# Hierarchical Spatio-Temporal-Spectral Multiple Hybrid Attention Mechanisms for Seizure Prediction

Sunan Ge, Wenjing Wang, Xin Shi, and Meng Wang

*Abstract*—**Epilepsy is one of the most prevalent diseases of the nervous system in the field of modern medicine, with abrupt and repetitive characterization. Seizure prediction provides a way to alleviate the lesions of the disease for refractory epilepsy patients, which can significantly improve the living standards of sufferers while safeguarding their physical and mental well-being. However, most of the deep learning networks that have been broadly applied in the field of epileptic seizure prediction in recent years restrict their attention to the interchannel information of the electroencephalogram (EEG) and neglect the highly complex spatial and temporal interactions of the EEG signals, so that the insufficiency of cross-channel correlation features becomes an obstacle to achieve high-precision prediction. In order to overcome the above problems, this paper proposed the hierarchical spatio-temporal-spectral multiple hybrid attention mechanism (HSTS-MHAM) to extract spatial and temporal features of transchannel. Capturing cross-domain multiscale correlation features with distinguishability by using the hierarchical spatio-temporal-spectral fusion attention network (STS-HFANet), and constructing multichannel spatial relationships of multidomain fusion features in three dimensions via the multiple hybrid attention convolutional network (MHAConvNet). Then, the spatial dynamic graph convolutional network (SDGConvNet) is used to coordinate the multirhythmic time-space dynamics in epileptic activity to construct the brain cortical graph information to capture the best biomarkers of epileptic seizures. The results show that the approach we proposed can effectively reduce information redundancy and achieve complementary multilevel feature fusion, with excellent performance in terms of accuracy, sensitivity, specificity, FPR/h and AUC.**

*Index Terms*—**epilepsy prediction, EEG signals, hierarchical spatio-temporal-spectral fusion features, mixed attention mechanisms.**

Manuscript received September 13, 2023; revised January 24, 2024. This work is supported in part by the Natural Science Basic Research Program of Shaanxi under Grant No.2021JQ-694, Grant No.2022JM-347, and Grant No.2023-JC-YB-558; and in part by the Key Research and Development Program of Shaanxi Province under Grant No.2023-YBGY-404, and Grant No.2023-ZDLGY-48; and in part by the Scientific Research Program Funded of Shaanxi Provincial Education Department under Grant No. 23JS028.

Sunan Ge is a lecturer at the Computer Science School, University of Xi'an Polytechnic, Shaanxi, 710048, P. R. China. (phone: +86-18189133761; e-mail: gesunan@ xpu.edu.cn).

Wenjing Wang is a postgraduate student at the Computer Science School, University of Xi'an Polytechnic, Shaanxi, 710048, P. R. China. (phone: +86-18706790895; e-mail: 210721056@stu.xpu.edu.cn).

Xin Shi is a lecturer at the Computer Science School, University of Xi'an Polytechnic, Shaanxi, 710048, P. R. China. (phone: +86-18691630806; e-mail: shixin@ xpu.edu.cn).

Meng Wang is a lecturer at the Computer Science School, University of Xi'an Polytechnic, Shaanxi, 710048, P. R. China. (phone: +86-15319768072; e-mail: wangmeng@ xpu.edu.cn).

## I. Introduction

Epilepsy is a non-infectious chronic neurological disorder caused by abnormal neuronal discharges in the brain, and recurrent multiplicity is the main identifier of the syndrome. Seizures occur without aura and last from a few seconds to several minutes, accompanied by severe disturbances of cerebral and organism dysfunction, showing typical symptoms such as limb spasms, whole-body convulsions, abnormal staring, and loss of consciousness [1]-[2].

Recent studies by the World Health Organization (WHO) indicate that at least 50 million patients all over the world are suffering from various types of seizures and that people with epilepsy are more likely than the general population to be subjected to psychological disorders and chronic illnesses [3]-[5], including depression, anxiety, cardiovascular disease and other comorbidities [6]-[7]. People with epilepsy are three to five times more likely to commit suicide and 24 times more likely to die prematurely than the general population, with sudden death in epilepsy accounting for the highest proportion of seizure-related causes of death [8]-[10]. Due to the disease being so demanding in terms of avoiding risk factors, patients and their families are confronted with limitations in their daily activities, which makes it difficult to ensure their long-term living standards [11]-[13]. In addition, the stigma of epilepsy, social discrimination, and ostracization over the centuries have led to patients being prone to strong self-blame and negative self-perceptions, and they often choose to conceal or evade the disease in order to alleviate the spiritual burden brought about by the stigma of the disease, which poses a formidable obstacle to the treatment of the disease [14]. With the ever-increasing development of modern medicine, 70% of patients can be cured with appropriate treatment, but 30% still suffer from intractable epilepsy and can only be prevented or reduced symptomatic seizures by medication [15]. Therefore, an effective seizure prediction technique will not only provide a robust guarantee for the vitality and well-being of patients, but also relieve their physical agony and psychological burden, thus improving the quality of life of the patients and their families, and enhancing the well-being of society as a whole.

A variety of epilepsy screening techniques are available, such as positron emission tomography (PET) [16], magnetic resonance imaging (MRI) [17], and EEG [18]. Among these methods, the identification of ictal episodes by EEG signals has become a mainstream operation in epilepsy diagnosis. Generally, EEG is categorized into two types: scalp

electroencephalography (sEEG) and intracranial electroencephalography (iEEG) [19], among which sEEG has been more widely used in the field of seizure prediction techniques because of its unique advantages such as the possibility of noninvasive acquisition of neuronal signals directly, the ability to continuously capture the electrical activity, and implicit features of neurological disorders in the nervous system [20].

Clinically, it is difficult to effectively determine the preseizure and interseizure periods based on the different phases of seizures, therefore, the main goal of seizure prediction is to effectively distinguish the differences in neurodynamic characteristics between preseizure and interseizure EEG signals. Earlier, conventional methods of visual analysis of EEG required domain-specific expertise, which was not only tremendously time-consuming but also impractical in terms of processing the information, hence researchers have combined various algorithms of signal processing with machine learning to come up with various strategies for predicting epileptic seizures. For example, Xu *et al*. [21] proposed an epilepsy seizure prediction method based on nonlinear features of EEG signals and a gradient-boosting decision tree, which classified the EEG signals into two categories: seizure-onset seizures and seizure-free seizures. After removing the noises of the EEG signals using complementary ensemble empirical pattern decomposition and wavelet threshold denoising, the nonlinear features of the two types of EEG signals were extracted and classified with a random forest classifier, in which the nonlinear features include approximation entropy [22], sample entropy [23], permutation entropy [24], spectral entropy [25], and wavelet entropy [26]. Kapoor *et al*. [27] proposed a hybrid seek optimization tuned ensemble classifier consisting of an AdaBoost classifier, random forest classifier, and decision tree classifier for automated analysis of EEG signals for prediction of epileptic seizures. The features such as statistical information, wavelet features, and entropy-based features of the signal are extracted by a hybrid optimization search algorithm. These extracted features are fed into the proposed ensemble classifier to yield the ultimate output forecasts. Husseina *et al*. [28] proposed a multistage seizure prediction algorithm to achieve seizure prediction by simulated annealing for feature selection. However machine learning methods are associated with the best biomarkers to extract, i.e., they depend on patient-specific features, annotations of relevant seizures, and features that are not discriminative generate more redundant information. Furthermore, models that combine manually designed feature extraction and classifiers tend to have lower generalization capabilities in terms of sensitivity and classification performance that are inextricably linked to the feature extraction process.

In recent years, deep neural networks have received considerable attention in the field of epilepsy seizure prediction due to their superior generalization ability and end-to-end output mode. For example, Lu *et al.* [29] proposed a multiframe network model for predicting epileptic seizures based on sEEG and iEEG datasets, which can be directly applied to the raw data, without including complicated preprocessing procedures, and perform automatic feature extraction and classification. Georgis-Yap *et al*. [30] combined convolutional neural networks (CNN), long short-term memory (LSTM), and temporal convolutional networks (TCN) using grid search to adjust the sliding window size and preseizure time duration to identify preseizure state by EEG and preseizure variability, which does not require preseizure data for training, thus reducing the data acquisition problem such as the effort and time spent on labelling data. Alizadeh *et al*. [31] proposed an epileptic seizure prediction and diagnosis method based on identical and non-identical integral participation methods, i.e., using the integral signal, the derivative signal, and the main signal itself to form a hybrid function prior to the CNN, and then realizing the final classification after selecting the corresponding coefficients to amplify the changes of the EEG signals in different periods using the hybrid function. Saemaldahr *et al*. [32] employed the federated learning (FL) technique by merging a spiking encoder with a graph convolutional network (GCN) into a local model that was trained using a bi-timescale methodology. Each local model determines the preictal probability through FL and determines the risk level of the preictal state from the preictal odds ratios determined from the EEG signals.

Although most of the above methods obtained satisfactory prediction performance comparatively, these models do not applicable to real-time scenarios because the channels in the spatial domain are heterogeneous and discontinuous, and because the structure of the EEG determines that the functional connectivity network is considered to be non-Euclidean data, where there are multiscale information interactions between each EEG channel. Nevertheless, CNN makes it difficult to describe the correlation and positional relationship between signal features, and the lack of spatial-temporal information makes it hard to recover the interchannel correlation of the EEG signals even if reconstructed to generate the data [33].

To overcome the aforementioned challenges, this paper proposes a novel hierarchical spatio-temporal-spectral multiple hybrid attention mechanism (HSTS-MHAM) for seizure prediction. HSTS-MHAM utilizes multiple fitness subnetworks to explore the spatio-temporal-spectral three-dimensional multichannel of the brain waveforms $\delta$ (0.5-4 Hz), $\theta$ (4-8 Hz), $\alpha$ (8-13 Hz), $\beta$ (13-30 Hz), and $\gamma$ (30-50 Hz) in five different scales. It is able to adequately capture and fuse extremely sophisticated spatial-temporal interactive EEG features, acquire the most discriminative ictal symbols, and effectively distinguish different periods of seizures. Therefore, it solves the existing technical problems such as the lack of multiscale features, insufficient information fusion and low prediction accuracy.

