Hierarchy Graph Convolution Network and Tree Classification for Epileptic Detection on Electroencephalography Signals

Difei Zeng, *Student Member, IEEE*, Kejie Huang, *Senior Member, IEEE*, Cenglin Xu, Haibin Shen, and Zhong Chen

***Abstract*—The epileptic detection with electroencephalography (EEG) has been deeply studied and developed. However, previous research gave little attention to the physical appearance and early onset warnings of seizure. When a seizure occurs, electrodes near the epileptic foci will exhibit significantly fluctuating and inconsistent voltages. In this article, a novel approach to epileptic detection based on the hierarchy graph convolution network (HGCN) structure is proposed. Multiple features of time or frequency domains extracted from the raw EEG signals are taken as the input of HGCN. The topological relationship between every single electrode is utilized by HGCN. The tree classification (TC) and preictal fuzzification (PF) are proposed to adapt both multiclassification tasks and refine-classification tasks. Experiments are performed on the CHB-MIT and TUH data sets. Compared with the state of the art, our proposed model achieves a 5.77% improvement of accuracy on the CHB-MIT data set, and an improvement of 2.43% and 19.7% for sensitivity and specificity on the TUH data set, respectively.**

***Index Terms*—Electroencephalography (EEG), epilepsy, hierarchy graph convolution network (HGCN), preictal fuzzification (PF), tree classification.**

I. Introduction

EPILEPSY is classified as a neurological disease that could have a severe impact on epileptic patients. Not only would uncontrolled epileptic seizure causes impairment to the normal brain functions but it can also result in premature death. Epilepsy is a long-term, frequent-onset disease, and the patients suffer excruciating pain every time a seizure occurs. The detection of seizure onset will be conducive to timely therapeutic interventions.

Manuscript received February 3, 2020; revised June 14, 2020; accepted July 22, 2020. Date of publication July 27, 2020; date of current version December 10, 2021. This work was supported by the National Natural Science Foundation of China under Grant U19B2043. *(Corresponding authors: Kejie Huang; Haibin Shen.)*

Difei Zeng, Kejie Huang, and Haibin Shen are with the College of Information Science Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: zjuzdf95@gmail.com; huangkejie@zju.edu.cn; shen_hb@zju.edu.cn).

Cenglin Xu and Zhong Chen are with the Department of Pharmacology, School of Medicine, Zhejiang University, Hangzhou 310058, China (e-mail: xucenglin5zz@zju.edu.cn; chenzhong@zju.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCDS.2020.3012278.

Digital Object Identifier 10.1109/TCDS.2020.3012278

A seizure is caused by abnormal repetitive neural firing in the seizure focus. Clinicians record electroencephalography (EEG) through scalp or invasive electrodes. Depending on the different manifestations of epileptic EEG, it can be classified into ictal, preictal, and interictal courses. The interictal course or the so-called non-seizure course refers to the circumstance where no seizure occurs, whereas the ictal course or the so-called seizure course means that a seizure occurs. The preictal is a brief period of time before the seizure onset which plays a crucial role in seizure prediction. In general, EEG exhibits different characteristics in the preictal course, for which this course can be categorized into preictal-I, preictal-II, and preictal-III courses. The detection of seizure on sEEG signals will significantly save the time of manual identification and make the classification on the whole day.

In recent years, various methods based on machine learning have been applied to detect the course of EEG among epileptic patients. In [1] and [2], a support vector machine (SVM) was used to analyze the input of features extracted from sEEG. In [2]–[17], the EEG signals were described in two dimensions and the convolution neural network (CNN) was applied for classification purpose. The axes of the input in the above-mentioned works could be time segments, multiple frequency bands, EEG channels, raw sampling points, etc. In [9] and [16]–[22], time-related information was considered to be the crucial feature and long short-term memory (LSTM) or other improved versions was used, including bidirectional LSTM (BiLSTM) and attention LSTM. In [23], the generative adversarial network (GAN) was applied to process the unlabeled data even in the absence of a state-of-the-art result. In [24], graph neural network (GNN) was used by aggregating the information from adjacent channels.

In some aforementioned works based on SVM and LSTM, barely consideration has been paid to the positional relationship between the electrodes on the scalp. The relationship of features from different electrodes is short of prior guidance. In some of the other works based on CNN, GAN, and GNN, the structure cannot ensure effective utilization of the adjacent information about each electrode. In order to address the above-mentioned issues, we first apply the graph convolution network (GCN) based on the adaptive adjacency matrix to extract higher level epilepsy discharge features. The hierarchy GCN (HGCN) model is proposed to improve the GCN by subjecting different adjacent relationships to exceptional

treatment. The loss function is redesigned using the dependencies between the categories with tree classification (TC). For the five-class task that refines the classification of the preictal course, preictal fuzzification (PF) is conducted to enhance the robustness of the model. The proposed model produces the most desirable results in multiple classification tasks.

The remainder of this article is structured as follows. Section II introduces the relevant works to the preprocessing, features extraction, and machine learning models for classification. Section III elaborates on the methods referred to in this article, including multiple feature processing, the use of GCN, TC, and PF, and the proposal of HGCN. Section IV details the experiments and discussions, covering the experiment environment, patient-specific experiments, ablation experiments based on cross-validation, and comparison with the state of the art. Finally, the conclusion is drawn in Section V.

II. Related Works

The prediction made for different courses of epileptic patients with EEG has experienced fast-paced development in the most recent years. There are four types of classification tasks, which are 2-class-seizure, 2-class-predict, 3-class, and 5-class in relation to the detection of seizure. The 2-class-seizure in [6], [10], [11], [15], [18], [20], [23], and [25] refers to the classification for nonseizure and seizure courses. The nonseizure course can be further classed into interictal and preictal courses. The 2-class-predict in [22] refers to the classification for interictal and preictal courses. It gives no consideration to seizure courses and aims at the prediction of seizures before the onset. The 3-class in [10] means to detect interictal, preictal, and ictal courses. The preictal course can be further divided into three equal and nonoverlapping time segments, preictal-I, preictal-II, and preictal-III courses. The 3-class task can be transformed into 5-class in [2]and [10] for the detection of interictal, preictal-I, preictal-II, preictal-III, and ictal courses.

The implementation of the above-mentioned classification tasks using EEG is primarily split into two phases, which are preprocessing and classification using a deep learning model. Furthermore, preprocessing can be divided into two phases, including normalization and feature extraction.

A. Normalization

In the normalization phase, Craley *et al.* [11], Yao *et al.* [20], and Singh and Malhotra [26] filtered out the power line noise from the original EEG. Their models were constructed on the input of the original EEG without normalization. In [6], [15], [23], and [27], the original EEG was normalized to [0, 1] according to the peak-to-peak value. In [5], [7], [13], and [14], the original signal was normalized employing with Z-score standardization. The aforementioned normalization methods are applied to rescale the original EEG signals to around zero, which allows the deep learning model to converge quickly. On the other hand, the normalization could be effective in reducing the difference between EEG signals from different patients and different channels, which is conducive to improving the robustness of the model.

B. Feature Extraction

In [16], [17], [21], and [28], the linear frequency cepstral coefficients (LFCCs) were extracted. The LFCCs used a linear frequency scale rather than the mel scale in the mel frequency cepstral coefficients (MFCCs). In [22], [29], and [30], the features were extracted using continuous wavelet transform (CWT) or discrete wavelet transform (DWT). In [1], [2], [9], [10], [12], [15], [18], [22], [23], and [25], spectrogram was extracted by short-time Fourier transform (STFT) or power spectral density (PSD). These methods featured the description of EEG in every single frequency point. In addition to the frequency domain, Hosseini *et al.* [12] and Tsiouris *et al.* [22] extracted features in the time domain, including mean, variance, skewness, kurtosis, zero-crossing, etc. These methods reflected some statistical characteristics of the EEG from the perspective of amplitude, stability, frequency, etc. In [6], [25]–[27], and [29], the features were extracted using such unsupervised methods as stacked autoencoder (SAE) or the improvement versions. In [25], the SAE on the EEG from each channel was used independently and the global information was integrated by the attention module. In [6], the 4096 sampling points in 16 s were defined as the input and the SAE based on one-dimensional CNN (1D-CNN) was used for classification purposes.

C. Classification Models

In the phase of classification with a machine learning model, SVM was used after the feature extraction in [1] and [31]. In [2], multiple binary SVM classifiers were used after CNN to make the model more robust in multiple-classification tasks. However, SVM is incapable of solving the time-related information by treating them as multiple features that may be related. In [2]–[4], [7], [10]–[12], and [15]–[17], common CNN models were applied. The features were propagated by CNN in two various dimensions, including continuous-time segments, multiple frequency bands, electrode channels, raw sampling points, etc. Cao *et al.* [10] defined a 2-s long segment as the input and used the stacked CNN (SCNN) model for classification. In [11], a 1-s long segment was defined as the input, while CNN on the EEG from each channel was used independently and the global information was integrated by the coupled hidden Markov model (HMM). In [4], [5], [9], and [14], 1-D-CNN was used. It is considered that there is no uniform extractable feature between electrode channels whose size in the convolution kernel of 1-D-CNN is one. Hao *et al.* [8] and Lu and Triesch [13] combined the 1-D-CNN with the residual network (ResNet). In doing so, the model is made deeper to achieve a better result. However, compared with the recurrent neural network (RNN), CNN is not suitable for the processing of time-related information restricted by the size of the convolution kernel. In [9], [19], and [22], the LSTM was applied to extract time-related information from multiple consequent-time segments. In [22], a 75–250-s long segment was defined as the input and the LSTM was used to classify multiple features extracted from the EEG. In [20], the attention module was added into LSTM to distinguish the significance attached to different time segments. In [16]–[18]

