CLEP: Contrastive Learning for Epileptic Seizure Prediction Using a Spatio-Temporal-Spectral Network

Lianghui Guo, Tao Yu, Shijie Zhao, *Member, IEEE*, Xiaoli Li, Xiaofeng Liao, *Fellow, IEEE*, and Yang Li, *Senior Member, IEEE*

***Abstract*— Seizure prediction of epileptic preictal period through electroencephalogram (EEG) signals is important for clinical epilepsy diagnosis. However, recent deep learning-based methods commonly employ intra-subject training strategy and need sufficient data, which are laborious and time-consuming for a practical system and pose a great challenge for seizure predicting. Besides, multi-domain characterizations, including spatio-temporal-spectral dependencies in an epileptic brain are generally neglected or not considered simultaneously in current approaches, and this insufficiency commonly leads to suboptimal seizure prediction performance. To tackle the above issues, in this paper, we propose Contrastive Learning for Epileptic seizure Prediction (CLEP) using a Spatio-Temporal-Spectral Network (STS-Net). Specifically, the CLEP learns intrinsic epileptic EEG patterns across subjects by contrastive learning. The STS-Net extracts multi-scale temporal and spectral representations under different rhythms from raw EEG signals. Then, a novel triple attention layer (TAL) is employed to construct inter-dimensional interaction among multi-domain features. Moreover, a spatio dynamic graph convolution network (sdGCN) is proposed to dynamically model the spatial relationships between electrodes and aggregate spatial information. The proposed CLEP-STS-Net achieves a sensitivity of 96.7% and a false prediction rate of 0.072/h on the CHB-MIT scalp EEG database. We also validate the proposed method on clinical intracranial EEG (iEEG) database from our Xuanwu Hospital of Capital Medical University, and the predicting system yielded a sensitivity of 95%, a false prediction rate of 0.087/h. The experimental results outperform the state-of-the-art studies which validate the efficacy of our method. Our code is available at https://github.com/LianghuiGuo/CLEP-STS-Net.**

***Index Terms*— EEG, contrastive learning, spatio-temporal-spectral dependencies, dynamic graph convolution, triple attention, seizure prediction.**

I. Introduction

EPILEPSY is one of the most common brain functional diseases, which is caused by sudden neurological disorders [1] and affects more than 40 million people in the world [2]. It is estimated that approximately 25% of patients have no suitable treatments to alleviate seizure symptoms [3], and the epilepsy surgery remains challenging because of multiple seizure foci, poor localization of seizure focus [4]. Especially, patients whose seizures arisen from important brain functional regions are not candidates for resection surgery, since it may cause the decline of brain functions such as motor and language [5]. Therefore, the study of epileptic seizure predicting, which can provide early warning of the incoming seizure, is becoming increasingly important. The majority of previous seizure prediction methods made the assumption that epileptic EEG signals can be divided into four consecutive brain activity states: interictal, preictal, ictal and postictal [2]. The goal of a seizure prediction method is to accurately classify interictal and preictal states and warn the patients before the seizure onset.

Recently many deep learning-based approaches are studied for automatic EEG seizure prediction. Most of these methods have been proposed for EEG feature extracting in time domain [2], [6], [7], [8], frequency domain [9], [10], [11], [12] and spatial domain [13], [14], [15], [16]. However, most studies mainly focused on intra-subject EEG pattern learning, which require a collection of sufficient data within one subject. For instance, convolution neural network (CNN) was commonly applied on the wavelet transformation of EEG which learned quantitative signatures for EEG classification [6], [11]. The short-time Fourier transform (STFT) was used to capture time-frequency characteristics from EEG signals and CNN was

Manuscript received 7 March 2023; revised 12 August 2023; accepted 26 September 2023. Date of publication 5 October 2023; date of current version 11 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62325301 and Grant 62201023, in part by the Beijing Natural Science Foundation under Grant Z220017, in part by the Beijing Municipal Education Commission-Natural Science Foundation under Grant KZ202110025036, in part by the China Post-Doctoral Science Foundation under Grant 2023M730175, and in part by the Beijing United Imaging Research Institute of Intelligent Imaging Foundation under Grant CRIBJQY202103. *(Corresponding author: Yang Li.)*

Lianghui Guo is with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100083, China (e-mail: guolianghuibuaa@buaa.edu.cn).

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

Shijie Zhao is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: shijiezhao666@gmail.com).

Xiaoli Li is with the National 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 Department of Automation Sciences and Electrical Engineering and the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100083, China (e-mail: liyang@buaa.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3322275

adopted for classification [12], [17]. Ozcan et al. [2] further investigated 3D CNN for evaluating the spatio-temporal correlations in EEG classification. In addition, recurrent neural network (RNN) was also introduced in many recent studies [18]. The above approaches are time-consuming during training process which becomes a major obstacle for the clinical use of EEG-based seizure predicting system. Therefore, developing the good cross-subject generalization ability is desirable for a practical seizure prediction application especially in the cases of new patients. Therefore, the substantial inter-subject variabilities of epilepsy-related EEG activities should be considered. Recently, contrastive learning for the EEG feature extraction has found that contrastive objectives can learn better representations than the supervised learning [19]. For example, Shen et al. [20] used contrastive learning to maximize the similarity in EEG representations for cross-subject emotion recognition. Banville et al. [21] further utilized the contrastive learning in EEG-based sleep staging and pathology and outperformed the supervised learning. Therefore, we explore contrastive pretraining for seizure predicting, which learns generic representations of epileptic EEG from source subjects and can be easily adapted to a target subject.