The remainder of the paper is as follows. Section II details the preparation of the data preprocessing process and the definition of dataset division. Section III describes the algorithms used and the structure of the model designed. Section IV presents the evaluation metrics used in this study and the performance comparison of the models. Finally, Section V summarizes our work.

## II. Preliminary

In this study, the data set is denoised by a short-time Fourier transform and low-pass filter. Additionally, a 5s sliding window is set for segmental sampling to obtain data

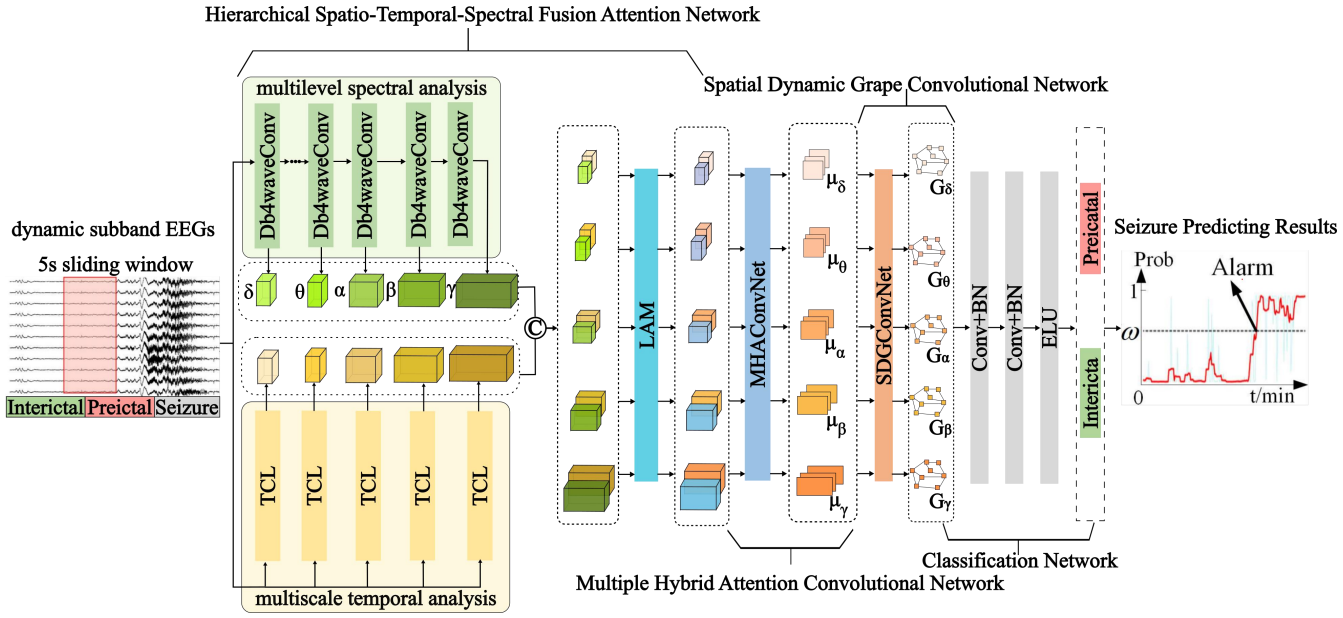


Fig. 2. The framework of HSTS-MHAM for seizure prediction

segments. Using 19 recording electrodes and two reference electrodes in a single reference montage based on the international 10-20 system to ensure the versatility of our model. For seizure prediction, the data are divided into seizure and nonseizure periods according to the patient's seizure situation. The nonseizure periods are then further divided into preseizure and interseizure periods. The preseizure periods are defined as the signal data between 15 minutes and 1 hour before the seizure time point. The interictal periods are defined as the signal data 2 hours before and 2 hours after the seizure. This allowed for the compilation of a list of interictal, preictal, and ictal period data corresponding to each patient. For seizure detection, we use seizure and nonseizure segments from the official dataset.

The leave-one-out cross-validation (LOOCV) method is used to divide the train and test sets to provide an unbiased evaluation of the seizure prediction performance of the method we proposed on the dataset. Specifically, it is assumed that there are a total of $N$ patients in the dataset and each patient has a total of $M$ seizures. For the *i-th* patient, all of its data are defined as the target field data, and the data of the remaining $N-1$ patients are defined as the source field data. The test set consists of the continuous data of the interictal, preictal, and ictal periods divided by the *j-th* seizure of the *i-th* patient, while the training set consists of the continuous data of the remaining $M-1$ seizures and the source field data.

In this paper, dynamic subband EEGs obtained from the temporal embedding network (TENet) are used as inputs to STS-HFANet to remove coupling factors without changing the signal size. Specifically, the data segments obtained after preprocessing are input to the TENet, which consists of a temporal convolutional layer (TCL) and a standard residual block, and the TCL and the standard residual block are serially concatenated as shown in Fig. 1. In particular, the TCL includes a sequentially concatenated convolutional layer with kernel size of $1\times3$, stride of 1, padding of 1, group of 4, batch normalization (BN), and an ELU activation function. The residual block includes a sequentially concatenated convolutional layer with a kernel size of $1\times3$, group of 8, BN, a convolutional layer, BN, an ELU activation function, and a skip connection layer with a kernel size of $1\times1$. Nine temporal embedding EEGs are obtained after the processing of TENet.

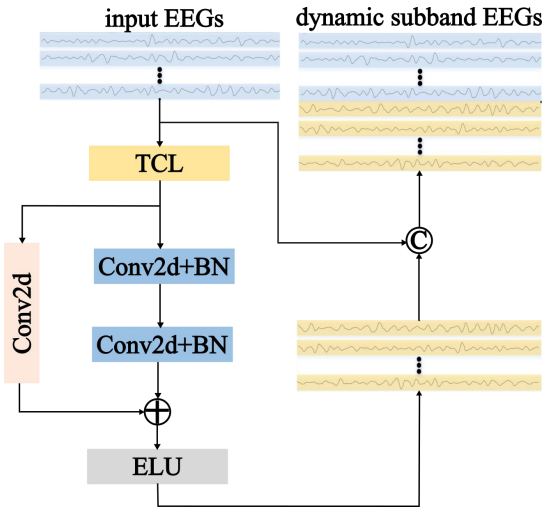


Fig. 1. The framework of TENet

## III. Methodology

The framework of the HSTS-MHAM-based seizure prediction method is shown in Fig. 2, which involves three modular basic subnetworks, i.e., (1) hierarchical spatio-temporal-spectral fusion attention network (STS-HFANet); (2) multiple hybrid attention convolutional network (MHAConvNet); (3) spatial dynamic graph convolutional network (SDGConvNet). Detailed information on each step is given in the following sections.

### A. STS-HFANet

The framework of the mentioned STS-HFANet consists of three submodules, i.e., (1) multilevel spectral analysis; (2) multiscale temporal analysis; (3) layered attention module (LAM). The subsections below provide details of each step.

*1) Design of the multilevel spectral analysis*

Spikes in epileptic pathological waves are transient signals with spiking characteristics, and neuronal spiking activity can reflect the obvious and widespread physiological state of focal epilepsy. It has been shown that the spike counts of EEG signals increased progressively with the advancement of the seizure in all three conditions: interictal, preictal, and ictal, and the differences in the mean spike rate in the preictal periods compared with the interictal periods were statistically significant [34]. Furthermore, the correlation coefficients between Daubechies wavelet convolution and the spike signals are quite high, whereas the transform coefficients are widely independent of each other [35]-[36]. Therefore, in this paper, the fourth-order Db4 waveConv is used to construct a multilevel spectral analysis module, which consists of $L$-layers serially concatenated Db4 waveConv, as shown in Fig. 3. The wavelet decomposition is carried out by taking the nine dynamic subband EEGs obtained after TENet processing as input for continuous $L$-layer periodic padding wavelet decomposition [37], which is performed in a similar way to the discrete wavelet transform. Further, the kernel size and stride in the convolutional layer are set to 8 and 2, respectively, and $L$ is defined by the signal sampling rate $f_s$: $L = [\log_2(f_s)] - 3$, $f_s$ is the sampling rate of the EEG signal, and $[\cdot]$ denotes the rounding operation.

Since the wavelet transform process contains a low-pass coefficient and a high-pass coefficient, the module extracts the low-frequency information $S_{w\circ} = (s_{\delta\circ}, s_{\theta\circ}, s_{\alpha\circ}, s_{\beta\circ}, s_{\gamma\circ})$ and the high-frequency information $S_{w\bullet} = (s_{\delta\bullet}, s_{\theta\bullet}, s_{\alpha\bullet}, s_{\beta\bullet}, s_{\gamma\bullet})$ of the input dynamic subband EEGs in the bands corresponding to $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$, respectively, where $w \in \{\delta, \theta, \alpha, \beta, \gamma\}$. $s_{\delta\circ}$, $s_{\theta\circ}$, $s_{\alpha\circ}$, $s_{\beta\circ}$, and $s_{\gamma\circ}$ are the low-frequency information under the frequency bands corresponding to $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$, respectively. $s_{\delta\bullet}$, $s_{\theta\bullet}$, $s_{\alpha\bullet}$, $s_{\beta\bullet}$, and $s_{\gamma\bullet}$ are the high-frequency information under the frequency bands corresponding to $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$, respectively. Then, the output of low-frequency information and high-frequency information is concatenated along dimension 1 to obtain the frequency domain feature with shape $S_w = (s_\delta, s_\theta, s_\alpha, s_\beta, s_\gamma)$.

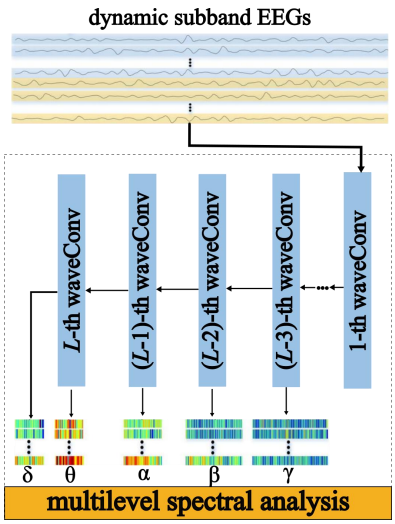


Fig. 3. The structure of multilevel spectral analysis

*2) Design of the multiscale temporal analysis*

The multiscale temporal analysis module is implemented by five parallel TCLs. The kernel size of the five convolutional layers is $\{k/8, k/4, k/2, k, k\}$, and the corresponding stride of the five convolutional layers is set to $\{k/8, k/4, k/2, k, k\}$, where $k = 2^L$, as shown in Fig. 4, the TCL includes convolutional layers with trainable kernel parameters, BN, and ELU activation functions concatenated serially in sequence.

