain networks between individuals. Additionally, methods based on graph-structured data do not consider the temporal dependency information of brain networks. To address these issues, we propose a deep learning algorithm that explores adaptive graph topologies and temporal graph networks for EEG-based depression detection. Specifically, we designed an Adaptive Graph Topology Generation (AGTG) module that can adaptively model the real-time connectivity of the brain networks, revealing differences between individuals. In addition, we designed a Graph Convolutional Gated Recurrent Unit (GCGRU) module to capture the temporal dynamical changes of brain networks. To further explore the differential features between depressed and healthy individuals, we adopt Graph Topology-based Max-Pooling (GTMP) module to extract graph representation vectors accurately. We conduct a comparative analysis with several advanced algorithms on both public and our own datasets. The results reveal that our final model achieves the highest Area Under the Receiver Operating Characteristic Curve (AUROC) on both datasets, with values of 83% and 99%, respectively. Furthermore, we perform extensive validation experiments demonstrating our proposed method's effectiveness and advantages. Finally, we present a comprehensive discussion on the differences in brain networks between healthy and depressed individuals based on the outputs of our final model's AGTG and GTMP modules.**

***Index Terms*— Depression detection, EEG, adaptive graph topology, graph neural network.**

Manuscript received 21 April 2023; revised 25 July 2023 and 2 September 2023; accepted 26 September 2023. Date of publication 29 September 2023; date of current version 13 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62102133, in part by the Kaifeng Science and Technology Major Project under Grant 21ZD011, in part by the Jiangxi Province Science and Technology Fund under Grant 20232BCJ25026, in part by the Jiangxi Province Applied Research and Cultivation Program Fund Project under Grant 20212BAG70025, and in part by the Ji'An Finance and Science Foundation under Grant 20211085454, Grant 20222151704, and Grant 20222151746. *(Corresponding authors: Panfeng An; Hong Rao; Shengbo Chen; Ruiyun Hong.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Ethics Committee of Ji'an Third People's Hospital.

Gang Luo is with the School of Mathematics and Computer Science, Nanchang University, Nanchang, Jiangxi 330031, China, and also with the School of Computer and Information Engineering, Henan University, Kaifeng, Henan 475001, China (e-mail: luog@email.ncu.edu.cn).

Hong Rao and Panfeng An are with the School of Software, Nanchang University, Nanchang, Jiangxi 330047, China (e-mail: raohong@ncu.edu.cn; panfengan@whu.edu.cn).

Yunxia Li and Wenwu Chen are with the Department of Neurology, The First Affiliated Hospital of Henan University, Kaifeng, Henan 475001, China (e-mail: hdyfynd@126.com; chenww1629@126.com).

Ruiyun Hong is with the Psychology Department, Ji'an Third People's Hospital, Jian, Jiangxi 343000, China (e-mail: lu31061hongruiyun@163.com).

Shengbo Chen is with the School of Computer and Information Engineering, Henan University, Kaifeng, Henan 475001, China (e-mail: ccb02kingdom@gmail.com).

Digital Object Identifier 10.1109/TNSRE.2023.3320693

I. Introduction

MAJOR depressive disorder (MDD), as a prevalent psychiatric condition, is often accompanied by symptoms such as anxiety, irritability, and insomnia. Moderate to severe depression has been associated with self-harm behaviors, including suicide [1], [2], [3]. The global prevalence of depression has been increasing steadily over the years, with the number of cases rising from 171 million in 1990 to 258 million in 2017, indicating a growth rate of 49.86% [4]. Moreover, the outbreak of the COVID-19 pandemic has exacerbated the situation, leading to a surge in depression cases worldwide [4], [5], [6], [7]. Currently, clinical diagnosis of depression relies primarily on specialized physician interviews and various psychiatric rating scales, such as the Patient Health Questionaire-9 items (PHQ-9), Back Depression Inventory (BDI), and Hamilton Depression Scale (HAMD). However, specialized physician interviews are time-consuming and labor-intensive, while rating scale assessments are susceptible to deception. Consequently, the quest for a more objective and efficient diagnostic modality for depression is imperative.

Electroencephalography (EEG) boasts high temporal resolution, non-invasiveness, and low data collection costs, rendering it a commonly used tool in clinical settings for disease diagnosis. Psychological and cognitive sciences have evinced that most psychological and cognitive functions can be reflected via EEG signals [8], [9], [10]. Previous studies have employed techniques such as dimensionality reduction or extraction of frequency band signals in advance as features, followed by applying machine learning methods as classifiers to perform relevant tasks [11], [12], [13], [14]. However, the effectiveness of such methods heavily relies on the accuracy of the selected features, and there is no correlation between the classifier and the pre-extracted features. Furthermore, EEG

signals, characterized by high randomness and nonlinearity, pose challenges for traditional methods in effectively extracting features. Presently, deep learning technology has showcased remarkable performance in image recognition, natural language processing, and other fields, leveraging its powerful nonlinear fitting capacity and end-to-end learning benefits [15], [16]. Existing research endeavors have sought to apply deep learning techniques to EEG signals for tasks such as emotion recognition, motor imagery, and disease diagnosis [9], [17], [18]. Various studies have attempted to mine effective biomarkers from EEG for depression detection [19], [20], [21]. Among them, the used deep learning methods can be broadly categorized into CNN(Convolutional Neural Network)-based and GNN(Graph Neural Network)-based. CNN-based algorithms view EEG signals as images and design various convolutional kernels to extract features from them [22], [23], [24], [25], but do not entirely consider the relationship between different channels. On the other hand, GNN-based algorithms map EEG signals into graph-structured data and use pre-calculated graph adjacency matrices to represent the relationships between different channels. These methods holistically contemplate the potential spatial structural relationships among diverse channels [26], [27], [28], thereby facilitating the extraction of cross-channel-related features. However, pre-calculated methods can only generate fixed adjacency matrices, which cannot reflect the differences in individual brain networks and the dynamic changes in network connections. For instance, during rapid eye movement sleep, small-world networks change in depressed patients [29]. Yao et al. [30] found that the cross-hemisphere functional connectivity of corresponding brain regions in patients with mental illness was abnormally reduced, indicating differences in brain networks between depressed and healthy individuals. Moreover, the connectivity of brain networks may be time-varying [31], implying dynamic changes in brain network connections.

Therefore, the existing GNN-based algorithms still face the following challenges. Firstly, there are variations in the brain networks among different individuals, and the neural mechanisms of the human brain are highly complex. The current methods fail to accurately construct comprehensive brain network topological structures, particularly in simulating the dynamic changes of the brain network. Secondly, the existing research has not integrated the temporal dependency information of brain networks. To address these two challenges more effectively, we propose the following two improvement strategies:

1). In light of the correlation between the strength of functional links among brain hemispheres and distance, we introduce a universal graph adjacency matrix based on the distance to depict the universal connectivity status of the subject's brain network employing the spatial distance amid various electrodes. To consider the individual differences in brain networks and the variability of brain network connections, we propose using learnable parameters to compute the correlation matrix among various channels and mutually correct the correlation matrix and universal adjacency matrix. Finally, during training, the model adaptively adjusts the learnable parameters to construct the final graph adjacency matrix, thereby generating corresponding adjacency matrices for different EEG slices.

2). We integrate GNN and Gated Recurrent Unit (GRU) to capture spatiotemporal dependencies in EEG signals. We employ the GNN to aggregate node features in graph data for obtaining spatial correlations, and the GRU is utilized to capture the dynamic changes of brain networks and obtain time series correlations.

Furthermore, we introduce the self-attention graph pooling idea [32] and propose a Graph Topology-based Max-Pooling (GTMP) module for further enhancement. Pooling layers are commonly utilized in deep learning models to reduce the number of model parameters, enhance computational efficiency, and improve the robustness of feature extraction. However, the conventional pooling function is too simplistic and overlooks the rich topological structure information present in the graph. Differing from conventional pooling functions, we calculate node importance scores founded on graph topology and generate a node mask, retaining the feature vector of the highest-scoring node as the representation vector of the graph.

