A Spatiotemporal Graph Attention Network Based on Synchronization for Epileptic Seizure Prediction

Yao Wang ⬤, Yufei Shi, Yinlin Cheng, Zhipeng He, Xiaoyan Wei ⬤, Ziyi Chen ⬤, and Yi Zhou ⬤

Abstract—Accurate early prediction of epileptic seizures can provide timely treatment for patients. Previous studies have mainly focused on a single temporal or spatial dimension, making it difficult to take both relationships into account. Therefore, the effective properties of electroencephalograms (EEGs) may not be fully evaluated. To solve this problem, we propose a spatiotemporal graph attention network (STGAT) based on synchronization. The spatial and functional connectivity information between EEG channels was extracted by using the phase locking values (PLVs) first, which allowed multichannel EEG signals to be modeled as graph signals. Afterward, the STGAT model was used to dynamically learn the temporal correlation properties of EEG sequences and explore the spatial topological structure information of multiple channels. Experimental results demonstrated that the STGAT model was able to obtain spatiotemporal correlations and achieve good results on two benchmark datasets. The accuracy, specificity and sensitivity were 98.74%, 99.21% and 98.87%, respectively, on the CHB-MIT dataset. Moreover, all evaluation indices of the private dataset had reached more than 98.8%, with the area under the curve (AUC) reaching 99.96%. The proposed method is superior or comparable to the state-of-the-art models. Extensive experiments demonstrate that our end-to-end automatic seizure prediction model can be extended to design clinical assistant decision systems.

Index Terms—Epilepsy, graph attention network, seizure prediction, spatiotemporal correlation, synchronization.

I. INTRODUCTION

EPILEPSY is a brain disorder caused by abnormal synchronous discharges of neurons. Seizures affect the entire central nervous system, manifesting as conscious, motor, vegetative and psychiatric disturbances. The disability-adjusted life years (DALYs) of epilepsy rank second among neurological disorders [1]. Approximately 50 million people worldwide are currently suffering from epilepsy [2], which severely affects all aspects of life, including their physiological and psychological function. A major challenge that results from this disease is the unpredictability and uncontrollability of seizure attacks. If a reliable seizure prediction algorithm can be designed, it can enable health care workers or patients to take effective measures in advance to reduce or even avoid the damage caused by seizures. Furthermore, it may promote the improvement of automatic diagnosis, prediction and analysis systems of epilepsy, which is of practical significance to reduce patient distress and improve their quality of life.

EEG is a recording of voltage fluctuations caused by electrical activity on the surface of the scalp. As one of the oldest neurophysiological techniques, it facilitates the identification of normal and abnormal events occurring in the human brain and reveals the dynamic changes caused by seizures. Hence, it is widely used in studies related to seizure detection [3], [4], [5], [6], seizure prediction [7], [8], [9] and the localization of epileptic focus [10], [11], [12].

The seizure process of epilepsy is progressive, with distinct state changes in EEG signals at different periods [13], [14]. Researchers have divided the seizure process into four distinct periods: the interictal, preictal, ictal and postictal stages. The focus of seizure prediction is the accurate identification of the preictal stage. It is achieved by defining the length of preictal data for different tasks and discriminating between interictal and preictal data to select the preictal signal.

Traditional machine learning methods are based on a manual feature extraction approach, which usually can only focus on a single dimension of information and can only classify a single channel or focused channels [15], [16], [17]. Since the brain is a global structure, this undoubtedly results in the loss of

important EEG information. EEG signals have millisecond temporal resolutions as well as spatial resolutions on the centimeter scale. Therefore, to achieve accurate seizure prediction, both the spatial correlation and temporal dependence of EEGs must be considered. Benefiting from the development of deep learning, researchers have started to use it for the integrated analysis of EEG spatiotemporal information and have obtained good results [18], [19].

Noting that current studies of EEG signals have focused on time-frequency analyse and ignored the influence of spatial factors, Ma et al. [20] introduced the channel and spatial attention (CASA) into batch normalization long short-term memory (BNLSTM) [21] to preserve the spatial and temporal information of EEGs. Sun et al. [22] proposed the channel attention dual-input convolutional neural network, which can synthesize the time-domain, frequency-domain and spatial information of EEG data to achieve accurate seizure prediction of real, representable EEG signals.

However, the electrode distribution properties of EEGs present a non-Euclidean topological structure. Therefore, representing the EEG as a two-dimensional signal may cause a loss of information about the connectivity between functional brain regions. EEG studies based on graph theory analysis have been applied in different psychiatric disorders [23], [24], [25], [26] to explore the functional connectivity patterns of the brain. Due to the high requirement of time series information for seizure prediction tasks, graph analysis techniques are currently focused on seizure detection tasks [27], [28], [29] and relatively few prediction studies.

Zhao et al. [30] used a graph attention network (GAT) combined with the focal loss to solve the problem of unbalanced data classes in a seizure detection task. In the work of Chen et al. [31], a multi-dimensional enhanced seizure prediction model from the multi-level consideration of "frequency + space + time' based on graph convolutional networks (GCN) was constructed. Lian et al. [32] built a global-local graph convolutional neural network that uses a data-driven approach to optimize the learned patient-specific graph and feature representations. These methods have explored the rules of epileptic EEG signals in spatial and temporal relationships to a certain extent. However, the current graphical representation process for EEG data is mainly based on Pearson correlation analysis to construct the adjacency matrix, which cannot show the information specific to EEG.

Epileptic seizures involve different areas of the cerebral cortex, resulting in interactions between different electrodes. This influence changes dynamically with the seizure process and is highly dynamic [33]. The correlation shows the differences in EEG among different electrodes. This suggests that synchronization of the brain can be measured at a local scale spatially. Significant changes in synchronization occur 5-30 minutes before a seizure [34]. Thus, synchronization can also be investigated on different time scales. Phase synchronization analysis methods have high sensitivity and have been widely used to analyze the interaction of brain regions during seizures [35], [36], [37]. In this article, we use PLV as a measure of EEG synchronization to construct the adjacency matrix.

The traditional graph neural network (GNN) cannot represent the dynamic change information of the brain network during epileptic seizures well in the global training process. To solve the above problems, a synchronization-based STGAT model is proposed in this study. Considering the functional connectivity information implied by EEGs, it combines the degree of EEG temporal correlation and the spatial topology information of multi-channel electrodes. The main contributions of this study are summarized as follows.

1) A new deep learning model is proposed to implement the seizure prediction task based on EEGs. We construct a prior graph from the perspective of synchronization and extract EEG information using the STGAT model. This method can directly process full-lead EEG signals and achieve end-to-end seizure predictions.

2) The spatial information and functional connectivity information implied between EEG channels are extracted by using PLV. Then we model the multi-channel EEG as graph signals, thus obtaining real and identifiable synchronization relationships of EEG signals.

3) Our framework incorporates the temporal and spatial correlation of EEGs by combining GAT and gate recurrent unit (GRU). It can capture the spatial relationship between multi-channel electrodes and learn the complex topology and dynamic changes of EEGs, which enhances the ability to utilize spatiotemporal information.

4) We conduct extensive experiments on the private dataset and the CHB-MIT dataset to validate the model using different validation methods. Experimental results show that the STGAT algorithm performs significantly better than other studies using the same dataset.

The organization of the rest of this article is as follows. In Section II, the synchronization-based STGAT model is introduced in terms of the overall model structure, synchrony analysis, spatiotemporal information extraction and algorithm description. Section III describes the private dataset and CHB-MIT dataset and provides a detailed analysis of the results under different experimental methods. Analysis and discussion are given in Section IV, and the conclusion is provided in Section V.

II. METHODS

A. General Structure

In this section, we propose a synchronization-based seizure prediction model, which consists of three steps. Fig. 1 illustrates the main workflow of our method. The raw EEG signal is first processed to calculate its PLV to obtain the synchronization information of different channels and complete the construction of the graph. The contents are given in detail in Section II-B. Then, the feature vector and electrode correlation matrix are used as the input of the STGAT layer to obtain the spatiotemporal information of EEG, which is introduced in Section II-C. Finally, the recognition results are acquired on a fully connected layer with feeding the output result of the fusion model. Section II-D is the algorithm description of the proposed seizure prediction model. For brevity, the variables used in this section are summarized in Table I.

