

# Dynamic Multi-Graph Convolution-Based Channel-Weighted Transformer Feature Fusion Network for Epileptic Seizure Prediction

Yifan Wang<sup>1</sup>, Weigang Cui<sup>1</sup>, Tao Yu<sup>1</sup>, Xiaoli Li<sup>1</sup>, Xiaofeng Liao<sup>1</sup>, *Fellow, IEEE*,  
and Yang Li<sup>2</sup>, *Senior Member, IEEE*

**Abstract**—Electroencephalogram (EEG) based seizure prediction plays an important role in the closed-loop neuromodulation system. However, most existing seizure prediction methods based on graph convolution network only focused on constructing the static graph, ignoring multi-domain dynamic changes in deep graph structure. Moreover, the existing feature fusion strategies generally concatenated coarse-grained epileptic EEG features directly, leading to the suboptimal seizure prediction performance. To address these issues, we propose a novel multi-branch dynamic multi-graph convolution based channel-weighted transformer feature fusion network (MB-dMGC-CWTFNet) for the patient-specific seizure prediction with the superior performance. Specifically, a multi-branch (MB) feature extractor is first applied to capture the temporal, spatial and spectral representations from the epileptic EEG jointly. Then, we design a point-wise dynamic multi-graph convolution network (dMGCN) to dynamically

learn deep graph structures, which can effectively extract high-level features from the multi-domain graph. Finally, by integrating the local and global channel-weighted strategies with the multi-head self-attention mechanism, a channel-weighted transformer feature fusion network (CWTFNet) is adopted to efficiently fuse the multi-domain graph features. The proposed MB-dMGC-CWTFNet is evaluated on the public CHB-MIT EEG and a private intracranial sEEG datasets, and the experimental results demonstrate that our proposed method achieves outstanding prediction performance compared with the state-of-the-art methods, indicating an effective tool for patient-specific seizure warning. Our code will be available at: <https://github.com/Rockingsnow/MB-dMGC-CWTFNet>.

**Index Terms**—Seizure prediction, multi-graph convolution network, transformer, EEG, epilepsy.

## I. INTRODUCTION

EPILEPSY is one of the most common brain diseases of nervous system, producing recurrent seizures and threatening the patients' life [1]. Recently, more than 50 million people worldwide suffer from epilepsy, and there are approximately 30% of patients deteriorating into refractory epilepsy, despite both drug and surgical treatment [2]. Fortunately, the seizure prediction based on Electroencephalography (EEG) provides an additional solution for these refractory epilepsy patients, who can give early warning for advanced neuromodulation treatments [3], so as to suppress seizures effectively. The previous studies divided the long-term recorded epileptic EEG signals into four neurophysiological periods: inter-ictal, pre-ictal, ictal and post-ictal periods [4], [5]. Therefore, the core problem for the epileptic seizure prediction is how to accurately distinguish the pre-ictal period from inter-ictal period, promoting intelligent warning before seizure onset for patients and clinicians [6].

For automatic EEG seizure prediction, the primary challenge is to extract discriminative EEG features of the epileptic activity. Due to the high temporal resolution of EEG, the long short-term memory (LSTM) [7], [8] was introduced to the seizure prediction models to capture the temporal information of the epileptic EEG. In addition, to exploit the spectral representation in epileptic rhythms, the wavelet transformation [9]

Manuscript received 1 April 2023; revised 23 August 2023; accepted 29 August 2023. Date of publication 2 October 2023; date of current version 1 November 2023. This work was supported in part by the Beijing Natural Science Foundation under Grant Z220017, in part by the National Natural Science Foundation of China under Grant 62325301 and Grant 62201023, in part by the China Postdoctoral Science Foundation under Grant 2023M730175, in part by the Beijing Municipal Education Commission-Natural Science Foundation under Grant KZ202110025036, and in part by the Beijing United Imaging Research Institute of Intelligent Imaging Foundation under Grant CRIB-JQY202103. (Yifan Wang and Weigang Cui contributed equally to this work.) (Corresponding author: Yang Li.)

Yifan Wang is with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: by2003156@buaa.edu.cn).

Weigang Cui is with the School of Engineering Medicine, Beihang University, Beijing 100191, China (e-mail: cwg1994@buaa.edu.cn).

Tao Yu is with the Beijing Institute of Functional Neurosurgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China (e-mail: yu\_tao@sinacn.com).

Xiaoli Li is with the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China (e-mail: xiaoli@bnu.edu.cn).

Xiaofeng Liao is with the College of Computer Science, Chongqing University, Chongqing 400715, China (e-mail: xfliao@cqu.edu.cn).

Yang Li is with the State Key Laboratory of Virtual Reality Technology and Systems, Beijing Advanced Innovation Center for Big Data and Brain Computing, Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: liyang@buaa.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3321414

and the short-time Fourier transform [10] were combined with the convolution neural network (CNN), which can learn quantitative time-frequency characteristics to facilitate the classification of inter-ictal and pre-ictal periods. Moreover, Ahmet et al. [4] proposed a 3D-CNN seizure prediction framework to evaluate the spatio-temporal evolution correlation from multi-channel EEG time series. Zhang et al. [11] designed a spatial filter of common spatial pattern to extract distinguishing spatial features from epileptic EEG, which further fed into a shallow CNN to discriminate between the pre-ictal and inter-ictal states. However, these methods just obtained the coarse-grained EEG features in single or multi domains by a fixed mode, without taking full advantage of the patient-specific temporal, spectral and spatial signatures simultaneously, which may lead to the loss of essential epileptic activity information. Thus, a multi-branch feature extractor is needed to capture the multi-level fine-grained representations from the epileptic EEG in multiple domains.

Another existing issue is that the CNN framework in seizure prediction task can only learn low-dimensional spatial correlations among EEG channels, due to its regular convolution operation and the local receptive field [12]. It is difficult to track the complex non-Euclidean structure in the epileptic seizures [13]. To deal with this problem, the graph convolutional network (GCN) were investigated in recent studies of the seizure prediction field [14], [15]. The common procedure in GCN is to define the prior adjacency matrix for constructing the graph structure among channels, which helps to convert the epileptic EEG signals to a graph representation with graph nodes and edges [16]. For example, Wang et al. [17] employed phase locking value (PLV) in EEG data to construct the adjacency matrix of graph edges. The differential entropy (DE) was applied in the inference of the spatial coupling in network topology to calculate the temporal correlations of EEG and yield the graph nodes [18]. Unfortunately, these GCN methods based on the information theory depended on the handcrafted features to generate EEG graphs, neglecting the dynamic changes in patient-specific graph construction. Lian et al. [14] developed a joint graph structure and representation learning network (JGRN) to predict seizures, where the graph structures can be jointly optimized with patient-specific connection weights of temporal channels. A similar study proposed a subject-independent seizure predictor by using geometric deep learning, realizing the seizure prediction from LSTM EEG graph synthesis [15]. It is notable that most of these models ignored the spatial position relationship among EEG channels, and only focused on the single and shallow static EEG graph construction without the spatial position guidance, which cannot fully represent the dynamic changes of individualized channel connectivity in multiple domains. Therefore, a novel GCN is highly required to jointly characterize high-level multi-domain features, and map patient-specific dynamic EEG graph representations.

Additionally, in order to integrate comprehensive feature information for the precise seizure prediction, some feature fusion strategies were designed to fuse the EEG features from different scales and domains. For example, Li et al. [19] adopted a temporal-spectral squeeze-and-excitation scheme

to fuse the hierarchical multi-domain representations of epileptic EEG, which reduced the information redundancy of high-dimensional features. Gao et al. [20] combined the attention mechanism with dilated convolution to aggregate spatio-temporal multi-scale features, providing a promising solution for EEG-based seizure prediction. Although these feature fusion methods obtained a comprehensive feature, they only considered the general fusion of low-level features in Euclidean space [21]. High-level EEG graph node features, embedded in non-Euclidean graph structures, urgently need a specific fusion approach to enable robust seizure prediction.

The main motivation of our study aims to break through limitations of the existing prediction methods, including coarse-grained EEG features in single domain, shallow static EEG graph construction without spatial position guidance and difficulties in high-level graph feature fusion. Thus, we propose a novel multi-branch dynamic multi-graph convolution based channel-weighted transformer feature fusion network (MB-dMGC-CWTFNet), for the patient-specific seizure prediction. First, a multi-branch (MB) feature extractor is used to capture multi-level fine-grained features from epileptic EEG in multiple domain. Second, in order to extract multi-domain graph features, a point-wise dynamic multi-graph convolution network (dMGCN) is constructed to adaptively learn three-view dynamic graph structure with spatial position guidance. Finally, we investigate a channel-weighted transformer feature fusion network (CWTFNet) to efficiently fuse the multi-domain graph features, which introduces the channel-weighted self-attention mechanism to map discriminative fused representations for the seizure prediction. The proposed MB-dMGC-CWTFNet is evaluated on two kinds of epileptic datasets, i.e., CHB-MIT EEG dataset and our Xuanwu intracranial stereo-electroencephalography (sEEG) dataset, and achieves the promising performance compared with the state-of-the-art methods, which validates its outstanding capability in seizure prediction task.