In summary, we propose a deep learning method for depression detection based on EEG signals. Our contributions are as follows:

- We adroitly model the connectivity status of brain networks based on EEG signals to acquire a supple and accurate depiction of the graph structure.
- We integrate the GRU architecture to grasp the temporal variation information of brain network connectivity. To the best of our knowledge, we are the first to consider the connectivity changes of brain networks in depression detection.
- To bolster the efficacy of our proposed method, we introduce a GTMP module that can acquire the feature representation of the graph more accurately.

The rest of the paper is organized as follows: Section II briefly delineates related research. In Section III, we elaborate on our proposed methodology and specifics. Section IV include experimental setup and results. We discuss the brain network structure and pooling nodes in Section V. Finally, the conclusion of the paper is presented in Section VI.

II. Relate Work

A. EEG-Based Brain Network Topology

Many studies have mapped EEG signals as graph-structured data, and the adjacency matrix is the explicit core feature of graph-structured data. Existing research on the topological relationship of brain networks can be roughly divided into two categories, adjacency matrices based on predefined and adjacency matrices based on neural network generation. Adjacency matrices based on predefined methods calculation formula is designed according to prior knowledge, such as biomedicine. The adjacency matrix is explicitly defined before deep learning model training. For example, Chen et al. [28] designed the adjacency matrix of the brain network according to the local correlation and global correlation between multiple electrodes and designed three kinds of connections across hemispheres. Tang et al. [27] designed the adjacency matrix of the two graphs according to the spatial position of the

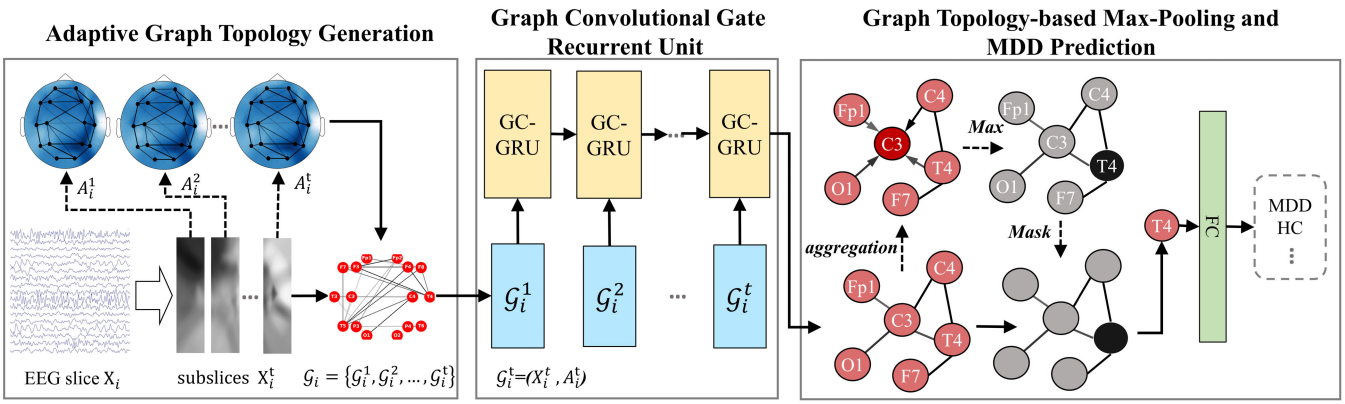

Fig. 1. Overall model architecture diagram.

electrodes and the normalized cross-correlation of different channels. However, none of these methods considers individual brain network connectivity differences.

The method based on neural network generation adopts the design of learnable modules inside the model and adaptively generates the adjacency matrix. For example, GraphSleepNet adaptively learns the internal connection between different EEG channels according to the input features, uses rectified linear unit to ensure the non-negative adjacency matrix, and defines the loss function to control the sparsity of graph topological connections [33]. Song et al. [34] divided the original EEG signal into frequency bands to calculate the adjacency matrix under different frequency bands to obtain a more accurate representation of the topological structure of the brain network. Different from modelling the adjacency matrix and then calculating the laplacian matrix, the work [35] proposes to learn the embedding representation of the nodes and then define the graph structure through the node similarity. This method directly generates a symmetric normalized Laplacian matrix to avoid unnecessary double calculations. Wang et al. [36] have employed temporal convolutional blocks to extract temporal features of channels as node attributes and subsequently represented the connectivity between nodes using parameterized adjacency matrices. Although the existing research methods have achieved acceptable results, they have not explored the dynamic variability of network topology.

B. EEG-Based Temporal Feature Fusion

As natural high-time resolution data, the EEG signal has a time dimension. Many studies are constantly exploring the time series dependence in the EEG signal, which is used to realize tasks such as emotion recognition and disease diagnosis, and have achieved acceptable results. Existing research can be roughly divided into two categories: models with stacked time series modules and time series hybrid models. The model with a stacked time series module can mine the timing relationship in features by adding individual timing modules in the processing flow. For example, Hu et al. [37] adopts the tunable Q-factor wavelet transform preprocess of the epileptic seizure features extracted by the deep convolutional variational autoencoder, then input into the stacked bidirectional LSTM to detect epilepsy. Xing et al. [38] proposed using the linear EEG mixed model of stacked autoencoder to decompose the source EEG signal, extract serialized features, and then input LSTM to capture time-series related information to realize emotion classification.

Time series hybrid models that integrate multiple network structures directly fuse the time series model structure and other network structures to construct a new hybrid model to extract features in multiple dimensions. For example, Xu et al. [39] proposes a hybrid deep learning model GRU-Conv based on GRU and CNN to extract sleep classification features and perform well on SEED and DEA datasets. ST-GCLSTM is equipped with a spatial graph convolutional network module and an attention-enhanced bidirectional LSTM module, which extracts crucial spatiotemporal features from continuous EEG signal data for emotion recognition [40]. Similarly, Wu et al. [41] designed a spatiotemporal convolution block based on the Graph Convolution (GConv) block and Time Convolution (TConv), and cyclically applied the GConv block and TConv blocks are used to capture both spatial and temporal features. ATDD-LSTM is integrated into the timing module LSTM to extract timing features while considering the nonlinear channel relationship [42]. In addition, Kumar et al. [43] also used the time-series features based on EEG signals to predict the trend of EEG changes in patients with depression at the next moment. Although existing studies have achieved competitive results in many EEG tasks, they cannot be directly applied to EEG-based MDD detection tasks. As mentioned earlier, in the case of variable brain networks in patients with depression, the challenge of simultaneously fusing spatial structure and temporal dependence information is still the focus of our research.

III. Method

This section introduces the overall architecture of the model. Fig. 1 illustrates the general framework of the model, which consists of three stages. First, we explore the brain network connectivity status reflected by EEG and calculate the adjacency matrix. Then, we use multi-layer GCN and GRU to fuse multi-dimensional information, exploring the spatial correlation between different channels and the temporal dependency of brain network changes. Finally, we compute node importance based on graph topology and retain the most representative nodes for predicting depression.

A. Adaptive Graph Topology Generation

Diagnosing depression using GNN typically necessitates pre-calculating the adjacency matrix, which overlooks the distinctions in brain network connectivity between depressed and healthy subjects. To tackle this issue, we propose an Adaptive Graph Topology Generation (AGTG) module that constructs the topological connections of a graph based on EEG signals in an adaptive manner, i.e., adjacency matrix generation. Our method aims to obtain more flexible and precise connectivity by integrating spatial distance and correlations between different channels.

Firstly, we introduce a distance-based universal graph adjacency matrix A_{com} that characterizes the universal connectivity state of brain networks. We calculate it using the spatial distance between various electrodes and express it as follows:

$$A_{com} = \{A|A_{ij} = norm(dis(i,j))\}, \quad (i,j = 1,2,\ldots,E) \tag{1}$$

Here, $norm$ denotes the normalization function, and $dis(i,j)$ represents the Euclidean distance between electrodes i and j according to the EEG electrode placement. E denotes the number of nodes in the graph structure, that is, the number of EEG channels.