Additionally, in order to extract supplementary information in temporal-spectral domain for the seizure prediction, some recent studies introduced attention mechanisms into the feature extraction. Specifically, Li et al. [15] used the squeeze-and-excitation network (SENet) to capture correlations between EEG channels. Similarly, an additional branch was used to generate individual dependencies among EEG channels [22]. Moreover, Yang et al. [17] adopted attentions in both spectral and channel domain, which built global dependencies on spectrums and interdependence on EEG channels. The above attention methods have been proved helpful in boosting the seizure prediction performance. However, most of them mainly focused on building the channel attention of EEG signals, where the spatial attention was not considered. We take inspiration from the Convolutional Block Attention Module (CBAM) [23], which successfully demonstrated the importance of building spatial attention along with the channel attention. However, the attention method in CBAM did not account for the cross-dimension interaction, which ignored the relation between channel dimension and spatial dimension, and thus may degrade the performance. Motivated by this, we attempt to introduce a comprehensive attention mechanism to capture the cross-dimension interaction, which can characterize both inter-channel and spatial dependencies while building inter-dimensional interaction between channel and spatial attentions.

It should be noticed that, although the CNN has shown the promising performance in EEG classification tasks [24], it can only captured the spatial information between EEG channels in a short range due to its regular operation and the local receptive field [25]. Moreover, the non-Euclidean structure of EEG electrodes cannot be fully represented by the standard convolution operation [26]. Therefore, in order to mitigate the above disadvantages of the CNN, a graph convolutional network (GCN) was investigated, which viewed the EEG signals as graph representations and captured spatio-temporal features from EEG [25]. Specifically, EEG graphs were built by associating signals' spatial and temporal properties with graph nodes and edges [27], and then fed into the GCN for feature extraction. For example, Wang et al. [28] applied the phase locking value (PLV) to capture spatial information between EEG channels. Variational Instance-adaptive graph (V-IAG) was used to characterize the dependencies among EEG channels [22]. Zhong et al. [29] adopted the differential entropy (DE) to represent temporal correlations in EEG signals and build graph nodes. However, these methods merely relied on handcrafted features to represent EEG graphs, and the priori indicators probably ignored the heterogeneities between different epileptic patients, which affected the generalization ability of the GCN. Thus, in this study, we focus on building a dynamic GCN framework which can infer a patient-specific EEG graph and extract spatio-temporal responses with the graph convolution jointly.

In summary, to deal with the above issues, in this paper, we propose Contrastive Learning for Epileptic seizure Prediction (CLEP) using a Spatio-Temporal-Spectral Network (STS-Net). Specifically, our CLEP strategy pretrained the EEG Encoder using the contrastive learning, which optimizes an EEG contrastive (EC) loss to learn generic EEG representations. Then, the proposed STS-Net serves as the EEG Encoder and includes three subnets as follows. The pyramid convolution net first captures multi-scale temporal-spectral evolutions from raw EEG signals under five rhythms. Second, the triple attention fusion net is followed, including fives parallel branches, each of which takes a certain group of temporal-spectral evolutions as inputs and fuses the features by consecutive triple attention layers (TAL). Third, the spatial embedded net is applied which embeds spatial information between electrodes into the refined feature maps through the proposed spatio dynamic graph convolution network (sdGCN). Experiment results on two epileptic EEG datasets show that our CLEP-STS-Net can predict seizures accurately and the performance is better than the state-of-the-art studies.

The main contributions of this study are summarized below:

1) A novel CLEP-STS-Net scheme is proposed for the seizure prediction, which explores multi-scale spatio-temporal-spectral features from EEG signals through our STS-Net. Besides, the contrastive learning strategy CLEP is proposed to minimize the similarity of inter-class epileptic EEG patterns and maximize the similarity of inner-class patterns across subjects, which improves the generalization ability and benefits the patient-specific seizure predicting;
2) A triple attention layer (TAL) is introduced for building inter-dimensional interaction between input feature maps, which encodes inter-channel and spatial dependencies through a triplet attention structure;
3) We propose a spatio dynamic graph convolution net (sdGCN) to better capture the preictal transitions in EEG, which embeds channel-information into temporal-spectral features under different rhythms dynamically by using the graph convolution;

II. Methodology

The proposed seizure prediction framework is shown in Fig. 1, and summarized as follows: (1) The proposed CLEP