In general, the main contributions of our study are summarized as follows:

- 1) A novel MB-dMGC-CWTFNet is proposed to predict seizures for the individual epilepsy patient, which can efficiently fuse multi-domain graph features, yielding the highest prediction performance on both CHB-MIT and Xuanwu datasets, respectively.

- 2) We design a MB feature extractor, including three parallel sub-branches in temporal, spatial and frequency domains respectively, to capture the multi-level fine-grained features jointly, which offsets the inadequate representation of coarse-grained EEG features in traditional feature extractors.

- 3) A dMGCN is constructed by point-wise dynamic graph neural network, which can learn dynamic changes of three-view graph structures with spatial position guidance, and extract deep multi-domain graph features, and thus overcomes insufficient expression of spatial connectivity in shallow static EEG graph.

- 4) A CWTFNet is developed by introducing both the local and the global channel-weighted self-attention into the transformer network. The local graph edge weights are

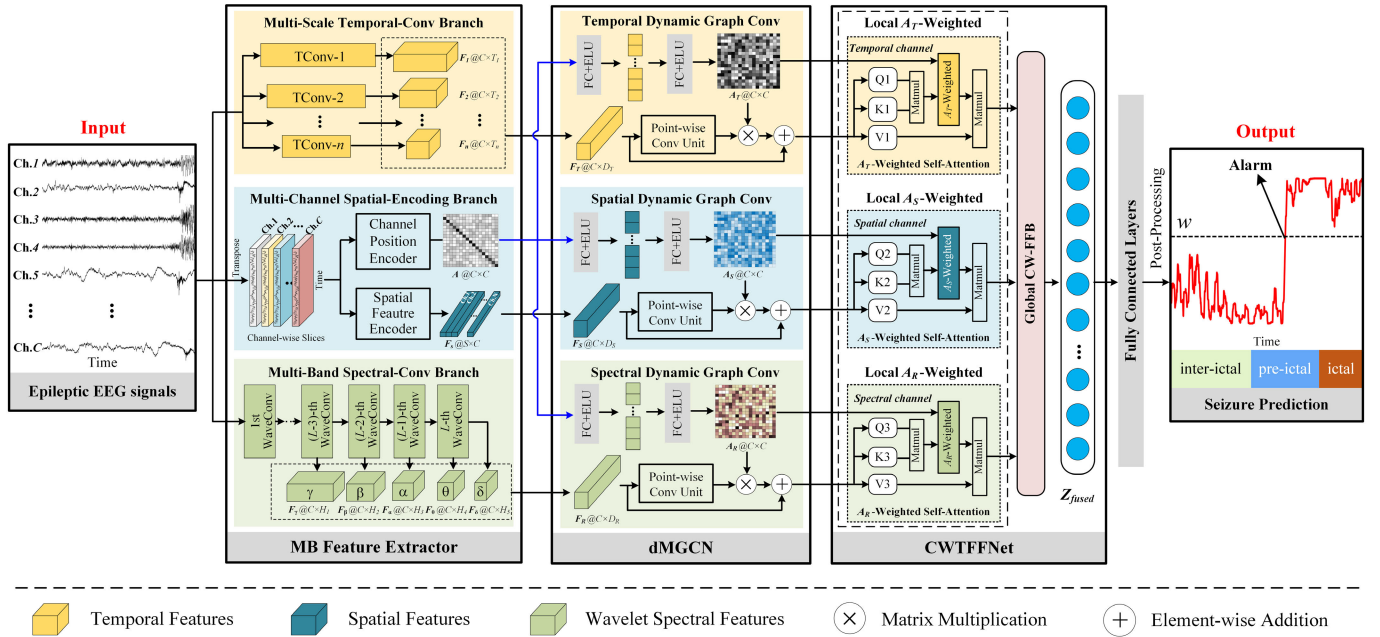


Fig. 1. The framework of the proposed MB-dMGC-CWTFFFNet for the seizure prediction.

complementary to the global channel position information, which can implement efficient fusion of high-level graph features against current feature fusion strategies.

## II. METHODOLOGY

The seizure prediction framework of our proposed MB-dMGC-CWTFFFNet is displayed in Fig. 1. The overall architecture mainly consists of the multi-branch feature extractor, the point-wise dynamic multi-graph convolution network and the channel-weighted transformer feature fusion network, summarized as follows: 1) The MB feature extractor is primarily designed to extract the multi-domain temporal-spatial-spectral features from EEG signals. 2) The dMGCN is further employed to transform the temporal-spatial-spectral features into high-level graph representations from temporal, spatial and spectral views. 3) The CWTFFFNet is adopted to obtain the fused feature maps, and the fully connected layers are utilized to generate the recognition results ultimately. The well-trained MB-dMGC-CWTFFFNet is then transformed into a practical seizure warning system by a post-processing strategy. Details of each step are given in following subsection.

### A. Multi-Branch Feature Extractor

The epileptic EEG signals are defined as  $E = (x_i, y_i) | i = 1, 2, \dots, N$ , where  $x_i \in R^{C \times S}$  represents the  $i$ -th EEG trial with  $C$  channels and  $S$  sampling points.  $N$  is the total number of EEG trials.  $y_i$  is a binarized label of pre-ictal or inter-ictal state corresponding to  $x_i$ .

Considering the individualized differences of epileptic activities in both time domain, frequency domain and spatial domain, we firstly construct the MB feature extractor to capture the temporal-spatial-spectral representations from epileptic EEG signals. In Fig. 1, the MB feature extractor includes three sub-branches: the multi-scale temporal-conv branch,

the multi-band spectral-conv branch and the multi-channel spatial-encoding branch.

1) *Multi-Scale Temporal-Conv Branch*: Epileptic seizure recordings involved the critical electrophysiological fluctuation from inter-ictal period to pre-ictal period [22]. In order to capture the comprehensive temporal information of EEG with its higher time resolution in time domain, a multi-scale temporal-conv branch is first designed with  $n$  parallel temporal convolution (TConv) layers. Thus, we can gain the multi-scale temporal features with different sizes from TConv-1 to TConv- $n$ , denoted as  $F_k \in R^{C \times T_k}$ ,  $k = 1, 2, \dots, n$ , where  $T_k$  is the output scale of the temporal feature from the  $k$ -th TConv. Additionally, the batch normalization and exponential linear unit (ELU) are also applied in the each TConv of the multi-scale temporal-conv branch to accelerate the training and convergence of the proposed model. Accordingly, these multi-scale temporal features are concatenated to generate the overall feature map in time domain:  $F_T \in R^{C \times D_T}$ , where  $D_T$  is the sum of  $T_k$ ,  $k = 1, 2, \dots, n$ .

2) *Multi-Band Spectral-Conv Branch*: Previous studies have proven that the epileptic activities may be of different frequencies for epilepsy patients [23]. Thus, according to the clinical five frequency sub-bands:  $\delta$  band (0-4Hz),  $\theta$  band (4-8Hz),  $\alpha$  band (8-13Hz),  $\beta$  band (13-30Hz),  $\gamma$  band (30-50Hz) [24], the multi-band spectral-conv branch is adopted to contain the hierarchical wavelet convolutions (WaveConv) based on Daubechies order-4 (Db4) wavelet [25]. The wavelet decomposition can be accomplished on the EEG trials due to its high correlation coefficients with the epileptic signal [26] to obtain the wavelet spectral features in the five sub-bands. The hierarchical WaveConv layers perform successive spectral analysis by means of  $L$ -level iteration, where  $L = \lfloor \log_2(f_s) \rfloor - 3$ , determined by the EEG sampling rate  $f_s$ , and  $\lfloor \cdot \rfloor$  represents the rounding-down operation [27]. Then, the frequency boundaries

of the  $l$ -th WaveConv are  $(0, f_s/2^{(l+1)})$  and  $(f_s/2^{(l+1)}, f_s/2^l)$ , respectively, where  $l = 1, 2, \dots, L$ . After inputting the EEG trial  $x_i \in R^{C \times S}$  into the multi-band spectral-conv branch, we can obtain the multi-band wavelet spectral features:  $F_\delta \in R^{C \times H_1}$ ,  $F_\theta \in R^{C \times H_2}$ ,  $F_\alpha \in R^{C \times H_3}$ ,  $F_\beta \in R^{C \times H_4}$ ,  $F_\gamma \in R^{C \times H_5}$  corresponding to five standard physiological frequency sub-bands, where  $H = S/2^l$  is the output dimension of the wavelet spectral features generated from the  $l$ -th WaveConv. Additionally, due to the similar of time-frequency analysis with the discrete wavelet transform, the WaveConv operators have no learnable parameters in the processing of the feature extraction, which weights are fixed and given by Db4 wavelet filter. Then, the five-band wavelet spectral features are concatenated into the integral spectral feature map in frequency domain:  $F_R \in R^{C \times D_R}$ , where  $D_R$  is the sum of  $H_1, H_2, \dots, H_5$ .