Next, for the input EEG signals X, we left-multiplies the matrix P to capture the correlations between different channels and right-multiplies the mapping matrix Q to map the correlations onto an E-dimensional matrix A_{cor}. We use the absolute value function to capture the mutually exclusive or synchronous correlations between different channels and obtain the correlation matrix A_{cor}, which we express as follows:

$$A_{cor} = abs(PXQ + b) \tag{2}$$

Here, abs denotes the absolute value function, and $P \in \boldsymbol{R}^{E\times E}$ and $Q \in \boldsymbol{R}^{F\times E}$ are learnable parameter matrices, where F denotes the feature dimension and b is the bias value.

Finally, we allow the universal graph adjacency matrix and the correlation matrix to mutually correct each other to determine the final graph adjacency matrix, defined as the following formula:

$$\boldsymbol{A} = Relu(A_{cor} + A_{com} * d) \tag{3}$$

Here, d denotes the learnable weight parameter of the universal graph adjacency matrix, and the Relu function ensures that all elements in the matrix are non-negative.

B. Graph Convolutional Gate Recurrent Unit

In this section, we present the Graph Convolutional Gate Recurrent Unit (GCGRU) module, as illustrated in Fig. 2. The GCGRU module combines the GCN structure and the GRU structure to perform spatiotemporal feature extraction on the graph structure generate by the AGTG module and capture the dynamic changes in brain network connectivity. Specifically, for each EEG sample X_i, we partition it into a series of fixed-length EEG sub-slices X_i^t. We then utilize the AGTG module to generate a graph adjacency matrix G_i^t for each sub-slice, with the original time-domain data acting as the node features of the graph. Finally, the EEG sample X_i was transformed

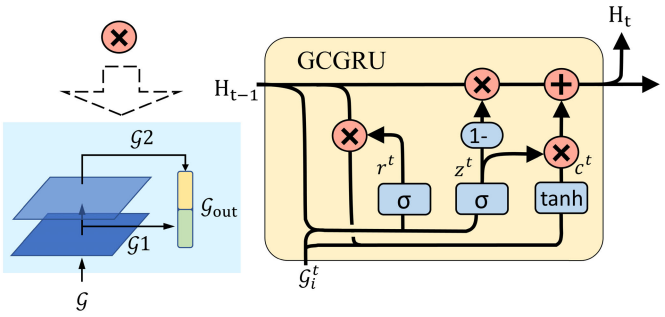

Fig. 2. GCGRU structure.

into a series of graphs $G_i = (G_i^1, G_i^2, \ldots, G_i^t)$, and we stack multiple GCGRU to capture the spatiotemporal dependency features in G_i. Specifically, at each time step t, we input the corresponding G_i^t into the GCGRU and use the last GCGRU output H_t as the input for the next module. The design of the GCGRU module is depicted below:

$$\begin{aligned} r^t &= \sigma(W_r\left[F(X_i^t), H^{t-1}\right] + b_r) \\ u^t &= \sigma(W_u\left[F(X_i^t), H^{t-1}\right] + b_u) \\ c^t &= tanh(W_c\left[F(X_i^t), (r^t \odot H^{t-1})\right] + b_c) \\ H^t &= u^t \odot H^{t-1} + (1-u^t) \odot c^t \end{aligned} \tag{4}$$

where H^{t-1} denotes the hidden state of the previous time step, r^t and u^t represent the reset gate and update gate of the GCGRU, respectively, and c^t represents how much information from the input X_i^t will be retained in the hidden state vector H^t. The function $F()$ represents the multiple layers of GCN, and W_r, W_u, W_c, b_r, b_u, and b_c are the parameter matrices and bias values for each calculation formula.

To further improve the model's feature extraction capability, we introduce the idea of residual networks [15] and design a cross-layer concatenation operation in the multiple layers of GCGRU to fuse outputs from different layers. The formula for the two-layer GCN is shown below:

$$\begin{aligned} F(X) &= cat\left[f_1(X), f_2(X)\right] \\ f_1(X) &= Relu(\hat{A}XW_1) \\ f_2(X) &= Relu(\hat{A}Relu(\hat{A}XW_1)W_2) \end{aligned} \tag{5}$$

where X symbolizes the feature matrix, A stands for the adjacency matrix engendered by the AGTG module. $\hat{A} = \widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}$ represents the pre-processing step. Here, $\widetilde{A} = A + I_N$ is the matrix of self-connection structures. $\widetilde{D}$ refers to the degree matrix, and $\widetilde{D} = \sum_j \widetilde{A}_{ij}$. W_0 and W_1 denote the parameter matrices for the first and second-layer graph convolutions, respectively. Furthermore, $Relu()$ is the activation function, and $cat()$ denotes the feature concatenation function.

C. Graph Topology-Based Max-Pooling and MDD Prediction

To further enhance the model's performance, we introduce the idea of self-attention graph pooling [32] and devise a Graph Topology-based Max-Pooling (GTMP) module that

TABLE I
COMPARE WITH OTHER EXISTING MODELS ON PUBLIC DATASET. USE BOLD FOR BEST RESULTS. P-VALUE OF OUR METHOD OVER THE METHOD: * INDICATING (P < 0.05), ** INDICATING (P < 0.01), *** INDICATING (P < 0.001)

Model	ACC(%)	F1-score(%)	REC(%)	PRE(%)
ShallowConvNet [23]	68.76***	77.95***	92.82***	68.21***
DeepConvNet [23]	63.17	76.08*	**99.22**	61.71*
EEGNet [24]	65.62**	77.03**	97.38**	63.82***
TSception [22]	74.53	81.73	96.30***	71.04**
T-GCN [47]	65.87***	77.06***	97.10***	63.87***
DCGRU [27]	65.34**	76.42**	94.93*	64.02*
LGGNet [44]	77.70	82.39	88.47	**77.22**
Our	**77.78**	**82.75**	90.23	76.46

TABLE II
COMPARED WITH OTHER EXISTING MODELS ON OUR OWN DATASET. USE BOLD FOR BEST RESULTS. P-VALUE OF OUR METHOD OVER THE METHOD: * INDICATING (P < 0.05), ** INDICATING (P < 0.01), *** INDICATING (P < 0.001)

Model	ACC(%)	F1-score(%)	REC(%)	PRE(%)
ShallowConvNet [23]	89.49	87.68	91.20	85.31
DeepConvNet [23]	97.40	96.86	97.58	**96.18**
EEGNet [24]	93.16	91.81	94.42	89.43
TSception [22]	93.13	91.77	93.10	90.54
T-GCN [47]	78.73	75.80**	81.61***	70.89**
DCGRU [27]	92.33	90.34	88.36	92.48
LGGNet [44]	**97.77**	**97.30**	**98.65**	96.00
Our	95.61	94.61	94.85	94.39

considers both node features and graph connectivity, thereby accurately extracting the representations of the graph.

In this module, we first evaluate the significance of each node in the graph based on the output H_t of the GCGRU module. We assume that if a node is involved in the feature aggregation of several other nodes, it indicates a higher degree of importance. Therefore, we execute feature aggregation based on the graph topology structure to compute the importance score S_{node} of each node, as per the following expression:

$$S_{node} = Relu(\hat{A} H_t W + b) \tag{6}$$

where $H_t \in \boldsymbol{R}^{E \times h}$ represents the graph feature matrix produced by the GRU module, where h is the hidden state dimension. $\hat{A}$ is the symmetrically normalized laplacian matrix corresponding to the graph feature H_t; W and b are the parameter matrix and bias value, respectively.

Next, we sort the nodes based on their importance scores and return the indices of the top num nodes, denoted by N_{idx}. Since we employ max-pooling, $num = 1$. N_{idx} is defined as follows:

$$N_{idx} = top - rank(S_{node}, num) \tag{7}$$

where $top - rank()$ is a function that returns the indices of the top num nodes with the highest scores.