B. Synchronization Analysis of EEG

The clinical signs of epilepsy and the formation of epileptic foci are produced by the combined drive of different areas of the cerebral cortex. Peripheral nerve discharge is abnormal and EEGs in certain areas of the brain show significant changes during seizures. Because the brain is a complexly connected
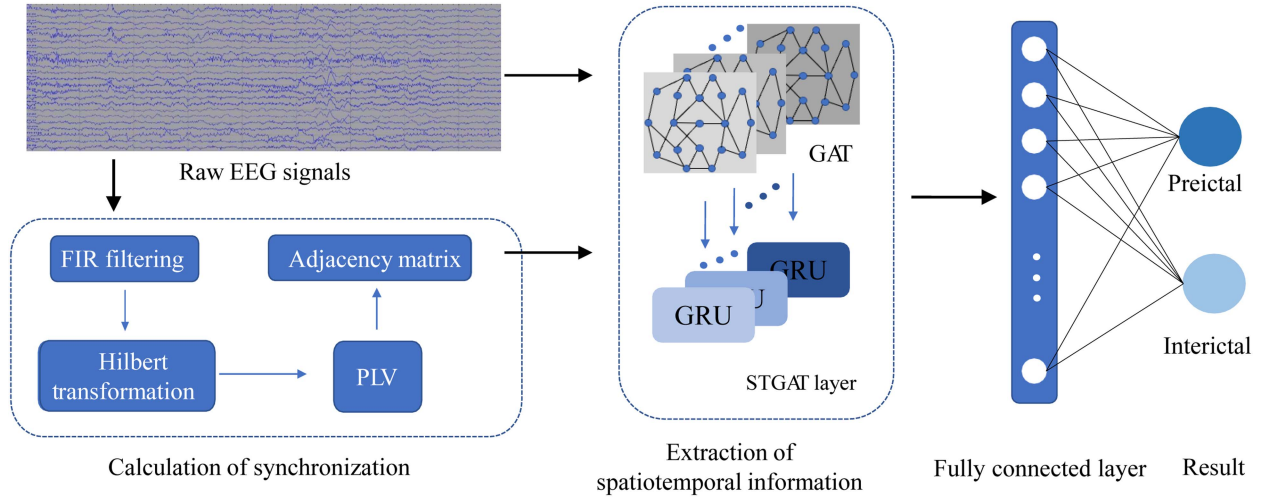
Fig. 1. The overall architecture of proposed model.

TABLE I
IMPORT NOTATIONS IN OUR MODEL

Variables	Description
$x(t)$	The input EEG signal at the time t.
$\hat{x}(t)$	The EEG signal after the Hilbert transform.
$x_{an}(t)$	The analytic signal of x(t).
$A(t), \varphi(t)$	The instantaneous amplitude and instantaneous phase of x(t).
V, E	The set of nodes/edges.
A	The adjacency matrix of V.
$\vec{a}$	The learnable weight vector.
$\mathbf{W}$	The learnable weight matrix.
$\vec{h_i}, \vec{h_i'}$	The input/output feature vectors of Node i in the graph attention layer.
e_{ij}	The influence degree of Node i on Node j.
a_{ij}	The attention coefficient.
$z(t)$	The update gate.
$r(t)$	The reset gate.
$\widetilde{h}(t)$	The current candidate state.
$h(t)$	The current state.

organ, this discharge will surely continue to spread, resulting in strong correlations between adjacent regions.

The PLV measures the absolute value of the average phase difference between any two signals [38]. It can reflect the strength of the association between channels from the perspective of the instantaneous phase to be used to study the changes in EEG data caused by epilepsy. Higher regional synchronization indicates a closer fluctuation pattern between two signals, which in turn indicates a relatively strong association between the two signals. Conversely, it indicates that the fluctuation mode is more independent, and the coupling is poorer as well.

The bandpass filtered time series $x(t)$ is subjected to the Hilbert transform as shown in (1).

$$\hat{x}(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau \tag{1}$$

where $P.V.$ denotes the Cauchy principal value. By constructing its analytic signal $x_{an}(t)$, the complex signal can be decomposed into two parts, the amplitude $A(t)$ and phase $\varphi(t)$.

$$x_{an}(t) = x(t) + i x_H(t) = A(t) e^{i\varphi(t)} \tag{2}$$

where $A(t) = \sqrt{x(t)^2 + x_H(t)^2}$ represents the instantaneous amplitude, and $\varphi(t) = \arctan \frac{x_H(t)}{x(t)}$ denotes the instantaneous phase. For two Nodes i and j, their PLVs can be defined according to (3), thus excluding the effect of instantaneous amplitudes and assessing the degree of phase synchronization.

$$PLV = \frac{1}{N} \left| \sum_{n=1}^{N} e^{i(\Delta\varphi_{ij}(t))} \right| \tag{3}$$

where $\Delta\varphi_{ij}(t) = (\varphi_i(t) - \varphi_j(t))$ means the instantaneous phase difference between i and j at the moment t, and N is the length of the selected time window. Afterward, we use the correlation between multi-channel EEGs to establish the edges of the graph nodes and then construct the adjacency matrix A, which can be denoted as:

$$\begin{cases} A_{ij} = 1, & \text{if } PLV \leq \tau \\ A_{ij} = 0, & \text{others} \end{cases} \tag{4}$$

where τ is a fixed threshold. A_{ij} denotes the simultaneity relationship between Nodes i and j.

In conclusion, this model uses PLV to describe the coupling relationship of EEG signals. By calculating the connectivity between electrodes, it can assess the information interaction between brain regions and represent the synchronization relationship between EEG channels. We combine the PLV and EEG feature matrix as the input of the STGAT layer to study the relationship between the seizures and the synchronization of brain regions.

C. Extraction of the Spatiotemporal Information

Fig. 2 depicts the process of EEG-based spatiotemporal information extraction. We input an EEG sequence of the duration T into the model, and then the spatiotemporal correlation of this sequence is extracted through the GATs layer and GRUs layer. At the end of the model, the period of the sequence is judged using the fully connected layer, and the recognition result is subsequently output.

We set up T GATs in the GATs layer to extract the topological information of T-second EEG data. The GATs layer can calculate
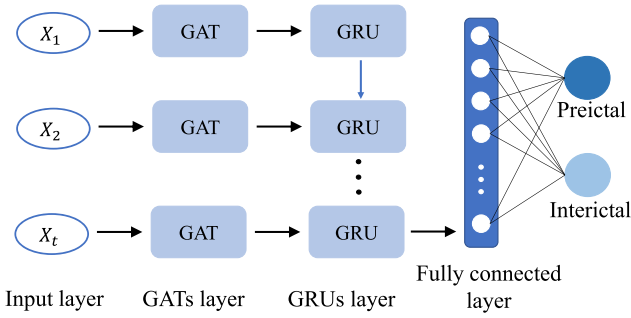
Fig. 2. The process of extracting spatiotemporal information.

the features of any two different nodes of the graph domain. Thus, we can obtain the internal relationship of each cell. After converting all the above information into label vectors, the GRUs layer is used to attain the time domain information and then memorize important temporal dynamic characters in EEGs. Finally, the conversion of signal dimensions is completed in the fully connected layer, and the classification results are output.

1) Multi-Channel Spatial Information Extractions: In recent years, GNNs have been widely used in text classifications [39], traffic flow predictions [40], image analyses [41], emotion recognitions [42], and molecular predictions [43] due to their powerful topological structure analysis capabilities. EEG signals are spontaneous electrophysiological activities performed in multiple functional regions of the brain that are not located in Euclidean space. GAT is a new neural network structure that organically combines the attention mechanism and graph convolution [44], which can eliminate the high dependence of GCN on the Laplacian matrix. It can handle the nodes and graph data of arbitrary topology and solve the problems caused by different structures of graphs.

Seizures are often accompanied by cross-area brain activity. EEG signals have different propagation patterns in different periods, which leads to a complicated evolution process. Therefore, a GAT model is used in this study to obtain the dynamic spatial features of EEGs. The idea is to perform the high aggregation of feature information by calculating the hidden state of each node and output it as label vectors to improve modeling flexibility. The details of the t-th GAT structure are shown in Fig. 3.

$G = (V, E, A)$ is used to represent a connectivity graph, where V represents the set of N nodes [27]. Each node in the diagram represents an EEG channel. E denotes the set of edges. The value of edge indicates the closeness between two channels. Meanwhile, a larger edge value indicates stronger synchronization between the channels. $A \in R^{N \times N}$ is the adjacency matrix of V, which is calculated through PLV in this article.