3) *Multi-Channel Spatial-Encoding Branch*: Apart from the multi-scale temporal-conv branch and multi-band spectral-conv branch, we also propose a multi-channel spatial-encoding branch to excavate the representations of channel mapping in spatial domain. Specially, the multi-channel EEG trials are transposed to the channel-wise slices, which are imported into the channel position encoder and spatial feature encoder to complete the channel correlation construction and the spatial feature extraction, respectively. For the channel position encoder, a distance set  $U$  is foremost established by  $U = \{u_{ij} | i, j \in [1, C], i \neq j\}$ , where  $u_{ij}$  represents the Euclidean distance of the  $i$ -th channel and the  $j$ -th channel,  $C$  is the total number of channels. we can get the  $u_{ij}$  from the international standard electrode system [28]. Then the initialized channel adjacency matrix  $A \in R^{C \times C}$  is generated by the following position embedding method:

$$a_{ij} = \begin{cases} \frac{1}{u_{ij}} & \text{if } u_{ij} < M(U) \\ 0 & \text{if } u_{ij} \geq M(U) \\ \frac{1}{M(\{u_{ij} | u_{ij} < M(U)\})} & \text{if } i = j \end{cases} \quad (1)$$

where  $M(\cdot)$  is the mean operation,  $a_{ij}$  is the element of the  $i$ -th row and the  $j$ -th column of adjacency matrix  $A \in R^{C \times C}$ . Therefore, the channel adjacency matrix  $A$  contains the global position information of the multi-channel relationship, which will be used to construct the dynamic graph in the following point-wise dMGCN. Additionally, we adopt the spatial feature encoder based on the channel-wise spatial convolution [29], [30] to extract the multi-channel spatial characteristics, which are ultimately concatenated and reshaped as  $F_S \in R^{C \times D_S}$ , where  $D_S$  is the output dimension of integrated spatial feature map.

### B. Point-Wise Dynamic Multi-Graph Convolution Network

To further learn deep dynamic connectivity of different brain regions for the individual epilepsy patient, in this subsection, a novel point-wise dMGCN is proposed to extract multi-domain graph features. Three synchronized dynamic graph convolution networks are involved by temporal, spatial and spectral views. Three views constitute the point-wise

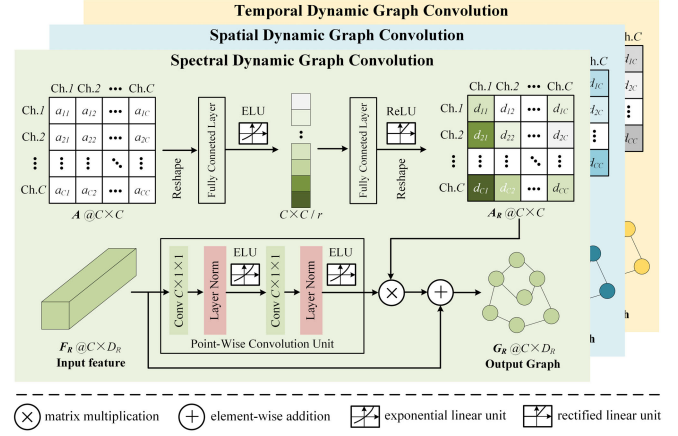


Fig. 2. The architecture of the point-wise dynamic multi-graph convolution network (dMGCN).

dMGCN in Fig. 2, which can explore the deep channel relationship of the temporal feature map  $F_T$ , the spatial feature map  $F_S$  and the wavelet spectral feature map  $F_R$ , respectively. For each graph convolution view, the initialized adjacency matrix  $A \in R^{C \times C}$ , depicting original distance between any two channels, has been calculated by the channel position encoder of the MB feature extractor. To further guide the dynamic evolution process of the channel relationship from three kinds of views, a self-gating strategy is employed in initialized adjacency matrix  $A$  as follows:

$$\tilde{A}_T = \sigma(W_{11}\delta(W_{12}(\tilde{A}_1))) \quad (2)$$

$$\tilde{A}_S = \sigma(W_{21}\delta(W_{22}(\tilde{A}_2))) \quad (3)$$

$$\tilde{A}_R = \sigma(W_{31}\delta(W_{32}(\tilde{A}_3))) \quad (4)$$

where  $\tilde{A}_1, \tilde{A}_2, \tilde{A}_3 \in R^{(C \times C) \times 1}$  are reshaped from  $A \in R^{C \times C}$ ,  $W_{12}, W_{22}, W_{32} \in R^{((C \times C)/r) \times (C \times C)}$  and  $W_{11}, W_{21}, W_{31} \in R^{(C \times C) \times ((C \times C)/r)}$  are weight matrixes of fully-connected layers,  $r$  is the reduction ratio,  $\delta(\cdot)$  and  $\sigma(\cdot)$  are the ELU activation function and the rectified linear unit (ReLU). Hence the three dynamic adjacency matrixes  $A_T, A_S, A_R \in R^{C \times C}$  corresponding to temporal, spatial and spectral graph convolution nets are acquired by reshaping  $\tilde{A}_T, \tilde{A}_R, \tilde{A}_S \in R^{(C \times C) \times 1}$  into  $R^{(C \times C)}$ .

After constructing the dynamic connectivity of epileptic activities from three views, the operations of the dynamic graph convolution are performed on the temporal feature map  $F_T \in R^{C \times D_T}$ , the spatial feature map  $F_S \in R^{C \times D_S}$  and the wavelet spectral feature map  $F_R \in R^{C \times D_R}$ , respectively, which are formulated by:

$$G_T = \delta(D_T^{-1}A_T\delta(F_T\Theta_{11})\Theta_{12} + F_T) \quad (5)$$

$$G_S = \delta(D_S^{-1}A_S\delta(F_S\Theta_{21})\Theta_{22} + F_S) \quad (6)$$

$$G_R = \delta(D_R^{-1}A_R\delta(F_R\Theta_{31})\Theta_{32} + F_R) \quad (7)$$

where  $G_T \in R^{C \times D_T}$ ,  $G_S \in R^{C \times D_S}$ ,  $G_R \in R^{C \times D_R}$  are the dynamic graph features corresponding to  $F_S, F_T, F_R$  with the hidden non-Euclidean topology in epileptic activities,  $D_T^{ii} = \sum_j A_T^{ij}$ ,  $D_S^{ii} = \sum_j A_S^{ij}$ ,  $D_R^{ii} = \sum_j A_R^{ij}$  are the degree matrixes corresponding to  $A_T, A_S, A_R$  respective,

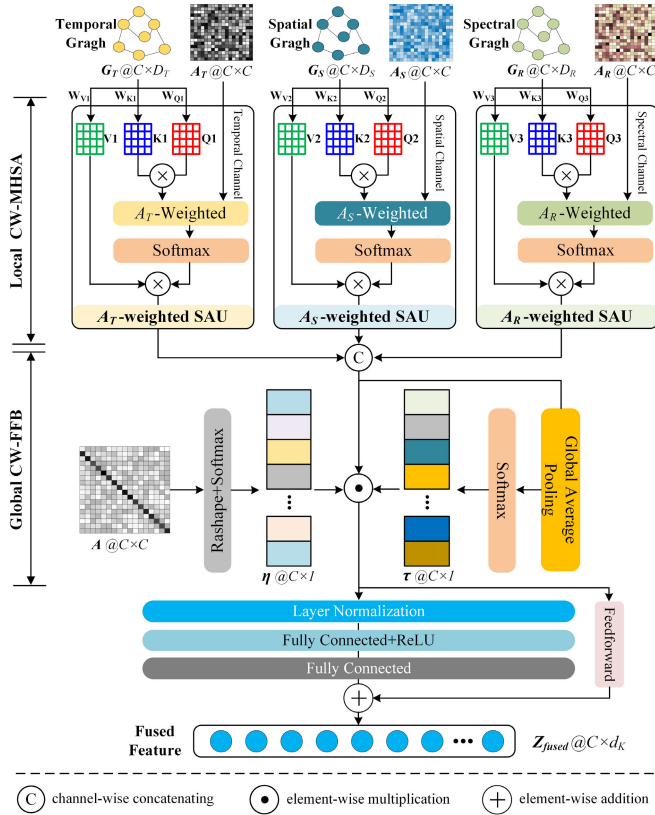


Fig. 3. The outline of channel-weighted transformer feature fusion network (CWTFNet).

$\Theta_{11}, \Theta_{12} \in R^{D_T \times D_T}$ ,  $\Theta_{21}, \Theta_{22} \in R^{D_S \times D_S}$ ,  $\Theta_{32}, \Theta_{33} \in R^{D_R \times D_R}$  represent the weight matrixes of convolution kernels in the point-wise convolution unit [31]. Therefore, we obtain the dynamic multi-domain graph features  $G_T, G_S, G_R$  with their corresponding dynamic adjacency matrix  $A_T, A_S, A_R$ , which will be fed into the CWTFNet to conduct the final feature fusion in the next subsection.