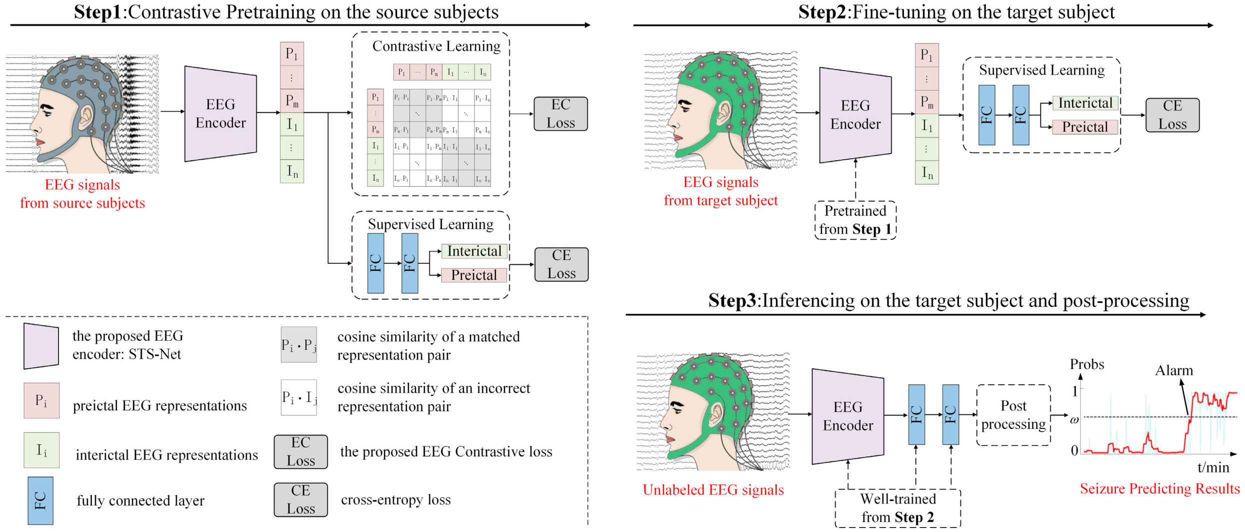
Step1:Contrastive Pretraining on the source subjects
EEG signals from source subjects
EEG Encoder
Contrastive Learning
EC Loss
Supervised Learning
FC
Interictal
Preictal
CE Loss
Step2:Fine-tuning on the target subject
EEG signals from target subject
Pretrained from Step 1
Step3:Inferencing on the target subject and post-processing
Unlabeled EEG signals
Well-trained from Step 2
Post processing
Probs
Alarm
t/min
Seizure Predicting Results
the proposed EEG encoder: STS-Net
preictal EEG representations
interictal EEG representations
fully connected layer
cosine similarity of a matched representation pair
cosine similarity of an incorrect representation pair
the proposed EEG Contrastive loss
cross-entropy loss

Fig. 1. The illustration of the proposed CLEP-STS-Net, where the proposed EEG Encoder: STS-Net can be found in Section II-B.

strategy is applied to pretrain EEG Encoder on the source subjects using the contrastive learning; (2) The proposed STS-Net contains three modular subnets, i.e., the pyramid convolution net, the triple attention fusion net, the spatio embedded net, performs as the EEG encoder and is fine-tuned on target subject; (3) The optimized CLEP-STS-Net is then transformed into a practical seizure warning system through a post-processing scheme.

A. CLEP: Contrastive Learning for Epileptic Seizure Prediction

Deep-Learning-based seizure prediction methods commonly need sufficient data and train the model in a patient-specific way, which means that extensive EEG signals are required for one patient for seizure prediction. However, it usually takes a long time and cost to prepare enough data clinically. Besides, the patient-specific model has little generalization ability due to heterogeneities among patients. Therefore, to overcome the challenges of data insufficiency and low generalization ability, the proposed CLEP performs contrastive pretraining on source subjects and fine-tunes the model on the target subject, which transfers relevant epileptic EEG patterns to help the learning task for a new patient. The proposed CLEP framework is shown in Fig. 1, which includes the following steps. First, the CLEP pretrains the EEG encoder on the source subjects using both the contrastive learning and the supervised learning. Given a batch of N EEG trial, suppose there are m preictal representations and n interictal representations in a batch. We decide that every two preictal representations form a matched pair, every two interictal representations form a matched pair. Therefore, we get $m^2 + n^2$ matched pairs and $2 \times m \times n$ incorrect pairs in each batch. In the contrastive learning, the CLEP is trained to predict which of the $N \times N$ possible pairs match and which do not match. Therefore, the CLEP trains the EEG Encoder to maximize the cosine similarity of the $m^2 + n^2$ matched representation pairs in the batch while minimizing the cosine similarity of the $2 \times m \times n$ incorrect representation pairs.