Finally, we generate a node mask based on the node's index N_{idx} and apply it to the graph feature matrix H_t, retaining the most important node feature vector as the graph representation vector V_{graph}, defined as follows:

$$V_{graph} = mask(N_{idx}) \odot H_t \tag{8}$$

where $mask()$ is a mask generation function that constructs a node mask vector based on the node indices.

After that, we input the graph representation vector V_{graph} produced by the GTMP module into a fully connected layer to calculate the depression probability.

IV. EXPERIMENT AND RESULTS

A. Datasets

1) Dataset 1: Resting-State Depression Dataset: The University of New Mexico collected the resting-state depression EEG dataset [45]. The participants were recruited from a broad survey of the BDI taken in an introductory psychology course at Arizona State University. All the participants provided their written consent after the approval of Arizona State University. The dataset comprises 122 subjects, of which 46 are depressed or have high scores on the BDI. The participants' BDI scores were kept stable throughout the testing and experimental evaluation phase. Information such as age, gender, BDI, Spielberger Trait Anxiety Inventory, and cognitive and affective subscale scores were recorded during the experiment. EEG data were collected using 66 Ag/AgCl electrodes, with bandpass filtering set between 0.5Hz-100Hz, and a sampling rate of 500Hz, with impedance $< 10k\Omega$. We selected 63 subjects as experimental data, including 33 depressed patients and 30 healthy subjects. The dataset can be found at: https://openneuro.org/datasets/ds003478/versions/1.1.0

2) Dataset 2: Our Depression EEG Dataset: We collaborated with the Third People's Hospital of Jian City, China, improved the HAMD, and gathered EEG data from subjects. The Third People's Hospital of Ji'an City recruited subjects and obtained Written Informed Consent Forms.

- Inclusion criteria for subject: age between 18 and 60 years old; right-handedness. Depressed patients who meet the International Classification of Diseases (ICD-10) diagnostic criteria.
- Exclusion criteria for subject: the presence of severe organic diseases such as malignant tumors and a history of cardiovascular or cerebrovascular diseases, pregnant women, individuals with neurological disorders or a family history of mental illness, and individuals with communication impairments.

The dataset was recorded using the international 10-20 system, comprising 16 effective recording electrodes and two reference electrodes, with a sampling rate of 500Hz. Each subject's EEG data consisted of two eyes-open and closed periods, typically between 6 and 8 minutes. After collecting EEG data, the subjects completed the HAMD questionnaire. We collected data from 40 subjects, including 18 depression patients and 22 healthy subjects.

3) Data Processing: In order to mitigate the impact of data quality and the number of EEG channels on the model performance assessment, we extract the common EEG data from 16 channels across two datasets. These channels include Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, and T6. We apply identical preprocessing steps to both datasets

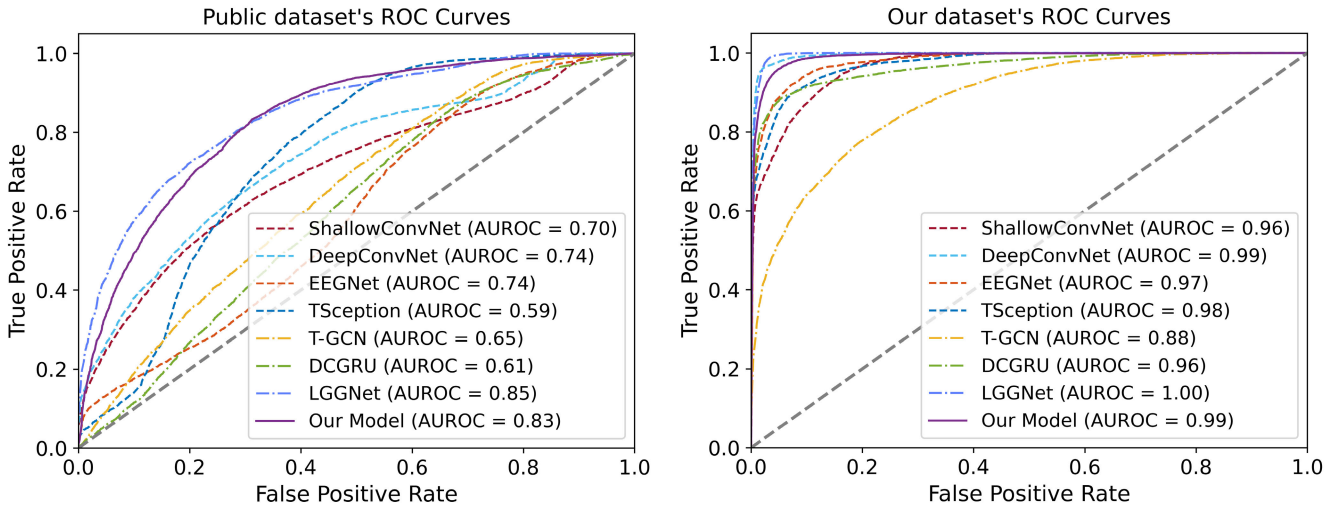

Fig. 3. ROC curves of all models on different datasets. The left figure is a comparison chart of the ROC curve on the public dataset, and the right is the comparison chart of the ROC curve on our own dataset. Each ROC curve plotted therein takes the mean of a five-fold cross-validation.

as follows: considerate the potential of the gamma rhythm within the frequency range of 40Hz-100Hz as a promising biomarker for MDD, and to retain the integrity of the common EEG frequency bands, we utilize a bandpass filter to extract EEG signals ranging from 0.5Hz to 100Hz. Additionally, we utilize a notch filter at 50Hz or 60Hz to remove power line interference. To reduce the number of data points while preserving the information within the signal, we resample the EEG signals to 200Hz. Lastly, we employ the Faster algorithm [46] to automatically remove artifacts such as eye movement and muscle activity from the EEG signals.

B. Implementation Details

In order to ensure that the model evaluation is not affected by class imbalance, we select a nearly equal number of individuals with depression and healthy individuals for the experiment dataset. We construct the class labels based on the depression scale scores of the participants, considering participants with scores below seven as healthy individuals and those with scores greater than or equal to 17 as individuals with depression. We use a sliding window strategy to segment the clean EEG data to build the dataset. Since deep learning models typically require a large number of data samples for training, we utilize an overlapping sliding window technique to augment the sample size. Specifically, we employed a window length of 6 seconds and slid it over the participants' EEG data with a step size of one-third of the window length, which is 2 seconds. Each EEG slice was assigned a binary label based on the participant's class label. It is important to note that since we employ an individual-based train-test split, EEG slices from the same individual belonged to the same class. Finally, we obtained 15,150 and 14,462 EEG slices from the public and our own datasets.

We determine the following hyperparameters based on the best performance on the validation set: (a) EEG slice length and sub-slice length; (b) dimension of the GRU hidden state; (c) the number of GCGRU layers. The remaining hyperparameters of the model were set as follows: initial learning rate of 2e-4 with automatic adjustment using the cosine annealing algorithm. We used a batch size of 40 EEG clips, training epochs of 50, and a dropout probability of 0.7. We utilize the Binary Cross Entropy with Logits Loss as the loss function to train the depression detection model. All experimental results in this study were based on five-fold cross-validation.

C. Comparison With Other Methods

To validate the efficacy of our proposed method, we conducted a comparative analysis with various existing methods on both public and our datasets, including ShallowConvNet [23], DeepConvNet [23], EEGNet [24], TSception [22], T-GCN [47], DCGRU [27], and LGGNet [44]. We first provide a brief overview of the similarities and dissimilarities between our proposed method and the other methods.

1) CNN-Based: Conventional CNN algorithms treat EEG signals as images and apply diverse convolutional kernels to extract features. We compared four CNN-based methods. The study [23] proposed a universal convolutional architecture based on EEG and implemented two cases. DeepConvNet comprises five convolutional layers and a softmax function, while ShallowConvNet consists of temporal and spatial convolutional layers and then squaring function, an average pooling layer, and a logarithm function. EEGNet leverages deep and separable convolutions for feature extraction and demonstrates exceptional robustness across various BCI paradigms. TSception employs multi-scale convolutional kernels within the designated time and spatial learners to acquire distinct time or spatial feature representations.