### C. Channel-Weighted Transformer Feature Fusion Network

To further fuse the high-level graph features  $G_T, G_S, G_R$ , the CWTFNet is proposed by combining the dynamic adjacency matrix  $A_T, A_S, A_R$  with multi-head self-attention mechanism. In Fig. 3, the CWTFNet can be divided into a local channel-weighted multi-head self-attention (Local CW-MHSA), a global channel-weighted feature fusion block (Global CW-FFB) and the multi-layer perception (MLP).

The Local CW-MHSA consists of three heads: the  $A_T$ -weighted self-attention unit (SAU), the  $A_S$ -weighted SAU and the  $A_R$ -weighted SAU. For each self-attention head, three kinds of weight matrixes, denoted as  $W_Q \in R^{D \times d_K}$ ,  $W_K \in R^{D \times d_K}$ ,  $W_V \in R^{D \times d_V}$ , are initially introduced to encode the input graph features, where  $d_K$  and  $d_V$  both represent the hyperparameters. So, the query  $Q$ , the key  $K$  and the value  $V$  are calculated via:

$$Q = GW_Q \quad (8)$$

$$K = GW_K \quad (9)$$

$$V = GW_V \quad (10)$$

where the  $G \in \{G_T, G_S, G_R\}$  indicates three kinds of graph features. We introduce the local channel-weighted strategy by applying the dynamic adjacency matrixes in multi-head self-attention mechanism. Then, the local features  $Z_{local} \in R^{C \times d_K}$  from Local CW-MHSA is obtained by:

$$Z_T = A_T \text{softmax}\left(\frac{Q_1 K_1^T}{\sqrt{d_K}}\right) V_1 \quad (11)$$

$$Z_S = A_S \text{softmax}\left(\frac{Q_2 K_2^T}{\sqrt{d_K}}\right) V_2 \quad (12)$$

$$Z_R = A_R \text{softmax}\left(\frac{Q_3 K_3^T}{\sqrt{d_K}}\right) V_3 \quad (13)$$

$$Z_{local} = \text{Concat}(Z_T, Z_S, Z_R) \quad (14)$$

where  $Z_T, Z_S, Z_R$  are the outputs from the three self-attention heads respectively, and  $\cdot \text{Concat}()$  is the concatenation function. To further capture the global information, we employ the Global CW-FFB and MLP in local features  $Z_{local}$  according to the following equations:

$$Z_{global} = \text{softmax}(A) Z_{local} \text{softmax}(GAP(Z_{local})) \quad (15)$$

$$Z_{fused} = \mathcal{F}_{fc}^2(\text{ReLU}(\mathcal{F}_{fc}^1(\text{LN}(Z_{global})))) + \text{FM}(Z_{global}) \quad (16)$$

where  $Z_{global} \in R^{C \times d_K}$  is the global features,  $A \in R^{C \times C}$  is the initialized adjacency matrix from the channel position encoder,  $\cdot GAP()$  represents the global average pooling,  $\cdot LN()$  means the layer normalization,  $\cdot \text{ReLU}()$  is the rectified linear unit,  $\mathcal{F}_{fc}^1$  and  $\mathcal{F}_{fc}^2$  denote the FC layers, and  $\cdot \text{FM}()$  is a feedforward module (including two feedforward layers and an ELU activation). Therefore, we acquire the fused features  $Z_{fused} \in R^{C \times d_K}$  through the constructed CWTFNet.

Finally, two fully-connected layers are used to conduct the decoding for the fused features. They are flattened into a 1-dimensional tensor to feed into the fully connected layers, then the classification probabilities of the pre-ictal and the inter-ictal states are estimated by the Softmax function, and the index corresponding to the maximum of probabilities represents the final result of seizure prediction. Moreover, a cross-entropy loss function is employed for the patient-specific model training of proposed MB-dMGC-CWTFNet, the cross-entropy loss  $L_{CE}$  between the prediction result and the label is minimized by:

$$L_{CE} = \sum_{i=1}^N \sum_{j=1}^M -\log(p_i) \varphi(y_i = l_j) + \lambda \|\theta\| \quad (17)$$

where  $p_i$  is the conditional probability of the  $i$ -th EEG trial outputted by the proposed MB-dMGC-CWTFNet,  $l_j$  is the class from the label set,  $\varphi()$  represents the indicator function,  $N$  is the total number of samples,  $M = 2$  is the number of classes.  $\lambda \|\theta\|$  belongs to the trade-off regularization term of Eq. (17), and aims to alleviate the overfitting problem during the model training, where  $\lambda$  is the regularization parameter and the  $\theta$  denotes the updatable parameters of the model. As a result, a personalized well-trained model of the proposed MB-dMGC-CWTFNet is generated, and will be performed the individual seizure prediction by the following post-processing [32].

TABLE I  
PATIENTS' DETAILS OF THE XUANWU INTRACRANIAL sEEG DATASET

Patient ID	Age (years)	Gender	No. of seizures	Recording duration/h
1	16	M	4	11.0
2	17	F	3	7.0
3	18	F	2	7.0
4	31	F	4	11.0
5	26	M	3	6.0
Total	-	-	16	42.0

#### D. Post-Processing Strategy

Eventually, the well-trained MB-dMGC-CWTFNet is then transformed into a practical seizure warning system by a post-processing strategy [33]. Specifically, after inputting the consecutive EEG signals into the well-trained MB-dMGC-CWTFNet, the probability series  $P(i)$  belonging to pre-ictal class from  $i$ -th epoch is generated. Then we employ a moving average filter on  $P(i)$  to reduce the oscillation and obtain the smoothed probability series  $P_s(i)$  over time [32]. The lengths of the moving average filter are configured to 15s and 25s for CHB-MIT and our Xuanwu dataset respectively, which will be discussed by the experimental results in Section IV-B.

### III. EXPERIMENTAL RESULTS

#### A. Dataset Description

The performance of the proposed MB-dMGC-CWTFNet is evaluated on two epileptic datasets, which is given as follows:

1) **CHB-MIT Scalp EEG Dataset [34]**: The CHB-MIT dataset contains the scalp EEG signals from 23 patients, which were recorded with 18 common electrodes and sampled at 256Hz in the Children's Hospital Boston. In this study, there were at least two seizures and three-hour inter-ictal recordings from each patients, who were selected for the patient-specific model evaluation of seizure prediction [4]. In addition, the neural recordings within two hours after a seizure are removed to exclude the effect of post-ictal period [32]. Specially, if several seizures cluster within two hours, only the first seizure prediction is considered as an effective evaluation, because a successful warning depends on whether the model can predict the leading seizure [35].

2) **Xuanwu Intracranial sEEG Dataset**: The Xuanwu dataset is collected by the Xuanwu Hospital of Capital Medical University, Beijing, China, which consists of sEEG recordings on the intracranial depth electrode for 5 focal epilepsy patients, which sampled at 256Hz with 15 channels. From Table I, there are total 16 seizures, and the recording duration of sEEG from these patients is 42 hours. The labels of the inter-ictal, pre-ictal and ictal states were marked by the professional clinicians. This study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University (LYS2018041) in Beijing and complied with the ethical standards of the Declaration of Helsinki. Informed consent was obtained from all patients.

#### B. Experimental Settings and Evaluation Metrics

In this study, based on the recent research [4], [11], the EEG signals from CHB-MIT and Xuanwu datasets are both

cropped into 5-second clips before fed into the proposed MB-dMGC-CWTFNet. Additionally, the pre-ictal period was popularly defined by 15 minutes before seizure onset in the latest methods [4], [32]. Thus, we adopt the identical setting to the pre-ictal period, and the inter-ictal period is defined at least 2 hours away prior to seizure onset and after seizure ending [10].

To conduct a comprehensive performance evaluation of the proposed MB-dMGC-CWTFNet, the patient-specific leave-one-out cross-validation (LOOCV) [20] is employed in this study. Since the inter-ictal period is much larger than the pre-ictal period in the model training stage, the inter-ictal clips are randomly down-sampled to the same number of the pre-ictal clips [8]. Then, assuming that there are total  $N_i$  seizures for the  $i$ -th patient, in each leave-one-out loop,  $N_i - 1$  seizures are utilized for training while the left one is for testing, during the training stage, the cross-validation is used to divide training set and validation set. It is repeated with  $N_i$  loops until the proposed model completes the prediction evaluation of all seizures for the  $i$ -th patient. Since the number of seizures and the recording duration are both different for each patient in two datasets, for each leave-one-out loop, the number of training data samples varies from 3028 to 18241, the validation data size varies from 572 to 3543, the testing data size varies from 3131 to 7508, and the total data size varies from 6731 to 29292. The proposed MB-dMGC-CWTFNet is evaluated via four metrics, including area under curve (AUC), sensitivity ( $S_n$ ), false prediction rate (FPR/h) and the  $p$ -value. AUC mainly reflects the classification performance for the inter-ictal and the pre-ictal states.  $S_n$  denotes the ratio of successfully predicted seizures to the total number of seizures. FPR/h indicates the number of false alarms per hour, and the  $p$ -value represents the significance of an improvement over chance-level, which is used to evaluate statistically significance whether the seizure warning system is better than a random predictor [4].