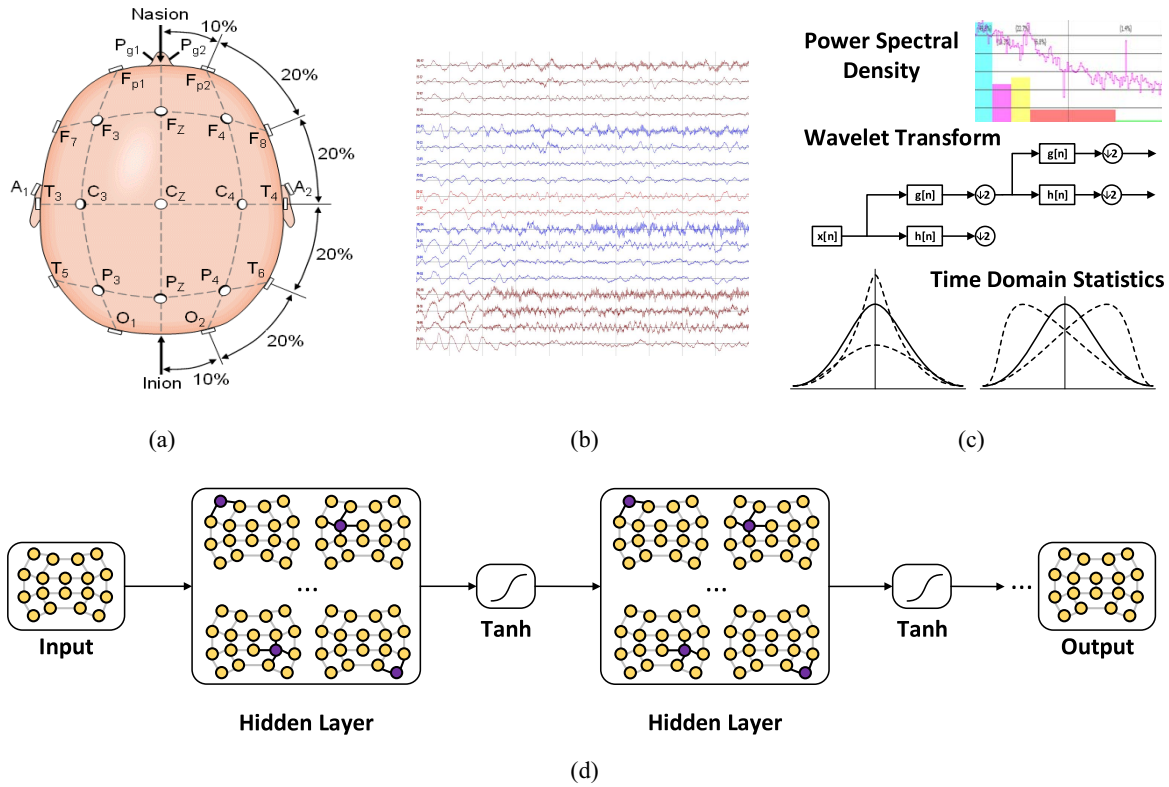


Fig. 1. (a) Environment when the EEG is sampled. (b) EEG collected in the International 10–20 system. (c) Features in multiple domains, including PSD, wavelet transform, and time-domain statistics. (d) Deep learning model based on GCN.

and [20], BiLSTM was used from sequential and reverse time segments. Golmohammadi *et al.* [16], Shah *et al.* [17], and Thodoroff *et al.* [18] used the CNN module to extract deep features before the processing of multiple time segments by BiLSTM. Deficiently, models based on RNN do not consider the locational relationship between electrode channels. Thodoroff *et al.* [18] defined a 30-s long segment as the input and used the recurrent CNN (RCNN) model for classification purpose. In [24], GNN was first applied to utilize structure information. Unfortunately, it only aggregated information from adjacent electrodes through mean-pooling or max-pooling that lost a great number of neighborhood features.

III. Methods

As shown in Fig. 1, the entire flowchart of the proposed scheme consists of four parts. The EEG signals are sampled with the electrodes on the scalp as shown in Fig. 1(a). In this article, the sEEG in the following experiments was collected using the International 10–20 system as shown in Fig. 1(b). The EEG signals are preprocessed to extract features from both frequency domain and time domain as shown in Fig. 1(c). The model based on GCN and the TC is proposed for the detection of seizures as illustrated in Fig. 1(d). The GCN is applied to explore the intrinsic locational relationship between electrode pairs. Improved from the GCN model, the HGCN model is proposed to treat longitudinal and transverse adjacent electrode pairs in a different way. The TC is proposed for the classification task whose categories have dependencies with each other.

A. Preprocessing and Multiple Features Extraction

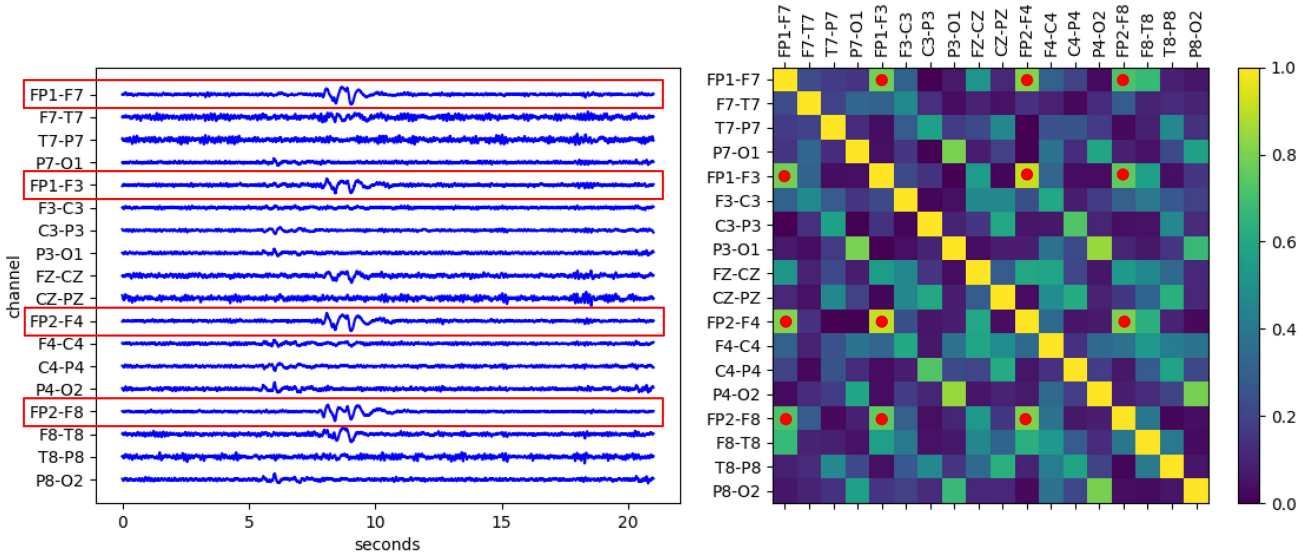
This section elaborates on the works to be undertaken before the deep learning model, including preprocessing and multiple feature extraction.

1) Preprocessing: EEG signals are usually coupled with the noises from muscle, eye movement, sleep spindle wave, power line, and environment. To ensure the robustness of the whole system, nonphysiologically slow drifts, power line noise, and high-frequency environmental noise must be removed [32]. Since both CHB-MIT data set and TUH data set used in the experiments have 60-Hz power line noise mixed during sampling, the raw EEG signals will be filtered with a band stop to remove the frequency range between 57 and 63 Hz and a bandpass to keep a frequency range between 0.5 and 128 Hz.

There is a significant difference in the EEG signals at each onset. It is possible that the rectified average of all the EEG signals in the interictal periods from different patients shows a four-times difference. For the generation of the model, EEG is normalized prior to feature extraction. A 1-h length EEG signal is denoted as $X \in \mathbb{R}^{N*60*60}$, where N represents the sampling frequency of the EEG signals. In this article, a common normalization method Z-score standardization is adopted

$$X_{\text{Norm}} = \frac{X - \overline{X}}{\text{std}(X)} \tag{1}$$

where $\overline{X}$ and $\text{std}(X)$ indicate the average and the standard deviation of X, respectively.



(a) (b)

Fig. 2. (a) Raw EEG signal of all 18 channels. The four channels marked with red boxes are adjacent in the International 10–20 system. (b) Correlation matrix between every two channels.

2) Multiple Features Extraction: The features extracted from the time domain and frequency domain could be used to describe the appearances of the signals throughout the segment time and in a specific frequency, respectively. In the process of feature extraction, the appropriate segment length should be used to make a tradeoff between the time resolution and information integrity. Therefore, the EEG signals are analyzed in a 2-s long segment with 50% overlapping as the processing methods in [1] and [10].

The features extracted from time domain include average, rectified average, standard deviation, peak-to-peak value, kurtosis, skewness, and the crossing frequency of the EEG signal in a 2-s long analyzing segment or $2N$ sampling points. The average of all the sampling points refers to the amplitude of the EEG signal, which will be large and erratic in the ictal courses. However, the wave which is symmetry about x-axis appears on a frequent basis. The rectified average is unaffected by the problem as mentioned above. The standard deviation and the peak-to-peak value are the degree of confusion and the difference between the minimum value and the maximum value, respectively. These two features tend to be higher in the ictal course than in interictal or preictal courses. The kurtosis and the skewness indicate the sharpness of the peak and the asymmetry in all sampling points, respectively. These two features are the description for the waveform composed of all sampling points. The crossing frequency reflects the whole frequency from a view of time-domain statistics. The above-mentioned time-domain features extracted from seven directions are denoted as $F^{\text{time}} \in \mathbb{R}^7$.

The features are extracted from the frequency domain by two methods, which are PSD and CWT. The PSD describes the signal power in a frequency distribution, which is achieved through the Fourier transform. However, the Fourier transform encounters the problem that the time- and frequency-domain information of the signal is incapable to be localized simultaneously. The CWT is adopted to address this problem. The average energy in multiple frequency bands is computed as well, including δ band (0.5–4 Hz), θ band (4–8 Hz), α band (8–13 Hz), β band (13–30 Hz), low γ band (33–55 Hz), and high γ band (65–110 Hz), which contain majority signals of brain activities.

The aforementioned features extracted by PSD and CWT from the frequency domain are denoted as $F^{\text{fre}} \in \mathbb{R}^{267}$. The features extracted from time and frequency domains are combined together as $F = [F^{\text{time}}; F^{\text{fre}}] \in \mathbb{R}^{274}$.