The multiscale temporal analysis module takes nine dynamic subband EEGs obtained after TENet processing as the input data and extracts the time-domain features under the frequency bands corresponding to $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$ in the input dynamic subband EEGs as the output data $T_w = (t_\delta, t_\theta, t_\alpha, t_\beta, t_\gamma)$, in which $t_\delta$, $t_\theta$, $t_\alpha$, $t_\beta$, and $t_\gamma$ are the time-domain features under the frequency bands corresponding to $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$, respectively.

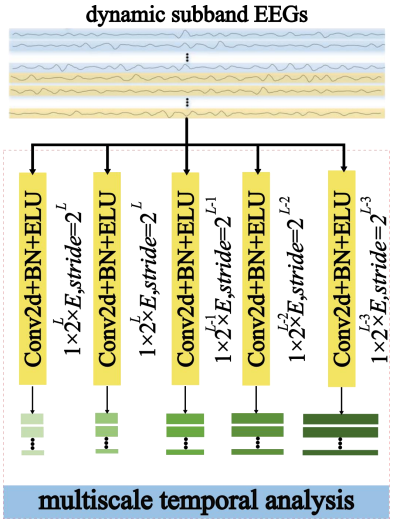


Fig. 4. The structure of multiscale temporal analysis

*3) Design of the layered attention module*

Considering the heterogeneity of transdomain features, this paper designs the layered attention module (LAM) to capture both spectral features $\dot{S}$ and temporal features $\dot{T}$ in-depth simultaneously, thus enhancing the multidomain channel response jointly.

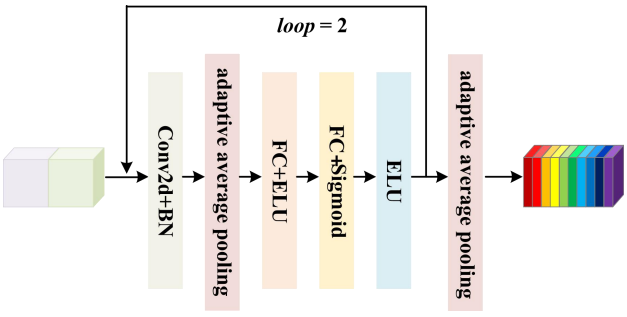


Fig. 5. The structure of LAM

As shown in Fig. 5, the LAM consists of a sequentially concatenated convolutional layer, BN, variant squeeze-and-excitation network (SENet), an ELU activation function, a convolutional layer, BN, variant SENet, an ELU activation function, a convolutional layer, BN, variant SENet, an ELU activation function, and adaptive average pooling. In this case, the kernel size of the convolutional layer in LAM is

set to 1×4, stride is set to 1, padding is set to 2, and group is set to 3. The variant SENet consists of sequentially serially concatenated adaptive average pooling, a fully concatenated layer (FC) with an ELU activation function, an FC layer with a sigmoid activation function, and finally, an ELU activation function that further enhances the robustness of the structure.

Firstly, LAM concatenates the frequency-domain feature output data and time-domain feature output data along dimension 1 to obtain the multidomain fusion feature $U_{ST} = (s_\delta \oplus t_\delta,\ s_\theta \oplus t_\theta,\ s_\alpha \oplus t_\alpha,\ s_\beta \oplus t_\beta,\ s_\gamma \oplus t_\gamma)$, where $\oplus$ is the element-by-element summation method. Subsequently, $U_{ST}$ is input into the variant SENet after convolution and BN. The variant SENet [38]-[40] is a computational unit that contains two steps: squeeze and excitation. Squeeze is implemented by adaptive average pooling of the variant SENet, which is achieved by compressing the global spatial information after convolution and BN of $U_{ST}$ into a channel descriptor $z \in R^{1\times1\times(E+E)}$, which fully captures the channel dependence, where $E$ is the number of channels corresponding to the EEG signal data, and $z$ is computed from the descriptive statistic, defined as the statistic of the *E-th* channel for $z$ :

$$z_E = \frac{1}{\rho}\sum_{l=1}^{\rho} H\left(s_w^E, t_w^E\right)(E,1,l) \tag{1}$$

where $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$ correspond to each $s_\delta$, $s_\theta$, $s_\alpha$, $s_\beta$, and $s_\gamma$ have $E$ EEG channels, then $s_w = \{s_w^1, s_w^2,..., s_w^E\}$ is the set of frequency-domain features of $E$ channels inside the brainwave $w$, and $s_w^E$ is the frequency-domain feature of the *E-th* channel within the brainwave $w$. $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$ correspond to each $t_\delta$, $t_\theta$, $t_\alpha$, $t_\beta$, and $t_\gamma$ have $E$ EEG channels, then $t_w = \{t_w^1, t_w^2,..., t_w^E\}$ is the set of time-domain features of the *E-th* channel within brainwave $w$, and $t_w^E$ is the time-domain feature of the *E-th* channel within brainwave $w$. $H(\cdot)$ is the channel cascade function, and $\rho$ is the spatial dimension.

The excitation is implemented by the FC layer with an ELU activation function, as well as the FC layer with a sigmoid activation function, i.e., channel dependencies are modelled by a simple self-gating mechanism to obtain the channel weight $\hat{z}$ adapted to a specific channel descriptor $z$. The weight parameter for the *E-th* channel of $\hat{z}$ is defined as follows:

$$\hat{z}_E = \alpha\left(W_2\beta\left(W_1 z_E\right)\right) \tag{2}$$

where $W_1 \in R^{((E+E)/r)\times(E+E)}$ and $W_2 \in R^{(E+E)\times((E+E)/r)}$ are two fully concatenated layer weights, $\alpha(\cdot)$ and $\beta(\cdot)$ represent the ELU activation function and sigmoid activation function, respectively. $E$ is the number of EEG channels, and $r$ is the compression ratio parameter indicating the bottleneck of the self-gating mechanism.

The distinguishability of the features in multidomain is jointly improved by adaptively recalibrating the importance of the input features $U_{ST}$ under the different channels corresponding to each brain waveform by means of channel-by-channel operations. The computational formulae are as follows:

$$\begin{aligned} \dot{S}_w &= \left[\hat{z}_1 s_w^1, \hat{z}_2 s_w^2, ..., \hat{z}_E s_w^E\right] \\ \dot{T}_w &= \left[\hat{z}_{E+1} t_w^1, \hat{z}_{E+2} t_w^2, ..., \hat{z}_{E+E} t_w^E\right] \end{aligned} \tag{3}$$

Finally, the distinguishable multidomain fusion features with shape $U_{S\dot{T}} = (\dot{S}_\delta \oplus \dot{T}_\delta,\ \dot{S}_\theta \oplus \dot{T}_\theta,\ \dot{S}_\alpha \oplus \dot{T}_\alpha,\ \dot{S}_\beta \oplus \dot{T}_\beta,\ \dot{S}_\gamma \oplus \dot{T}_\gamma)$ corresponding to the five brain waveforms are obtained after adaptive average pooling operation.

### B. MHAConvNet

Although enriched feature extraction facilitates EEG classification, inappropriate fusion methods may involve redundant information [41]. Traditional methods compute channel attention using a global average pooling method to decompose the spatial dimension of the input tensor into individual pixels in order to compute the channel weights, which can lead to a significant loss of spatial information. As a result, the interdependence between channel and space is also absent when computing attention on a single-pixel channel. The later proposed spatial and channel-based convolutional block attention module (CBAM) [42] alleviates the problem of spatial interdependence, but the computation of channel attention and spatial attention are independent of each other. In this paper, we design a mixed attention module (MAM) by jointly improving the dual attention module (DA) [43] and triple attention module (TA) [44]. Consequently, we propose the MHAConvNet, which can effectively capture the interactions between spatial dimensions and channel dimensions, fully utilize the spatio-temporal-spectral fusion information of EEG signals, enhance the capability to characterize the multiscale features of complex EEG signals and improve the learning ability and robustness of the model.

MHAConvNet consists of a sequentially concatenated convolutional layer, BN, MAM, and ELU activation function. The output of the ELU activation function is then returned as an input, resulting in a total of three cycles of data processing between the convolutional layer, BN, MAM, and ELU activation function. Finally, adapted average pooling and TCL are concatenated in series. The network framework is shown in Fig. 6.