For each node in the graph, the GAT layer performs a self-attention mechanism, which assigns different weights. Then it aggregates the nodes and adjacent nodes according to the weight magnitude. For two nodes with the input feature vectors $\vec{h}_i$ and $\vec{h}_j$, the influence degree of Node i on Node j, e_{ij}, can be expressed as follows:

$$e_{ij} = \mathrm{a}\left(\boldsymbol{W}\vec{h}_i, \boldsymbol{W}\vec{h}_j\right) \qquad (5)$$

where $\boldsymbol{W}$ denotes the weight matrix of the graph, which can achieve the relational mapping between the input features and the output features by completing the feature transformation of

the nodes. The attention mechanism $\mathrm{a}(\cdot)$ is a feedforward neural network that maps features to a real number. Using a softmax function to perform a regularization of attention, the attention correlation coefficient a_{ij} can be expressed as:

$$a_{ij} = \mathrm{softmax}_j\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k \in N_i} \exp\left(e_{ik}\right)} \qquad (6)$$

To avoid a certain value of the attention coefficient being much larger than the other values which is not easy to train, and to improve the generalization ability of the model, the nonlinear activation function $\mathrm{LeakyReLU}(\cdot)$ is used to activate a_{ij} with a slope of 0.2.

$$a_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\vec{\boldsymbol{a}}^T\left[\boldsymbol{W}\vec{h}_i \| \vec{\boldsymbol{W}}\vec{h}_j\right]\right)\right)}{\sum_{k \in N_i} \exp\left(\mathrm{LeakyReLU}\left(\vec{\boldsymbol{a}}^T\left[\boldsymbol{W}\vec{h}_i \| \vec{h}_k\right]\right)\right)} \qquad (7)$$

where $[\cdot\|\cdot]$ denotes the feature vector after connecting Vertices i and j, and $\vec{\boldsymbol{a}}^T$ represents the transpose.

The ability of single-layer attention to learn the surrounding nodes is relatively weak. To obtain more accurate features, a multiple attention mechanism is introduced in this study [45]. Using multiple attention mechanisms to calculate the attention coefficients of surrounding nodes can make the learning process of the model more stable. For the calculation results under k independent attention mechanisms, the k-average is used instead of the connection. The final feature output result $\vec{h}_i'$ after fusing the information of the neighboring nodes can be expressed as:

$$\vec{h}_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} a_{ij}^k \boldsymbol{W}^k \vec{h}_j\right) \qquad (8)$$

where $\sigma(\cdot)$ is the activation function, and a_{ij}^k is the result of normalizing the attention coefficient through the attention mechanism of the k-th order.

In this section, the attention mechanism is introduced to automatically capture the spatial features with the highest correlation based on the input data to obtain the topological relationship between the central node and surrounding electrodes. Furthermore, the structural features of the channel dimension of EEG data are fully utilized to obtain the spatial dependence between EEG signals.

2) Time Correlation Modeling: With a powerful internal storage and the ability to analyze sequential data, recurrent neural networks (RNN) have been shown to be well suited for learning, classifying and predicting the time-series data from the experience to the process. GRU [46] and long short-term memory (LSTM) [47] belong to the same RNN variant. Both can solve the long dependency problem in RNN and the gradient problem in backpropagation. In comparison, GRU has a simpler unit structure and faster training speed. Hence, we choose GRU to learn the dynamic dependencies within EEG data to obtain the temporal information.

The internal structure of the GRU is shown in Fig. 4. $r(t)$ denotes the reset gate. It determines whether to ignore the previous hidden state. The smaller its value, the less state information is written. The update gate is represented by $z(t)$, which mainly determines what information is forgotten and retained. The smaller its value, the less state information is brought in from
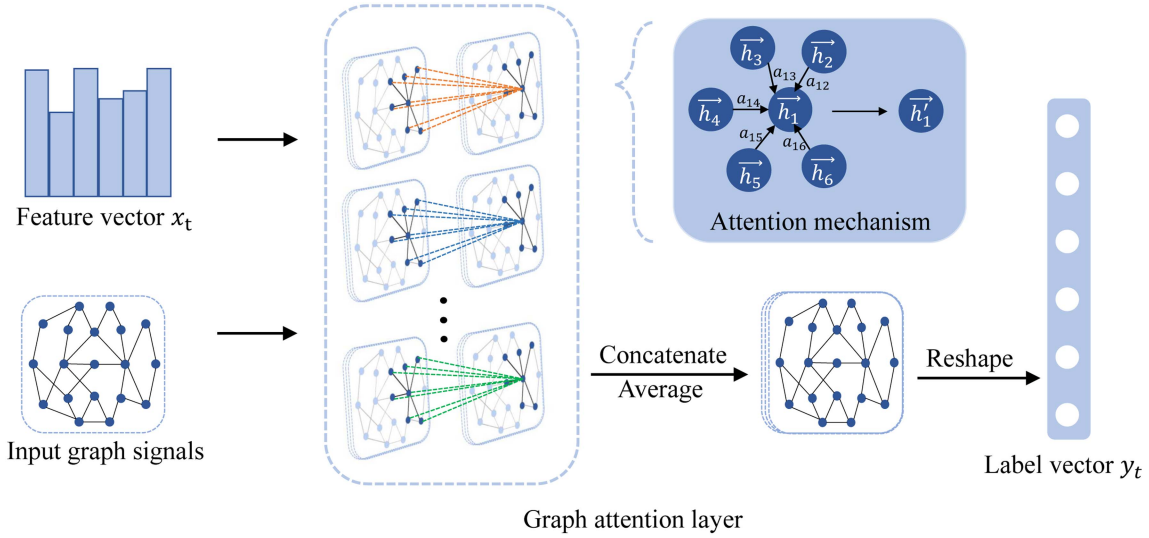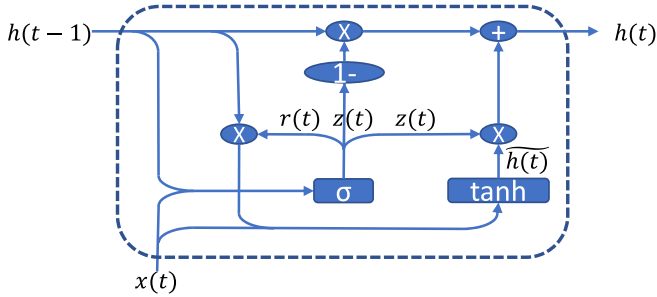
Fig. 3. The structure of graph attention network.

Fig. 4. The structure of GRU.

the previous moment. The equations of the GRU are as follows.

$$z(t) = \sigma\left(\boldsymbol{W_z} \cdot (h(t-1), x(t)) + b_z\right) \qquad (9)$$

$$r(t) = \sigma\left(\boldsymbol{W_r} \cdot (h(t-1), x(t)) + b_r\right) \qquad (10)$$

$$\widetilde{h(t)} = \tanh\left(\boldsymbol{W_{\tilde{h}}} \cdot (r(t) * h(t-1), x(t)) + b_{\tilde{n}}\right) \qquad (11)$$

$$h(t) = (1 - z(t)) * h(t-1) + z(t) * \widetilde{h(t)} \qquad (12)$$

where $\widetilde{h(t)}$ is the candidate state stored at the moment t. The model inputs the hidden state at the moment $t-1$ and the multi-lead EEG signal at the moment t to obtain the information at the moment t. In the meantime, it is able to maintain the trend of EEG changes at the historical moment and finally output the EEG information at the moment t.

D. Description of the STGAT Algorithm

We define Y_t as the actual label vector, and $\hat{Y}_t$ denotes the predicted label. In (13) shows the loss function of the model, where θ is the model parameter, and α indicates the regularization weight. $\alpha\|\theta\|_2^2$ is used to avoid the problem of model parameter overfitting. The purpose of the cross entropy function $cross\_entropy(\cdot)$ is to calculate the inconsistency between the true label of the model and the predicted value.

$$\text{Loss} = \text{cross\_entropy}\left(Y_t, \hat{Y}_t\right) + \alpha\|\theta\|_2^2 \qquad (13)$$

Algorithm 1: The Description of STGAT Based on Synchronization.