B. GCN

The essence of epilepsy is a sudden, abnormal, and synchronized discharge of neurons in the brain. The electrodes around the epileptic foci show obvious voltage fluctuations in EEG, including the increase of high-frequency activities, increased amplitude, etc. The raw EEG signals under 18 channels are recorded in Fig. 2(a), and the correlation analysis between every two different channels is shown in Fig. 2(b). It can be seen from Fig. 2(a) that the EEG channels, i.e., FP1-F7, FP1-F3, FP2-F4, and FP2-F8, which are adjacent in the International 10–20 system, have the similar fluctuation and strong correlation. The correlation coefficients between these four channels are marked with red dots in Fig. 2(b). The average correlation coefficient between these four channels is 0.85 while the average correlation coefficient between all channels is only 0.2625. Therefore, it can be judged that there is an apparent relationship between these four adjacent channels. The discharging of neurons in epileptic foci is expressed on EEG through the surrounding adjacent electrodes, so the phenomenon of grasping the neuronal discharges by a convolution between adjacent electrodes and extracting features becomes the key to the work.

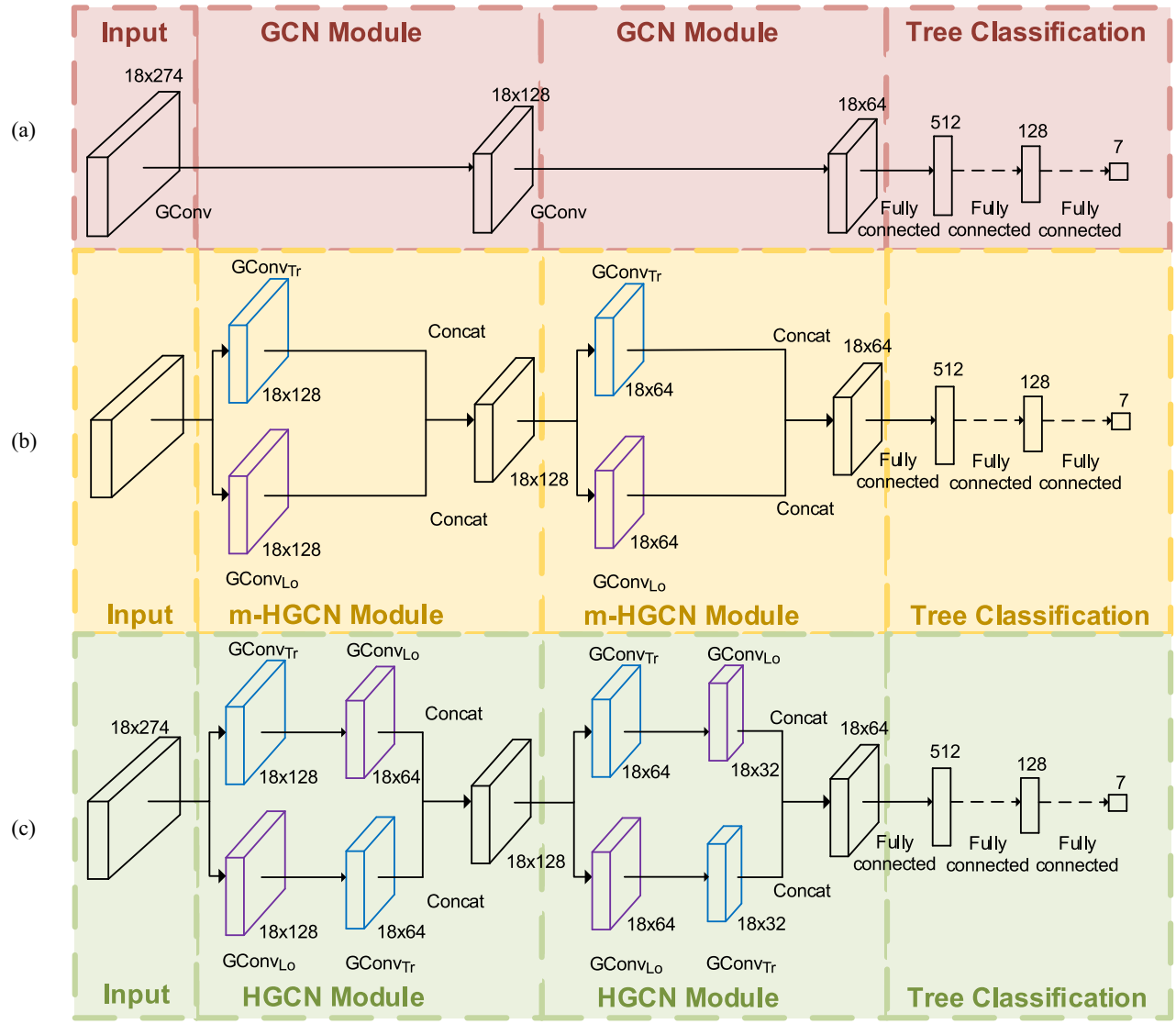


Fig. 3. (a) GCN model. (b) m-HGCN model. (c) HGCN model. The structure consists of four parts, which are Input, two GCN/m-HGCN/HGCN Modules, and TC framed with red/orange/green dotted lines. In (b) or (c), the purple and blue layers represent the $\mathrm{GConv_{Lo}}$ and $\mathrm{GConv_{Tr}}$, respectively.

With regard to physical position, the relationship between the electrode pairs in the International 10–20 system is utilized with the model based on GCN.

The spectral convolution on the graph is expressed as follows:

$$L = I_N - D^{-\frac{1}{2}}\mathrm{A}D^{-\frac{1}{2}} \tag{2}$$

$$L = U\Lambda U^T \tag{3}$$

$$g_\theta \star x = U g_\theta U^T x \tag{4}$$

where $x \in \mathbb{R}^N$ indicates all nodes in graph, $A \in \mathbb{R}^{N\times N}$ represents the adjacency matrix of graph, D refers to the degree matrix of A with $D_{ii} = \sum_j A_{ij}$, I_N denotes an identity matrix, L stands for the normalized graph Laplacian, and U and Λ refer to the eigenvectors and eigenvalues of L, respectively. g_θ indicates a function of Λ, i.e., $g_\theta(\Lambda)$, and $U^T x$ means the graph fourier transform of x. The calculation of U is a costly operation. By applying the method proposed in [33], $g_\theta(\Lambda)$ can be well approximated by a truncated expansion in terms of Chebyshev polynomials $T_k(x)$ up to Kth order

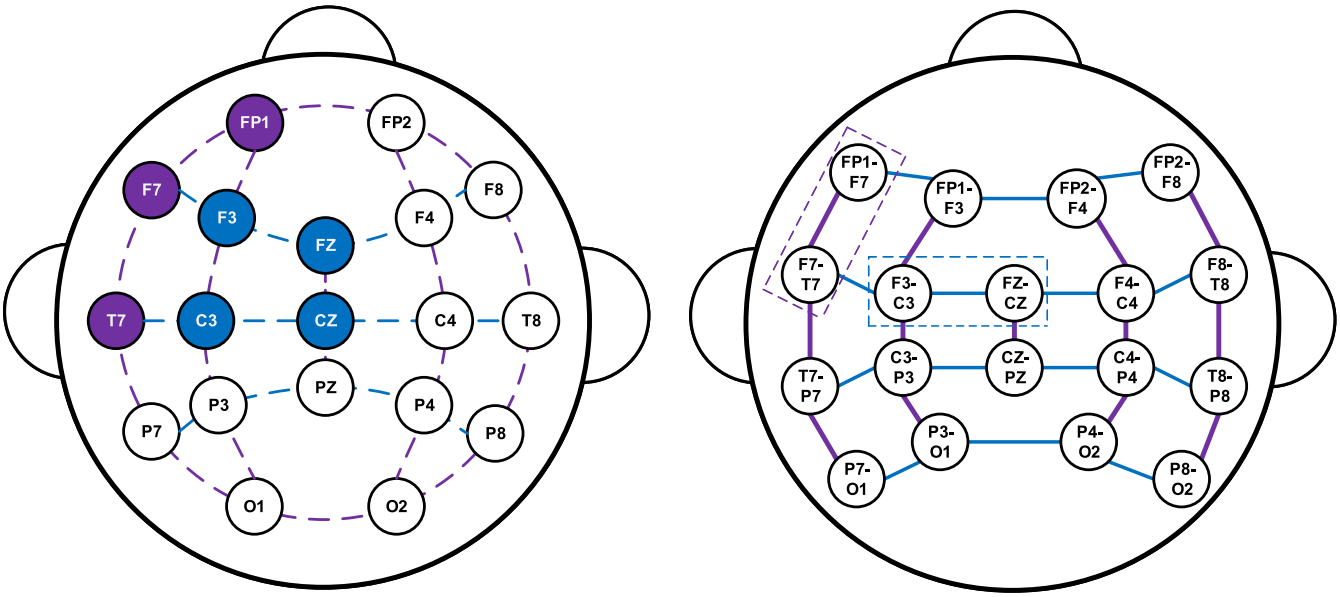
$$g_{\theta'}(\Lambda) \approx \sum_{k=0}^{K} \theta'_k T_k\left(\tilde{\Lambda}\right) \tag{5}$$

with a rescaled $\tilde{\Lambda} = (2/\lambda_{\max})\Lambda - I_N$, where $\lambda_{\max}$ indicates the maximum eigenvalue of L. The role of $\tilde{\Lambda}$ is to normalize the eigenvector matrix to $[-1, 1]$. $\theta' \in \mathbb{R}^K$ indicates the vector of Chebyshev coefficients. Equation (4) can be simplified by

$$g_{\theta'} \star x \approx \sum_{k=0}^{K} \theta'_k T_k(\tilde{L}) x \tag{6}$$

where $\tilde{L} = (2/\lambda_{\max})L - I_N$. With the condition of $\lambda_{\max} \approx 2$, (6) can be simplified as follows:

$$g_{\theta'} \star x \approx \theta'_0 x + \theta'_1 (L - I_N)x = \theta'_0 x - \theta'_1 D^{-\frac{1}{2}}\mathrm{A}D^{-\frac{1}{2}}x. \tag{7}$$



(a) (b)

Fig. 4. (a) Electrodes in the International 10–20 system leading in common reference coupling. (b) Electrode pairs in the International 10–20 system leading in longitudinal montage coupling. The blue and purple lines between the nodes indicate the transverse and longitudinal adjacent relationship, respectively.