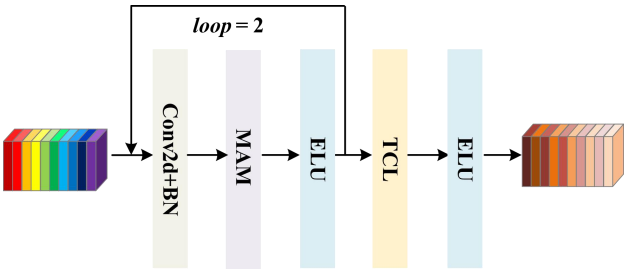


Fig. 6. The framework of MHAConvNet

MHAConvNet inputs the output data of the LAM module into the MAM after sequential convolution and BN processing, where the structure of the MAM is shown in Fig. 7. Among them, the spatial attention module aims to establish cross-dimensional dependencies between the height $H$ and width $W$ contained in the spatial dimension and the channel dimension $C$ through branches with different orientations for the five brain waveforms of $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$, respectively. The first branch, $X_{HW}$, establishes the spatial dependence between dimension $H$ and dimension $W$. The second branch, $X_{CW}$, is obtained by dimensional permutation of $X_{HW}$, which establishes the spatial-channel dependence between dimension $C$ and dimension $W$. The third branch, $X_{HC}$, is obtained by dimensional permutation of $X_{HW}$, which establishes the spatial-channel dependence between dimension $H$ and dimension $C$. The channel attention module

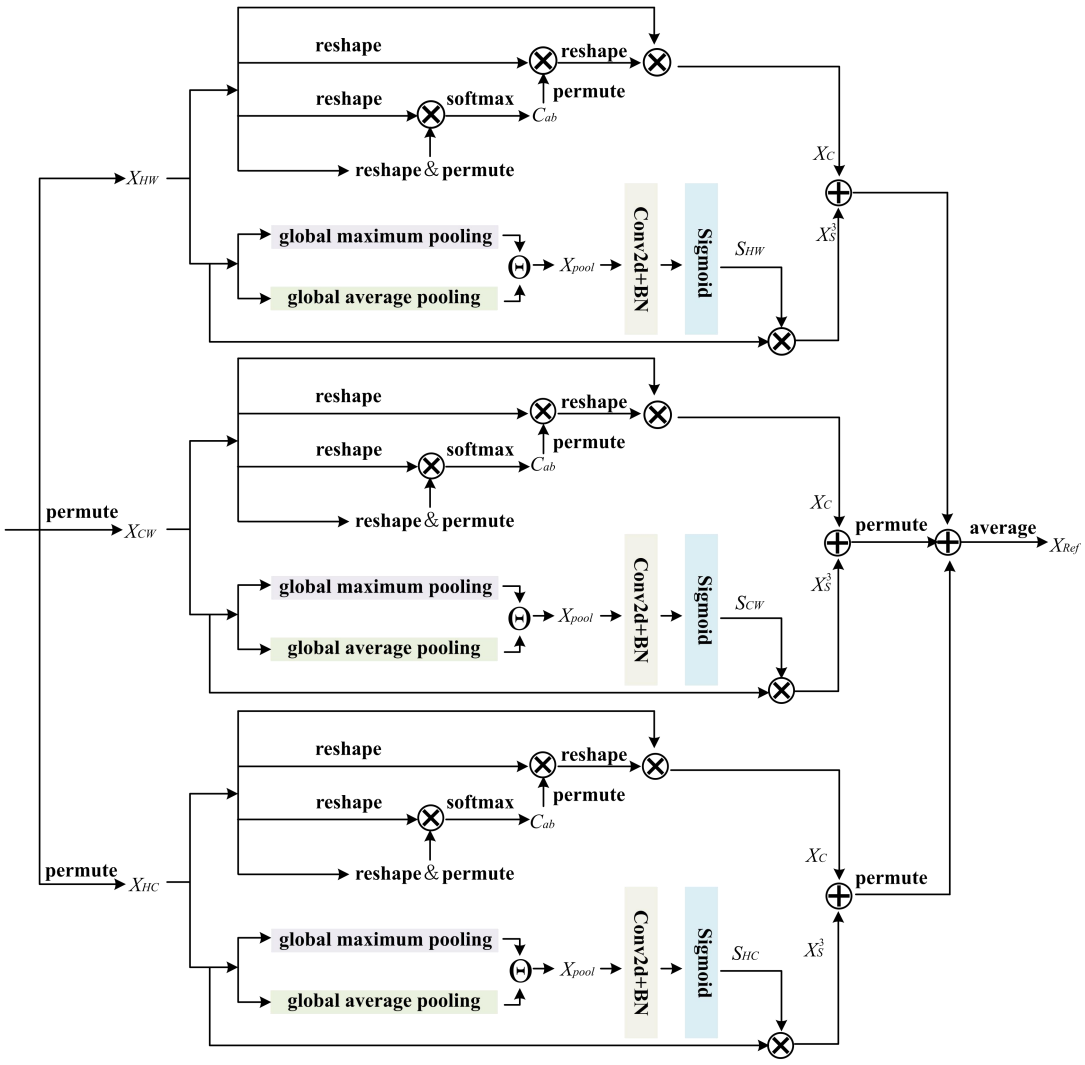


Fig. 7. The structure of MAM

utilizes the output features after BN to establish channel dependencies between dimension $C$ and dimension $C$ under three different directions of branching for each of the five brain waveforms: $\delta$, $\theta$, $\alpha$, $\beta$, and $\gamma$.

The spatial attention module first performs pooling operations on the data $X_{pq}$ in the corresponding dimensions through channel-based global maximum pooling and global average pooling, respectively, and then connects the feature aggregates to generate the feature $X_{pool}$, which is defined as follows:

$$X_{pool} = \mathrm{MaxPool}\left(X_{pq}\right) \Theta \mathrm{AvgPool}\left(X_{pq}\right) \tag{4}$$

where $X_{pq}$ ($p \in \{H, C\}$, $q \in \{H, C\}$, $p \neq q$), and $X_{pool}$ is the pooling feature. MaxPool(·) and AvgPool(·) refer to the channel maximum pooling and average pooling, and Θ is the channel connection.

Subsequently, splicing and convolution operations are performed on $X_{pool}$, and the features after BN are subject to the sigmoid activation function to compute the generative spatial attention $S_{pq}$ with the following formula:

$$S_{pq} = \beta\left(h^{7\times7} X_{pool}\right) \tag{5}$$

where $h^{7\times7}$ denotes a convolutional layer with kernel size 7× 7 and $\beta(\cdot)$ is the sigmoid activation function.

Finally, the generated attention weight $S_{pq}$ is multiplied by $X_{pq}$ to get the output $X_S^v$ ($v = 1, 2, 3$), where $v$ refers to the $v$-*th* branch.

The channel attention module first reshapes the feature $X_{pq}$ to obtain the reshaped feature $X_{re}$, then performs matrix multiplication between $X_{re}$ and the substitution of $X_{re}$. The weights are normalized by the softmax function to obtain the channel attention $C_{ab}$:

$$C_{ab} = \frac{e^{X_{re}^a \cdot X_{pq}^b}}{\sum_{a=1}^{\sigma} e^{X_{re}^a \cdot X_{pq}^b}} \tag{6}$$

where $\sigma$ is the number of channels included in dimension $C$, and $C_{ab}$ measures the influence of the $a$-*th* channel on the $b$-*th* channel.

A matrix multiplication operation is then performed between the reshaped feature $X_{re}$ and the permutation of the channel attention $C_{ab}$, and the result is multiplied by the scale parameter $\alpha(\cdot)$ to reshape it into the original input shape. This is followed by an element-by-element weighted summation operation with $X_{pq}$, where the final feature for each channel is the weighted sum $X_C$ of all the channel features and the original feature $X_{pq}$:

$$X_C = \alpha \sum_{a=1}^{\sigma} \left(C_{\mathrm{ab}} X_{re}^a\right) + X_{pq}^a \tag{7}$$

where the $\alpha(\cdot)$ is initialized to 0.

Ultimately, the outputs of the three branches are aggregated by averaging to generate the most discriminating

refinement feature, $X_{Ref}$. This can be expressed as follows:

$$X_{\mathrm{Re}f} = \frac{1}{3}\left(X_S^1 \oplus X_C + \left\|X_S^2 \oplus X_C\right\| + \left\|X_S^3 \oplus X_C\right\|\right) \quad (8)$$

where $\oplus$ is the element-by-element summation method, and $\|\cdot\|$ denotes the permutation operation.

Finally, the refined features corresponding to the different brain waveforms obtained are sequentially fed into adaptive mean pooling and TCL to yield the most discriminative fusion features with shape $U = (\mu_\delta, \mu_\theta, \mu_\alpha, \mu_\beta, \mu_\gamma)$.

### C. SDGConvNet

In order to obtain the most discriminative epileptic seizure markers, the SDGConvNet adopted in this paper applies a self-gating method to dynamically simulate the dependencies between electrodes, fully considers the spatial synchronization and connectivity information inside the brain, and extends the geometric graph-based deep learning methods to improve the prediction accuracy. The SDGConvNet consists of the sequentially serially concatenated channel embedding network (CENet), dynamic neighbourhood network (DNNet), dynamic graph convolutional network (DGConvNet), and graph attention network (GANet).

CENet aims to calculate the adjacency matrix $Z_{ds}$ for EEG electrodes 'AF7', 'FT7', 'TP7', 'PO7', 'AF3', 'FC3', 'CP3', 'PO3', 'FCz', 'CPz', 'AF4', 'FC4', 'CP4', 'PO4', 'AF8', 'FT8', 'TP8', 'PO8'. DNNet applies a self-gating mechanism to the adjacency matrix $Z_{ds}$ in order to establish the dependency between electrodes and obtain the dynamic adjacency matrix $\hat{Z}_{ds}$. DNNet consists of sequentially concatenated adaptive average pooling, the FC layer with an ELU activation function, the FC layer with a Tanh activation function, and a ReLU activation function. DGConvNet obtains a feature map $P$ characterizing the dynamic relationships of the five brain waveforms by graph convolution of the degree matrix, $L^{\phi\phi}$, of the adjacency matrix $Z_{ds}$ and the dynamic adjacency matrix $\hat{Z}_{ds}$. DGConvNet consists of a sequentially serially concatenated graph convolutional layer, followed by BN, an ELU activation function, another graph convolutional layer, another BN, and a final ELU activation function. GANet includes a serially concatenated MAM and an FC layer, and the output of DGConvNet is filtered using MAM to select the most representational feature map $P^*$ as the final output.

#### 1) Design of the channel embedding network

CENet uses MNE-Python to create a list of electrode locations for the 10-20 system. From the list, the distance from each electrode to the reference electrode is obtained to create the distance matrix $L_{dis}$. $L_{dis}$ is a square matrix containing the distances between all pairs of electrodes. In this study, we define the set $Z$ as the distances between any two EEG electrodes, represented as $Z = \{d_{\phi\varphi} \mid \phi, \varphi \in (1, \lambda), \phi \neq \varphi\}$. Here, $d_{\phi\varphi}$ denotes the Euclidean distance between electrode $\phi$ and electrode $\varphi$, and $\lambda$ is the number of electrodes. Two electrodes are considered adjacent if their distance, $d_{\phi\varphi}$, is less than the average value of $Z$, $\bar{Z}$. The distance between an electrode and itself is defined as the average distance from all neighbouring electrodes. The spatial location information of the 3D EEG, constructed using the electrodes, is embedded into $Z_{ds} \in R^{\lambda\times\lambda}$. The adjacency matrix, $Z_{ds}$, is created by calculating the average distance between all pairs of electrodes. $Z_{ds}$ represents the connectivity information between electrodes, where $Z_{ds}$ is $1/d_{\phi\varphi}$ if $d_{\phi\varphi} < \bar{Z}$, and $Z_{ds}$ is 0 if $d_{\phi\varphi} \geq \bar{Z}$.