2) GNN-Based: Similar to our model, T-GCN accounts for both temporal and spatial dependent features. It constructs graph-structured data founded on regional location and traffic velocity, which is employed for spatiotemporal traffic prediction. The study [27] devised two graph adjacency matrices based on prior knowledge, utilizing EEG temporal features or logarithmic amplitude after Fourier transformation as the feature of each graph node. DCGRU combines diffusion convolution and gated recurrent units and attains exceptional seizure detection and classification performance. Ding et al. [44] introduced neuroscientific prior knowledge and defined

TABLE III

ABLATION EXPERIMENTS FOR AGTG, GCGRU, AND GTMP MODULES. AMONG THEM, "✓" MEANS TO INCLUDE THE POLICY, AND THE SYMBOL "-" MEANS TO DELETE THE POLICY. TO IMPROVE THE CLARITY AND READABILITY OF THE MODELS, WE NUMBER THE VARIOUS MODELS ACCORDING TO DIFFERENT EXPERIMENTAL SETTINGS: S1–S8. THE BEST RESULTS ARE INDICATED IN BOLD. P-VALUE OF OUR METHOD OVER THE METHOD: * INDICATING (P < 0.05), ** INDICATING (P < 0.01), *** INDICATING (P < 0.001)

Index	AGTG	GCGRU	GTMP	ACC (%)	F1-score (%)	REC (%)	PRE (%)	AUROC (%)
S1.	-	-	-	68.60***	77.95***	93.93	66.66***	71.98**
S2.	-	-	✓	68.39***	78.06***	**95.30*** **	66.11***	65.34***
S3.	-	✓	-	70.65***	78.46***	90.56	69.23***	75.14**
S4.	-	✓	✓	69.44***	78.31***	93.47**	67.39***	66.33***
S5.	✓	-	-	71.07**	79.40**	94.46***	68.50***	73.43**
S6.	✓	-	✓	72.42**	80.52*	96.58**	69.08***	70.35**
S7.	✓	✓	-	71.05**	78.34***	88.93	70.55	74.45**
S8.	✓	✓	✓	**77.78**	**82.75**	90.23	**76.46**	**83.27**

three types of local-global graph structures, which were utilized for feature fusion in the graph learning block based on local and global connections.

We implement identical data preprocessing and cross-validation strategies to those of other models and conduct experiments on two datasets. The results are reported in Table I and Table II, indicating that our proposed method outperforms other models in detecting depression in both datasets. Furthermore, the paired T-test is utilized for the statistical analysis of the public and our own datasets. We also report the p-value in Table I and Table II and illustrate that the performance improvements of our proposed method are statistically significant. Specifically, on the public dataset, ShallowConvNet, DeepConvNet, EEGNet, and TSception achieves accuracies of 68.76%, 63.17%, 65.62%, and 74.53%, respectively. In contrast, our proposed method achieves an accuracy of 77.78%, demonstrating the effectiveness of using graph-structured data to represent EEG signals. Furthermore, our proposed method outperforms T-GCN and DCGRU by 11.91% ($p < 0.001$) and 12.44% ($p < 0.01$), respectively, confirming the superiority of our proposed model. On our own dataset, our model achieves an accuracy of 95.61%, which is only slightly lower than that of DeepConvNet in other classification metrics. The accuracy of ShallowConvNet and TSception on our own dataset are 89.49% and 93.13%, respectively, which are lower than our method by 6.12% and 2.48%. Although DCGRU also achieved an accuracy of 92.33%, it is still lower than our method. It is worth noting that although DeepConvNet achieved the highest accuracy on our own dataset, it performed extremely poorly on the public dataset, indicating its susceptibility to the influence of data quality. In contrast, our proposed method demonstrates more stable and excellent MDD detection performance on both datasets. Additionally, our method achieves the comparable depression detection performance on two datasets compared to the LGGNet model. Notably, our model params (56,388) is only one-fifteenth of the LGGNet model params (882,334), significantly reducing computational complexity and deployment costs. We also show the Receiver Operating Characteristic (ROC) curves of different methods on two datasets. As shown in Fig. 3, our method achieved a True Positive Rate (TPR) of 81.36% at a low False Positive Rate of 25% on the public dataset, while the TPR of other methods is all lower than 70%. Similarly, our method exhibits excellent performance on our own dataset, achieving a TPR of 99.83%. Furthermore, the area under the ROC curve of our model on the two datasets is 83% and 99%, respectively, indicating that our proposed method shows the best depression detection performance.

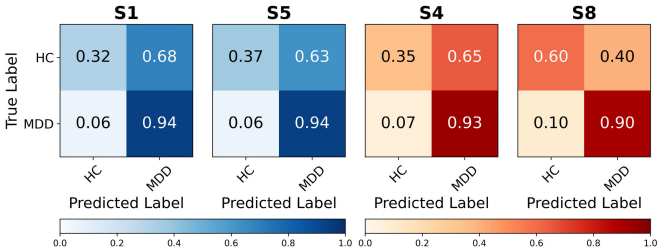

Fig. 4. Confusion matrix comparison of AGTG module.

D. Ablation Experiment

We conduct extensive ablation experiments on public dataset to demonstrate the efficacy of our proposed three optimization strategies.

1) Adaptive Graph Topology Generation: we compare the models utilizing AGTG module (S5.) with those utilizing predefined graph topology (S1.), under the same conditions. As per the experimental results present in Table III, the AGTG module (S5.) model shows a higher accuracy of 71.07% compared to the model with the predefined graph topology (S1). In addition to the comparison above, all models utilizing the AGTG module always outperform the models with predefined graph topology, i.e., $S5 > S1$, $S6 > S2$, $S7 > S3$, and $S8 > S4$. Furthermore, we plot confusion matrices to analyze the experimental results further. In order to facilitate the analysis of the confusion matrix, we assign different colors to different control groups and normalize each row of the confusion matrix. Moreover, we visualize confusion matrices for two groups based on the experimental results to further assess the model's performance, as depicted in Fig. 4. Our proposed method can achieve a TPR of over 90%, indicating that our method can detect most depression patients. Additionally, Fig. 4 shows that the AGTG module can further reduce the False Negative Rate (FNR). For instance, compared

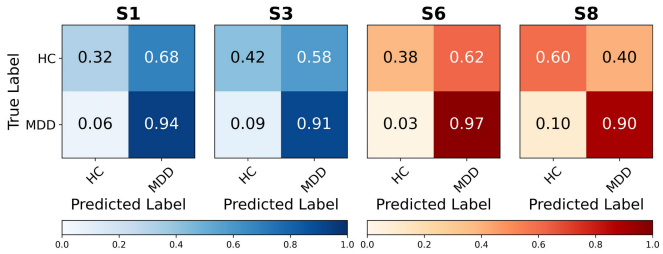

Fig. 5. Confusion matrix comparison of GCGRU module.

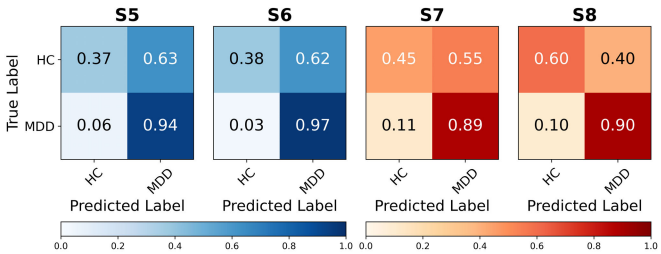

Fig. 6. Ablation experiments for the GTMP module.

to S4, S8 exhibits a 25% reduction in FNR, while S5 achieves a 5% lower FNR than S1. These experimental results suggest that the AGTG module can accurately depict the real-time connectivity status of the brain network, reveal the differences in brain networks between different individuals, and enable the model to more accurately aggregate graph node features. Furthermore, unlike pre-calculated adjacency matrices, the AGTG module can automatically learn an optimal adjacency matrix without human intervention.