With a parameter constraint $\theta = \theta_0' = -\theta_1'$, (7) can be simplified to

$$g_\theta \star x \approx \theta\left(I_N + D^{-\frac{1}{2}}\mathrm{A}D^{-\frac{1}{2}}\right)x = \theta S(A)x \tag{8}$$

where $S(A) = I_N + D^{-(1/2)}\mathrm{A}D^{-(1/2)}$ is denoted as the propagation matrix of A. In summary, Kipf and Welling [33] presented the layerwise linear formula of GCN

$$Z = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X\Theta \tag{9}$$

where $\tilde{A} = I_N + A$, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, and $S(A) = \tilde{D}^{-(1/2)}\tilde{A}\tilde{D}^{-(1/2)}$. In this article, $X \in \mathbb{R}^{N\times M}$ is all electrode pairs in the International 10–20 system, where N and M refer to the number of electrode pairs and features, respectively. The adjacency matrix is denoted as $A \in \mathbb{R}^{N\times N}$. The adjacency matrix is

$$A^{ij} = \begin{cases} 1, & \text{for } x_i \oplus x_j & \text{(10a)} \\ 0, & \text{for } !x_i \oplus x_j & \text{(10b)} \end{cases}$$

where $\oplus$ indicates that the two nodes are adjacent. $\Theta \in \mathbb{R}^{M\times F}$ represents the weight matrix of one GCN layer and $Z \in \mathbb{R}^{N\times F}$ indicates the layerwise linear output.

In one GCN layer, the propagation matrix is

$$H^{l+1} = \sigma\left(S\left(A^l\right)H^l W^l\right) \tag{11}$$

where l represents the layer number, H indicates the hidden state, and W denotes the weight matrix. In this article, σ is tanh, W is added with $L2$ normalization to prevent from overfitting, and A is added with $L1$ normalization to ensure that the maximum eigenvalue is approximately equal to 2.

The overall network model is illustrated in Fig. 3(a). Two GCN layers are applied to propagate the characteristics between adjacent electrode pairs. Two fully connected layers are used to summarize the characteristics of all electrode pairs. A dropout layer is added between the two fully connected layers to prevent overfitting, and a softmax layer is applied to the classification of a time segment. The classifications obtained from multiple time segments are voted for the whole classification.

C. Hierarchy Graph Convolution Network

In CHB-MIT or TUH data sets, the bipolar electrodes in the International 10–20 system lead to longitudinal montage coupling as shown in Fig. 4(b). For instance, the EEG signal in the FP1-F7 channel represents the voltage difference between the FP1 and F7 electrodes. It can be seen that there are two kinds of edges between the nodes in Fig. 4(b), the longitudinal edge indicated by the purple line and the transverse edge indicated by the blue line. For example, the two nodes F3-C3 and FZ-CZ connected by a transverse edge in a blue dotted frame in Fig. 4(b) are comprised of four blue nodes in Fig. 4(a), which are F3, C3, FZ, and CZ. The two nodes FP1-F7 and F7-T7 linked by another kind of longitudinal edge in a purple dotted frame indicated in Fig. 4(b) are consisted of three purple nodes in Fig. 4(a), including FP1, F7, and T7. It means that the transverse and longitudinal edges in Fig. 4(b) are two kinds of edges with different information and implication, thus it is inappropriate to represent both edges with 1 in the adjacency matrix of A in Section III-B. A natural idea is to represent the transverse and longitudinal edges with two independent adjacency matrixes, which are denoted as A_{Lo} and A_{Tr}.

A_{Lo} means the longitudinal adjacency matrix constructed by the longitudinal edges in Fig. 4(b) with

$$A_{\mathrm{Lo}}^{ij} = \begin{cases} 1, & x_i \oslash x_j & \text{(12a)} \\ 0, & !x_i \oslash x_j & \text{(12b)} \end{cases}$$

where $\oslash$ means the nodes are longitudinal adjacent in Fig. 4(b). The longitudinal graph convolution layer performs graph convolution with A_{Lo} is denoted as $\mathrm{GConv}_{\mathrm{Lo}}$. A_{Tr} means the transverse adjacency matrix constructed by the transverse edges in Fig. 4(b) with

$$A_{\mathrm{Tr}}^{ij} = \begin{cases} 1, & x_i \ominus x_j \quad (13a) \\ 0, & !x_i \ominus x_j \quad (13b) \end{cases}$$

where $\ominus$ means the nodes are transverse adjacent in Fig. 4(b). The transverse graph convolution layer performs graph convolution with A_{Tr} is denoted as $\mathrm{GConv}_{\mathrm{Tr}}$. In this article, the problem of information difference in the edges of Fig. 4(b) will be solved by $\mathrm{GConv}_{\mathrm{Lo}}$ and $\mathrm{GConv}_{\mathrm{Tr}}$ instead of the ordinary graph convolution layer in Section III-B.

HGCN and middle-HGCN (m-HGCN) are used to show the benefits obtained by applying $\mathrm{GConv}_{\mathrm{Lo}}$ and $\mathrm{GConv}_{\mathrm{Tr}}$ to the network. The difference between m-HGCN and HGCN is illustrated in Fig. 3(b) and (c). Both m-HGCN and HGCN consist of two similar base modules stacked together. In each of the base module, features are propagated through two branches. Each branch in the m-HGCN base module contains only one layer of $\mathrm{GConv}_{\mathrm{Lo}}$ or $\mathrm{GConv}_{\mathrm{Tr}}$ and uses a full connection layer to fuse the two-way branch. The m-HGCN design idea is to use $\mathrm{GConv}_{\mathrm{Lo}}$ and $\mathrm{GConv}_{\mathrm{Tr}}$ instead of normal GConv, but each branch of m-HGCN uses only partial graphical structure information when propagating features. The feature maps of the above-mentioned two paths are combined at the end of the m-HGCN module without the graphical structure information. The propagation formulas of m-HGCN module are expressed as

$$H^{l+1,1} = \sigma\left(S\left(A_{\mathrm{Tr}}^{l}\right)H^{l}W_{\mathrm{Tr}}^{l}\right) \quad (14)$$

$$H^{l+1,2} = \sigma\left(S\left(A_{\mathrm{Lo}}^{l}\right)H^{l}W_{\mathrm{Lo}}^{l}\right) \quad (15)$$

$$H^{l+1} = \sigma\left(\left[H^{l+1,1}; H^{l+1,2}\right]W^{l+1}\right) \quad (16)$$

where H^{l} and H^{l+1} are the input and the output of the lth m-HGCN layer, respectively. The activation function σ and the penalty function on A and W are the same as the GCN module in Section III-B.

The proposed HGCN structure utilizes all graph structure information for feature propagation hierarchically in each branch. In one branch, the features are transmitted to $\mathrm{GConv}_{\mathrm{Lo}}$, and then to $\mathrm{GConv}_{\mathrm{Tr}}$. In the other branch, the features are transmitted to $\mathrm{GConv}_{\mathrm{Tr}}$, and then to the $\mathrm{GConv}_{\mathrm{Lo}}$. The propagation formulas of HGCN module are expressed as

$$h^{l+1,1} = \sigma\left(S\left(A_{\mathrm{Lo}}^{l,1}\right)H^{l}W_{\mathrm{Lo}}^{l,1}\right) \quad (17)$$

$$H^{l+1,1} = \sigma\left(S\left(A_{\mathrm{Tr}}^{l,1}\right)h^{l+1,1}W_{\mathrm{Tr}}^{l,1}\right) \quad (18)$$

$$h^{l+1,2} = \sigma\left(S\left(A_{\mathrm{Tr}}^{l,2}\right)H^{l}W_{\mathrm{Tr}}^{l,2}\right) \quad (19)$$

$$H^{l+1,2} = \sigma\left(S\left(A_{\mathrm{Lo}}^{l,2}\right)h^{l+1,2}W_{\mathrm{Lo}}^{l,2}\right) \quad (20)$$

$$H^{l+1} = \sigma\left(\left[H^{l+1,1}; H^{l+1,2}\right]W^{l+1}\right) \quad (21)$$

where H^{l} and H^{l+1} are the input and the output of the lth HGCN layer, respectively. The activation function σ and the penalty function on A and W are the same as the GCN module in Section III-B.

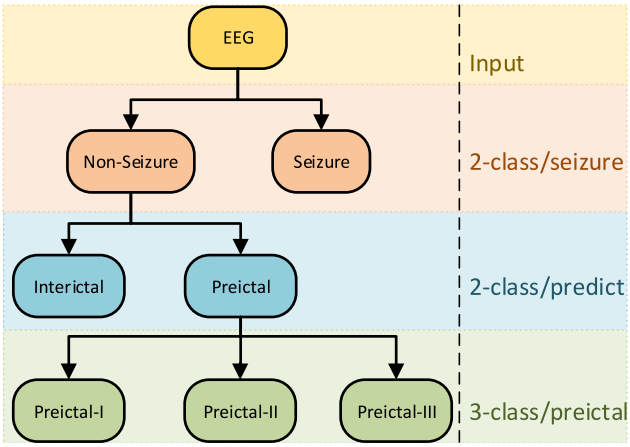


Fig. 5. Dependencies of categories are described by means of TC. The four kinds of classification tasks are transformed by three new classification tasks framed with orange, blue, and green lines. The yellow frame represents the input. The orange frame indicates the 2-class-seizure classification task. The blue frame denotes the 2-class-predict classification task. The green frame refers to the 3-class/preictal classification task.

The entire m-HGCN or HGCN model consists of two basic modules transmitting the input by different kinds of relationships in the structure of the International 10–20 system. The two fully connected layers are applied to summarize the characteristics of all electrode pairs. A dropout layer is added between the fully connected layers to prevent overfitting, and a softmax layer is added at the end of the model to classify a time segment classification. The processing of classification results obtained from multiple time segments is the same as the GCN model in Section III-B.

D. Tree Classification

In some classification tasks, including 2-class-predict task, 3-class task, and 5-class task, there is a relationship between the categories. Both interictal and preictal courses are classed into nonseizure courses that show some common characteristics compared with the seizure courses. Preictal-I, preictal-II, and preictal-III courses are all preictal course with some common characteristics compared with interictal or ictal courses. It is thus inappropriate to consider each category separately with each other in the aforementioned classification tasks.