#### 2) Design of the dynamic neighbourhood network

DNNet first performs adaptive average pooling of $Z_{ds}$ to spread it into a one-dimensional vector, followed by the FC layer with an ELU activation function and the FC layer with a Tanh activation function. It retains the nonnegative elements using the ReLU activation function to obtain the distance weight $\hat{Z}^* \in R^{(\lambda\times\lambda)\times 1}$. By reshaping the $\hat{Z}^*$ to its original dimensions, we obtain the dynamic adjacency matrix $\hat{Z}_{ds} \in R^{\lambda\times\lambda}$. After the aforementioned self-gating mechanism, the distance weight $\hat{Z}^*$ containing the interelectrode dependencies is defined as follows:

$$\hat{Z}^* = \alpha\left(W_2\beta\left(W_1\hat{Z}\right)\right) \quad (9)$$

where $\hat{Z} \in R^{(\lambda\times\lambda)\times 1}$ is reshaped by $Z$, $W_1 \in R^{((\lambda\times\lambda)/r)\times(\lambda\times\lambda)}$ and $W_2 \in R^{(\lambda\times\lambda)\times((\lambda\times\lambda)/r)}$ are the weight matrices of two FC layers. The compression ratio parameter $r$ represents the bottleneck of the gating mechanism, $\alpha(\cdot)$ and $\beta(\cdot)$ denote the Tanh activation function and the ELU activation function, respectively.

#### 3) Design of the dynamic graph convolutional network

DGConvNet divides $\hat{Z}_{ds}$ and $L_{dis}$ into five groups. For each group, it calculates the weighted sum of the neighbouring node features cyclically, with the weights determined by $L_{dis}$. Subsequently, graph convolution is performed on each of the five groups of features in $U = (\mu_\delta, \mu_\theta, \mu_\alpha, \mu_\beta, \mu_\gamma)$ and $\hat{Z}_{ds}$, as described in the following formula:

$$P^w = \alpha\left(L_{dis}^{-1}\hat{Z}_{ds}\alpha\left(\mu_w\beta_1\right)\beta_2\right) \quad (10)$$

where $P^w$ represents the dynamic EEG feature map under the brainwave $w$, $\alpha(\cdot)$ is the ELU activation function, $L^{\phi\phi} = \Sigma_\varphi Z_{ds}^{\phi\varphi}$ is the degree matrix of $\hat{Z}_{ds}$, $F$ is the length of the corresponding feature vector for each brainwave, $\beta_1 \in R^{F\times F}$ and $\beta_2 \in R^{F\times F}$ denote the weight matrices of the convolution kernels in the $1\times 1$ convolutional layer, respectively.

The five output dynamic EEG feature maps were then concatenated into one, resulting in the output $P$ of DGConvNet. $P$ represents the epileptic brain cortex graph of the input data.

#### 4) Design of the graph attention network

GANet splits $P'$ into five tensors along dimension 1 and then joins the five tensors along the channel dimension to form a new brain cortex graph $P'$. Subsequently, the MAM module is used to filter out the refined feature cortical graph $P_{Ref}$, which is the most representative of $P'$ in terms of cross-dimensional dependencies. The output is then split into five tensors. The importance of neighbouring nodes within each graph is computed separately through a linear layer, and finally, the five outputs are concatenated along dimension 1 as output $P^*$.

## IV. Experimental Results

### A. Dataset Description

#### 1) CHB-MIT dataset

CHB-MIT is one of the few authoritative public datasets on continuous long-term seizures, created by researchers from the Massachusetts Institute of Technology (MIT) and Children's Hospital Boston (CHB). As shown in Fig. 8, EEG signals were collected from 21 electrodes at a sampling rate of 256 Hz with 16-bit resolution using the bipolar montage

technique of the international 10-20 system [45].

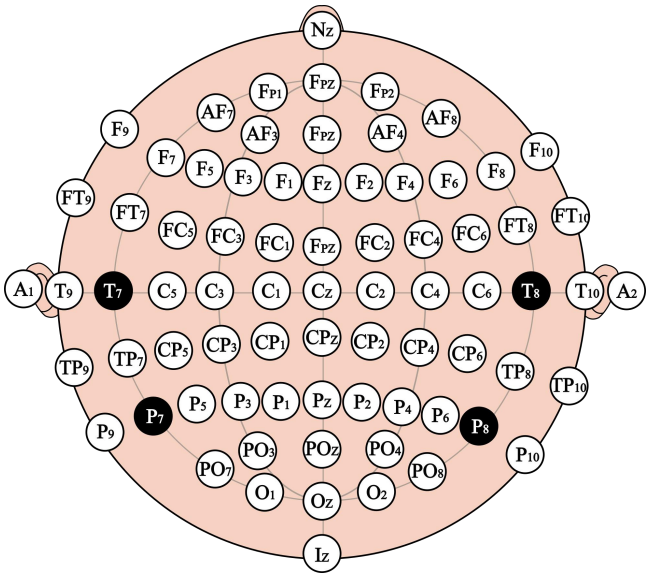


Fig. 8. Internationally standardized electrode systems

This dataset consists of 23 patients with refractory epilepsy, which includes a total of 24 cases (case chb21 and case chb01 are from the same patient, with a 1.5-year interval between case records; case chb24 was added to this collection in December 2010 and is not currently included in SUBJECT-INFO), with 5 male patients (aged 3-22 years), 17 female patients (aged 1.5-19 years), and 1 patient with missing sex and age data, totalling 967.55 hours of continuous EEG recordings and 198 seizures. The annotation file of the dataset provides information about the channel, seizure onset, and offset. For this paper, the preseizure interval was set to 15 minutes, and recordings of at least 2 seizures and a 3-hour interictal interval were used for seizure prediction evaluation to avoid the effect between different ictal times and to ensure that the prediction results could be evaluated by LOOCV in each patient. Specific information is shown in TABLE I.

TABLE I
EEG RECORDS FOR THE CHB-MIT DATASET

| ID | Gender | Age | Duration of record (h) | Duration of seizure (s) | Number of seizures | Number of used seizures |
|---|---|---|---|---|---|---|
| 1 | F | 11 | 40.55 | 449 | 7 | 7 |
| 2 | M | 11 | 175 | 175 | 3 | 3 |
| 3 | F | 14 | 28 | 409 | 7 | 7 |
| 5 | F | 7 | 39 | 563 | 5 | 5 |
| 6 | F | 1.5 | 66.7 | 147 | 10 | 7 |
| 7 | F | 14.5 | 68.1 | 328 | 3 | 3 |
| 8 | M | 3.5 | 20 | 924 | 5 | 5 |
| 9 | F | 10 | 67.8 | 280 | 4 | 4 |
| 10 | M | 3 | 50 | 454 | 7 | 6 |
| 11 | F | 12 | 34.8 | 809 | 3 | 3 |
| 13 | F | 3 | 33 | 547 | 12 | 5 |
| 14 | F | 9 | 26 | 117 | 8 | 6 |
| 16 | F | 7 | 19 | 94 | 10 | 8 |
| 17 | F | 12 | 21 | 296 | 3 | 3 |
| 18 | F | 18 | 36 | 323 | 6 | 6 |
| 20 | F | 6 | 29 | 302 | 8 | 8 |
| 21 | F | 13 | 33 | 203 | 4 | 4 |
| 22 | F | 9 | 31 | 207 | 3 | 3 |
| 23 | F | 6 | 28 | 431 | 7 | 7 |

*2) TUSZ dataset*

The Temple University Hospital (TUH) EEG Seizure Corpus (TUSZ) dataset [46] is the only open-source EEG dataset containing annotations for multiple epilepsy types. Considering the superiority of the TUSZ dataset for long-term maintenance and updating, we used TUSZ v2.0.1, updated on 4 October 2023, as the experimental data source for seizure detection. Table II summarizes the TUSZ data. For this paper, we resampled the EEG recordings to 256 HZ because of the heterogeneity of the sampling rate across patients in the TUSZ.

TABLE II
SUMMARY OF DATA IN TRAIN AND TEST SETS OF TUSZ V2.0.1

| | Number of patients | Number of seizure events | Duration of record (h) | Duration of seizure (h) |
|---|---|---|---|---|
| Train Set | 579 | 2,474 | 910.3 | 43.47 |
| Test Set | 43 | 469 | 129.9 | 7.56 |

## B. Evaluation Metrics

For comparison with other state-of-the-art methods, the overall performance of the different models on the test dataset was evaluated using five evaluation metrics, namely, accuracy (ACC), specificity (SPE), sensitivity (SEN), false positive rate (FPR), and area under the curve (AUC). Additionally, the assessment methodology uses LOOCV.

ACC denotes the ratio of the number of correct predictions to the total number of samples, with larger results indicating a better ability of the model to discriminate between categories. SPE denotes the proportion of correct predictions among all results with negative actual values, with larger results indicating a better identification of predictions of the interictal periods. SEN denotes the proportion of correct prediction of the counts of seizures. FPR/h denotes the number of incorrect predictions per hour. AUC is used to evaluate the classification performance of the model. The definition formula is as follows:

$$\mathrm{ACC}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FN}+\mathrm{FP}} \tag{11}$$

$$\mathrm{SPE}=\frac{\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}} \tag{12}$$

$$\mathrm{SEN}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{13}$$

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \tag{14}$$

$$\mathrm{AUC}=\frac{1}{2}\sum_{\tau=1}^{\xi-1}\left(x_{\tau+1}-x_{\tau}\right)\left(y_{\tau}-x_{\tau+1}\right) \tag{15}$$

where TP, FP, TN, and FN in (11) to (14) denote true positive, false positive, true negative, and false negative, respectively, and detailed explanations are listed in TABLE III. The $x$ and $y$ in (15) are consecutive coordinates on the ROC, denoted as $\{(x_1=0, x_\xi=1) \mid (x_1, y_1),(x_2, y_2),\ldots,(x_\xi, y_\xi)\}$.