We optimize a new EEG contrastive (EC) loss over these cosine similarity scores. Given the input EEG representations $X_{eeg} = \{(X_i, y_i)|i = 1, 2, \ldots, N\}$, where $X_i \in R^{1\times T}$ is the i-th EEG representation with a feature vector length of T, N is the batch size. y_i is the corresponding label of X_i, either preictal or interictal. The cosine similarity score of two input representations X_A and X_B is given by:

$$sim(X_A, X_B) = \frac{X_A \cdot X_B}{\|X_A\|\|X_B\|} \tag{1}$$

where $X_A \cdot X_B$ is the inner product of X_A and X_B. The EC loss is calculated by the following formula:

$$L_i^{EC} = -\frac{1}{N_{y_i}} \log \frac{\sum_{j=1}^{N} \mathbb{1}_{[y_i=y_j]} \exp(sim(X_i, X_j)/\tau)}{\sum_{k=1}^{N} \exp(sim(X_i, X_k)/\tau)} \tag{2}$$

where N_{y_i} is the number of samples that have the same label as y_i in a batch. $\mathbb{1}_{[y_i=y_j]} \in \{0, 1\}$ is an indicator function which is set to 1 if $y_i = y_j$. τ is the temperature and controls the range of the softmax, which is directly optimized during training to avoid turning as a hyperparameter. By minimizing EC loss in Eq. (2), the model will increase the similarities between representations that come from the same class and reduce the similarities between preictal and interictal representations. During pretraining, the CLEP also uses the supervised learning as a supplement, which predicts whether an EEG representation is preictal or interictal, and this is implemented by the cross-entropy loss:

$$L^{CE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log (p_i) + (1 - y_i) \log (1 - p_i)] \tag{3}$$

where p_i is the predicted probability of preictal and y_i is the label. By combing the contrastive learning and the supervised learning, the CLEP is pretrained by using the hybrid loss function:

$$L = \alpha L^{EC} + (1 - \alpha) L^{CE} \tag{4}$$

where α is a hyperparameter and set to 0.5. When pretraining is done on the source subjects, the CLEP finetunes the EEG Encoder on the target subject to eliminate interdomain differences. The EEG Encoder is then trained by using the supervised learning which adapts the extracted representations to the target domain. Finally, our CLEP-STS-Net is

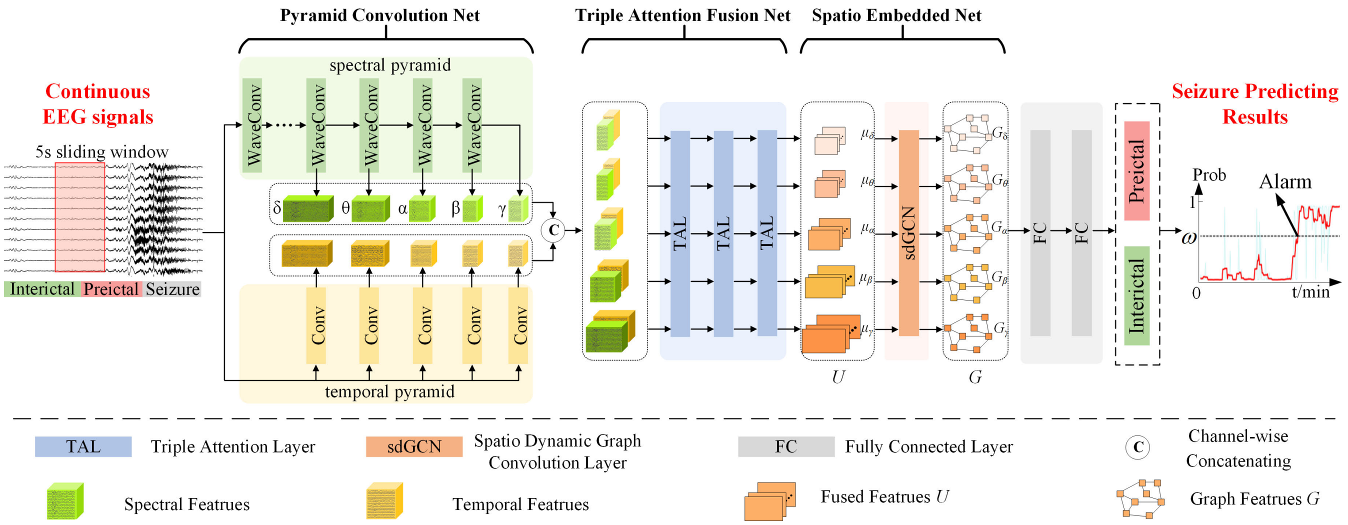

Fig. 2. The structure of the proposed EEG Encoder: Spatio-Temporal-Spectral Network (STS-Net).

well-trained to serve as a real-time seizure prediction system, which is able to perform inference on unlabeled EEG signals from target subject.

B. STS-Net: Spatio-Temporal-Spectral Network

In this subsection, we describe how the proposed EEG Encoder: STS-Net is built in details, which extracts temporal-spectral features from different rhythms while embedding dynamic spatial information to generate classification results of epileptic EEG signals. Fig. 2 shows the structure of STS-Net, which is demonstrated as follows:

C. Design of the Pyramid Convolution Net

EEG signals contain abundant temporal, spatial and spectral features, which are difficult to define manually [30], [31]. In order to capture significant temporal and spectral responses from epileptic EEG, the pyramid convolution net is implemented by two separate pyramids: spectral pyramid and temporal pyramid, extracting multi-level spectral features and multi-scale temporal features, respectively. The structure of the pyramid convolution net is shown in Fig. 2.