Input: Raw multichannel EEG data
Output: The recognition result of the model

1:   Calculate the PLV of the preictal and interictal EEG signals to obtain the appropriate graph structure A according to (1), (2), (3), and (4).
2:   Divide EEG fragments into 5 seconds as the input data X. Input X, A and the ground truth label vector Y_t into the model as the training data, D.
3:   $\hat{Y}_t \leftarrow \mathbf{0} \triangleright$ Initialize the predicted label vector $\hat{Y}_t$ to 0.
4:   **for** (X, A, Y_t) in D **do**
5:     $X \leftarrow W_{in}X + B_{in}$
6:     **for** j in time-step **do**
7:       Hidden-state $\leftarrow \text{GRU}(\text{GAT}(X[j], A), \hat{Y}_t)$
8:       $\hat{Y}_t \leftarrow$ Hidden-state
9:     **end for**
10:    $\hat{Y}_t \leftarrow \text{sigmoid}(W * \hat{Y}_t + B) \qquad \triangleright W$ and B are parameters to be trained
11:    Calculate Loss according to (13).
12:    $\theta \leftarrow \theta - \lambda\nabla\theta \triangleright \lambda$ is learning rate, and $\nabla\theta$ denotes gradient.
13:   **end for**

Algorithm 1 describes the detailed algorithm of the proposed seizure prediction model.

III. EXPERIMENTS AND RESULTS

A. Datasets

In this section, we conducted validation experiments on a private epileptic dataset and a public CHB-MIT EEG dataset to evaluate the classification performance of the proposed STGAT model. For the two different datasets, we used different methods to validate the accuracy of the model, which will be shown in detail in Sections III-B and III-C.

TABLE II
CHARACTERISTICS OF PRIVATE DATASET

Patient	Gender	Age	Consciousness	Time(hours)	Number
1	Male	62	Wake–Sleep	24	8
2	Male	19	Wake–Sleep	24	4
3	Female	16	Wake–Sleep	24	7
4	Female	15	Wake–Sleep	24	3
5	Female	15	Wake–Sleep	24	4
6	Female	22	Wake–Sleep	48	7
7	Female	40	Wake–Sleep	24	5
8	Female	16	Wake–Sleep	24	5
9	Female	26	Wake–Sleep	24	6
10	Female	31	Wake–Sleep	24	8
11	Male	20	Wake–Sleep	24	5
12	Male	46	Wake–Sleep	24	4
13	Female	15	Wake–Sleep	24	6

The private dataset used in this study was derived from the neurology EEG center of a tertiary care hospital. It contained 72 seizures from 13 patients (4 males, 9 females, aged from 15 to 62 years). The sampling frequency was 500 Hz. All EEG data were obtained using the international 10-20 system. Full-lead EEG signals were used because the dataset format is relatively uniform. Detailed EEG data information is shown in Table II.

The CHB-MIT dataset [48] was collected by Boston Children's Hospital and contained 198 seizures from 5 male (3-22 years) and 17 female (1.5-19 years) patients. The EEG of Chb21 was obtained by the re-monitoring of Chb01 patient after 1.5 years, and the information of Chb24 patient was unknown. The download link is https://physionet.org/content/chbmit/1.0.0/. All signals were sampled at a rate of 256 samples per second at a 16-bit sampling rate. All samples were collected by the International 10-20 system and stored in EDF format.

Because of the poor normality of the CHB-MIT dataset, we screened it to select a suitable subset for the experiment. The requirements for selecting subsets were as follows. 1) Seizure-free recordings for at least 35 min before seizures were included in the analysis. 2) The last consecutive EEG recording was used to supplement data that did not last 35 minutes. 3) If the interval between two EEG recordings exceeded 5 seconds, they were marked as discontinuous. After selection, we used data collected from 66 seizures for the following experiments.

In addition, most patients in this dataset used different electrodes, so we selected 18 identical channels, including FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ.

B. Implementation Details

Seizure prediction is a difficult task, and there are no clear criteria for the concept of the pre-seizure. It has been noted that patients with epilepsy show significant fluctuations in EEG signals 30 minutes before a seizure [49]. From the clinical perspective, 30 minutes is enough to draw the attention of patients and health care professionals to reserve time to make interventions as well. Therefore, EEG segments from 35 minutes to 5 minutes before each seizure in epileptic patients were extracted as the preictal period.

In addition, we selected a corresponding number of interictal EEG signals from the EEG recordings of each patient. The

TABLE III
SUMMARY OF DATASETS

Datasets	Categories	Number of EEG segments (interictal/preictal)
The private dataset	Training set	22680
	Validation set	5184
	Testing set	2592
The CHB-MIT dataset	Total	27838

The number of interictal and preictal segments was the same because we selected the interictal segments corresponding to the number of preictal segments.

TABLE IV
MAIN PARAMETERS OF STGAT MODEL

		Activation	Input size	Output size	Heads
Input layer		RELU	$(\sim,20,1280)$	$(\sim,20,12,1)$	-
GATs layer	GAT-1	RELU	$(\sim,20,12,1)$	$(\sim,20,12,64)$	4
	GAT-2	RELU	$(\sim,20,12,64)$	$(\sim,20,12,64)$	4
	Concatenate	-	$(\sim,20,12,64)$	$(\sim,20,12,128)$	-
	Pooling	-	$(\sim,20,12,128)$	$(\sim,12,128)$	-
GRUs layer		Tanh	$(\sim,12,128)$	$(\sim,128)$	-
Output layer		Sigmoid	$(\sim,128)$	$(\sim,1)$	-

$\sim$ represents the value of batch size.

remaining EEG data were excluded. We used a 5-second sliding window to analyze the EEG records. Considering the limited duration of seizures and the unbalanced sample size, a 20% window overlap rate was used to divide the preictal data.

The partitioned training, validation and testing segments of the private dataset with 72 seizures are listed in Table III. Since the CHB-MIT dataset was validated for different patients with leave-one-seizure-out and leave-one-subject-out experiments, only the total number is given here for convenience.

For the experimental environment, all the experiments were completed using the NVIDIA P6000, based on the Pytorch library. The hyperparameters of the model were set as follows: the learning rate was 1e-3, the dropout was 0.3 and the batch size was 256. Moreover, τ was set to 0.4. The Adam optimizer was employed in the experiments. We used 2 GATs for model training and connected them with the skip connection, where each layer had 4 head nodes. The main parameters of the model are shown in Table IV.

C. Subject-Independent Experiments on the Private Dataset

Subject-independent means that the EEG fragments of all patients are randomly divided into training set, validation set and testing set. In this kind of experiment, the model is able to perform as a global classifier to achieve seizure prediction for all patients in the dataset. To ensure the randomness of the dataset when segmenting it in the ratio of 7:2:1, a 5-fold cross-validation was used in this article. All results are given as the mean $\pm$ standard deviation.

We chose the 2-layer GCN, 2-layer GAT and GRU as baseline methods for seizure prediction. The treatments were kept uniform for all experiments. Table V illustrates the test results on different models. It can be seen that our proposed STGAT model achieved more than 98% for all evaluation metrics, with the AUC reaching 99.96%. The F1-score, as the summed average of precision and recall, is able to systematically evaluate the

TABLE V
SUBJECT-INDEPENDENT EXPERIMENTS USING PRIVATE DATASET

	Accuracy(%)	Specificity(%)	Sensitivity(%)	AUC(%)	F1-score(%)	Kappa	Time (seconds/epoch)
GCN***	66.56±3.65	79.89±3.41	87.06±3.55	72.17±3.41	67.11±3.39	0.2071±0.0093	**1.30**
GRU***	66.87±2.37	77.65±2.28	88.22±2.34	78.56±1.98	70.77±2.52	0.1099±0.0074	1.37
GAT**	97.03±1.57	97.62±1.26	98.25±0.53	98.83±1.12	96.44±1.57	0.5158±0.0062	1.43
STGAT	**99.01±0.35**	**98.82±0.43**	**99.10±0.22**	**99.96±0.02**	**99.98±0.01**	**0.9737±0.0086**	1.61