TC is proposed to utilize the relationship as referred to above. The old four classification tasks 2-class-seizure, 2-class-predict, 3-class, and 5-class are transformed into three new classifications as shown in Fig. 5: 1) judge whether it is a seizure or nonseizure (2-class-seizure); 2) judge whether it is an interictal or preictal (2-class-predict); and 3) judge whether it is preictal-I, preictal-II, or preictal-III (3-class-preictal). The prediction probabilities of the above-mentioned three classification tasks are denoted as $P_{\mathrm{sei}} \in \mathbb{R}^2$, $P_{\mathrm{pred}} \in \mathbb{R}^2$, and $P_{\mathrm{pre}} \in \mathbb{R}^3$, respectively. To be specific, P_{sei}^{0} and P_{sei}^{1} are the probabilities of seizure and nonseizure category in task 2-class-seizure, respectively. P_{pred}^{0} and P_{pred}^{1} represent the probabilities of interictal and preictal category in task

TABLE I
PROBABILITIES OF THE 2-CLASS-SEIZURE, 2-CLASS-PREDICT, 3-CLASS, AND 5-CLASS TASK ARE TRANSFORMED BY THE PROBABILITIES OF THE 2-CLASS-SEIZURE, 2-CLASS-PREDICT, AND 3-CLASS-PREICTAL, WHICH ARE DEFINED IN TC

Task	**Probability**				
2-class-seizure	**NonSeizure**				**Seizure**
	P^0_{sei}				P^1_{sei}
2-class-predict	**Interictal**	**Preictal**			/
	P^0_{pred}	P^1_{pred}			/
3-class	**Interictal**	**Preictal**			**Ictal**
	$\mathrm{P}^0_{sei}P^0_{pred}$	$\mathrm{P}^0_{sei}P^1_{pred}$			P^1_{sei}
5-class	**Interictal**	**Preictal-I**	**Preictal-II**	**Preictal-III**	**Ictal**
	$\mathrm{P}^0_{sei}P^0_{pred}$	$\mathrm{P}^0_{sei}P^1_{pred}P^0_{pre}$	$\mathrm{P}^0_{sei}P^1_{pred}P^1_{pre}$	$\mathrm{P}^0_{sei}P^1_{pred}P^2_{pre}$	P^1_{sei}

2-class-predict, respectively. P^0_{pre}, P^1_{pre}, and P^2_{pre} indicate the probabilities of preictal-I, preictal-II and preictal-III category in task 3-class-preictal, respectively.

The prediction probabilities of the old four tasks are transformed by the three new tasks as shown in Table I.

In the training, the labels of samples in the three new classification tasks 2-class-seizure, 2-class-predict, and 3-class-preictal are also denoted as $L_{\text{sei}} \in \mathbb{R}^2$, $L_{\text{pred}} \in \mathbb{R}^2$, and $L_{\text{pre}} \in \mathbb{R}^3$, respectively. The loss function consists of three parts as follows:

$$\mathbb{L} = -\sum_{\text{task}}^{[\text{sei, pred, pre}]} \alpha_{\text{task}} \sum_{i}^{N_{\text{task}}} L^i_{\text{task}} * \log\left(P^i_{\text{task}}\right) \tag{22}$$

where $N_{\text{sei}} = 2$, $N_{\text{pred}} = 2$, and $N_{\text{pre}} = 3$. In this article, adjustment is made to weight for 2-class-seizure, 2-class-predict, and 3-class-preictal tasks with $\alpha_{\text{sei}} = 1.0$, $\alpha_{\text{pred}} = 1.5$, and $\alpha_{\text{pre}} = 2.0$, respectively.

E. Preictal Fuzzification

In the 5-class classification task as mentioned in this article, the preictal-I, preictal-II, and preictal-III courses are defined as the first 20 min, the middle 20 min, and the last 20 min of the entire one-hour preictal course, respectively. In this classification task, the more accurate categories in the preictal course are predictable. The EEG signals in preictal-I, preictal-II, and preictal-III courses from different onsets may behave in a similar way due to the interval length of the preictal course before the onset is unstable. In other cases, the EEG signals in preictal-I and preictal-II courses may behave similarly near their demarcation point, which is the same as preictal-II and preictal-III courses.

The proposed PF is intended to modify the label of one-hot preictal category representation with formula as

$$L_{\text{pre}} = \begin{cases} (1-\epsilon, \epsilon, 0), & Y = \text{Preictal}_{\text{I}} \quad (23a) \\ (0.5\epsilon, 1-\epsilon, 0.5\epsilon), & Y = \text{Preictal}_{\text{II}} \quad (23b) \\ (0, \epsilon, 1-\epsilon), & Y = \text{Preictal}_{\text{III}} \quad (23c) \end{cases}$$

where $\epsilon = 0.1$.

IV. EXPERIMENTS

A. Experiment Environment

Experiments are carried out on two epileptic EEG data sets, including CHB-MIT data set and TUH data set, for patient-specific validation, cross-validation, and comparison.

The CHB-MIT data set is collected by Boston Children's Hospital and is primarily used to record EEG in adolescent patients with controllable epilepsy. The subjects in the data set consist of 5 males (3–22 years) and 17 females (1.5–19 years). The scalp EEG signals are sampled at 256 Hz and represented as 16-b floating-point type. Each recording is about 1-h long. The data set has a total of over 958-h long recording which contains 198 seizures.

The TUH data set in [34] is collected by Temple University Hospital, and the subset for 2-class-seizure classification task of the TUH data set v1.1.0 is adopted. The data set consists of 6724 subjects at all ages. The scalp EEG signals are sampled at 250 Hz and represented as 16-b floating-point type. Each recording in the TUH data set is about 20-min long. The data set has a total of over 3048-h long recording which contains 1421 seizures. The TUH data set consists of more subjects across all age groups, so the experiments on this data set are more relevant to the real-world scenario and more instructive for clinical application.

The International 10–20 system of EEG electrode positions and nomenclature was applied to make these recordings. The experiments adopt the common EEG signals captured from 18 bipolar electrode leads in Fig. 4, including FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FZ-CZ, CZ-PZ, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, and P8-O2.

The CHB-MIT data set is classed into five categories, including interictal, preictal-I, preictal-II, preictal-III, and ictal in four classification tasks, including 2-class-seizure, 2-class-predict, 3-class, and 5-class. The interictal segmentation is defined as 4 hr before or after a seizure. The preictal segmentation means a time segmentation in 1 hr before a seizure. The preictal-I, preictal-II, and preictal-III courses refer to the first 20 min, middle 20 min, and last 20 min in an hour-long preictal segmentation, respectively. Only the leading seizure is recognized in the case of multiple consecutive and compact seizures within an hour minutes time interval. Considering the limited total duration of seizure, the EEG signals in the seizure

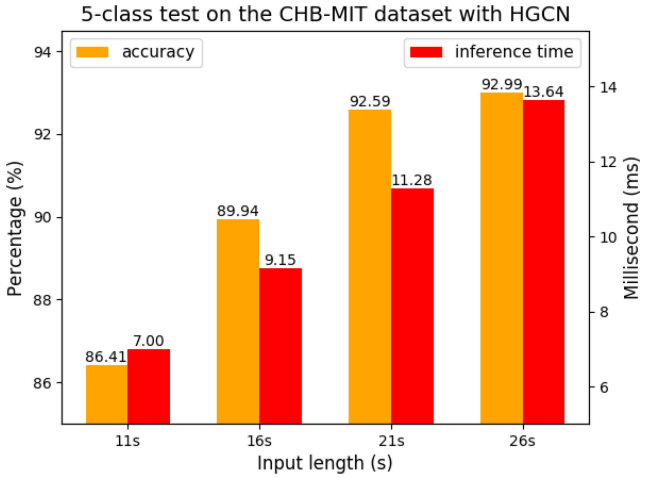


Fig. 6. Accuracy and inference time of the HGCN model for 5-class cross-validation on the CHB-MIT data set.

course are segmented with a 50% overlapping. Since the TUH data set was recorded in less than 1 hr, it is not possible to subdivide the preictal course every 20 min when labeling the data. The TUH data set is only classed into three categories, including intericatl, preictal, and ictal in three classification tasks, including 2-class-seizure, 2-class-predict, and 3-class.

The input length of EEG should be long enough for classification. However, too long EEG length may affect the detection time. Therefore, the input lengths in [15]–[18], [20], [21], and [23] are from 16 to 30 s to analyze the recording in both CHB-MIT and TUH data sets. The lengths of 11, 16, 21 and 26 s are tested on the CHB-MIT data set. As shown in Fig. 6, the length of 21 s is the most cost effective, which guarantees both accuracy and time consumption. In the rest of this article, the length of an EEG input is 21 s.

The experiments are conducted on the benchmark of TensorFlow 1.12.0 with a GeForce RTX 2080 Ti NVIDIA GPU.

B. Patient-Specific Experiments

In this section, the data collected from every single patient in the CHB-MIT data set will be used to assess the performance of the proposed model in detail. The recordings of subjects were collected from 22 subjects (5 males, ages 3–22; and 17 females, ages 1.5–19). When splitting the seizure samples of chb06, chb12, chb14, chb16, and chb20, the actual seizure duration in the sample is allowed to be less than 21 s.

The patient-specific experiments consist of four classification tasks, including 2-class-seizure, 2-class-predict, 3-class, and 5-class. In the 2-class-seizure and 2-class-predict experiments, sensitivity and specificity will be tested to evaluate the capability of the proposed method in each category. In the 3-class and 5-class, accuracy will be tested to evaluate the capability in all categories. The results of the above-mentioned metrics are presented in the Table II. It can be seen from the table that the metrics are almost 100%, and the mean values are also higher than 99.79% with low standard deviation, which suggests that the proposed model is applicable to each patient with high accuracy and durable stability.