Since neural activities in an epileptic seizures may be of different frequencies, the spectral pyramid is designed to obtain spectral feature responses under different scales, which also correspond to the clinical frequency subbands: δ rhythm (0-4Hz), θ rhythm (4-8Hz), α rhythm (8-13Hz), β rhythm (13-30Hz), γ rhythm (30-50Hz) [32]. Specifically, the spectral pyramid net consists of hierarchical wavelet convolutions (waveConv), which implements wavelet decomposition on the input EEG sample through the convolution layer. Inspired by the Daubechies order-4 (Db4) wavelet, which is useful in spectral analysis due to its high correlation coefficients with the epileptic spikes [33], [34], Db4 is applied in the waveConv in this study. Specifically, for a given input EEG representation x at time sample t, the waveConv performs in a way analogous to the discrete wavelet transform, and is defined by:

$$x_A(t) = \sum_{r=0}^{R} x(s \times t - k) \times u(r)$$
$$x_D(t) = \sum_{r=0}^{R} x(s \times t - k) \times v(r) \tag{5}$$

where u and v denote the approximation filter and the detail filter, respectively. x_A and x_D are approximation coefficients and detail coefficients. r and s refer to the kernel size and stride, set to 8 and 2, both of which are consistent with the order of Db4 wavelet filter [35]. The approximation coefficients x_A is then fed into the next waveConv layer, and the consecutive waveConv layers perform spectral analysis iteratively through L pyramid level, where L is defined by the signal sampling rate f_s: $L = \lfloor \log_2(f_s) \rfloor - 3$, and $\lfloor\cdot\rfloor$ is the rounding-down operation [30]. Through the above hierarchical waveConv layers, the frequency boundaries of x_A and x_D of the l-th pyramid level are $(0,\ f_s/2^{l+1})$ and $(f_s/2^{(l+1)},\ f_s/2^{l})$, respectively, where l=1, 2, ..., L. Moreover, since we set the strider to 2 in each layer, we get spectral feature map of shape $((E+1)\ @\ C\times\ T/2^{l})$ from the l-th pyramid level. As a result, a set of pyramid spectral features under five standard physiological subbands are captured. Note that the waveConv involves no learnable parameters, whose weights are preloaded from Db4 wavelet filter.

The pyramid spectral analysis is essential for EEG feature extraction [36], the previous studies was also proved to be helpful by using temporal patterns for EEG classification [37], a temporal pyramid net is thus implemented through several parallel temporal convolution layers. Specifically, the pyramid level is set to 5 in order to generate temporal representations consistent with 5 pyramid spectral features [35]. The kernel sizes and strides are empirically set to $\{k/8,\ k/4,\ k/2,\ k,\ k\}$, where $k = 2^{L}$, to get pyramid temporal features with the consistent sizes as pyramid spectral features above. Note that each temporal convolution layer is followed by batch normalization and exponential linear unit (ELU). Next, in order to combine the above spectral and temporal analysis, the pyramid spectral features and temporal features are concatenated channel-wisely, resulting in five groups of temporal-spectral features.

D. Design of the Triple Attention Fusion Net

Although abundant feature extraction is proved helpful in EEG classification, inappropriate fusion method may involve redundant information and produce a poor performance.

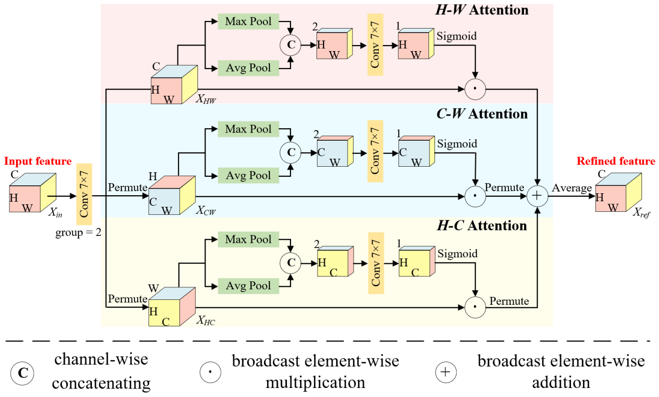

Fig. 3. The architecture of the triple attention layer (TAL).

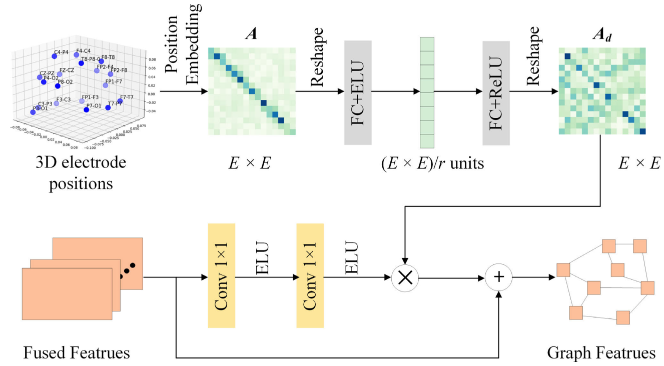

Fig. 4. A schematic illustration of the sdGCN.