The differences in accuracy between STGAT and other methods are highlighted with stars ($***p<0.001, **p<0.01$).
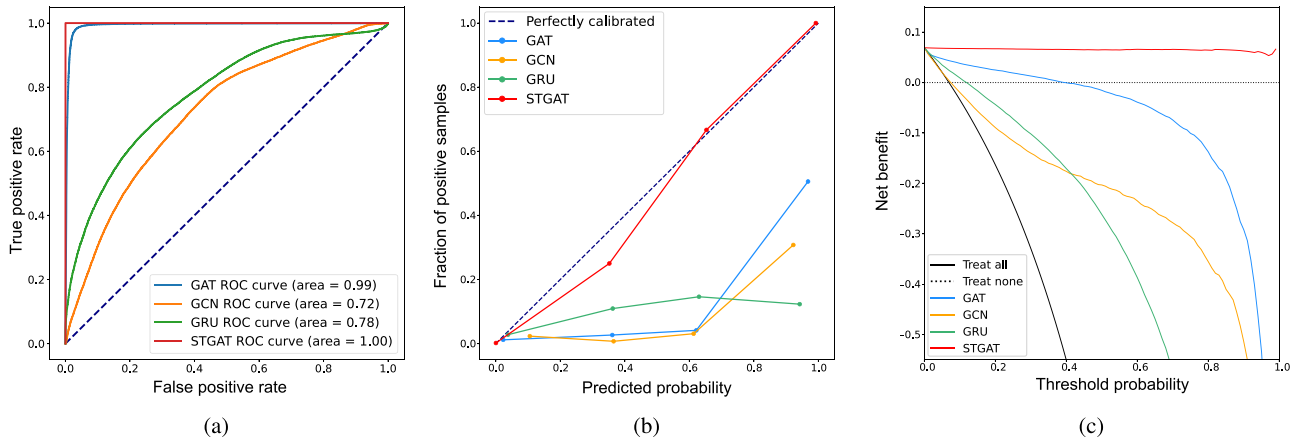

Fig. 5. Comparison of STGAT with baseline models. (a) The area under the ROC curves, (b) calibration curves and (c) decision curve analysis of prediction models.

overall metrics of the classifier. Furthermore, we also used Kappa coefficients to evaluate the classification results of the method more comprehensively. Among all models, the STGAT had the highest F1-score and Kappa coefficient, which proved that our model had better performance.

GAT shows superior performance in capturing the topological structure of non-European space, but it cannot make full use of the time information carried by EEG signals. Therefore, there is room to improve GAT in seizure prediction. We introduced GRU into the GAT mechanism to make full use of the time domain characteristics of EEGs. Under the same experimental conditions, the accuracy of our model was increased by 1.98% compared with that of GAT. The specificity and sensitivity of the model were improved by 1.2% and 1.13%, respectively.

To observe the effect of the spatiotemporal characteristics of EEG on the computational efficiency, we designed experiments to calculate the computational time of the model. More specifically, we calculated the time needed to iterate through each epoch during training. Table V shows the average training time. The GCN, GRU and GAT only extracted separate temporal or spatial features, while our model covered the spatiotemporal information of EEGs. The multilayer architecture of our model inevitably increased the training time. However, considering the performance and computational cost of the model, our model reached an acceptable training time.

Fig. 5(a) shows the area under the curve for several different methods. We also plotted calibration curves to avoid overfitting the model, as shown in Fig. 5(b). The STGAT model exhibited a higher consistency between the actual results and the predicted probabilities. This provided further validation of the discriminative and calibration capabilities of the model. A decision curve analysis was able to determine the clinical utility of the model by quantifying the net benefit at all threshold probabilities. Fig. 5(c) shows that the STGAT model has the best net benefit.

Our proposed model achieved better results in different areas compared to baseline algorithms. The experimental results demonstrate that our model has greater potential to process the classification of EEGs. The superiority of the fusion model is mainly reflected in that it can make full use of the information carried by EEGs by integrating the spatial characteristics and temporal information.

D. Patient-Specific Experiments on the CHB-MIT Dataset

Patient-specific validation means that each experiment is conducted for only one patient, with none of the remaining patients participating in the experiment. In this section, we conducted leave-one-seizure-out experiments and leave-one-subject-out experiments using the CHB-MIT dataset. Specifically, in the leave-one-seizure-out setting, we trained our model using one seizure data of the patient as the testing set and the remaining seizure signals as the training set. For the leave-one-subject-out experiment, all EEG signals from 22 patients were used to train the model and the data of one patient were used for testing to check the generalization performance of the model. The results of the experiments are shown in Table VI. The results in the leave-one-subject-out table are the results of the experiment whenever that patient is used in the test set.

As seen from the table, in leave-one-seizure-out experiments, the average accuracy of the model was 98.74%, the specificity was 99.21%, and the sensitivity was 98.87%. For most patients, the accuracy was above 98%, and the sensitivity reached 100% in 9 patients.

Since new models were not trained using EEG data specific to the test datasets, the variability of EEG signals poses a significant challenge to leave-one-subject-out experiments. The results of the leave-one-subject-out experiments will undoubtedly be

TABLE VI
PATIENT-SPECIFIC EXPERIMENTS USING CHB-MIT

	Leave-one-subject-out			Leave-one-seizure-out			
	Accuracy(%)	Specificity(%)	Sensitivity(%)	Accuracy(%)	Specificity(%)	Sensitivity(%)	Seizure prediction time (minutes)
Chb01	90.56	89.36	92.47	99.00	100.00	98.00	28.82
Chb02	90.68	92.38	94.52	99.00	99.00	99.00	35.34
Chb03	94.25	91.62	92.58	100.00	100.00	100.00	35.48
Chb04	90.38	88.86	86.34	98.77	99.17	99.36	20.83
Chb05	92.36	88.46	89.68	99.18	99.34	100.00	33.42
Chb06	97.42	94.38	96.56	100.00	100.00	100.00	15.12
Chb07	94.54	92.27	96.31	98.56	100.00	98.38	28.57
Chb08	94.83	94.90	96.80	99.63	99.90	99.80	24.51
Chb09	93.53	97.12	95.48	97.53	99.17	96.48	46.08
Chb10	87.63	88.78	89.56	99.88	99.87	100.00	24.09
Chb11	88.25	89.64	85.32	98.15	98.89	97.22	33.25
Chb13	90.49	84.35	89.68	98.15	98.89	98.68	23.78
Chb14	92.56	90.75	92.45	100.00	100.00	100.00	31.69
Chb15	91.53	90.56	87.36	99.38	99.44	100.00	29.98
Chb16	83.64	89.17	90.25	98.20	99.27	99.56	15.92
Chb17	94.59	92.85	93.71	99.59	99.80	99.71	39.41
Chb18	94.67	92.76	91.58	99.69	99.74	100.00	28.62
Chb19	91.57	90.37	92.38	100.00	100.00	100.00	59.20
Chb20	96.38	93.91	91.49	96.30	97.78	97.29	32.48
Chb21	95.36	94.17	97.58	100.00	100.00	100.00	19.80
Chb22	88.38	87.78	90.15	98.77	99.72	97.95	17.13
Chb23	90.03	91.32	92.16	95.06	95.56	94.44	5.20
Chb24	91.45	92.34	93.48	96.14	96.24	98.12	50.44
Mean	**91.96**	**91.22**	**92.08**	**98.74**	**99.21**	**98.87**	**29.53**

TABLE VII
A COMPARISON OF THE PROPOSED METHOD WITH THE STATE-OF-THE-ART

Author	Year	Dataset	Method	Accuracy(%)	Specificity(%)	Sensitivity(%)
Duan et al. [50]	2019	CHB-MIT	MT-CRNN	94.80	97.70	91.70
Zhang et al. [51]	2020	CHB-MIT	CSP-CNN	90.00	-	92.20
Ma et al. [20]	2021	CHB-MIT	BNLSTM-CASA	95.60	91.50	96.20
Sun et al. [22]	2021	CHB-MIT	CADCNN	-	95.60	97.10
Chen et al. [31]	2021	CHB-MIT	MESPF	**98.78**	98.95	98.61
Proposed model	2022	CHB-MIT	STGAT	98.74	**99.21**	**98.87**

slightly worse than those of the leave-one-seizure-out experiments. Even so, the average sensitivity still reached 92.08%. This proved that our model has a strong generalization performance for every patient, which greatly benefits the supplementary diagnosis of the physician.

To be applicable in a clinical environment, it is required that our proposed seizure prediction method be able to predict seizures in a reasonable time horizon with as high an accuracy as possible. In the leave-one-seizure-out experiments, we calculated the actual prediction time of the model. The model was validated by using all periods of EEG recordings, and the prediction results are presented in Table VI.