TABLE II
PATIENT-SPECIFIC EXPERIMENTS IN 2-CLASS-SEIZURE, 2-CLASS-PREDICT, 3-CLASS, AND 5-CLASS CLASSIFICATION TASKS. THE METRICS CONSIST OF SENSITIVITY (SENS), SPECIFICITY (SPEC), AND ACCURACY (ACCU). THE TABLE LISTS THE METRICS OF ALL PATIENTS AS WELL AS THEIR MEAN AND STANDARD DEVIATION

Case	2-class-seizure		2-class-predict		3-class	5-class
	Sens(%)	Spec(%)	Sens(%)	Spec(%)	Accu(%)	Accu(%)
chb01	100.0	100.0	100.0	100.0	100.0	100.0
chb02	100.0	100.0	100.0	100.0	100.0	100.0
chb03	100.0	100.0	100.0	100.0	100.0	100.0
chb04	100.0	100.0	100.0	100.0	100.0	100.0
chb05	100.0	100.0	100.0	99.5	99.7	99.5
chb06	100.0	100.0	98.6	98.1	98.7	97.9
chb07	100.0	100.0	100.0	100.0	100.0	100.0
chb08	100.0	100.0	100.0	100.0	100.0	100.0
chb09	100.0	100.0	100.0	100.0	100.0	100.0
chb10	100.0	100.0	98.8	100.0	99.7	99.7
chb11	100.0	100.0	100.0	100.0	100.0	100.0
chb12	100.0	100.0	100.0	100.0	100.0	99.8
chb13	100.0	100.0	100.0	100.0	100.0	100.0
chb14	100.0	100.0	100.0	100.0	100.0	100.0
chb15	100.0	100.0	100.0	100.0	100.0	98.5
chb16	100.0	100.0	100.0	100.0	100.0	100.0
chb17	100.0	100.0	100.0	100.0	100.0	100.0
chb18	100.0	100.0	100.0	100.0	100.0	100.0
chb19	100.0	100.0	100.0	100.0	100.0	100.0
chb20	100.0	100.0	100.0	100.0	100.0	100.0
chb21	100.0	100.0	99.3	100.0	99.7	99.7
chb22	100.0	100.0	100.0	100.0	100.0	100.0
chb23	100.0	100.0	100.0	100.0	100.0	100.0
mean	**100.00**	**100.00**	**99.86**	**99.90**	**99.90**	**99.79**
std	**0.00**	**0.00**	**0.39**	**0.41**	**0.28**	**0.53**

C. Ablation Experiments Based on Cross-Validation

The three contributions, including TC, PF, and HGCN will be validated by conducting cross-validation experiments, in which fivefold cross-validation is adopted to test the accuracy of four classification tasks including 2-class-seizure, 2-class-predict, 3-class, and 5-class.

1) Results on CHB-MIT Data Set: The comparison of the GCN and GCN+TC models is shown in Table III. The accuracy of 2-class-seizure shows only an insignificant improvement by 0.05%. The accuracy of 2-class-predict, 3-class, and 5-class shows a noticeable improvement by 0.66%, 0.79%, and 1.90%, respectively, which indicates that the classification tasks whose categories have a dependency on each other are improved by TC.

The comparison of the GCN+TC and GCN+TC+PF models is shown in Table III. The accuracy of 2-class-seizure, 2-class-predict, and 3-class shows only an insignificant improvement by 0.17%, 0.03%, and 0.16%, respectively. The accuracy of 5-class achieves a significant improvement by 1.03%, which evidences that the 5-class classification task is improved by PF, to reflect the phenomenon of noise in actual preictal labels from different epileptic seizures.

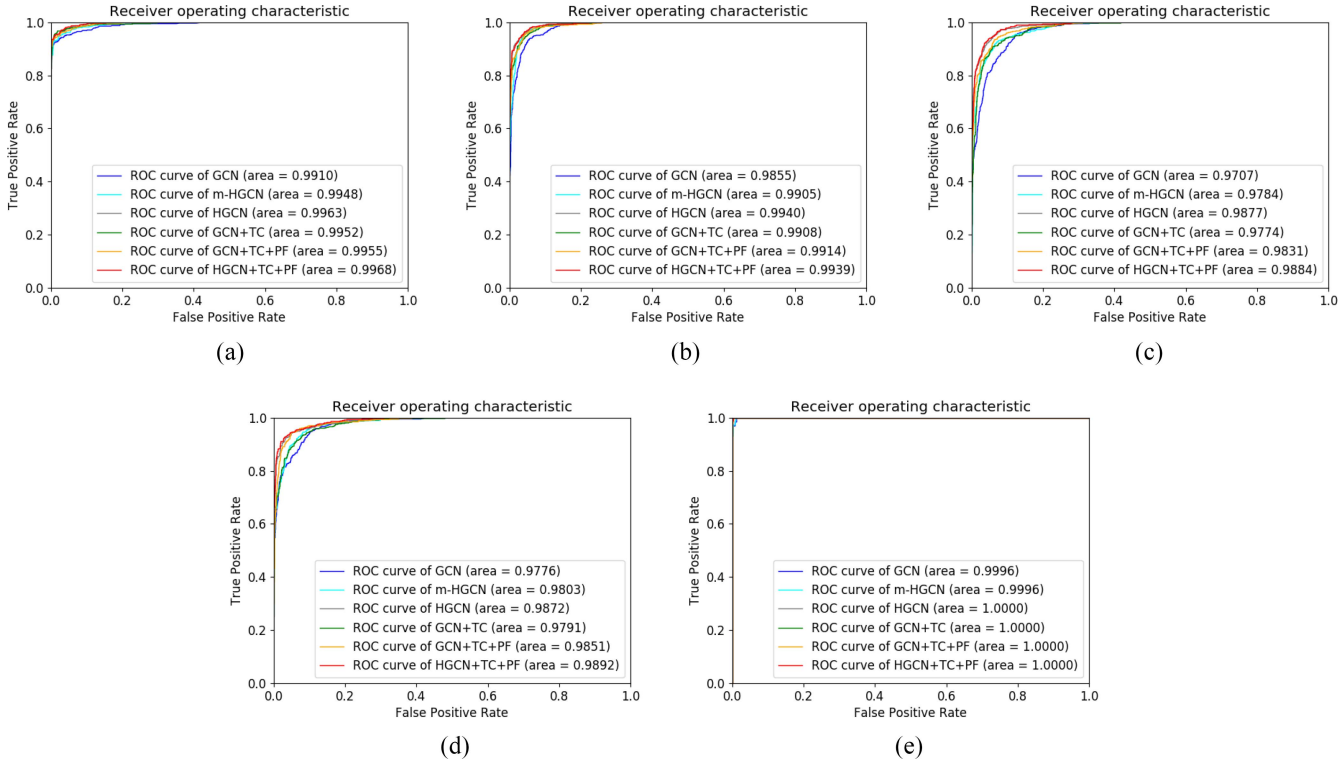


Fig. 7. ROC curves and ROC-AUCs of (a) Interictal, (b) preictal-I, (c) preictal-II, (d) preictal-III, and (e) ictal from different models in the 5-class classification task on the CHB-MIT data set.

TABLE III
CROSS-VALIDATION EXPERIMENTS IN 2-CLASS-SEIZURE, 2-CLASS-PREDICT, 3-CLASS, AND 5-CLASS CLASSIFICATION TASKS ON THE CHB-MIT DATA SET

Model	Accuracy(%)			
	2-class-seizure	2-class-predict	3-class	5-class
GCN	99.08	97.04	96.98	89.00
m-HGCN	99.13	97.96	97.51	91.07
HGCN	99.40	98.02	97.81	92.59
GCN+TC	99.13	97.70	97.77	90.90
GCN+TC+PF	99.30	97.73	97.93	91.93
HGCN+TC+PF	**99.72**	**98.76**	**98.64**	**93.72**

TABLE IV
CROSS-VALIDATION EXPERIMENTS IN 2-CLASS-SEIZURE, 2-CLASS-PREDICT, AND 3-CLASS CLASSIFICATION TASKS ON THE TUH DATA SET

Model	Accuracy(%)		
	2-class-seizure	2-class-predict	3-class
GCN	92.66	90.22	86.14
m-HGCN	94.52	91.12	88.64
HGCN	95.72	93.22	91.24
HGCN+TC	**96.15**	**94.73**	**92.68**

The results of the GCN, m-HGCN, and HGCN models are shown in Table III. It can be seen that both m-HGCN and HGCN outperform GCN on several of the classification tasks mentioned above. The HGCN model shows better performance than the typical GCN model or imperfect m-HGCN model by utilizing the locational topological relationship in electrodes on the scalp. Compared with GCN, HGCN improves the accuracy of the 2-class-seizure, 2-classpredict, 3-class, and 5-class classification tasks by 0.32%, 0.98%, 0.83%, and 3.59%, respectively. HGCN+TC+PF shows 0.42%, 1.03%, 0.71%, and 1.79% accuracy improvement over GCN+TC+PF in 2-class-seizure, 2-classpredict, 3-class, and 5-class, respectively.

The receiver operating characteristic (ROC) curves and their area under the curve (AUC) for each class on the 5-class of the CHB-MIT data set are shown in Fig. 7. In all five categories, the GCN+TC model has a significant improvement over the GCN model in ROC-AUC. In the three categories preictal-I, preictal-II, and preictal-III, the GCN+TC+PF model obtained a 0.0006, 0.0057, and 0.0060 improvement over the GCN+TC model in ROC-AUC, respectively. In the comparison of the three graph convolution-based models, including GCN, m-HGCN, and HGCN, the m-HGCN has a significant improvement from GCN, and HGCN obtains the best ROC-AUC in all five categories. The proposed HGCN+TC+PF model outperforms the other models with 0.9968, 0.9939, 0.9884, 0.9892, and 1.0000 of ROC-AUC in interictal, preictal-I, preictal-II, preictal-III, and ictal, respectively.

Compared with the patient-specific experiments in Section IV-B, the accuracy of tasks 2-class-seizure, 2-class-predict, 3-class, and 5-class reaches as high as 99.72%, 98.76%, 98.64%, and 93.72%, respectively, which suggests that the proposed model HGCN+TC+PF possesses an outstanding capability of generalization in different classification tasks.

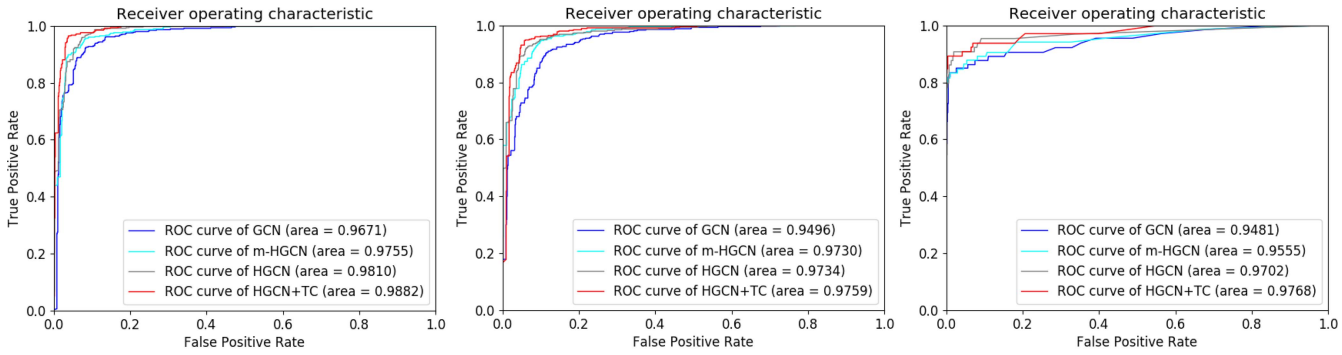


Fig. 8. ROC curves and ROC-AUCs of (a) Interictal, (b) preictal, and (c) ictal from different models in the 3-class classification task on the TUH data set.