#### C. Overall Performance

In order to illustrate the patient-specific prediction efficiency of the proposed MB-dMGC-CWTFNet, we compare our proposed model with the following state-of-the-art methods in the same chance-level, which are tested on two datasets.

- 1) **DCNN-Bi-LSTM [8]**: This is a typical deep learning method by combining the deep convolutional network with a bidirectional long short-term memory, extracting the spatial and temporal features of epileptic EEG signals respectively, which were used for the seizure prediction.
- 2) **CE-stSENet [19]**: This method used a temporal-spectral squeeze-and-excitation network to capture hierarchical multi-domain representations, which introduced the attention mechanism into the epileptic seizure detection task and improved the recognition performance.
- 3) **TS-MS-DCNN [20]**: This advanced model encoded the multi-scale EEG features by designing temporal and spatial multi-scale stages, and a dilated convolution block was constructed to further expand the feature receptive and achieving the EEG-based seizure prediction.

TABLE II  
THE PATIENT-SPECIFIC OVERALL COMPARISON OF PERFORMANCE ON CHB-MIT DATASET

Patient ID	DCNN-Bi-LSTM [8]				CE-stSENet [19]				TS-MS-DCNN [20]				Our MB-dMGC-CWTFNet			
	AUC	$S_n$ (%)	FPR/h	$p$	AUC	$S_n$ (%)	FPR/h	$p$	AUC	$S_n$ (%)	FPR/h	$p$	AUC	$S_n$ (%)	FPR/h	$p$
1	0.989	100.0	0.000	0.070	0.999	100.0	0.000	<0.001	0.999	100.0	0.000	<0.001	0.999	100.0	0.000	<0.001
2	0.693	100.0	0.000	0.004	0.799	66.7	0.000	<0.001	0.861	100.0	0.000	<0.001	0.834	100.0	0.452	<0.001
3	0.838	100.0	0.000	<0.001	0.892	83.3	0.070	<0.001	0.733	66.7	0.163	<0.001	0.913	100.0	0.021	<0.001
5	0.845	100.0	0.000	<0.001	0.824	100.0	0.052	<0.001	0.923	100.0	0.000	<0.001	0.926	100.0	0.000	<0.001
6	0.824	85.7	5.616	<0.001	0.832	100.0	0.274	<0.001	0.924	100.0	0.000	<0.001	0.931	100.0	0.000	<0.001
7	0.712	33.3	0.000	0.004	0.743	66.7	0.000	0.006	0.623	66.7	0.367	<0.001	0.827	100.0	0.181	0.001
8	0.930	80.00	0.000	0.001	0.927	100.0	0.000	0.003	0.976	100.0	0.000	<0.001	0.986	100.0	0.000	<0.001
9	0.919	100.0	0.000	0.002	0.712	50.0	0.000	<0.001	0.964	100.0	0.000	<0.001	0.933	75.0	0.025	<0.001
10	0.929	100.0	0.000	<0.001	0.961	83.3	0.000	<0.001	0.848	100.0	0.000	<0.001	0.984	100.0	0.000	<0.001
11	0.938	100.0	0.060	<0.001	0.897	66.7	0.000	<0.001	0.964	100.0	0.061	<0.001	0.921	100.0	0.037	<0.001
13	0.847	80.00	1.084	<0.001	0.943	100.0	0.000	<0.001	0.975	80.0	0.000	<0.001	0.986	100.0	0.000	<0.001
14	0.707	83.3	0.105	<0.001	0.708	66.7	0.210	<0.001	0.761	100.0	1.232	<0.001	0.854	83.3	0.234	<0.001
16	0.926	100.0	0.000	<0.001	0.916	100.0	0.000	<0.001	0.980	100.0	0.000	<0.001	0.952	100.0	0.000	<0.001
17	0.892	100.0	0.059	0.001	0.882	100.0	0.059	0.001	0.832	66.7	0.847	0.002	0.947	100.0	0.000	<0.001
18	0.833	100.0	0.000	<0.001	0.817	83.3	0.036	<0.001	0.846	83.3	0.053	<0.001	0.959	100.0	0.000	<0.001
20	0.994	100.0	0.000	<0.001	0.982	100.0	0.000	<0.001	0.996	100.0	0.000	<0.001	0.982	100.0	0.000	<0.001
21	0.765	75.0	3.438	<0.001	0.778	100.0	2.181	<0.001	0.875	75.0	0.000	<0.001	0.863	100.0	0.175	<0.001
22	0.916	100.0	0.884	0.015	0.679	66.7	4.133	0.016	0.834	100.0	0.000	<0.001	0.980	100.0	0.000	<0.001
23	0.943	85.7	0.000	<0.001	0.998	100.0	0.000	<0.001	0.999	100.0	0.000	<0.001	0.991	100.0	0.000	<0.001
Aver	0.865	90.7	0.592	-	0.857	86.0	0.369	-	0.890	91.5	0.143	-	<b>0.935</b>	<b>97.8</b>	<b>0.059</b>	-

where the bold values indicate the best average results.

TABLE III  
THE PATIENT-SPECIFIC OVERALL COMPARISON OF PERFORMANCE ON XUANWU DATASET

Patient ID	DCNN-Bi-LSTM [8]				CE-stSENet [19]				TS-MS-DCNN [20]				Our MB-dMGC-CWTFNet			
	AUC	$S_n$ (%)	FPR/h	$p$	AUC	$S_n$ (%)	FPR/h	$p$	AUC	$S_n$ (%)	FPR/h	$p$	AUC	$S_n$ (%)	FPR/h	$p$
1	0.664	50.00	0.993	0.024	0.819	75.0	1.780	0.003	0.878	75.0	1.156	<0.001	0.972	100.0	0.047	<0.001
2	0.982	100.0	0.000	0.013	0.817	66.67	0.000	0.003	0.895	100.0	0.355	0.019	0.984	100.0	0.022	0.001
3	0.984	100.0	0.329	0.015	0.984	100.0	0.000	0.359	0.912	100.0	0.000	0.013	0.988	100.0	0	0.011
4	0.996	100.0	0.000	0.004	0.940	100.0	0.733	0.004	0.996	75.0	0.000	0.003	0.976	100.0	0.326	0.002
5	0.995	100.0	0.000	0.018	0.999	100.0	0.000	0.019	0.999	100.0	0.000	0.018	0.999	100.0	0	0.014
Aver	0.924	90.0	0.264	-	0.912	88.3	0.503	-	0.936	90.0	0.302	-	<b>0.984</b>	<b>100.0</b>	<b>0.079</b>	-

where the bold values indicate the best average results.

The experimental results of the patient-specific comparison on public CHB-MIT and our Xuanwu datasets are listed in Table II and Table III respectively. From Table II, we can observe that the DCNN-Bi-LSTM, CE-stSENet and TS-MS-DCNN gain the average AUC of 0.865, 0.857 and 0.890 respectively on CHB-MIT dataset, while our proposed MB-dMGC-CWTFNet reaches the highest average AUC of 0.935. Especially for Patient 1, 8, 13 and 23, which the AUC are all greater than 0.985, indicating an excellent performance of our method in distinguishing between the inter-ictal state and the pre-ictal state. In the seizure prediction scenario, the average sensitivity of our proposed model achieves an ideal 97.8% as well, which outperforms three baseline methods with 7.1%, 11.8% and 6.3% respectively. The distinct advancement on  $S_n$  demonstrates that our proposed model, which is transformed into the seizure warning system, performed a more successful seizure warning for an individual patient. In addition, our proposed model yields the lowest average FPR/h of 0.059, which is at least 58.7% improvement against other methods. Meanwhile, the  $p$ -value of our seizure warning system is less than or equal to 0.001 for all patients, implying the improvement-over-chance of our seizure predictor is statistically significant with 99.9% confidence interval. It indicates that our proposed MB-dMGC-CWTFNet has the significantly patient-specific capability for the epileptic seizure prediction.

Additionally, to further validate the effectiveness of our proposed method, Table III lists the prediction results for five focal epilepsy patients on Xuanwu dataset. It is obvious that our proposed model achieves more excellent performance on AUC and  $S_n$  with average 0.984 and 100% respectively, which are at least 5.1% and 10.0% higher than that of the state-of-the-art models. The average FPR/h of our method is 0.079, which is lower than other methods. These encouraging experimental results demonstrate the remarkable performance ( $p < 0.05$ ) of our MB-dMGC-CWTFNet framework in the subject-independent intracranial seizure prediction task, which makes it possible to predict seizure by implanting intracranial deep electrode, and it enables more convenient treatment for refractory epilepsy patients [36].