TABLE III
CONFUSION MATRIX

| Forecast Category | True Category | |
|---|---|---|
| | True | False |
| True | TP | FP |
| False | FN | TN |

TABLE IV
PERFORMANCE COMPARISON OF SEIZURE DETECTION EXPERIMENTS ON TRAIN SET OF TUSZ

| Method | ACC(%) | SPE(%) | SEN(%) |
|---|---|---|---|
| STS-HGCN | 81.44 | 87.34 | 66.32 |
| TA-STS-ConvNet | 83.76 | 89.15 | 68.17 |
| **Our Method** | **85.47** | **92.37** | **69.01** |

Where bold fonts indicate the best average results.

TABLE V
PERFORMANCE COMPARISON OF SEIZURE DETECTION EXPERIMENTS ON TEST SET OF TUSZ

| Method | ACC(%) | SPE(%) | SEN(%) |
|---|---|---|---|
| STS-HGCN | 77.31 | 83.10 | 61.57 |
| TA-STS-ConvNet | 79.48 | 85.53 | 62.23 |
| **Our Method** | **82.60** | **90.12** | **65.33** |

Where bold fonts indicate the best average results.

An effective feature extraction process tends to have a significant impact on seizure prediction results. The use of datasets with richer types is also favourable for enhancing the generalisability of the model. Therefore, in order to verify the superiority of our method in the feature extraction process, in this paper, we use the TUSZ dataset for seizure detection experiments and the CHB-MIT dataset for seizure prediction experiments.

The duration of seizures, characteristics, and dynamics of EEG signals vary widely between patients depending on the age, gender, seizure type, and the epileptogenic zone of the subjects. Therefore, the typical characteristics of seizures in some patients may not be applicable to other patients. Most of the existing studies use only a limited quantity and a single type of dataset, i.e., the training and test data are from the same patient, which is only applicable to patient-specific scenarios, making it difficult to establish a generalized approach to achieve high prediction performance for each patient [47]. Considering the aforementioned issues, we separately evaluated the validity of the model for seizure prediction using the single-subject experiments and the cross-subject experiments.

### C. *Seizure Detection Experiments*

We performed the detection of seizure events on the TUSZ dataset with the aim of distinguishing different types of seizures and overcoming the heterogeneity among patients.

TABLE IV shows the performance of three different models for seizure detection on the train set of TUSZ. It can be seen that the average ACC of STS-HGCN [40] and TA-STS-ConvNet [36] is 81.44% and 83.76%, respectively, and the average SPE is 87.34% and 89.15%, respectively, whereas the average ACC of the model implemented in this paper is 85.47%, and the average SPE is 92.37%. Furthermore, the average SEN of the baseline methods is 66.32% and 68.17%, respectively, and 69.01% for the method used in this paper.

TABLE V shows the performance of three different models for seizure detection on the test set of TUSZ. It can be seen that the average ACC of STS-HGCN and TA-STS-ConvNet is 77.31% and 79.48%, respectively, and the average SPE is 83.10% and 85.53%, respectively, whereas the average ACC of the model implemented in this paper is 82.60%, and the average SPE is 90.12%. Furthermore, the average SEN of the baseline methods is 61.57% and 62.23%, respectively, and 65.33% for the method used in this paper.

Competitive results show that our method extracts more informative biomarker features.

### D. *Single-Subject Experiments*

In a single-subject experiment, the effectiveness of the method is validated by comparing its performance with different existing state-of-the-art methods. Fig. 9 shows the comparison of the AUC results of the CNN-LSTM [30], the TCN [30], and the method used in this paper. In the vast majority of patients, the AUC of our method is higher than that of CNN-LSTM and TCN.

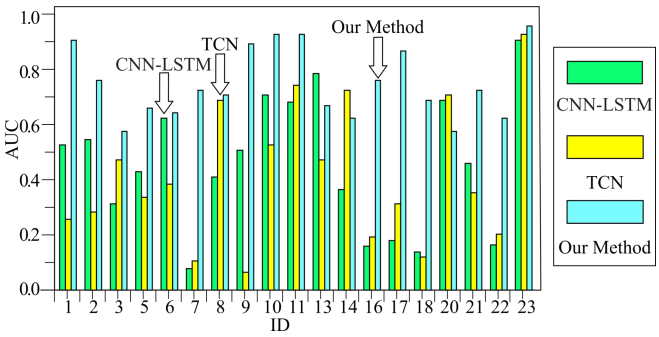


Fig. 9. AUC comparison of CNN-LSTM, TCN and our method

Fig. 10 shows the FPR/h results of Spiking-GCNN [32] versus the model used in this paper. In the vast majority of patients, the FPR/h of our method is higher than that of Spiking-GCNN. Better FPR/h implies more false alarms in the interictal periods, which is undesirable in clinical applications. The significant improvement in FPR/h suggests that the model employed in this paper does not suffer from simple errors in the interictal periods, and captures the preictal signals more accurately.

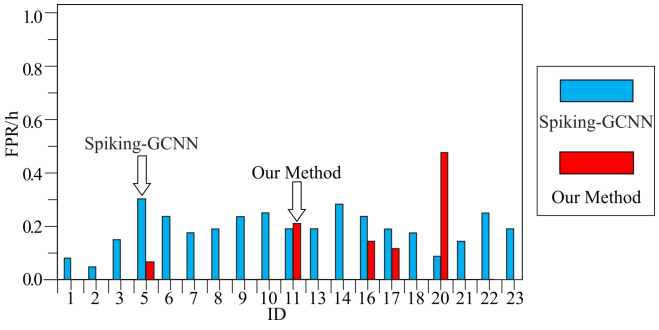


Fig. 10. FPR/h comparison of Spiking-GCNN and our method

TABLE VII
PERFORMANCE COMPARISON OF SINGLE-SUBJECT EXPERIMENTS ON THE CHB-MIT DATASET

| Authors | Method | No. of patients | ACC(%) | SPE(%) | SEN(%) | FPR/h |
|---|---|---|---|---|---|---|
| Xu *et al*. [21] | RF+GBDT | - | 91.76 | - | 91.87 | 0.083 |
| Kapoor *et al*. [27] | AdaBoost+DT+RF | - | 93.81 | 88.57 | 91.68 | - |
| Husseina *et al*. [28] | PIL-SEG+SA | 10 | 80.61 | 80.71 | 79.32 | 0.239 |
| Lu *et al*. [29] | CNN-LSTM | 19 | 80.96 | 68.80 | 78.04 | - |
| Alizadeh *et al*. [31] | KP-Ki-KD+CNN | 16 | 97.1 | 96.3 | 97.5 | - |
| Gao *et al*. [41] | MSPPNet | 16 | - | - | 93.8 | 0.054 |
| **Our Method** | **HSTS-MHAM** | **19** | **95.31** | **99.08** | **94.67** | **0.027** |

Where bold fonts indicate our proposed method, No. of patients are total number of patients involved in the evaluation.

TABLE VI shows the performance of two different models for seizure prediction on the CHB-MIT dataset. It can be seen that the average ACC and average SPE of Spiking-GCNN are 90.71% and 91.18%, respectively, whereas the average ACC of the model implemented in this paper is 95.31%, and the average SPE is 99.08%. Furthermore, the average FPR/h of the baseline method is 0.091/h, whereas the average FPR/h of the proposed method is 0.027/h.

TABLE VI
PERFORMANCE COMPARISON OF SPIKING-GCNN AND OUR METHOD ON THE CHB-MIT DATASET

| ID | Spiking-GCNN | | | **Our Method** | | |
|---|---|---|---|---|---|---|
| | ACC(%) | SPE(%) | FPR/h | ACC(%) | SPE(%) | FPR/h |
| 1 | 96.28 | 96.14 | 0.032 | 96.21 | 99.70 | 0.000 |
| 2 | 92.18 | 98.31 | 0.025 | 97.65 | 99.90 | 0.000 |
| 3 | 96.09 | 93.89 | 0.071 | 95.34 | 99.79 | 0.000 |
| 5 | 91.58 | 88.84 | 0.125 | 95.58 | 98.76 | 0.028 |
| 6 | 87.61 | 89.28 | 0.112 | 94.33 | 98.35 | 0.000 |
| 7 | 89.69 | 91.25 | 0.086 | 98.01 | 99.32 | 0.000 |
| 8 | 86.49 | 90.21 | 0.098 | 95.04 | 99.95 | 0.000 |
| 9 | 90.51 | 88.89 | 0.112 | 98.11 | 99.76 | 0.000 |
| 10 | 89.78 | 89.26 | 0.113 | 96.63 | 99.80 | 0.000 |
| 11 | 91.39 | 90.34 | 0.095 | 97.91 | 99.12 | 0.120 |
| 13 | 89.89 | 90.27 | 0.098 | 93.82 | 98.52 | 0.000 |
| 14 | 89.01 | 85.96 | 0.142 | 91.15 | 98.26 | 0.000 |
| 16 | 89.98 | 89.97 | 0.112 | 88.87 | 98.81 | 0.069 |
| 17 | 89.12 | 90.28 | 0.098 | 96.40 | 99.40 | 0.053 |
| 18 | 92.05 | 91.64 | 0.085 | 95.65 | 98.77 | 0.000 |
| 20 | 92.85 | 95.61 | 0.044 | 98.84 | 98.79 | 0.259 |
| 21 | 91.34 | 93.14 | 0.069 | 91.56 | 97.87 | 0.000 |
| 22 | 89.85 | 89.04 | 0.114 | 96.69 | 99.37 | 0.000 |
| 23 | 87.89 | 90.16 | 0.099 | 93.19 | 98.32 | 0.000 |
| Avg | 90.71 | 91.18 | 0.091 | **95.31** | **99.08** | **0.027** |

Where bold fonts indicate the best average results.