Therefore, the triple attention fusion net is employed to emphasize the most discriminative representations by the proposed triple attention layer (TAL), which is shown in Fig. 2. Unlike SENet [38] and CBAM [23] which required a certain amount of learnable parameters, the goal of TAL is to model channel attention and spatial attention cheaply but effectively while not involving dimension reduction.

The channel attention in SENet focuses the model to learn more on certain channels, while the CBAM introduced a spatial attention as a complementary module telling the model which channel should be emphasized. However, the channel attention and the spatial attention are considered separately, and the relation between channel dimension and spatial dimension is ignored. Motivated by the above attention mechanism, for the input tensor $X_{in} \in R^{C\times H\times W}$, a cross-dimension attention is considered, named triple attention layer, which uses three branches to capture dependencies between (H, W), (C, W) and (H, C) dimensions of the input tensor, respectively.

The structure of the proposed TAL is given in Fig. 3. Considering the existing of heterogeneity between multi-domain feature maps, the TAL first applies group convolution on the input temporal-spectral features $X_{in} \in R^{C\times H\times W}$. The parameter *group* is set to 2, which reduces the computational cost [39] and the alleviates aliasing effect as well [40]. Next, the feature maps are passed into three branches respectively, and the first branch is designed to capture interaction between (H, W) dimensions and performed on $X_{HW} \in R^{C\times H\times W}$. At the beginning, the channel-wise max-pooling and average-pooling are applied to reduce the channel dimension and then the pooled features are concatenated. This shrinks the feature maps to make further computation lightweight and results in a rich representation X_{pool}, which is defined by:

$$X_{pool} = MaxPool(X_{HW}) \copyright AvgPool(X_{HW}) \quad (6)$$

where $X_{pool} \in R^{2\times H\times W}$ is the pooled feature, $MaxPool(\cdot)$ and $AvgPool(\cdot)$ refer to the channel-wise max-pooling and average-pooling, © is the channel-wise concatenating. X_{pool} is then fed into a standard convolution with kernel of 7 × 7 and a sigmoid activation layer, which provides the intermediate attention weights $X_w \in R^{1\times H\times W}$. The generated attention weights are then applied to the input X_{HW} through element-wise multiplication. The second branch acts similarly, and a permuting operation is added to permute the input tensor into $X_{CW} \in R^{H\times C\times W}$, which aims at capturing dependencies between (C, W) dimensions. Moreover, another permuting operation at the end of the second branch transforms the tensor into the original input shape. The third branch is built in the same way, and the first permuting operation outputs the tensor $X_{HC} \in R^{W\times H\times C}$ to build dependencies between (H, C) dimensions. The outputs of the three branches are then aggregated by averaging to generate the refined feature. In summary, to obtain the refined feature X_{ref} from an input tensor $X_{in} \in R^{C\times H\times W}$, the operation of the proposed TAL can be represented by:

$$X_{ref} = \frac{1}{3}(X_{HW} \odot X_w^1 + \overline{X_{CW} \odot X_w^2} + \overline{X_{HC} \odot X_w^3}) \quad (7)$$

where X_{HW} equals to the input tensor X_{in}, X_{CW} and X_{HC} are the permuted results from X_{in}, X_w^i $(i = 1, 2, 3)$ is the attention weighs generated from the i-th attention branch, $\odot$ is the broadcast element-wise multiplication, $\overline{}$ denotes the permuting operation.

As a result, the triple attention fusion net branches into five subnets, and the five groups of temporal-spectral features are fed into the subnets respectively. Each subnet adopts consecutive TAL operations to generate coarser representations from multi-domain and refine the feature maps through triple attention. The outputs of five subnets are then passed into an average pooling layer and the fused temporal-spectral feature $U = [\mu_\gamma, \mu_\beta, \mu_\alpha, \mu_\theta, \mu_\delta]$ is obtained, where $\mu_i \in R^{E\times F}$, and F is the length of feature vector under each rhythm.

E. Design of the Spatio Embedded Net

From the above subnets, multi-scale temporal-spectral features are captured from the EEG signals and refined, but the interdependencies between different electrodes have not been considered yet. Therefore, we build EEG graphs by using a proposed sdGCN, which explicitly explores spatial relationships between EEG channels through dynamic graph convolution. The structure of the proposed sdGCN is given in Fig. 4.