From the results, the model was able to predict seizures within 29.53 minutes with 90.91% accuracy out of 66 seizures. The longest prediction time was 59.20 minutes and the shortest time was 5.20 minutes, proving that our model is trustworthy.

At the same time, we also compared the performance of other algorithms that also use the CHB-MIT dataset, and the results are presented in Table VII. As seen from the table, our model had a better stability and a more balanced performance in all metrics, compared to the outstanding ability of other models in a certain metric.

Sensitivity and specificity can reflect the accurate classification ability of the predictor, which are meaningful indicators for clinical diagnoses. Compared with other models using the same dataset, our model has better performance, and it is more conducive to clinically assisted decision-making. Specific patient validation experiments show that the present model is a practicable solution for seizure prediction.

IV. ANALYSIS AND DISCUSSION

A. Influence of the Prior Graphs

1) Visualization of the EEG Synchronization: The temporal asynchrony of the hypersynchronous discharges in each region of the brain results in a weaker correlation of the interictal stage. When a seizure is impending, the random discharges of the EEGs in the normal state will turn into rhythmic discharges in the epileptic state. There was a significant increase in this correlation during the seizure period. The correlation reaches a maximum when the brain reaches global hypersynchronous discharges, signaling the imminent end of the seizure. Therefore, there were significant differences in the synchronization between periods during seizures. We visualized the EEGs of some patients for validation.

We calculated the PLV during the interictal and preictal periods in private dataset of Patients 1, 3 and 6. Fig. 6(a)–(c) show their PLV box plots. The EEGs of Patient 1 is taken as an example, which contains 22 channels of EEG signals. As shown in Fig. 6(a), there is a more significant increase in the values of the preictal period compared to the PLV of
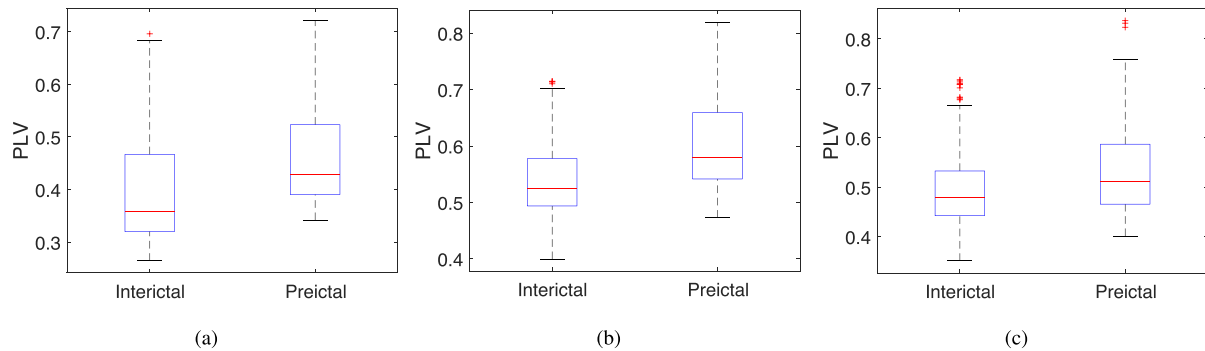
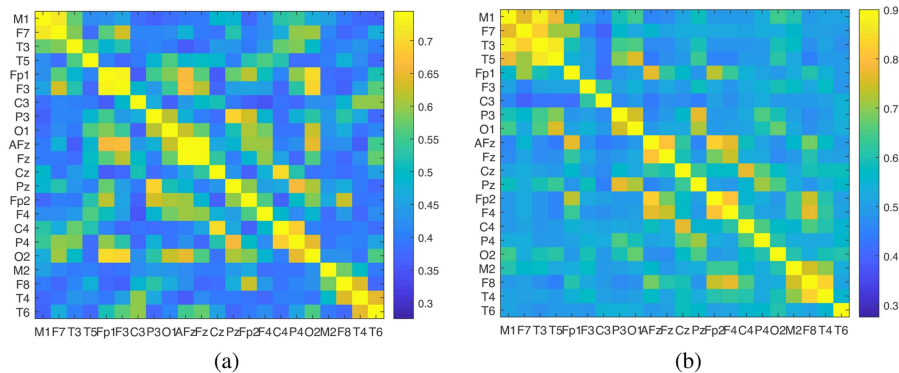Fig. 6. Box plots of the PLV of the Patients 1, 3 and 6.

Fig. 7. PLV correlation matrices of the Patient 1. The horizontal and vertical coordinates indicate the EEG leads respectively, and the different colors represent the magnitude of the PLVs between the electrodes.

the interictal period. There was a significant difference in the PLV between the two periods ($p = 0.0001 < 0.05$). This is because there is already a high degree of synchronization in neuronal activity before the seizure. This can also be seen in the correlation matrix plot between the different channels, as shown in Fig. 7. Fig. 7(a) represents the interictal period of the Patient 1 and Fig. 7(b) is the preictal period. We also found that the degree of synchronization of the EEG signals was also related to the relative distance of the electrodes. Synchronization between closer electrodes was higher than that between more distant electrodes. This reflected the spatial and temporal variability of EEG synchronization at different periods of seizures.

2) Selection of the Prior Graphs: To demonstrate the importance of phase synchronization of EEGs in seizure prediction, we calculated the Pearson correlation coefficients as the comparison experiment. As shown in Fig. 8, the model based on the PLV showed some improvement compared to the Pearson correlation. A 1.34% rise in the accuracy and F1-score of the model were enhanced by 2.74%, which proved that the phase synchronization index of EEGs had some superiority for the graph representation process.

The full-lead EEG data were used to calculate the PLV to construct the adjacency matrix as a representation of the connection relationship between different electrode channels. The adjacency matrix was set to 1 when the correlation between the electrode channels was larger than the threshold value, while the other was set to 0.
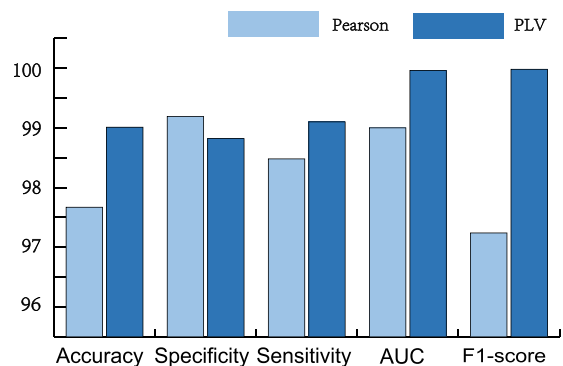

Fig. 8. Results based on Pearson and PLV.

To evaluate the impact of the selected threshold on the model performance, we conducted correlation experiments. Fig. 9 illustrates that our model achieved the best results when the threshold was set to 0.40, so we used 0.40 as the threshold value for subject-independent experiments. Considering the differences in brain synchronization caused by individual differences, we used the optimal thresholds for different patients in patient-specific experiments.

We visualized the connectivity based on the degree of correlation between electrodes, using 0.40 as the threshold. Fig. 10 shows the brain connectivity maps of the Patient 1 at different stages, where Fig. 10(a) represents the interictal period and
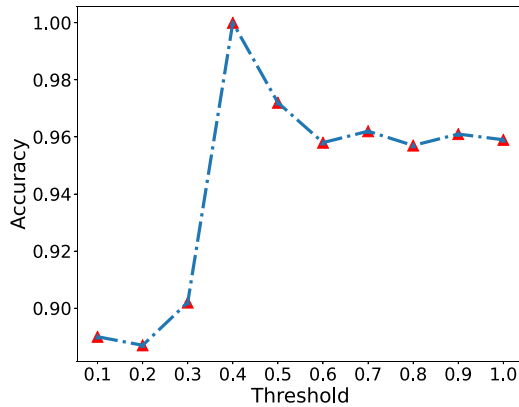
Fig. 9. Effect of different thresholds on model accuracy when the threshold value is 0.1-1.0.

TABLE VIII
IMPACT OF PARAMETERS ON ACCURACY

Layers of GAT	Heads of multi-attention	Hidden size	Accuracy(%)
1	4		98.62±0.15
2	4,4	64	99.87±0.04
3	4,4,4		99.91±0.02
	4,4		99.01±0.35
2	4,8	64	99.89±0.03
	8,8		99.94±0.05
		2	64.37±3.11
		4	78.72±0.66
		8	97.31±1.31
2	4,4	16	98.82±0.51
		32	99.61±0.08
		64	99.87±0.04
		128	99.86±0.05

Fig. 10(b) is the preictal period. The strength of connectivity was significantly lower in the interictal period than in the preictal period. The differences between the connectivity maps of the two different periods were visible.