## IV. DISCUSSIONS

### A. Ablation Studies

To prove the innovation of each component of our proposed MB-dMGC-CWTFNet, the ablation studies are conducted on both CHB-MIT and Xuanwu datasets. In this subsection, we discuss the efficacy of each innovation by comparing the proposed method with and without this component, which contributes to justifying the positive influence. The overall experimental result of ablation studies is presented

TABLE IV  
ABLATION STUDIES ON TWO DATASETS

Dataset	Methods	AUC	$S_n$ (%)	FPR/h
CHB-MIT	without Temporal-Conv	0.911	94.8	0.294
	without Spatial-Encoding	0.907	92.3	0.337
	without Spectral-Conv	0.892	93.4	0.515
	without dMGCN	0.871	89.7	0.326
	without CWTFNet	0.880	90.3	0.194
	<b>Our MB-dMGC-CWTFNet</b>	<b>0.935</b>	<b>97.8</b>	<b>0.059</b>
Xuanwu	without Temporal-Conv	0.954	85.0	0.146
	without Spatial-Encoding	0.948	88.3	0.243
	without Spectral-Conv	0.962	95.0	0.127
	without dMGCN	0.915	83.3	0.651
	without CWTFNet	0.936	86.7	0.306
	<b>Our MB-dMGC-CWTFNet</b>	<b>0.984</b>	<b>100.0</b>	<b>0.079</b>

where the bold values indicate the proposed method.

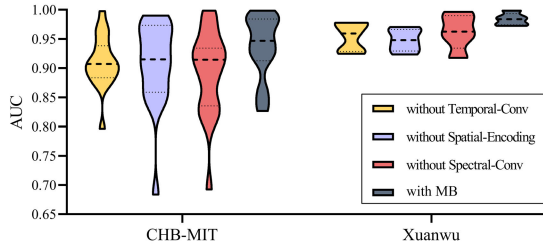


Fig. 4. Performance comparison of AUC between the models with and without MB feature extractor on CHB-MIT and Xuanwu datasets.

in Table IV, and the impacts of MB feature extractor, point-wise dMGCN and CWTFNet are demonstrated respectively as follows:

1) *Impact of MB Feature Extractor*: In order to give a comprehensive assessment for the proposed MB feature extractor, we compare our MB-dMGC-CWTFNet with three simplified sub-models: a) the model without temporal-conv; b) the model without spatial-encoding; c) the model without spectral-conv. From Table IV and Fig. 4, firstly, when using the temporal-conv to extract the multi-scale temporal features on two datasets, the AUC of our MB-dMGC-CWTFNet increased by 2.6% and 3.1% respectively compared with the model without the temporal-conv. The  $S_n$  also get the improvements of 3.0% on CHB-MIT dataset, 15.0% on our Xuanwu dataset. Additionally, the FPR/h declines by 79.9% and 45.9% on two datasets respectively, which illustrates the availability of the temporal-conv in capturing multi-scale temporal evolution, and indicates the effectiveness of MB feature extractor in extracting fine-grained temporal features.

Meanwhile, the spatial-encoding also plays an important role in the feature extraction of multi-channel spatial features. For instance, compared to the model without the spatial-encoding, the AUC of the proposed model increase from 0.907 and 0.948 to 0.935 and 0.984 on two datasets respectively, and the  $S_n$  are improved from 94.8% and 85% into 97.8% and 100%. The FPR/h decrease from 0.337 and 0.243 to 0.059 and 0.079. It verifies that the spatial-encoding branch enables exact spatial expression with cortical multi-channel representations, which contributes to the seizure prediction with distinct improvements of performance metrics.

In addition to the above two branches proposed in MB feature extractor, the effect of the spectral-conv is

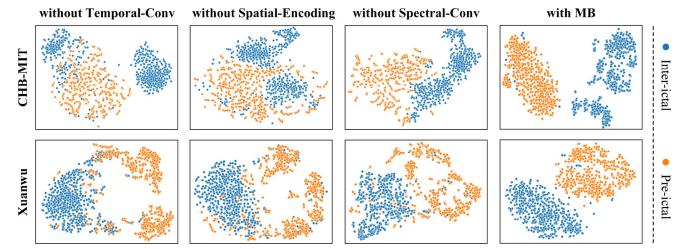


Fig. 5. The  $t$ -SNE visualization in 2D embedding space of inter-ictal and pre-ictal features by comparing the models with and without MB feature extractor on CHB-MIT and Xuanwu datasets.

further discussed. From Table IV, we can find that the proposed method with spectral-conv shows a better performance. For the CHB-MIT dataset, its evaluation metrics of AUC and  $S_n$  are 4.8% and 4.4% higher than that of the model without spectral-conv, and the FPR/h declines by 88.5%. Similarly, when using the spectral-conv on the Xuanwu dataset, our MB-dMGC-CWTFNet achieves the improvement of 2.3% AUC and 5.0%  $S_n$  over the model without spectral-conv, whose FPR/h are reduced by 37.8% accordingly. These evaluation results also prove that the designed spectral-conv can extract comprehensive spectrum characteristics in five clinical physiological rhythms, which facilitates the construction of the patient-specific MB feature extractor by combining with temporal-conv and spatial-encoding branches.

Moreover, to further validate the superiority of the proposed MB feature extractor intuitively, the  $t$ -SNE is applied to visualize the temporal-spatial-spectral features, which were extracted by the models with and without MB feature extractor. The  $t$ -SNE visualization in 2D embedding space of inter-ictal and pre-ictal features on two datasets is shown in Fig. 5. We can see that the binary-class feature distributions, learned by MB feature extractor, presents a better discrimination than the model without MB feature extractor on both CHB-MIT and Xuanwu datasets. Especially for the models without spatial-encoding, some inter-ictal and pre-ictal features are confused together. In contrast, the proposed model using MB feature extractor obtain more discriminative features, embodied in visible inter-class distance and dense intra-class distribution on both two datasets. These phenomena also explain that the best seizure prediction performance can be produced by combining the temporal-conv, the spatial-encoding and the spectral-conv branches simultaneously, which fully illustrates the innovation of MB feature extractor in extracting the multi-level fine-grained features jointly.

2) *Influence of Point-Wise dMGCN*: To judge the contribution of the proposed point-wise dMGCN, we perform an ablation experiments for our point-wise dMGCN to investigate its influence in the patient-specific seizure prediction. Fig. 6 shows the comparison of AUC between the proposed models with and without point-wise dMGCN for each patient. For CHB-MIT dataset, we can see that the average AUC of our MB-dMGC-CWTFNet is 7.3% higher than that of the model without point-wise dMGCN. Especially, a maximum AUC increase of 0.18 (about 22.5% improvement) occurs on Patient 13. For Xuanwu dataset, the AUC of our proposed model increased by about 7.5% compared to the model without



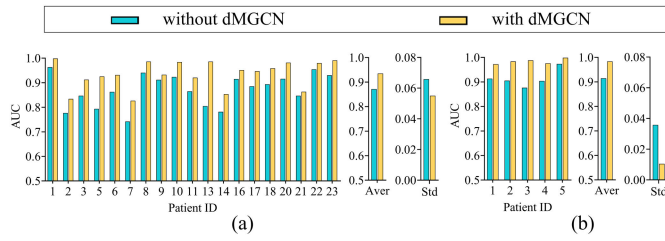


Fig. 6. Performance comparison of AUC between the proposed models with and without point-wise dMGCN on (a) CHB-MIT dataset and (b) Xuanwu dataset.

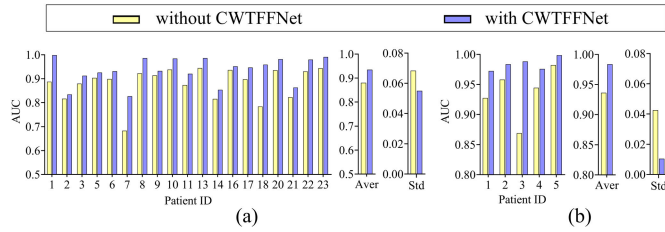


Fig. 7. Performance comparison of AUC between the proposed model with and without CWTFNet on (a) CHB-MIT dataset and (b) Xuanwu dataset.

point-wise dMGCN, where the AUC gets an encouraging ascending from 0.876 to 0.988 on Patient 3. Moreover, the relatively low standard deviations of 0.055 and 0.01 on CHB-MIT and Xuanwu demonstrates the robustness of our proposed model with point-wise dMGCN. These improvements of AUC on two datasets validate the outstanding capability of the proposed point-wise dMGCN in spatial representation learning. Besides, in Table IV, after employing point-wise dMGCN, the  $S_n$  of our model outperform that of the ablation model with 8.1% and 16.7% on two datasets respectively, and the FPR/h is decreased from 0.326 into 0.059 on CHB-MIT, from 0.651 into 0.079 on Xuanwu dataset. These enhancements give substantial evidences that point-wise dMGCN can better learn the three-view graph structures with spatial position guidance, and extract deep multi-domain graph features, which promotes the overall performance in seizure prediction warning.