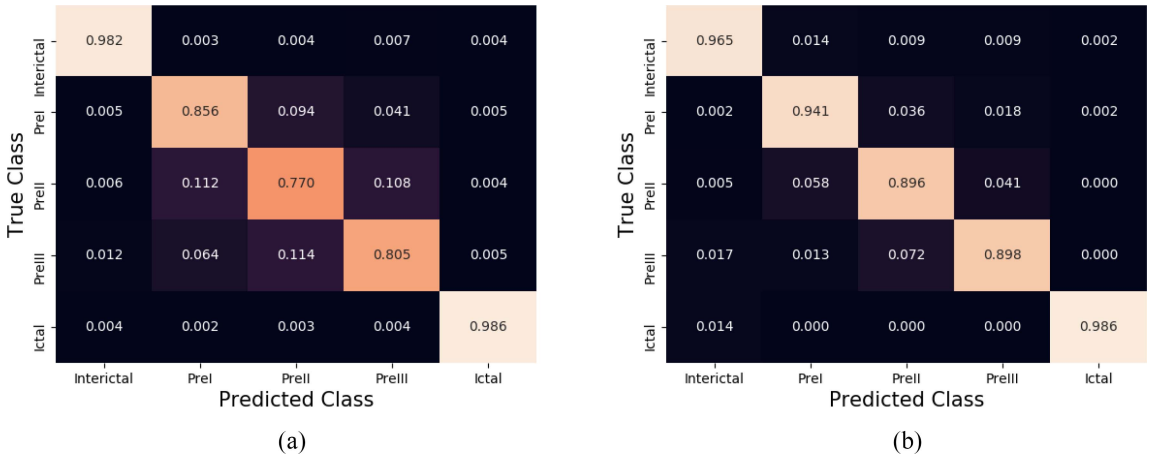


Fig. 9. (a) Confusion matrix of the state-of-the-art model SCNN and (b) confusion matrix of the proposed model HGCN+TC+PF in the 5-class cross-validation task.

2) Results on TUH Data Set: The GCN, m-HGCN, HGCN, and HGCN+TC models are compared in Table IV. It can be seen that both m-HGCN and HGCN outperform GCN on 2-class-seizure, 2-class-predict, and 3-class classification tasks. When compared with GCN, m-HGCN has a noticeable accuracy improvement of 1.86%, 0.90%, and 2.50% for the 2-class-seizure, 2-class-predict, and 3-class classification tasks, respectively. By utilizing all the structure information hierarchically in each branch, HGCN improves the accuracy of the 2-class-seizure, 2-class-predict, and 3-class classification tasks by 3.06%, 3.00%, and 5.10%, respectively, compared with GCN.

The proposed TC describes the initial dependencies of the categories in 2-class-predict and 3-class classification tasks. Therefore, the HGCN+TC model further improves the accuracy of the 2-class-seizure, 2-class-predict, and 3-class classification by an additional 0.43%, 1.51%, and 1.44%, respectively.

The ROC curves and ROC-AUC for each class on the 3-class of the TUH data set are shown in Fig. 8. The HGCN+TC model has a 0.0072, 0.0025,and 0.0066 improvement in ROC-AUC over the HGCN model in interictal, preictal, and ictal, respectively. In the comparison of the three graph convolution-based models, including GCN, m-HGCN, and HGCN, the m-HGCN has a significant improvement from GCN, and HGCN obtains the best ROC-AUC in all three categories. The proposed HGCN+TC model outperforms the other models with 0.9882, 0.9759, and 0.9768 of ROC-AUC in interictal, preictal, and ictal, respectively.

The accuracy of tasks 2-class-seizure, 2-class-predict, and 3-class reaches as high as 96.15%, 94.73%, and 92.68%, respectively, which suggests that the HGCN+TC model possesses an outstanding capability in such a TUH data set composed of thousands of subjects.

TABLE V
COMPARISON IN THE TUH DATA SET BETWEEN THE PROPOSED MODEL AND THE STATE OF THE ART IN 2-CLASS-SEIZURE CROSS-VALIDATION. THE BOLD POSITION REPRESENTS THE OPTIMAL RESULT OF EACH PART

2-class-seizure cross-validation			
Author	**Model**	**Performance(%)**	
		Sens	**Spec**
Shav *et al.* [17]	CNN-LSTM	90	61
Golmohammadi *et al.* [16]	CNN-LSTM	96.86	69.17
Golmohammadi *et al.* [21]	CNN-GRU	97.10	69.17
This work	**HGCN+TC**	**99.53**	**88.87**

TABLE VI
COMPARISON IN THE CHB-MIT DATA SET BETWEEN THE PROPOSED MODEL AND THE STATE OF THE ART IN FIVE PARTS, INCLUDING 2-CLASS-SEIZURE PATIENT SPECIFIC, 2-CLASS-SEIZURE CROSS-VALIDATION, 2-CLASS-PREDICT PATIENT SPECIFIC, 3-CLASS CROSS-VALIDATION, AND 5-CLASS CROSS-VALIDATION. THE BOLD POSITION REPRESENTS THE OPTIMAL RESULT OF EACH PART

2-class-seizure patient-specific					
Author	**Model**	**Performance(%)**			
		Accu	**Sens**	**Spec**	**Prec**
Thodoroff *et al.* [18]	RCNN	/	87	/	/
Truong *et al.* [15]	CNN	/	81.2	/	/
Truong *et al.* [23]	GAN	/	77.68	/	/
This work	**HGCN+TC+PF**	**100.0**	**100.0**	**100.0**	**100.0**
2-class-seizure cross-validation					
Author	**Model**	**Performance(%)**			
		Accu	**Sens**	**Spec**	**Prec**
Yuan *et al.* [25]	SAE	96.6	/	/	/
Yao *et al.* [20]	BiLSTM	/	88.63	87	88.6
Wen *et al.* [6]	AE-CDNN	91	/	/	/
Cao *et al.* [10]	SCNN	99.33	/	/	/
Craley *et al.* [11]	PGM-CNN	/	61	**98.7**	74
This work	**HGCN+TC+PF**	**99.72**	**99.93**	98.5	**99.6**
2-class-predict patient-specific					
Author	**Model**	**Performance(%)**			
		Sens		**Spec**	
Tsiouris *et al.* [22]	LSTM	99.63		99.78	
This work	**HGCN+TC+PF**	**99.86**		**99.90**	
3-class cross-validation					
Author	**Model**	**Performance(%)**			
		Accu			
Cao *et al.* [10]	SCNN	98.62			
This work	**HGCN+TC+PF**	**98.64**			
5-class cross-validation					
Author	**Model**	**Performance(%)**			
		Accu			
Cao *et al.* [10]	SCNN	87.95			
Hu *et al.* [2]	CNN	86.25			
This work	**HGCN+TC+PF**	**93.72**			

D. Comparison With Other Methods

1) Comparison on CHB-MIT Data Set: The CHB-MIT data set is a large public data set used in many works [2], [6], [10], [11], [15], [18], [20], [22], [23], [25] with two main kinds of validation cross-validation and patient specific.

The classification tasks on this data set include of 2-class-seizure, 2-class-predict, 3-class, and 5-class. Below is the comparison with the state of the art in all of the above-mentioned classification tasks.

Truong *et al.* [15], Thodoroff *et al.* [18], and Truong *et al.* [23] classified seizure and nonseizure courses of 2-class-seizure by patient-specific experiments. The comparison as performed between the proposed model and the above-mentioned works is shown in the 2-class-seizure patient-specific part of Table VI. It is demonstrated that the HGCN+TC+PF model achieves a 13.0% improvement of accuracy compared with the state of the art. The metrics in this task are invariably 100%.

Wen and Zhang [6], Cao *et al.* [10], Craley *et al.* [11], Yao *et al.* [20], and Yuan *et al.* [25] classified seizure and nonseizure courses of 2-class-seizure by cross-validation experiments. The comparison drawn between the proposed model and the aforementioned works is presented in the 2-class-seizure cross-validation part of Table VI. Compared with the state-of-the-art, the HGCN+TC+PF model shows an improvement by 0.39%, 11.3%, and 11.0% for accuracy, sensitivity, and precision, respectively.

Tsiouris *et al.* [22] classified interictal and preictal courses to predict seizure before the onset of 2-class-seizure by patient-specific experiments. The comparison made between the proposed model and the aforementioned work is shown in the 2-class-seizure patient-specific part of Table VI. Compared with the state of the art, the HGCN+TC+PF model achieves an improvement of 0.26% and 0.19% for sensitivity and specificity, respectively. The patient-specific experiments conducted on the 2-class-seizure classification task have produced a result of nearly 100%.

Cao *et al.* [10] classified interictal, preictal, and ictal courses of 3-class by means of cross-validation experiments. The comparison performed between the proposed model and the above-mentioned work is presented in the 3-class cross-validation part of Table VI. It is found out that the proposed HGCN+TC+PF model shows a 0.02% improvement in accuracy compared with state of the art.

Hu *et al.* [2] and Cao *et al.* [10] classified interictal, preictal-I, preictal-II, preictal-III, and ictal courses of 5-class by conducting cross-validation experiments. The comparison drawn between the proposed model and the above-mentioned works is presented in the 5-class cross-validation part of Table VI. The confusion matrices of the state-of-the-art model and the proposed model are illustrated in Fig. 6(a) and (b), respectively, which show that the accuracy of the preictal-I, preictal-II, and preictal-III category of the proposed model is significantly better than the state of the art with an improvement of 8.5%, 12.6%, and 9.3%, respectively. In summary, the proposed HGCN+TC+PF model shows a 5.77% improvement in accuracy compared with the state of the art.