B. Influence of the Key Parameters

To investigate the robustness of the STGAT, we presented the effects of the key parameters on the accuracy, including the number of GAT layers, the number of attention heads and the hidden size. Table VIII shows the effect of different parameters on the model performance.

First, the STGAT model is not quite sensitive to the number of layers of the GAT and attention heads. Choosing the right value can slightly improve the optimal results. From the results, the increase in GAT layers and attention heads in the model only had a weak effect on the accuracy. In the experiments, we used 2 layers of GAT, which provides a 1% improvement in the accuracy. In particular, the first layer of GAT mainly improved the data dimensionality and extracted the primary information, while the second layer of GAT extracted richer spatial information.

In the multi-head attention layer, more heads lead to higher accuracy. However, it reduces the gain increase and increases

the computational cost. Considering the performance and computational efficiency of the model, we finally decided to use the 4-head attention mechanism in each GAT layer.

The initial increase in the hidden units will greatly improve the performance, which then helps it to reach a relatively stable state. This is because when there are fewer hidden units, the model extracts less information with much information being lost. When the hidden layer size is large enough, the gain from increasing the hidden size is minimal while the computational cost increases exponentially. Given that there is a certain randomness in the training, we chose 64 hidden units for the experiments.

C. Visual Verification

A t-stochastic neighbor embedding (t-SNE) [52] algorithm was applied in visualization experiments to compare the feature representation capabilities of different methods. By converting the similarity between the data points into probability, the high-dimensional data were mapped into a two-dimensional map. Fig. 11(a)–(c) are the visualization figures that intuitively observe the feature distributions.

For many sequence modeling tasks, the feature extraction ability of GRU has been proven [53], [54]. EEG is generated by the interaction of multiple neurons, so it is not feasible to process multi-channel signals only in chronological order. As shown in Fig. 11(a), GRU is weak in classifying preictal data and interictal samples. GAT can automatically extract the multi-channel EEG correlation and assign weights to adjacent nodes. However, there is still some room for improvement, which can be verified from Fig. 11(b).

Compared to Fig. 11(a) and (b), the features in Fig. 11(c) show better separation ability through t-SNE visualization. The STGAT model can process the information of the intra- and inter-channel signals. Therefore, it further improves the performance of both models. This means that the combined neural network has superior performance to the single type of network approach.

D. Limitations and Prospects

The performance of STGAT relies on the construction of prior graphs to some extent. The more accurate graph learned by the model will theoretically lead to better results. However, we found from actual experiments that the model performed excellently in every subject-independent experiment, while the results of patient-specific trials varied from patient to patient. We analyzed the related possibilities, mainly because specificity experiments used less data, which limited the performance of model recognition. Considering that our previous work related to the construction of the epilepsy-specific database will acquire more patient information, we can compensate for this deficiency later by building an EEG database for each patient follow-up and updating the graph data regularly.

In actual clinical practice, the goal of seizure prediction is to alert the patient to prepare for an impending seizure. Therefore, studies that rely solely on offline EEG data cannot achieve this goal. Both the use of datasets and the construction of models in this article are considered based on offline data, and excellent results have been achieved. Although the computation time of the model is acceptable now, we will still look for algorithms to make improvements on the efficiency of the model in future research work. We will focus on the construction of online seizure prediction systems to achieve real-time seizure prediction on
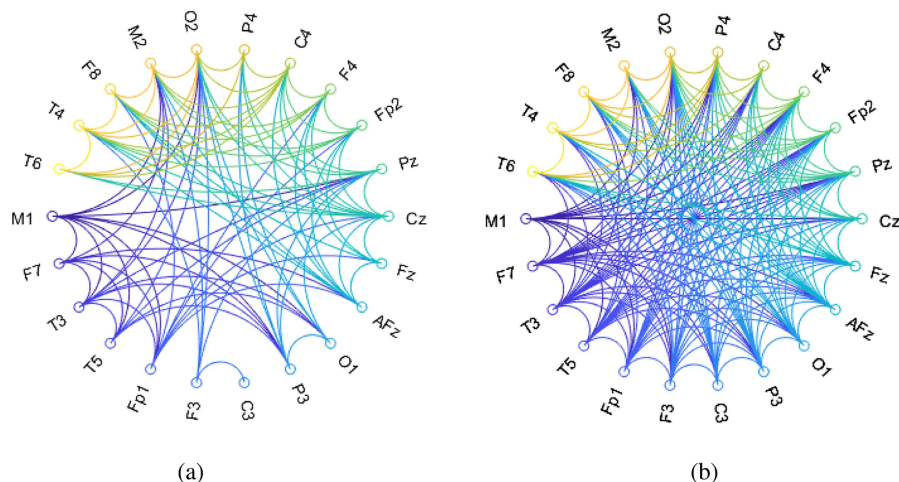
Fig. 10. Brain connectivity maps of the Patient 1 obtained through PLV for different periods. The connections between different nodes are represented by different colors.
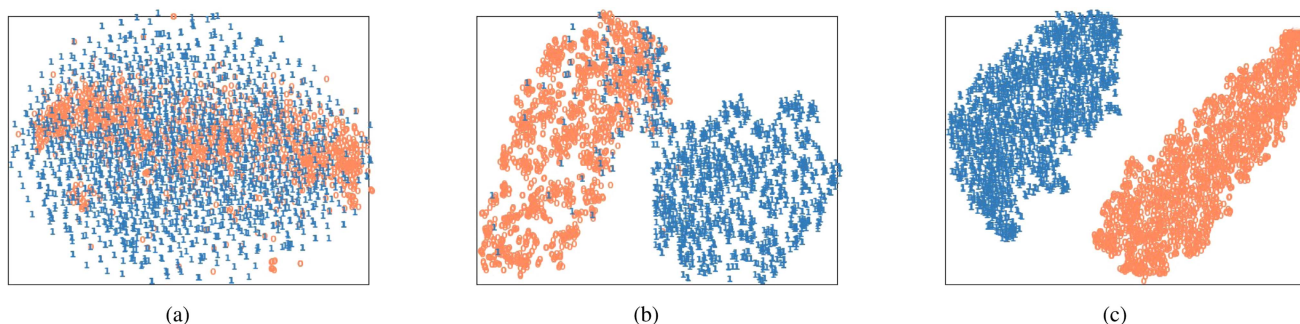

Fig. 11. Feature distributions of interictal and preictal using t-SNE algorithm: (a) Features extracted from GRU, (b) features extracted from GAT and (c) features extracted from STGAT. Orange represents preictal EEGs and blue denotes interictal samples.

EEG data to provide practical and effective help to patients. The combination of seizure prediction algorithms and wearable devices will also be considered to achieve clinically assisted diagnoses and treatment of epilepsy.

V. CONCLUSION

In this article, we presented a synchronization-based model for seizure prediction in spatiotemporal graph attention networks. This model took full advantage of the phase synchronization information between full-lead EEG signals and modeled each EEG channel as a node of the graph model, which highlighted the contribution of synchronicity to EEG classification in seizure prediction. To integrate the spatial characteristics and temporal information of signals, we fully exploited GAT and GRU in their professional fields, using fusion models to mine the spatiotemporal correlation of EEG sequences. Finally, the validity of this model was confirmed by taking different measures on public and private datasets.

ETHICS STATEMENT

The private dataset of this study involving human participants were reviewed and approved by Medical Ethics Committee of First Affiliated Hospital, Sun Yat-sen University. The patients provided their written informed consent to participate in this study.

REFERENCES

[1] C. J. Murray et al., "Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the global burden of disease study 2010," *Lancet*, vol. 380, no. 9859, pp. 2197–2223, 2012.

[2] "World health organization-epilepsy." Accessed: May 19, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[3] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.

[4] M. Zabihi, S. Kiranyaz, V. Jäntti, T. Lipping, and M. Gabbouj, "Patient-specific seizure detection using nonlinear dynamics and nullclines," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 543–555, Feb. 2020.

[5] A. Shoeibi et al., "A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals," *Expert Syst. Appl.*, vol. 163, 2021, Art. no. 113788.