Previous studies commonly use 2D position projection of electrodes to build EEG graph [41]. However, the 2D projection of EEG electrodes was merely a rough approximation which partially represented the EEG graph but ignored precise distance measurement [26]. Considering this, we propose a position embedding method to represent the correlation between two electrodes using 3D position. First, we define a set including distance between any two EEG electrodes by: $D_{is} = \{d_{ij}|i, j \in (1, E), i \neq j\}$, where d_{ij} denotes the Euclidean distance of node i and node j, E is the number of electrodes. The distance can be acquired from the international 10-20 system. Then we define two electrodes are neighbors if the distance d_{ij} is smaller than the average value of D_{is}. Moreover, the distance between an electrode and itself is defined as the average distance of all neighboring electrodes to this electrode. In summary, we embedded 3D positions into adjacent matrix $A \in R^{E\times E}$ through the proposed position embedding methods, which is defined by:

$$A_{ij} = \begin{cases} \dfrac{1}{d_{ij}} & if\ d_{ij} < M(D_{is}) \\ 0 & if\ d_{ij} \geq M(D_{is}) \\ \dfrac{1}{M(\{d_{iq}|d_{iq} < M(D_{is})\})} & if\ i = j \end{cases} \qquad (8)$$

where $M\,(D_{is})$ is the mean value of distance set D_{is}. However, the above position embedding method is subject-independent, and the spatial information embedded in electrodes is captured in a fixed mode, which is not able to precisely model the heterogeneity among different subjects. Therefore, the proposed sdGCN further applies a self-gating method on the adjacent matrix A. Specifically, the self-gating method forms a bottleneck with two fully-connected (FC) layers to perform the squeeze-and-excitation on the adjacent matrix A, where the first FC layer applies dimensionality-reduction and the second FC layer is for dimensionality-increasing. A rectified linear unit (ReLU) is further followed to prune negative correlations in A. Namely, the combination of the FC layers and ReLU nonlinearities turns A into a learnable dynamic matrix adapted to different subjects, which dynamically model the dependencies between electrodes. The self-gating method is defined by:

$$\tilde{A}_d = \sigma(W_2\delta(W_1(\tilde{A}))) \qquad (9)$$

where $\tilde{A} \in R^{(E\times E)\times 1}$ is reshaped from A, $W_1 \in R^{((E\times E)/r)\times(E\times E)}$ and $W_2 \in R^{(E\times E)\times((E\times E)/r)}$ are weight matrixes of the first and second FC layer, r denotes the reduction ratio, whose influence will be discussed in Section IV-F. $\delta\,(\cdot)$ is the ELU activation function and $\sigma\,(\cdot)$ is the ReLU activation function, which is adopted to get a sparse graph and suppress negative values. As results, a dynamic adjacent matrix A_d is obtained by reshaping $\tilde{A}_d \in R^{(E\times E)\times 1}$ into $R^{E\times E}$. Once the connection relationship between electrodes is built, the sdGCN applies graph convolution on the temporal-spectral feature U and the dynamic adjacent matrix A_d by:

$$G^i = \delta(D^{-1}A_d\delta(\mu_i\Theta_1)\Theta_2) \qquad (10)$$

where $G^i \in R^{E\times F}$ is the dynamic EEG graph under the i-th rhythm, δ is the ELU activation function, $D^{ii} = \sum_j A_d^{ij}$ is the degree matrix of A_d, $\mu_i \in R^{E\times F}$ is the fused temporal-spectral feature under the i-th rhythm, where $i = 1, 2, \ldots, 5$. $\Theta_1 \in R^{F\times F}$ and $\Theta_2 \in R^{F\times F}$ are the weight matrices of convolution kernels in the first and second 1×1 convolution layer. Therefore, the spatial information is dynamically embedded into temporal-spectral feature under five rhythms.

TABLE I
DATA INFORMATION OF THE PUBLIC CHB-MIT DATABASE

Patient ID	Age	Gender	Number. of seizures	Number. of used seizures	Recording duration/h
1	11	F	7	7	27.4
2	11	M	3	3	30.3
3	14	F	7	6	35.0
5	7	F	5	5	24.0
6	1.5	F	10	7	38.8
7	14.5	F	3	3	43.7
8	3.5	M	5	5	10.0
9	10	F	4	4	33.6
10	3	M	7	6	38.0
11	12	F	3	3	34.8
13	3	F	12	5	25.0
14	9	F	8	6	14.0
16	7	F	10	8	17.0
17	12	F	3	3	19.0
18	18	F	6	6	31.6
20	6	F	8	8	29.6
21	13	F	4	4	29.8
22	9	F	3	3	9.0
23	6	F	7	7	37.5
Total	-	-	115	99	528.0

Finally, consecutive fully connected (FC) layers at the end of the spatio embedded net map the multi-domain feature into the seizure predicting results.

F. Post-Processing and Implementing Details

In this section, the optimized CLEP-STS-Net is translated into a practical seizure warning system through a persistent post-processing scheme [42]. First, the proposed CLEP-STS-Net generates probability series $P(i)$ from the input EEG signals, where $P(i)$ denotes the probability the input signal belongs to preictal from the i-th EEG sample. Then a moving average filter is applied on $P(i)$ to alleviate the oscillation and get the smoothed probability series $P_s(i)$ [2], [15]. The lengths of the moving average filter are set to 15s for CHB-MIT and 25s for Xuanwu dataset, which will be discussed in Section IV-H.