2) Graph Convolutional Gate Recurrent Unit: Table III also demonstrates that temporal dependencies in brain networks can improve the model's performance. For instance, the model utilizing the GCGRU module (S8.) attains an accuracy of 77.78%, which is 5.36% ($p < 0.01$) higher than the model not employing the GRU structure (S6.). This trend holds for other variants of models that integrate the GRU module, namely $S3 > S1$, $S4 > S2$, and $S8 > S6$. Moreover, We plot two sets of confusion matrices for further analysis. According to the confusion matrices of groups (S1. - S3.) and (S6. - S8.) in Fig. 5, We can observe that models employing the GCGRU module demonstrate higher TPR and TNR. These demonstrate that integrating GRU can further explore differential features between depressed and healthy individuals while also validating the effectiveness of the temporal dependency information in brain networks.

3) Graph Topology-Based Max-Pooling: Furthermore, our model's performance is improved by utilizing the GTMP module. Table III shows that the model utilizing conventional max-pooling (S7.) achieves a 71.05% accuracy rate, which is 6.73% ($p < 0.01$) lower than the model employing the GTMP module (S8.). Similarly, the simultaneous application of the GTMP module alongside the AGTG module yielded significant performance improvements in the model, i.e., $S8 > S7$ and $S6 > S5$. The confusion matrix results in Fig. 6 also demonstrate that the GTMP module can improve both TPR and TNR. As a result, we believe there are differential features between depressed and healthy participants in the structural information, such as the graph topology information included in this strategy. Furthermore, we compare GTMP and average pooling under the same experimental settings, showing that the accuracy of employing the GTMP module was 2.23% higher than average pooling. These outcomes confirm the effectiveness of GTMP. Unlike traditional pooling, the GTMP module considers the topology information and node features of the graph and can preserve the useful information in the graph features to the fullest extent.

4) Graph Convolution Layers: In this section, we explore the impact of the number of internal graph convolution layers L

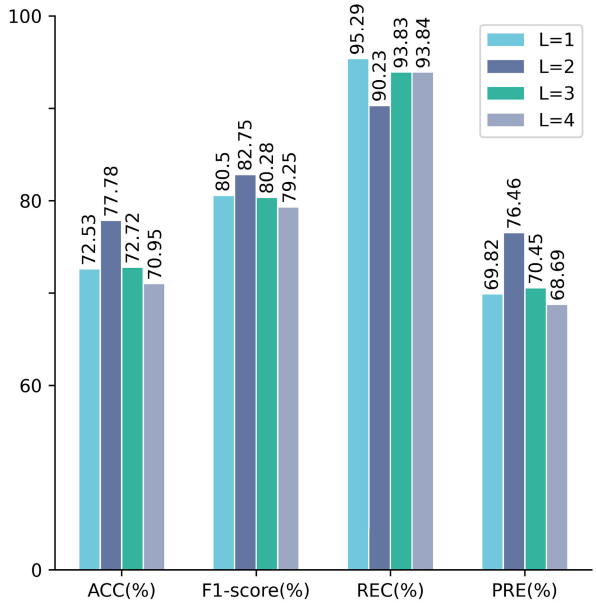

Fig. 7. Ablation experiments for the number of graph convolutional layers.

in GCGRU on the model's performance. As demonstrated in Fig. 7, the effect of L on the model performance displays a similar trend. When the number of layers is minimal, the model's capacity to extract features is insufficient. With an increase in the number of graph convolution layers to two, the model exhibits optimal performance. However, as L increases further, such as $L = 3$, we observe a significant decline in the model's performance. One plausible reason for this phenomenon is that the model's complexity is insufficient with fewer layers, making it challenging to fit the optimal classification function. With excessive L, the model becomes more challenging to train.

5) Ablation Experiment of EEG Slice Length and Sub-Slice Length: We also discuss the impact of different EEG slices and sub-slice lengths on the model's performance, as shown in Fig. 8. Firstly, we compare the effect of slice length on model performance with the same sub-slice length. The results indicate that longer slice lengths can slightly improve model performance but significantly increase the model training time. For example, under the setting of the 1s sub-slice, the accuracy of the model trained with a 6s slice is 6.06% lower than that trained with a 12s slice. Similarly, under the setting of a 2s sub-slice, the accuracy of the model trained with the 12s slices is 6.06% lower than that trained with the 6s slices. However, longer slice lengths mean the number of GCGRU iterations

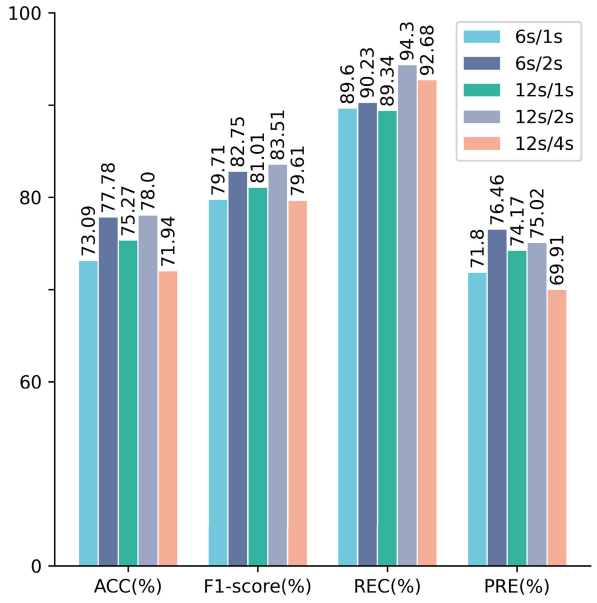

Fig. 8. Ablation experiments for EEG slice and sub-slice length selection.

increases, significantly increasing model training time. We also compare the performance of three sub-slice length settings under the 12s slice length setting. The results show that excessively long sub-slice lengths can lead to poorer model performance. For example, under the 12s slice setting, the accuracy of the model trained with the 4s sub-slice is 6.06% lower than that trained with the 2s sub-slice and 3.33% lower than the model trained with the 1s sub-slice. One possible reason for this phenomenon is that we used a bandpass filter of 0.5Hz-100Hz in the data preprocessing stage, and the 2s sub-slice length exactly contains a complete lowest frequency waveform.

6) GCGRU Hidden State Dimension: This section discusses the impact of the hidden state dimension H within GCGRU on the model's performance. As shown in Fig. 9, the model fails to achieve the desired fitting level when H is relatively small. The model demonstrates its optimal performance when H is set to 64. However, as H continues to increase, the model's performance suddenly deteriorates. Despite the upward trend in the metrics, they still fall short of the performance achieved with H set to 128. Additionally, the model's parameters drastically increase as H becomes more larger. We attribute this phenomenon to the concatenation of outputs from multiple graph convolutions within the model, which limits the output size of each GCGRU. Specifically, a small hidden state dimension forces the graph convolution to compress features into a smaller dimension, which leads to the loss of feature information. On the other hand, an excessively large hidden layer dimension leads to a sharp increase in the overall parameter, making the model more prone to overfitting.