TABLE VII shows the performance of different models for predicting seizures on the CHB-MIT dataset. It can be seen that our method provides superior seizure prediction ability and achieves higher ACC, SEN, SPE, and lower FPR/h.

### *E. Influence of the STS-HFANet*

Based on the above overall performance, STS-HFANet in HSTS-MHAM is able to capture spatio-temporal-spectral fusion information characterizing epileptic seizures and improve the prediction accuracy of epileptic seizures. STS-HFANet takes multichannel spatial relationships as a breakthrough to explore the contribution of connectivity and dependency between cross-domain information within the epileptic brain to the seizure prediction ability.

We demonstrated the efficacy of STS-HFANet in capturing the dependency between multiscale features of EEG by comparing the performance of HSTS-MHAM with and without STS-HFANet. Fig. 11 shows the AUC results for a single-subject experiment, and TABLE VIII shows the average performance results for the other four metrics.

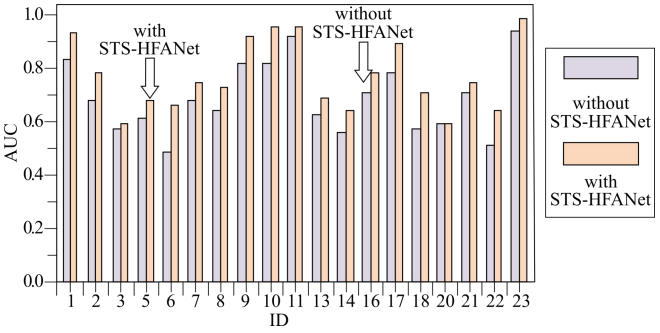


Fig. 11. AUC comparison of with and without STS-HFANet

TABLE VIII
PERFORMANCE COMPARISON OF INFLUENCE OF THE STS-HFANET ON THE CHB-MIT DATASET

| Method | ACC(%) | SPE(%) | SEN(%) | FPR/h |
|---|---|---|---|---|
| without STS-HFANet | 90.14 | 93.21 | 90.72 | 0.102 |
| with STS-HFANet | **93.82** | **96.70** | **91.34** | **0.045** |

Where bold fonts indicate the best average results.

### *F. Influence of the MHAConvNet*

In addition to the exploration of multidimensional feature dependencies via STS-HFANet, MHAConvNet is also used to efficiently model multichannel and spatial interactions, reducing information redundancy and fully integrating spatio-temporal-spectral multiscale EEG information.

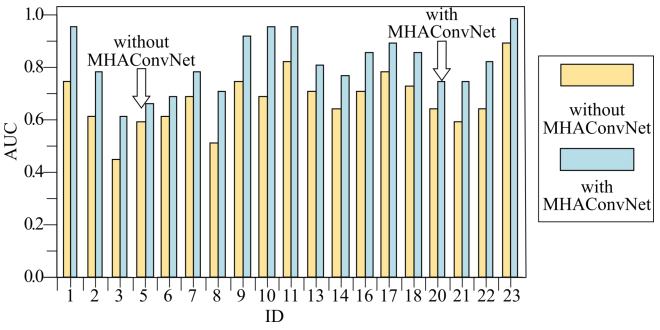


Fig. 12. AUC comparison of with and without MHAConvNet

TABLE IX
PERFORMANCE COMPARISON OF INFLUENCE OF THE MHACONVNET ON THE CHB-MIT DATASET

| Method | ACC(%) | SPE(%) | SEN(%) | FPR/h |
|---|---|---|---|---|
| without MHAConvNet | 91.47 | 95.38 | 89.01 | 0.148 |
| with MHAConvNet | **94.33** | **98.27** | **92.25** | **0.031** |

Where bold fonts indicate the best average results.

TABLE X
PERFORMANCE COMPARISON OF CROSS-SUBJECT EXPERIMENTS ON THE CHB-MIT DATASET

| Group | STS-HGCN | | | TA-STS-ConvNet | | | **Our Method** | | |
|---|---|---|---|---|---|---|---|---|---|
| | SEN(%) | FPR/h | AUC | SEN(%) | FPR/h | AUC | SEN(%) | FPR/h | AUC |
| A | 91.22 | 0.077 | 0.851 | 85.71 | 0.081 | 0.849 | 90.16 | 0.098 | 0.857 |
| B | 85.21 | 0.162 | 0.818 | 83.33 | 0.198 | 0.822 | 85.96 | 0.132 | 0.820 |
| C | 88.19 | 0.156 | 0.842 | 91.34 | 0.058 | 0.875 | 89.56 | 0.099 | 0.913 |
| D | 86.11 | 0.041 | 0.902 | 90.51 | 0.044 | 0.915 | 88.87 | 0.112 | 0.927 |
| E | 83.64 | 0.192 | 0.837 | 87.45 | 0.177 | 0.887 | 93.27 | 0.086 | 0.892 |
| Avg | 86.87 | 0.126 | 0.850 | 87.67 | 0.112 | 0.870 | **89.56** | **0.105** | **0.882** |

Where bold fonts indicate the best average results.

In order to evaluate the effectiveness of MHAConvNet in capturing the best biomarkers of seizures, we compared the performance of HSTS-MHAM with MHAConvNet to that of HSTS-MHAM without MHAConvNet. Fig. 12 shows the AUC results for the single-subject experiment, and TABLE IX shows the average performance results for the other four metrics.

### *G. Cross-Subject Experiments*

In the cross-subject experiments, we sequentially divided the data from 19 patients into five groups, specifically A: chb01, chb02, chb03, chb05; B: chb06, chb07, chb08, chb09; C: chb10, chb11, chb13, chb14; D: chb16, chb17, chb18, chb20; E: chb21, chb22, chb23.

TABLE X shows the performance of the three models in predicting seizures on the CHB-MIT dataset. It can be seen that the average SEN of STS-HGCN and TA-STS-ConvNet is 86.87% and 87.67%, respectively, and the average AUC is 0.850 and 0.870, respectively, whereas the average SEN of the model implemented in this paper is 89.56%, and the average AUC is 0.882. Furthermore, the average FPR/h of the baseline methods is 0.126/h and 0.112/h, respectively, and 0.105/h for the method used in this paper. From the experimental results, the model can effectively distinguish between interictal and preictal states.

## V. CONCLUSION

In this paper, we propose a new processing framework for EEG analysis that combines deep learning methods for geometric graphs to achieve automatic prediction of seizures. Specifically, HSTS-MHAM first extracts distinguishable cross-scale fused features from EEG via STS-HFANet. Further, MHAConvNet combines spatial attention and channel attention to enhance discriminative spatio-temporal-spectral features. Finally, SDGConvNet captures the most discriminative preseizure biomarkers by modeling spatiotemporal dependencies and maps the final results into a classification network. The method significantly improves the accuracy of seizure prediction.

The performance of our model is evaluated on two publicly available datasets, showing promising applications in common clinical scenarios through seizure detection and seizure prediction. Patient heterogeneity is overcome by designing single-subject experiments and cross-subject experiments. In addition, the impact of STS-HFANet and MHAConvNet on model robustness enhancement is validated. Our approach produced excellent results in ACC, SPE, SEN, AUC, and FPR/h demonstrating the feasibility of the system.

## REFERENCES


[1] E. Beghi, "The epidemiology of epilepsy," *Neuroepidemiology*, vol. 54, no.2, pp185-191, 2020.

[2] S. Balestrini, A. Arzimanoglou, I. Blümcke, I. E. Scheffer, S. Wiebe, J. Zelano, and M. C. Walker, "The aetiologies of epilepsy," *Epileptic Disorders*, vol. 23, no.1, pp1-16, 2021.

[3] B. Mesraoua, D. Deleu, A. H. Hassan, M. Gayane, A. Lubna, M. A. Ali, T. Tomson, B. A. Khalil, J. H. Cross, and A. A. Asadi-Pooya, "Dramatic outcomes in epilepsy: depression, suicide, injuries, and mortality," *Current medical research and opinion*, vol. 36, no.9, pp1473-1480, 2020.

[4] N. A. Shlobina and J. W. Sander, "Learning from the comorbidities of epilepsy," *Current Opinion in Neurology*, vol. 35, no.2, pp175-180, 2022.

[5] L. S. Gadelho and J. G. Marques, "Catatonia associated with epileptic seizures: a systematic review of case reports," *Epilepsy Research*, vol. 186, 2022.

[6] A. Shoeibi, P. Moridian, M. Khodatars, N. Ghassemi, M. Jafari, R. Alizadehsani, Y. Kong, J. M. Gorriz, J. Ramírez, A. Khosravi, S. Nahavandi, and U. R. Acharya, "An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: Methods, challenges, and future works," *Computers in Biology and Medicine*, vol. 149, 2022. [Online]. Available: https://arxiv.org/abs/2105.14278

[7] World Health Organization, "Improving the lives of people with epilepsy: a technical brief," 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240064072

[8] E. Gonzalez-Viana, A. Sen, A. Bonnon, and J. H. Cross, "Epilepsies in children, young people, and adults: summary of updated NICE guidance," *British Medical Journal*, vol. 378, 2022. [Online]. Available: https://www.bmj.com/content/378/bmj.o1446

[9] M. Marco, "Suicidality and antiepileptic drugs in people with epilepsy: an update," *Expert review of neurotherapeutics*, vol. 22, no.5, pp405-410, 2022.

[10] W. Robyn, S. Sharma, and R. Ramachandrannair, "Sudden unexpected death in epilepsy in children," *Developmental Medicine & Child Neurology*, vol. 65, no. 9, pp1150-1156, 2023.

[11] A. M. McIntosh, P. W. Carney, K. M. Tan, T. M. Hakami, P. Perucca, P. Kwan, T. J. O'Brien, and S. F. Berkovic, "Comorbidities in newly diagnosed epilepsy: Pre-existing health conditions are common and complex across age groups," *Epilepsy & Behavior*, vol. 138, 2023.

[12] C. Schwab, Nora-Elena Wadle, S. Knake, F. Podewils, K. Siebenbrodt, K. Kohlhase, J. Schulz, K. Menzler, C. Mann, F. Rosenow, C. Seifart, and A. Strzelczyk, "Patients' knowledge about epilepsy-related risks, morbidity, and mortality: A multicenter cohort study from Germany," *Epilepsy & Behavior*, vol. 124, 2021.