3) *Efficacy of CWTFNet*: In order to fuse the dynamic multi-domain graph features, the CWTFNet is adopted integrate the local and global representation based on the channel-weighted self-attention mechanism. Therefore, we further compare the efficacy between our MB-dMGC-CWTFNet and the model without CWTFNet. The results of the ablation experiment on two datasets are presented in Fig. 7. It can be noted that the CWTFNet increases the average AUCs from 0.880 and 0.936 to 0.935 and 0.984 on two datasets, respectively. The standard deviations achieve 0.013 and 0.032 lower than that of the model without CWTFNet, which indicate the better generalization performance of our proposed CWTFNet across multiple patients. Especially for Patient 7 from CHB-MIT and Patient 3 from Xuanwu, their AUCs achieve greater improvements of 21.0% and 13.7% respectively. In addition, compared to the model without CWTFNet in Table IV, the proposed model gains higher  $S_n$  of 7.5% on CHB-MIT and 13.3% on Xuanwu, and FPR/h values get decline of 0.135 and 0.227 after utilizing CWTFNet on two datasets. It proves the advantage of incorporating the

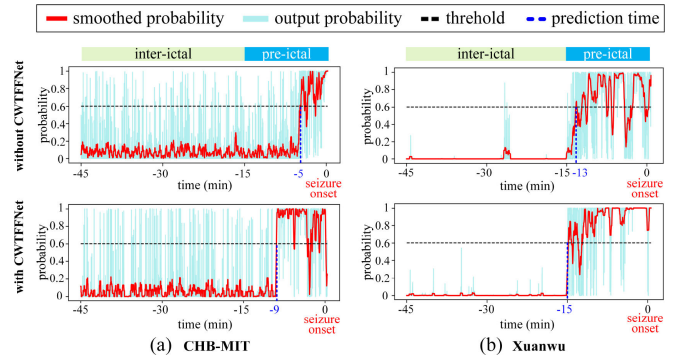


Fig. 8. Performance comparison of the seizure prediction time with and without CWTFNet from (a) CHB-MIT and (b) Xuanwu dataset.

multi-domain graph features by means of the CWTFNet. The local graph edges can be fully weighted into the multi-head attention, and complement global channel position information, which generates more distinguishable fused features for the seizure prediction.

To further evaluate the ability of the proposed CWTFNet contributing to the performance of the seizure prediction, we conduct the comparison of the prediction time with and without CWTFNet on CHB-MIT and Xuanwu respectively, and the results are unfolded in Fig. 8. The proposed models with and without CWTFNet both successfully implement the seizure prediction in pre-ictal periods. However, for the identical seizure from CHB-MIT dataset, the model without CWTFNet just achieves the seizure prediction with 5 minutes prior to the seizure onset, while our proposed model using CWTFNet obtain a 9-minute advance of prediction time. For Xuanwu dataset, the prediction time was improved from 13 minutes to 15 minutes before the seizure onset, and these improvements in seizure prediction time strongly embody the innovative contribution of CWTFNet. Specifically, the channel-weighted transformer in our CWTFNet can reinforce the learning of the model by multi-channel-weighted self-attention mechanism. Accordingly, the fused features, containing local multi-graph structures and global channel information, are more conducive to the seizure prediction.

### B. Parameters Analysis of Post-Processing

In this subsection, we mainly analyze the influence of two hyperparameters for our training model, the filter length and the threshold  $\omega$ , on the seizure warning system transformed by our MB-dMGC-CWTFNet. In the post-processing, the moving average filter can smooth the probability outputs of our model by filtering the outliers, resulting in practical seizure prediction results. Hence, the filter length in the moving average filter is set from 5 to 60 with a step of 5 (unit: second), and the corresponding variations of  $S_n$  and FPR/h on two datasets are shown in Fig. 9(a) respectively. Interestingly, in both two datasets, the larger filter length leads to the unsatisfactory  $S_n$ , while the smaller filter length causes the poor FPR/h. The main reason is that a large filter length may result in over-smoothed prediction results, and thus some short-duration warnings are probably missed. However, a small filter length may retain more predicted outliers, which greatly

TABLE V  
EXPERIMENTAL SETTINGS AND PERFORMANCE COMPARISON OF THE STATE-OF-THE-ART METHODS ON CHB-MIT

Authors	Year	Methods	No. of Patients	No. of Seizures	Validation Scheme	Interictal-Preictal Intervals (Min)	$S_n$ (%)	FPR/h
Khan et al. [9]	2018	WT-CNN	15	18	10-Fold CV	10-10	87.80	0.147
Truong et al. [10]	2018	STFT-CNN	13	64	LOOCV	240-30	81.20	0.160
Ozcan et al. [4]	2019	SP-3DCNN	16	77	LOOCV	120-60	87.01	0.186
Zhang et al. [11]	2020	CSP-CNN	23	156	LOOCV	NR-30	92.00	0.120
Yang et al. [35]	2021	STFT-RDANet	13	64	LOOCV	240-30	89.25	0.122
Gao et al. [20]	2022	TS-MS-DCNN	16	85	LOOCV	60-30	93.30	0.007
Dissanayake et al. [15]	2022	GDL	23	NR	10-Fold CV	NR-60	95.94	NR
Liang et al. [37]	2023	SSDA-SPM	13	64	LOOCV	240-30	88.80	0.182
<b>This Work</b>	<b>2023</b>	<b>Our MB-dMGC-CWTFNet</b>	<b>19</b>	<b>99</b>	<b>LOOCV</b>	<b>120-15</b>	<b>97.81</b>	<b>0.059</b>

where the bold values indicate the proposed method. NR is the nor reported values.

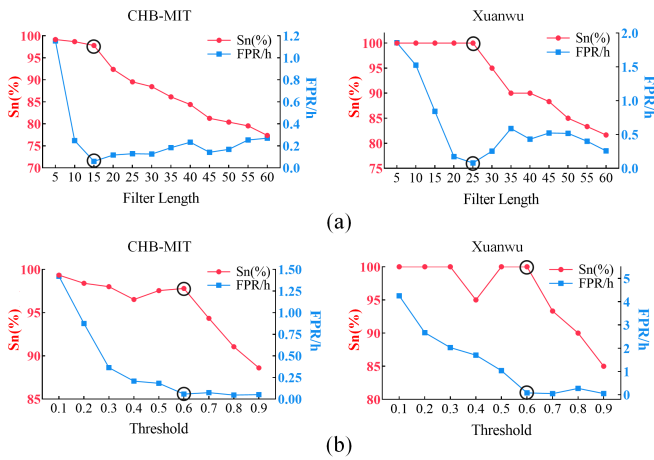


Fig. 9. Performance comparison of  $S_n$  and FPR/h with different post-processing parameters. (a) filter length; (b) threshold.

increases the probability of false alarm [32]. Consequently, to maintain the trade-off between  $S_n$  and FPR/h, the filter lengths are configured to 15 for CHB-MIT dataset and 25 for Xuanwu dataset.

Since the seizure warning depends on whether the predicted probability exceeds threshold  $\omega$ , we discuss this hyperparameter to evaluate its sensitivity on the proposed model. The threshold  $\omega$  is varied in the range from 0.1 to 0.9 with a stride of 0.1, and the performance trade-off between  $S_n$  and FPR/h is also displayed in Fig. 9(b). As can be observed, the variation trends of two evaluation metrics along with the threshold  $\omega$  are similar to that with the filter length. The best trade-off results between  $S_n$  and FPR/h both appear in the 0.6 threshold on two datasets. Thus, to achieve optimal seizure prediction after the post-processing, we set  $\omega$  to 0.6 as the final fixed threshold for CHB-MIT and Xuanwu datasets, which is consistent with existing studies [9], [32].

### C. Performance Comparison of the State-of-the Art Methods

The performance comparison of the state-of-the-art seizure prediction methods on CHB-MIT dataset is summarized

in Table V. In order to discuss the advantages of our proposed model, we conduct an objective comparative analysis among these methods. For example, Truong et al. [10] and Yang et al. [35] both employed the short-time Fourier transform (STFT) in the CNN of seizure prediction frameworks, which were tested on 13 patients and achieved the sensitivities of 81.2% and 89.25% respectively, lower than our MB-dMGC-CWTFNet. This is mainly because our proposed MB feature extractor can extract multi-level fine-grained features compared to traditional time-frequency feature extraction methods. Compared with two deep learning methods using spectral power [4] and common spatial pattern (CSP) [11] respectively, our proposed method applies the point-wise dMGCN learns three-view graph structures and captures deep multi-domain graph features, so it yields 10.8%, 5.81% higher in  $S_n$  and 0.127, 0.061 lower in FPR/h. Unlike the study [20] that fused the multi-scale temporal-spatial features by attention mechanism based dilated CNN, our MB-dMGC-CWTFNet introduces the local and global channel-weighted strategy into the multi-head self-attention units, which is beneficial to efficient feature fusion for complex graph structures and outperforms the TS-MS-DCNN with 4.51%  $S_n$ . Although some advanced methods [9], [15] gained the suboptimal performance in seizure prediction, their validation scheme using 10-Fold CV shuffled original EEG signals and destroyed the continuity of epileptic activity over time, and is not conducive to real-time seizure warning compared with our adopted LOOCV scheme. Additionally, compared with the model proposed by Liang et al. [37], our MB-dMGC-CWTFNet achieves 9.01% higher in  $S_n$  and 0.123 lower in FPR/h. Because our proposed method considers multi-domain variable information and constructs the multi-graph framework, it offsets the lack of partial domain information from the feature alignment strategy in SSDA-SPM. In summary, compared to most of existing studies, the main differences and advantages of our MB-dMGC-CWTFNet include that it can dynamically learn changes in multi-graph topologies with spatial position guidance. Meanwhile, it efficiently fuses multi-domain graph features by using channel-weighted multi-head self-attention mechanism.