7) Exploring EEG Personalized Information and Dataset Segmentation Strategy: In order to probe the effect of personalized information on the model, we conduct experiments on a public dataset by dividing the training set based on either the personal or slice. We report the experimental result in Table IV, which indicate that the slice-based division of the training set result

TABLE IV
COVERAGE SLICE AND THE VALIDATION OF PERSONALIZED INFORMATION

	ACC (%)	F1-score (%)	REC (%)	PRE (%)
Slice w/o overlapping	76.85	80.92	93.32	71.68
Slice w/ overlapping	88.83	89.85	94.14	85.93
Personal w/o overlapping	70.72	79.31	95.10	68.03
Personal w/ overlapping	77.78	82.75	90.23	76.46

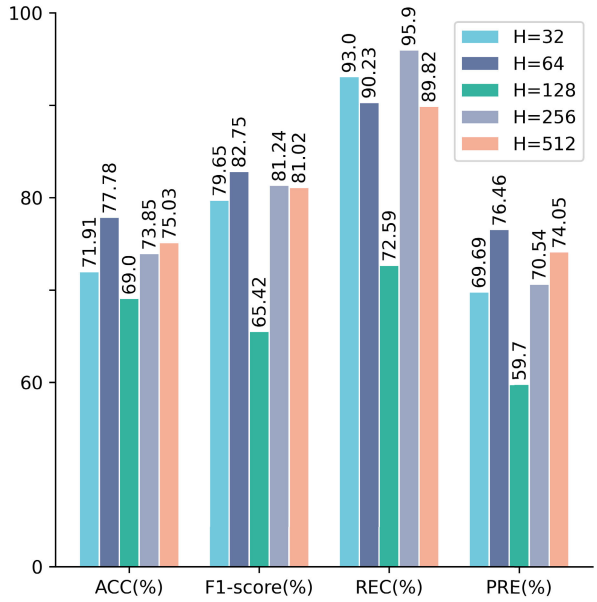

Fig. 9. Ablation experiments for GCGRU's hidden state dimension.

in higher model accuracy. Thus, we believe the subject's personalized information within the EEG signals exists. To be precise, when the training, validation, and testing sets are divided by slice, some EEG slices from an individual are included in both the training and testing sets, allowing the model to capture the subject's personalized information during training. However, while such personalized information can enhance the model's accuracy, it can also interfere with tasks such as disease diagnosis. Thus, we suggest dividing the dataset by individual to ensure the model's actual task effectiveness.

Furthermore, we compare the impact of overlapping and non-overlapping slice strategies on model performance during the sliding window EEG slicing period. The experimental findings in Table IV demonstrate that the overlapping slice strategy somewhat improves the model's performance. This improvement can be attributed to the fact that the overlapping slice strategy can obtain more training samples from the original EEG data, leading to better model performance.

V. DISCUSSION

We thoroughly discuss the practical significance of the AGTG and GTMP modules. To analyze the differences in brain network connectivity between individuals with depression and healthy individuals, we visualize and compare the adjacency matrices. To succinctly depict the connectivity, we compute the sum of the adjacency matrix and its transpose,

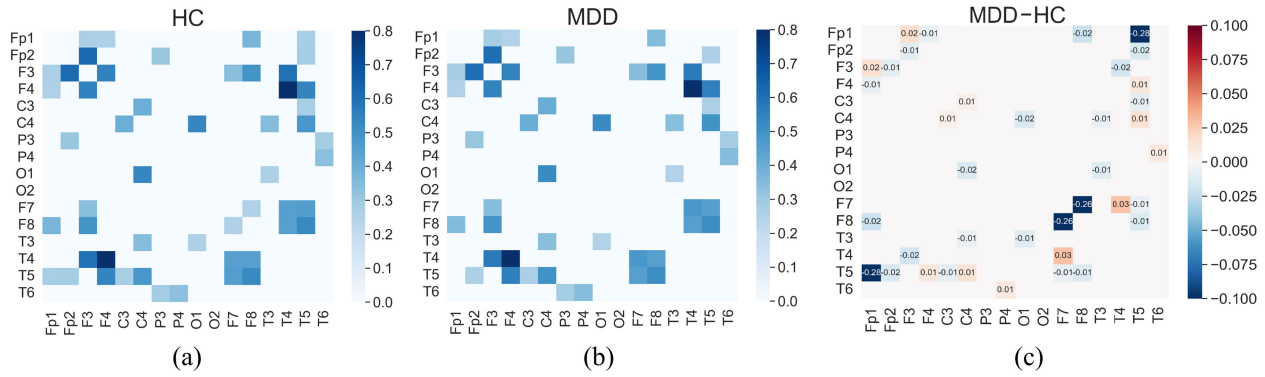

Fig. 10. Comparison of effective connectivity of brain networks in healthy and depressed subjects.

resulting in an undirected graph. We randomly select multiple EEG slices from the subject, generate adjacency matrices, and average them to represent the current subject's brain network. Previous research has indicated that the effective connectivity of brain networks accounts for approximately 20% [48]. To compare the differences in effective connectivity between individuals with depression and healthy subjects, we set a connection threshold of $q = 0.25$ to filter out weak connections in the adjacency matrix. We conducted comparative experiments and analysis by randomly selecting three healthy and three depressed subjects from the test dataset. In order to provide a more intuitive illustration of brain connectivity difference, we calculate the corresponding difference matrix (MDD-HC matrix) by subtracting healthy subjects' brain adjacency matrix from that of depressed subjects. One set of results from the visualization is shown in Fig. 10. Compared to healthy participants, we observe abnormal connectivity changes between the T5 node in the temporal lobe and the frontal nodes Fp1 and Fp2 in depressed patients. We conduct a comparative analysis of the graph connectivity patterns construct using the AGTG module and the predefined method, revealing a greater tendency of the AGTG module to establish interhemispheric connections. Consistent with the research findings by Yao et al. [30], we also found a widespread reduction in interhemispheric connectivity in the brain networks of depressed patients. For instance, in Figure 10, we observe weakened connections such as the frontal nodes Fp1-Fp2 and the central nodes C3-C4. Consequently, we believe that the abnormal connectivity between the temporal and frontal lobes and the reduced interhemispheric connections may be associated with depression.

Furthermore, in order to investigate the relationship between the positions of maximum pooling nodes and depression, we conduct a statistical analysis of the pooling node locations in both public and our own datasets. Our findings reveal that the pooling nodes are predominantly distributed in the parietal lobe at P3, the frontal lobe at Fp1, and the temporal lobe at T4. The selection of Fp1 as a critical node in the max-pooling indicates that the left frontal lobe contains more prominent discriminative information related to depression, which may be associated with abnormal activity in the left and right frontal cortices observed in individuals with depression [49]. Additionally, Kesebir et al. [50] demonstrated a correlation between depression scores and the activity of P3 and T4 delta, which explains why the GTMP module identifies these nodes as crucial. Further improvement in depression detection accuracy is still necessary to ensure the effectiveness and precision of analyzing brain networks and pooling nodes.

VI. Conclusion

This study presents a depression detection method based on EEG signals. We adaptively construct brain network structures for different individuals while preserving the dynamic changes in brain networks for depression detection. Our method relies on two main modules: the Adaptive Graph Topology Generation (AGTG) module and the Graph Convolutional Gate Recurrent Unit (GCGRU) module. The GCGRU module operates on the graph structure data generated by the AGTG module, extracting and fusing spatiotemporal dependency features. Additionally, we introduce a Graph Topology-based Max-Pooling (GTMP) module to extract the most representative features from the graph. Extensive experiments conducted on both public and our own datasets have demonstrated the effectiveness of our proposed method from various perspectives. The graph-based temporal method using the adaptive brain network modeling strategy can be more accurate and effective for EEG-based depression diagnosis in practical applications. As for our future work, we aim to optimize the brain network modeling method further and continue exploring the differences in brain networks among different categories of individuals.

References

[1] D. Nutt, S. Wilson, and L. Paterson, "Sleep disorders as core symptoms of depression," *Dialogues Clin. Neurosci.*, vol. 10, no. 3, pp. 329–336, Sep. 2008.

[2] M. W. M. Gijzen, S. P. A. Rasing, D. H. M. Creemers, F. Smit, R. C. M. E. Engels, and D. De Beurs, "Suicide ideation as a symptom of adolescent depression. A network analysis," *J. Affect. Disorders*, vol. 278, pp. 68–77, Jan. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032720327245

[3] N. H. Kalin, "The critical relationship between anxiety and depression," *Amer. J. Psychiatry*, vol. 177, no. 5, pp. 365–367, May 2020.

[4] O. Renaud-Charest et al., "Onset and frequency of depression in post-COVID-19 syndrome: A systematic review," *J. Psychiatric Res.*, vol. 144, pp. 129–137, Dec. 2021.