[6] S. Chakrabarti, A. Swetapadma, and P. K. Pattnaik, "A channel independent generalized seizure detection method for pediatric epileptic seizures," *Comput. Methods Programs Biomed.*, vol. 209, 2021, Art. no. 106335.

[7] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, 2018.

[8] X. Wei, L. Zhou, Z. Zhang, Z. Chen, and Y. Zhou, "Early prediction of epileptic seizures using a long-term recurrent convolutional network," *J. Neurosci. Methods*, vol. 327, 2019, Art. no. 108395.

[9] N. D. Truong et al., "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw.*, vol. 105, pp. 104–111, 2018.

[10] D. Chen, S. Wan, and F. S. Bao, "Epileptic focus localization using discrete wavelet transform based on interictal intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 5, pp. 413–425, May 2017.

[11] T. Mei et al., "Epileptic foci localization based on mapping the synchronization of dynamic brain network," *BMC Med. Inform. Decis. Making*, vol. 19, no. 1, pp. 1–11, 2019.

[12] H. Daoud and M. Bayoumi, "Deep learning approach for epileptic focus localization," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 209–220, Apr. 2020.

[13] C. A. Teixeira et al., "Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 324–336, 2014.

[14] Z. Zhang et al., "Construction of rules for seizure prediction based on approximate entropy," *Clin. Neurophysiol.*, vol. 125, no. 10, pp. 1959–1966, 2014.

[15] M. Z. Parvez and M. Paul, "Seizure prediction using undulated global and local features," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 208–217, Jan. 2017.

[16] E. V. Carrera and F. Quinga, "Analysis of epileptic seizure predictions based on intracranial EEG records," in *Proc. IEEE Colombian Conf. Commun. Comput.*, 2018, pp. 1–5.

[17] S. Elgohary, S. Eldawlatly, and M. I. Khalil, "Epileptic seizure prediction using zero-crossings analysis of EEG wavelet detail coefficients," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol.*, 2016, pp. 1–6.

[18] D. P. Dash, M. H. Kolekar, and K. Jha, "Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and hidden Markov model," *Comput. Biol. Med.*, vol. 116, 2020, Art. no. 103571.

[19] A. Bhattacharya, T. Baweja, and S. Karri, "Epileptic seizure prediction using deep transformer model," *Int. J. Neural Syst.*, vol. 32, no. 2, 2022, Art. no. 2150058.

[20] M. Ma et al., "Early prediction of epileptic seizure based on the BNLSTM-CASA model," *IEEE Access*, vol. 9, pp. 79600–79610, 2021.

[21] T. Cooijmans, N. Ballas, C. Laurent, Ç. Gülçehre, and A. Courville, "Recurrent batch normalization," 2016, *arXiv:1603.09025*. [Online]. Available: http://arxiv.org/abs/1603.09025

[22] B. Sun et al., "Seizure prediction in scalp EEG based channel attention dual-input convolutional neural network," *Physica A: Stat. Mechanics Appl.*, vol. 584, 2021, Art. no. 126376.

[23] C. J. Stam, "Modern network science of neurological disorders," *Nature Rev. Neurosci.*, vol. 15, no. 10, pp. 683–695, 2014.

[24] F. Vecchio, F. Miraglia, and P. M. Rossini, "Connectome: Graph theory application in functional brain network architecture," *Clin. Neurophysiol. Pract.*, vol. 2, pp. 206–213, 2017.

[25] J. Cao, Y. Li, H. Yu, X. Zhao, Y. Li, and S. Wang, "Investigation of brain networks in children with attention deficit/hyperactivity disorder using a graph theoretical approach," *Biomed. Signal Process. Control*, vol. 40, pp. 351–358, 2018.

[26] L. Rutter et al., "Graph theoretical analysis of resting magnetoencephalographic functional connectivity networks," *Front. Comput. Neurosci.*, vol. 7, 2013, Art. no. 93.

[27] I. C. Covert et al., "Temporal graph convolutional networks for automatic seizure detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2019, pp. 160–180.

[28] Y. Zhao et al., "EEG-based seizure detection using linear graph convolution network with focal loss," *Comput. Methods Programs Biomed.*, vol. 208, 2021, Art. no. 106277.

[29] D. Zeng, K. Huang, C. Xu, H. Shen, and Z. Chen, "Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 4, pp. 955–968, Dec. 2021.

[30] Y. Zhao, G. Zhang, C. Dong, Q. Yuan, F. Xu, and Y. Zheng, "Graph attention network with focal loss for seizure detection on electroencephalography signals," *Int. J. Neural Syst.*, vol. 31, no. 7, 2021, Art. no. 2150027.

[31] X. Chen, Y. Zheng, C. Dong, and S. Song, "Multi-dimensional enhanced seizure prediction framework based on graph convolutional network," *Front. Neuroinform.*, vol. 15, 2021, Art. no. 605729.

[32] Q. Lian, Y. Qi, G. Pan, and Y. Wang, "Learning graph in graph convolutional neural networks for robust seizure prediction," *J. Neural Eng.*, vol. 17, no. 3, 2020, Art. no. 035004.

[33] M. Ma, X. Wei, Y. Cheng, Z. Chen, and Y. Zhou, "Spatiotemporal evolution of epileptic seizure based on mutual information and dynamic brain network," *BMC Med. Inform. Decis. Making*, vol. 21, no. 2, pp. 1–11, 2021.

[34] P. Jiruska, M. D. Curtis, J. G. Jefferys, C. A. Schevon, S. J. Schiff, and K. Schindler, "Synchronization and desynchronization in epilepsy: Controversies and hypotheses," *J. Physiol.*, vol. 591, no. 4, pp. 787–797, 2013.

[35] H. Yu et al., "Investigation of phase synchronization of interictal EEG in right temporal lobe epilepsy," *Physica A: Stat. Mechanics Appl.*, vol. 492, pp. 931–940, 2018.

[36] H. Liu and P. Zhang, "Phase synchronization dynamics of neural network during seizures," *Comput. Math. Methods Med.*, vol. 2018, 2018, Art. no. 1354915.

[37] J.-J. Li, Y.-H. Li, H.-Q. Gong, P.-J. Liang, P.-M. Zhang, and Q.-C. Lu, "The spatiotemporal dynamics of phase synchronization during epileptogenesis in amygdala-kindling mice," *PLoS One*, vol. 11, no. 4, 2016, Art. no. e0153897.

[38] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for eeg-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, 1997.

[39] R. Liao, M. Brockschmidt, D. Tarlow, A. L. Gaunt, R. Urtasun, and R. Zemel, "Graph partition neural networks for semi-supervised classification," in *Proc. Int. Conf. Learn. Representations (Workshop Track)*, 2018.

[40] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 1720–1730.

[41] S. Qi, W. Wang, B. Jia, J. Shen, and S.-C. Zhu, "Learning human-object interactions by graph parsing neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 401–417.

[42] X. Li, J. Li, Y. Zhang, and P. Tiwari, "Emotion recognition from multi-channel EEG data through a dual-pipeline graph attention network," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2021, pp. 3642–3647.

[43] F. Capela, V. Nouchi, R. V. Deursen, I. V. Tetko, and G. Godin, "Multitask learning on graph neural networks applied to molecular property predictions," 2019, *arXiv:1910.13124*. [Online]. Available: http://arxiv.org/abs/1910.13124

[44] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *Stat*, vol. 1050, 2017, Art. no. 20.

[45] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6000–6010.

[46] K. Cho, B. V. Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*. [Online]. Available: http://arxiv.org/abs/1409.1259

[47] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin/Heidelberg, Germany: Springer, 2012, pp. 37–45.

[48] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.

[49] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H. U. Voss, A. Schulze-Bonhage, and J. Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Physica D: nonlinear phenomena*, vol. 194, no. 3/4, pp. 357–368, 2004.

[50] L. Duan, J. Hou, Y. Qiao, and J. Miao, "Epileptic seizure prediction based on convolutional recurrent neural network with multi-timescale," in *Proc. Int. Conf. Intell. Sci. Big Data Eng.*, 2019, pp. 139–150.

[51] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 465–474, 2019.

[52] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[53] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, 2017.

[54] S. Gao et al., "Short-term runoff prediction with gru and lstm networks without requiring time step optimization during sample generation," *J. Hydrol.*, vol. 589, 2020, Art. no. 125188.