[13] N. A. Shlobin and J. W. Sander, "Learning from the comorbidities of epilepsy," *Current Opinion in Neurology*, vol. 35, no. 2, pp175-180, 2022.

[14] J. Lang, S. Jeschke, B. Herziger, R. M. Müller, T. Bertsche, M. P. Neininger and A. Bertsche, "Prejudices against people with epilepsy as perceived by affected people and their families," *Epilepsy & Behavior*, vol. 127, 2022.

[15] E. Trinka and M. Leitinger, "Management of status epilepticus, refractory status epilepticus, and super-refractory status epilepticus," *CONTINUUM: Lifelong Learning in Neurology*, vol. 28, no. 2, pp559-602, 2022.

[16] H. Anwar, Q. U. Khan, N. Nadeem, I. Pervaiz, M. Ali, and F. F. Cheema, "Epileptic seizures," *Discoveries*, vol. 8, no. 2, pp14-32, 2020.

[17] UK National Guideline Alliance, "Magnetic resonance imaging scan to detectrelevant abnormalities in people with epilepsy," no. 217, 2022. Available: https://www.ncbi.nlm.nih.gov/books/NBK581149/

[18] D. Nieto, J. D. Martinez-Vargas, and E. Giraldo, "Multirate adaptive filter banks based on automatic bands selection for epilepsy detection," *Engineering Letters*, vol. 29, no. 1, pp43-48, 2021.

[19] R. Rahman, S. M. Varnosfaderani. O. Makke, N. J. Sarhan, E. Asano, A. Luat, and M. Alhawari, "Comprehensive analysis of EEG datasets for epileptic seizure prediction," *2021 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE*, pp1-5, 2021.
[20] G. Liu, R. Xiao, L. Xu, and J. Cai, "Minireview of epilepsy detection techniques based on electroencephalogram signals," *Frontiers in systems neuroscience*, vol. 15, 2021. [Online]. Available: https://doi.org/10.3389/fnsys.2021.685387
[21] X. Xu, M. Lin, and T. Xu, "Epilepsy seizures prediction based on nonlinear features of EEG signal and gradient boosting decision tree," *International Journal of Environmental Research and Public Health*, vol. 19, no. 18, 2022. [Online]. Available: https://doi.org/10.3390/ijerph191811326
[22] S. Parui, D. Samanta, N. Chakravorty, W. Mansoor, and U. Ghosh, "A study on seizure detection performance in an automated process by extracting entropy features," *2022 5th International Conference on Signal Processing and Information Security (ICSPIS). IEEE*, pp86-91, 2022.
[23] C. Canyurt and R. Zengin, "Epileptic activity detection using mean value, RMS, sample entropy, and permutation entropy methods," *The Journal of Cognitive Systems*, vol. 8, no. 1, pp16-27, 2023.
[24] M. L. Giudice, G. Varone, C. Ieracitano, N. Mammone, G. G. Tripodi, E. Ferlazzo, S. Gasparini, U. Aguglia, and F. C. Morabito, "Permutation entropy-based interpretability of convolutional neural network models for inter-ictal eeg discrimination of subjects with epileptic seizures vs. psychogenic non-epileptic seizures," *Entropy*, vol. 24, no. 1, 2022. [Online]. Available: https://doi.org/10.3390/e24010102
[25] K. Parashar, "Analyzing the efficiency of mutual information (MI) and a genetic algorithm (GA) for selecting spectral entropy features from an EEG signal," *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP). IEEE*, pp194-198, 2023.
[26] A. Humairani, A. Rizal, I. Wijayanto, S. Hadiyoso, and Y. N. Fuadah, "Wavelet-based entropy analysis on EEG signal for detecting seizures," *2022 10th International Conference on Information and Communication Technology (ICoICT). IEEE*, pp93-98, 2022.
[27] B. Kapoor, B. Nagpal, P. K. Jain, A. Abraham, and L. A. Gabralla, "Epileptic seizure prediction based on hybrid seek optimization tuned ensemble classifier using EEG signals," *Sensors*, vol. 23, no. 1, 2022.
[28] H. M. Husseina and K. K. Abdalla, "Seizure prediction algorithm based on simulated annealing and machine learning," *International Journal of Nonlinear Analysis and Applications*, vol. 14, no. 1, pp1499-1508, 2023.
[29] L. Lu, F. Zhang, Y. Wu, S. Ma, X. Zhang, and G. Ni, "A multi-frame network model for predicting seizure based on sEEG and iEEG data," *Frontiers in Computational Neuroscience*, vol. 16, 2022. [Online]. Available: https://doi.org/10.3389/fncom.2022.1059565
[30] Z. Georgis-Yap, M. R. Popovic, and S. S. Khan, "Supervised and unsupervised deep learning approaches for EEG seizure prediction," [Online]. Available: https://arxiv.org/abs/2304.14922
[31] G. Alizadeh, T. Y. Rezaii, and S. Meshgini, "Automatic epileptic seizure prediction based on convolutional neural network and EEG signal," *Preprints 2023*. [Online]. Available: https://www.preprints.org/manuscript/202306.0623/v1
[32] R. Saemaldahr and M. Ilyas, "Patient-specific pre-ictal pattern-aware epileptic seizure prediction with federated learning," *Sensors*, vol. 23, no. 14, 2023. [Online]. Available: https://doi.org/10.3390/s23146578
[33] Y. Hu, X. Zhang, J. Yang, and S. Fu, "A hybrid convolutional neural network model based on different evolution for medical image classification," *Engineering Letters*, vol. 30, no. 1, pp168-177, 2022.
[34] I. B. Slimen, L. Boubchir, and H. Seddik, "Epileptic seizure prediction based on EEG spikes detection of ictal-pre-ictal states," *Journal of biomedical research*, vol. 34, no. 3, pp162-169, 2020.
[35] Y. Wang, Q. Zhou, J. Luo, Y. Lu, H. Wang, Y. Pang, and Z. Huang, "VLSI design of multiclass classification using sparse extreme learning machine for epilepsy and seizure detection," *IEICE Electronics Express*, vol. 19, no. 2, 2022. [Online]. Available: https://doi.org/10.1587/elex.18.20210536
[36] L. Guo, T. Yu, S. Zhao, X. Li, X. Liao, and Y. Li, "Triple-attention-based spatio-temporal-spectral convolutional network for epileptic seizure prediction," [Online]. Available: https://www.techrxiv.org/doi/full/10.36227/techrxiv.20557074.v1
[37] X. Dang, G. Liu, X. Tang, S. Wang, T. Wang, and M. Zou, "Motor imagery EEG recognition based on generative and discriminative adversarial learning framework and hybrid scale convolutional neural network," *IAENG International Journal of Applied Mathematics*, vol. 52, no. 4, pp946-954, 2022.
[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp7132-7121, 2018.
[39] Y. Li, Y. Liu, W. G. Cui, Y. Z. Guo, H. Huang, and Z. Y. Hu, "Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp782-794, 2020.
[40] Y. Li, Y. Liu, Y. Z. Guo, X. F. Liao, B. Hu, and T. Yu, "Spatio-temporal-spectral hierarchical graph convolutional network with semisupervised active learning for patient-specific seizure prediction," *IEEE transactions on cybernetics*, vol. 52, no. 11, pp12189-12204, 2021.
[41] Y. Gao, X. Chen, A. Liu, D. Liang, L. Wu, R. Qian, H. Xie, and Y. Zhang, "Pediatric seizure prediction in scalp EEG using a multi-scale neural network with dilated convolutions," *IEEE journal of translational engineering in health and medicine*, vol. 10, pp1-9, 2022.
[42] X. Lu, A. Wen, L. Sun, H. Wang, Y. Guo, and Y. Ren, "An epileptic seizure prediction method based on CBAM-3D CNN-LSTM model," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp417-423, 2023.
[43] X. Yang, J. Zhao, Q. Sun, J. Lu, and X. Ma, "An effective dual self-attention residual network for seizure prediction," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp1604-1613, 2021.
[44] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, "Triple cross-domain attention on human activity recognition using wearable sensors," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 5, pp1167-1176, 2022.
[45] J. Malmivuo and R. Plonsey, "Electroencephalography," in *Bioelectromagnetism*, Ed. New York, NY, USA: Oxford University Press, 1995, ch. 13, pp247-264.
[46] G. Buckwalter, S. Chhin, S. Rahman, I. Obeid, and J. Picone, "Recent advances in the TUH EEG corpus: improving the interrater agreement for artifacts and epileptiform events," *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE*, pp1-3, 2021.
[47] Z. B. Tariq, A. Iyengar, L. Marcuse, H. Su, and B. Yener, "Patient-specific seizure prediction using single seizure electroencephalography recording," *International Workshop on Health Intelligence*, vol. 1013, pp295-308, 2021.

**Sunan Ge** received a PhD from the School of Electronic Information and Electrical Engineering of the Dalian University of Technology in 2016. She is a Lecturer in the School of Computer Science, at Xi'an Polytechnic University, and a Postdoc in the School of Mathematics, at Northwest University, China. Her current research interests include tensor decomposition, artefact removal of EEGs, and underdetermined blind source separation.

**Wenjing Wang** is a M.Eng Candidate from the School of Computer Science, at Xi'an Polytechnic University. She received a B.Eng degree from the School of Computer Science, at Xi'an Polytechnic University in 2017. Her current research interests include graph convolutional networks, seizure prediction, and EEG signal processing.

**Xin Shi** is a PhD Candidate from the Communication and Information School of Communication University of China. She is an engineer in the School of Computer Science, at Xi'an Polytechnic University. Her current research interests include computer vision and relevant applications.

**Meng Wang** received a MSc degree in software engineering from Xi'an Jiaotong University and a PhD degree from the School of Computer Science and Technology, Xidian University, China, respectively. He is currently affiliated with the School of Computer Science, Xi'an Polytechnic University, China. His research interests include spatiotemporal data, time-series analysis, data mining and management. He has published papers in major venues in these areas such as ICDE, TKDE, CIKM, EDBT, Inf.Sci., Bioinformatics and TIST.