## D. Limitations and Future Directions

Although our proposed prediction framework achieves satisfactory seizure prediction performance, two limitations still exist in our study. First, our MB-dMGC-CWTFFNet can realize the end-to-end seizure warning without complicated EEG pre-processing, but the artifacts in epileptiform discharges and potential bad channels may interfere with the predictor and cause some false positives in practical warning scenario. Therefore, we will devote to exploring the adaptive channel selection [38], [39] and unsupervised artifact removal algorithms [40], and further eliminating the redundant information in raw epileptic signals. Second, our proposed method conducts a patient-specific seizure prediction by training with the same patient's data, while it is difficult to complete the model fine-tuning across patients. Thus, we will combine the domain-adversarial transfer learning strategies [41], [42] with our seizure prediction framework in the future work, which aims to handle the drifting distribution between target domain and source domain, and contributes to the cross-patient seizure prediction.

## V. CONCLUSION

In this study, we propose a novel EEG-based MB-dMGC-CWTFFNet framework for patient-specific seizure prediction. The MB feature extractor is adopted to effectively capture the multi-level fine-grained representations in multiple domains. The designed point-wise dMGCN is further employed to dynamically learn the deep graph structures with spatial position guidance, which contributes to extracting the multi-domain graph features from temporal, spatial and spectral views. Finally, the CWTFFNet utilizes the local and global channel-weight strategy to facilitate the efficient fusion of high-level graph features. Furthermore, we conduct the comparative experiments on two epileptic datasets, and the results show our proposed MB-dMGC-CWTFFNet obtains a better evaluation metrics, whose AUC,  $S_n$ , FPR/h achieve 0.935 and 0.984, 97.8% and 100.0%, 0.059 and 0.079 on CHB-MIT and Xuanwu datasets respectively, outperforming the state-of-the-art methods. These findings prove the outstanding performance of our proposed MB-dMGC-CWTFFNet in patient-specific seizure prediction, and indicate its potential application prospect in neurostimulation treatment of refractory epilepsy patients.

## REFERENCES

- [1] E. Beghi et al., "Global, regional, and national burden of epilepsy, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016," *Lancet Neurol.*, vol. 18, no. 4, pp. 357–375, 2019.
- [2] *Epilepsy: A Public Health Imperative*, World Health Organization, Geneva, Switzerland, 2019.
- [3] N. Rincon, D. Barr, and N. Velez-Ruiz, "Neuromodulation in drug resistant epilepsy," *Aging Dis.*, vol. 12, no. 4, pp. 1070–1080, 2021.
- [4] A. R. Ozcan and S. Erturk, "Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2284–2293, Nov. 2019.
- [5] Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, and T. Tan, "Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach," *Knowl.-Based Syst.*, vol. 164, pp. 96–106, Jan. 2019.
- [6] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals," *IEEE Sensors J.*, vol. 21, no. 7, pp. 9377–9388, Apr. 2021.
- [7] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, Aug. 2018.
- [8] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 804–813, Oct. 2019.
- [9] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2109–2118, Sep. 2018.
- [10] N. D. Truong et al., "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw.*, vol. 105, pp. 104–111, Sep. 2018.
- [11] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 465–474, Feb. 2020.
- [12] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 379–388, Jan. 2022.
- [13] D. Zhang, L. Yao, K. Chen, S. Wang, P. D. Haghghi, and C. Sullivan, "A graph-based hierarchical attention model for movement intention detection from EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2247–2253, Nov. 2019.
- [14] Q. Lian, Y. Qi, G. Pan, and Y. Wang, "Learning graph in graph convolutional neural networks for robust seizure prediction," *J. Neural Eng.*, vol. 17, no. 3, Jun. 2020, Art. no. 035004.
- [15] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Geometric deep learning for subject independent epileptic seizure prediction using scalp EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 2, pp. 527–538, Feb. 2022.
- [16] S. Jang, S.-E. Moon, and J.-S. Lee, "EEG-based video identification using graph signal modeling and graph convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3066–3070.
- [17] M. Wang, H. El-Fiqi, J. Hu, and H. A. Abbass, "Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3259–3272, Dec. 2019.
- [18] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1290–1301, Jul. 2022.
- [19] 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 Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 782–794, Apr. 2020.
- [20] Y. Gao et al., "Pediatric seizure prediction in scalp EEG using a multi-scale neural network with dilated convolutions," *IEEE J. Transl. Eng. Health Med.*, vol. 10, pp. 1–9, 2022.
- [21] H. He and D. Wu, "Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 399–410, Feb. 2020.
- [22] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, "Epilepsy detection from EEG using complex network techniques: A review," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 292–306, 2023.
- [23] F. Bartolomei et al., "Defining epileptogenic networks: Contribution of SEEG and signal analysis," *Epilepsia*, vol. 58, no. 7, pp. 1131–1147, Jul. 2017.
- [24] Y. Li, X.-D. Wang, M.-L. Luo, K. Li, X.-F. Yang, and Q. Guo, "Epileptic seizure classification of EEGs using time-frequency analysis based multiscale radial basis functions," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 2, pp. 386–397, Mar. 2018.

- [25] Q. Li, L. Shen, S. Guo, and Z. Lai, "WaveCNet: Wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification," *IEEE Trans. Image Process.*, vol. 30, pp. 7074–7089, 2021.
- [26] K. Fukumori, N. Yoshida, H. Sugano, M. Nakajima, and T. Tanaka, "Epileptic spike detection using neural networks with linear-phase convolutions," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 1045–1056, Mar. 2022.
- [27] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534–1545, 2021.
- [28] J. N. Acharya and V. J. Acharya, "Overview of EEG montages and principles of localization," *J. Clin. Neurophysiol.*, vol. 36, no. 5, pp. 325–329, 2019.
- [29] Y. Chen et al., "Channel-UNet: A spatial channel-wise convolutional neural network for liver and tumors segmentation," *Frontiers Genet.*, vol. 10, p. 1110, Nov. 2019.
- [30] G. Xie, K. Yang, T. Zhang, J. Wang, and J. Lai, "Balanced decoupled spatial convolution for CNNs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3419–3432, Nov. 2019.
- [31] C. Liu et al., "RB-Net: Training highly accurate and efficient binary neural networks with reshaped point-wise convolution and balanced activation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 6414–6424, Sep. 2022.
- [32] 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 Trans. Cybern.*, vol. 52, no. 11, pp. 12189–12204, Nov. 2022.
- [33] D. E. Snyder, J. Echaz, D. B. Grimes, and B. Litt, "The statistics of a practical seizure warning system," *J. Neural Eng.*, vol. 5, no. 4, pp. 392–401, Dec. 2008.
- [34] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Harvard-MIT Health Sci. Technol., Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
- [35] X. Yang, J. Zhao, Q. Sun, J. Lu, and X. Ma, "An effective dual self-attention residual network for seizure prediction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1604–1613, 2021.
- [36] T. Yu et al., "High-frequency stimulation of anterior nucleus of thalamus desynchronizes epileptic network in humans," *Brain*, vol. 141, pp. 2631–2643, Jul. 2018.
- [37] D. Liang, A. Liu, Y. Gao, C. Li, R. Qian, and X. Chen, "Semi-supervised domain-adaptive seizure prediction via feature alignment and consistency regularization," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [38] J. Wang, L. Shi, W. Wang, and Z.-G. Hou, "Efficient brain decoding based on adaptive EEG channel selection and transformation," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 6, pp. 1314–1323, Dec. 2022.
- [39] A. Nagarajan, N. Robinson, and C. Guan, "Relevance-based channel selection in motor imagery brain-computer interface," *J. Neural Eng.*, vol. 20, no. 1, Feb. 2023, Art. no. 016024.
- [40] M. Wang et al., "Multidimensional feature optimization based eye blink detection under epileptiform discharges," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 905–914, 2022.
- [41] S. Tan, K. Wang, H. Shi, and B. Song, "A novel multiview predictive local adversarial network for partial transfer learning in cross-domain fault diagnostics," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [42] Y. Li, A. Zhang, J. Huang, and Z. Xu, "An approach based on transfer learning to lifetime degradation rate prediction of the dry-type transformer," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 1811–1819, Feb. 2023.