[5] N. Vindegaard and M. E. Benros, "COVID-19 pandemic and mental health consequences: Systematic review of the current evidence," *Brain, Behav., Immunity*, vol. 89, pp. 531–542, Oct. 2020.

[6] R. Lakhan, A. Agrawal, and M. Sharma, "Prevalence of depression, anxiety, and stress during COVID-19 pandemic," *J. Neurosci. Rural Pract.*, vol. 11, pp. 519–525, Sep. 2020.

[7] M. G. Mazza et al., "Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors," *Brain, Behav., Immunity*, vol. 89, pp. 594–600, Oct. 2020.

[8] E. L. Johnson, J. W. Y. Kam, A. Tzovara, and R. T. Knight, "Insights into human cognition from intracranial EEG: A review of audition, memory, internal cognition, and causality," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 051001.

[9] X. Li et al., "EEG based emotion recognition: A tutorial and review," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–57, Apr. 2023.

[10] T. Hernández-Del-Toro, C. A. Reyes-García, and L. Villaseñor-Pineda, "Toward asynchronous EEG-based BCI: Detecting imagined words segments in continuous EEG signals," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102351.

[11] A. O. Khadidos, K. H. Alyoubi, S. Mahato, A. O. Khadidos, and S. N. Mohanty, "Computer aided detection of major depressive disorder (MDD) using electroencephalogram signals," *IEEE Access*, vol. 11, pp. 41133–41141, 2023.

[12] S. Mahato, S. Paul, N. Goyal, S. N. Mohanty, and S. Jain, "3EDANFIS: Three channel EEG-based depression detection technique with hybrid adaptive neuro fuzzy inference system," *Recent Patents Eng.*, vol. 17, no. 6, pp. 32–48, Nov. 2023.

[13] R. Chatterjee, D. Guha, D. K. Sanyal, and S. N. Mohanty, "Discernibility matrix based dimensionality reduction for EEG signal," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2016, pp. 2703–2706.

[14] A. O. Khadidos, K. H. Alyoubi, S. Mahato, A. O. Khadidos, and S. N. Mohanty, "Machine learning and electroencephalogram signal based diagnosis of depression," *Neurosci. Lett.*, vol. 809, Jul. 2023, Art. no. 137313. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304394023002720

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[16] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[17] G. Sharma, A. Parashar, and A. M. Joshi, "DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression," *Biomed. signal Process. control*, vol. 66, Apr. 2021, Art. no. 102393.

[18] Z. Huang and M. Wang, "A review of electroencephalogram signal processing methods for brain-controlled robots," *Cognit. Robot.*, vol. 1, pp. 111–124, Jan. 2021.

[19] F. S. de Aguiar Neto and J. L. G. Rosa, "Depression biomarkers using non-invasive EEG: A review," *Neurosci. Biobehav. Rev.*, vol. 105, pp. 83–93, Oct. 2019.

[20] A. Khosla, P. Khandnor, and T. Chand, "Automated diagnosis of depression from EEG signals using traditional and deep learning approaches: A comparative analysis," *Biocybern. Biomed. Eng.*, vol. 42, no. 1, pp. 108–142, Jan. 2022.

[21] B. Zhang, G. Yan, Z. Yang, Y. Su, J. Wang, and T. Lei, "Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 215–229, 2021.

[22] Y. Ding et al., "TSception: A deep learning framework for emotion detection using EEG," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.

[23] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[24] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[25] S. Hashempour, R. Boostani, M. Mohammadi, and S. Sanei, "Continuous scoring of depression from EEG signals via a hybrid of convolutional neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 176–183, 2022.

[26] D. Wang et al., "Identification of depression with a semi-supervised GCN based on EEG data," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 2338–2345.

[27] S. Tang et al., "Self-supervised graph neural networks for improved electroencephalographic seizure analysis," 2021, *arXiv:2104.08336*.

[28] T. Chen, Y. Guo, S. Hao, and R. Hong, "Exploring self-attention graph pooling with EEG-based topological structure and soft label for depression detection," *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 2106–2118, Oct. 2022.

[29] M. Hein, J.-P. Lanquart, G. Loas, P. Hubain, and P. Linkowski, "Alterations of neural network organisation during rapid eye movement sleep and slow-wave sleep in major depression: Implications for diagnosis, classification, and treatment," *Psychiatry Res., Neuroimag.*, vol. 291, pp. 71–78, Sep. 2019.

[30] S. Yao and K. M. Kendrick, "Reduced homotopic interhemispheric connectivity in psychiatric disorders: Evidence for both transdiagnostic and disorder specific features," *Psychoradiology*, vol. 2, no. 4, pp. 129–145, Dec. 2022.

[31] A. Khanna, A. Pascual-Leone, C. M. Michel, and F. Farzan, "Microstates in resting-state EEG: Current status and future directions," *Neurosci. Biobehav. Rev.*, vol. 49, pp. 105–113, Feb. 2015.

[32] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3734–3743.

[33] Z. Jia et al., "GraphSleepNet: Adaptive spatial–temporal graph convolutional networks for sleep stage classification," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 1324–1330.

[34] T. Song, S. Liu, W. Zheng, Y. Zong, and Z. Cui, "Instance-adaptive graph for EEG emotion recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 3, 2020, pp. 2701–2708.

[35] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17804–17815.

[36] H. Wang, L. Xu, A. Bezerianos, C. Chen, and Z. Zhang, "Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[37] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, "Scalp EEG classification using deep bi-LSTM network for seizure detection," *Comput. Biol. Med.*, vol. 124, Sep. 2020, Art. no. 103919.

[38] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurorobot.*, vol. 13, p. 37, Jun. 2019.

[39] G. Xu, W. Guo, and Y. Wang, "Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-conv architecture," *Med. Biol. Eng. Comput.*, vol. 61, no. 1, pp. 61–73, Jan. 2023.

[40] L. Feng, C. Cheng, M. Zhao, H. Deng, and Y. Zhang, "EEG-based emotion recognition using spatial–temporal graph convolutional LSTM with attention mechanism," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 11, pp. 5406–5417, Nov. 2022.

[41] H. Wu, J. Liu, and Y. Zhao, "EEG-based depression identification using a deep learning model," in *Proc. IEEE 6th Conf. Inf. Commun. Technol. (CICT)*, Nov. 2022, pp. 1–5.

[42] X. Du et al., "An efficient LSTM network for emotion recognition from multichannel EEG signals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1528–1540, Jul. 2022.

[43] S. D. Kumar and D. Subha, "Prediction of depression from EEG signal using long short term memory(LSTM)," in *Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2019, pp. 1248–1253.

[44] Y. Ding, N. Robinson, C. Tong, Q. Zeng, and C. Guan, "LGGNet: Learning from local–global-graph representations for brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 24, 2023, doi: 10.1109/TNNLS.2023.3236635.

[45] J. F. Cavanagh, A. W. Bismark, M. J. Frank, and J. J. B. Allen, "Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: Evidence from computationally informed EEG," *Comput. Psychiatry*, vol. 3, p. 1, Jan. 2019.

[46] H. Nolan, R. Whelan, and R. B. Reilly, "FASTER: Fully automated statistical thresholding for EEG artifact rejection," *J. Neurosci. Methods*, vol. 192, no. 1, pp. 152–162, Sep. 2010.

[47] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[48] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral Cortex*, vol. 15, no. 9, pp. 1332–1342, Sep. 2005.

[49] Y. Zhang et al., "Theta oscillations: A rhythm difference comparison between major depressive disorder and anxiety disorder," *Frontiers Psychiatry*, vol. 13, Aug. 2022, Art. no. 827536.

[50] S. Kesebir, A. Yosmaoglu, and N. Tarhan, "A dimensional approach to affective disorder: The relations between Scl-90 subdimensions and QEEG parameters," *Frontiers Psychiatry*, vol. 13, Aug. 2022, Art. no. 651008.