2) Comparison on TUH Data Set: The TUH data set is a large public data set used in the previous works [16], [17], [21] with cross-validation of 2-class-seizure classification task. Below is the comparison drawn with the state of the art.

Golmohammadi *et al.* [16], Shah *et al.* [17], and Golmohammadi *et al.* [21] classified seizure and nonseizure courses of 2-class-seizure by cross-validation experiments. The comparison drawn between the proposed model and the above-mentioned work is presented in the 2-class-seizure cross-validation part of Table V. Compared with the state of the art, the HGCN+TC model achieves a significant improvement of 2.43% and 19.70% for sensitivity and specificity, respectively.

V. CONCLUSION

The proposed model has realized the full utilization of the locational topological relationship in electrodes on the scalp and treats different adjacent relationship by HGCN.

Moreover, the model has been adapted to different classification tasks by TC and PF. It has achieved the most desirable results from cross-validation or patient-specific experiments in all four classification tasks, including 2-class-seizure, 2-class-predict, 3-class, and 5-class. Compared with the state of the art, the HGCN+TC+PF has achieved a 5.77% improvement of accuracy in the 5-class task on the CHB-MIT data set, and the HGCN+TC has achieved an improvement of 2.43% and 19.7% for sensitivity and specificity, respectively, in the 3-class task on the TUH data set.

ACKNOWLEDGMENT

A team of investigators from Children's Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT) created and contributed this database to PhysioNet. The clinical investigators from CHB include Jack Connolly, REEGT; Herman Edwards, REEGT; Blaise Bourgeois, MD; and S. Ted Treves, MD. The investigators from MIT include Ali Shoeb, Ph.D., and Prof. John Guttag.

REFERENCES

[1] Z. Zhang and K. K. Parhi, "Low-complexity seizure prediction from IEEG/SEEG using spectral power and ratios of spectral power," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 3, pp. 693–706, Jun. 2016.

[2] W. Hu, J. Cao, X. Lai, and J. Liu, "Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks," *J. Ambient Intell. Humanized Comput.*, to be published.

[3] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," 2015. [Online]. Available: arXiv:1511.04306.

[4] A. Antoniades, L. Spyrou, C. C. Took, and S. Sanei, "Deep learning for epileptic intracranial EEG data," in *Proc. IEEE 26th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, 2016, pp. 1–6.

[5] I. Ullah, M. Hussain, E.-U.-H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Syst. Appl.*, vol. 107, pp. 61–71, Oct. 2018.

[6] T. Wen and Z. Zhang, "Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals," *IEEE Access*, vol. 6, pp. 25399–25410, 2018.

[7] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.

[8] Y. Hao, H. M. Khoo, N. von Ellenrieder, N. Zazubovits, and J. Gotman, "DeepIED: An epileptic discharge detector for EEG-fMRI based on deep learning," *NeuroImage Clin.*, vol. 17, pp. 962–975, Dec. 2018.

[9] M. Sun, F. Wang, T. Min, T. Zang, and Y. Wang, "Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm," *IEEE Access*, vol. 6, pp. 77596–77605, 2018.

[10] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cogn. Develop. Syst.*, early access, Aug. 20, 2019, doi: 10.1109/TCDS.2019.2936441.

[11] J. Craley, E. Johnson, and A. Venkataraman, "Integrating convolutional neural networks and probabilistic graphical modeling for epileptic seizure detection in multichannel EEG," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2019, pp. 291–303.

[12] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Optimized deep learning for EEG big data and seizure prediction BCI via Internet of Things," *IEEE Trans. Big Data*, vol. 3, no. 4, pp. 392–404, Dec. 2017.

[13] D. Lu and J. Triesch, "Residual deep convolutional neural network for EEG signal classification in epilepsy," 2019. [Online]. Available: arXiv:1903.08100.

[14] M. T. Avcu, Z. Zhang, and D. W. S. Chan, "Seizure detection using least EEG channels by deep convolutional neural network," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 1120–1124.

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

[16] M. Golmohammadi, S. Ziyabari, V. Shah, S. L. de Diego, I. Obeid, and J. Picone, "Deep architectures for automated seizure detection in scalp EEGs," 2017. [Online]. Available: arXiv:1712.09776.

[17] V. Shah, M. Golmohammadi, S. Ziyabari, E. Von Weltin, I. Obeid, and J. Picone, "Optimizing channel selection for seizure detection," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, 2017, pp. 1–5.

[18] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 178–190.

[19] R. Hussein, H. Palangi, R. Ward, and Z. J. Wang, "Epileptic seizure detection: A deep learning approach," 2018. [Online]. Available: arXiv:1803.09848.

[20] X. Yao, X. Li, Q. Ye, Y. Huang, Q. Cheng, and G.-Q. Zhang, "A robust deep learning approach for automatic classification of seizures against non-seizures," 2018. [Online]. Available: arXiv:1812.06562.

[21] M. Golmohammadi *et al.*, "Gated recurrent networks for seizure detection," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, 2017, pp. 1–5.

[22] 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, Aug. 2018.

[23] N. D. Truong, L. Zhou, and O. Kavehei, "Semi-supervised seizure prediction with generative adversarial networks," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, 2019, pp. 2369–2372.

[24] I. Covert *et al.*, "Temporal graph convolutional networks for automatic seizure detection," 2019. [Online]. Available: arXiv:1905.01375.

[25] Y. Yuan *et al.*, "A novel channel-aware attention framework for multi-channel EEG seizure detection via multi-view deep learning," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, 2018, pp. 206–209.

[26] K. Singh and J. Malhotra, "Stacked autoencoders based deep learning approach for automatic epileptic seizure detection," in *Proc. IEEE 1st Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, 2018, pp. 249–254.

[27] Q. Lin *et al.*, "Classification of epileptic EEG signals with stacked sparse autoencoder based on deep learning," in *Proc. Int. Conf. Intell. Comput.*, 2016, pp. 802–810.

[28] M. Golmohammadi, A. H. H. N. Torbati, S. L. de Diego, I. Obeid, and J. Picone, "Automatic analysis of EEGs using big data and hybrid deep learning architectures," *Front. Human Neurosci.*, vol. 13, p. 76, Mar. 2019.

[29] S. Gasparini *et al.*, "Information theoretic-based interpretation of a deep neural network approach in diagnosing psychogenic non-epileptic seizures," *Entropy*, vol. 20, no. 2, p. 43, 2018.

[30] N. Singh and S. Dehuri, "Usage of deep learning in epileptic seizure detection through EEG signal," in *Nanoelectronics, Circuits and Communication Systems*. Singapore: Springer, 2019, pp. 219–228.

[31] M.-P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, and D. Pompili, "Cloud-based deep learning of big EEG data for epileptic seizure prediction," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, 2016, pp. 1151–1155.

[32] W. Wu *et al.*, "An electroencephalographic signature predicts antidepressant response in major depression," *Nat. Biotechnol.*, vol. 38, no. 4, pp. 439–447, 2020.

[33] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016. [Online]. Available: arXiv:1609.02907.

[34] A. Harati, S. Lopez, I. Obeid, J. Picone, M. Jacobson, and S. Tobochnik, "The TUH EEG corpus: A big data resource for automated EEG interpretation," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, 2014, pp. 1–5.

Difei Zeng (Student Member, IEEE) received the bachelor's degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2017, where he is currently pursuing the M.A.Eng. degree with the College of Information Science Electronic Engineering.

His research interests include computer vision, deep learning, signal process, and data analysis.

Kejie Huang (Senior Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, National University of Singapore, Singapore, in 2014.

He spent five years in IC design industry, including Samsung, Seoul, South Korea, and Xilinx, San Jose, CA, USA, two years in the Data Storage Institute, Agency for Science Technology and Research, Singapore, and another three years in the Singapore University of Technology and Design, Singapore. He has been a Principal Investigator with the College of Information Science Electronic Engineering, Zhejiang University, Hangzhou, China, since 2016. He has authored or coauthored more than 30 scientific papers in international peer-reviewed journals and conference proceedings. He holds four granted international patents, and another eight pending ones. His research interests include low-power circuits and systems design using emerging nonvolatile memories, architecture and circuit optimization for reconfigurable computing systems and neuromorphic systems, machine learning, and deep learning chip design.

Dr. Huang currently serves as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: EXPRESS BRIEFS.

Cenglin Xu received the Ph.D. degree from the Department of pharmacology, School of Basic Medical Science, Zhejiang University, Hangzhou, China, in 2016.

He is currently a Postdoctoral Researcher with the School of Basic Medical Science, Zhejiang University. He has authored or coauthored more than 20 SCI papers about epilepsy on impacted journals. His main research interests include the circuitry mechanisms of pharmacoresistant epilepsy and the specific EEG biomarker for epilepsy.

Haibin Shen received the Ph.D. degree from the College of Computer Science and Technology, Zhejiang University, Hangzhou, China, in 2003.

He is currently a Professor with Zhejiang University, a Member of the second level of 151 talents project of Zhejiang Province, and a Member of the Key Team of Zhejiang Science and Technology Innovation. His research achievement has been used by many authority organizations. He has published more than 100 papers on academic journals, and he has been granted more than 30 patents of invention. His research interests include learning algorithm, processor architecture, and modeling.

Prof. Shen was a recipient of the First Prize of Electronic Information Science and Technology Award from the Chinese Institute of Electronics, and has won a second prize at the provincial level.

Zhong Chen received the Ph.D. degree from Okayama University, Okayama, Japan, in 1999.

He is currently the President of Zhejiang Chinese Medical University, Hangzhou, China. He is also a Cheung Kong Scholar, a Professor, and a Distinguished Professor with the NSF of China with the Department of Pharmacology, College of Pharmaceutical Science, School of Medicine, Zhejiang University, Hangzhou. He has published more than 100 papers on impacted journals. His research interests include the circuitry mechanisms and novel treatments of chronic neurological diseases including epilepsy, and the role of histamine and its receptors in the central nervous system.

Prof. Chen has won the First Prize of Natural Science of the Ministry of Education, the Second Prize of National Science and Technology Progress Award of China, the Third Prize of Scientific and Technological Progress Awards of Chinese Medical Association, and New Century Excellent Talents by Ministry of Education. He also received several prizes at the provincial level.