Next, when $P_s(i)$ exceeds a pre-defined threshold ω, a trigger $T_r(i)$ of duration τ_ω will start to warn the patient for an imminent seizure. The threshold ω for CHB-MIT and Xuanwu is set to 0.6, whose influence will be discussed in Section IV-H. τ_ω is the persistence parameter and is equal to the preictal period length [42]. For a true warning, $T_r(i)$ should start at least $\tau_{\omega 0}$ prior to the seizure onset, and remain activated until the seizure onset, otherwise it becomes a false warning. $\tau_{\omega 0}$ is the detection interval, which ensures the patient to be prepared for the incoming seizure [42]. Recent studies commonly define $\tau_{\omega 0}$ less than 1-minute [2], [15], and we set $\tau_{\omega 0}$ to 30-second in this study. At last, the seizure warning system is produced through the proposed CLEP-STS-Net.

III. EXPERIMENTAL RESULTS

A. Dataset Description

The effectiveness of the proposed CLEP-STS-Net is evaluated on two epileptic datasets described in this section.

1) CHB-MIT scalp EEG Dataset [43]: The CHB-MIT dataset consists of scalp EEG recordings from 23 pediatric patients, sampled at 256Hz from 18 common electrodes. Details about CHB-MIT dataset can be found in Table I, including

TABLE II
DATA INFORMATION OF OUR XUANWU DATABASE

Patient ID	Sampling rate/Hz	Number. of seizures	Number. of used seizures	Recording duration/h
1	1024	5	4	11.0
2	1024	3	3	7.0
3	1024	3	2	7.0
4	256	5	4	11.0
5	1024	3	3	6.0
Total	-	19	16	42.0

115 seizures in total. In this paper, patients with at least two seizures and three-hour interictal recordings are included for the seizure predicting evaluation [2]. Data within two-hour after a seizure are removed in order to eliminate effect of postictal period [15]. For the seizure prediction, we are interested in whether our proposed method can predict the leading seizure. Therefore, in the case where several seizures cluster within two-hour period, only the first seizure is used [12], [17], [42]. Finally, due to the above criterion, 99 out of 115 seizures are used in this paper.

2) Xuanwu iEEG Dataset: The Xuanwu dataset was recorded by the Xuanwu Hospital of Capital Medical University, Beijing, China. It contains multi-channel iEEG recordings from 5 patients. Totally there are 19 seizures and each patient has at least 2 seizures. The total duration of recordings is about 42 hours. The start time and end time of each seizure were labeled clearly according to expert judgments. All the experimental protocols have been approved by the Ethics Committee of Xuanwu Hospital, and informed consent was obtained from all patients participated in our study. Details about Xuanwu dataset can be found in Table II.

Deep learning-based seizure predicting methods usually crop EEG signals into clips through a sliding window ranging from 1s to 5s [2], [6], [10], [15]. In this study, for both datasets, the EEG signals are sampled into 5-second clips before fed into the proposed CLEP-STS-Net [15]. Moreover, previous studies commonly define the preictal period varying from 15 to 30 minutes [2], [15], [17]. Inspired by this, we define the 15-minute period before seizure onset as the preictal period, and define the interictal period at least 2-hour away before seizure onset and after seizure ending [12], [15].

B. Experimental Settings and Evaluation Metrics

In order to balance the interictal and preictal data, the interictal clips are randomly sampled to the same number of preictal clips [8]. Moreover, patient-specific leave-one-out cross-validation (LOOCV) is used to evaluate the performance of the proposed method. Specifically, suppose there are total N seizures for a specific patient, N-1 seizures are used for training and the left one is for testing. This procedure is repeated for N times, which ensures that all the N seizures are covered in testing. The performance for the specific patient is averaged across N times, and the overall performance is averaged across all patients. The performance of the proposed CLEP-STS-Net is evaluated by using four metrics: Area under curve (AUC), Sensitivity (S_n, the ratio of truly predicted seizures to the total number of seizures), False Predicting Rate (FPR/h), and p-value. The p-value is used to evaluate the significance of the improvement over a chance predictor [2], [42].

C. Overall Performance

The proposed CLEP-STS-Net is a more competitive method in the presence of the pyramid convolution net, the triple attention fusion net and the spatio embedded net. In this section, the patient-specific overall performance of the proposed CLEP-STS-Net is evaluated by comparing with the following baseline methods. All these methods are retested on two datasets.

1) **DCNN+Bi-LSTM [8]**: This method used deep convolutional network for EEG spatial feature extraction and applies a bidirectional LSTM to capture temporal features and perform the classification, which is a typical deep learning strategy for seizure predicting.
2) **STFT+CNN [12]**: This method first used short-time Fourier transform to generate spectrograms from EEG and CNN was adopted for further feature extraction and classification, which becomes the backbone architecture of many seizure predicting model.
3) **CE-stSENet [35]**: This method introduced attention mechanism into epileptic seizure classification task, which adopted squeeze-and-excitation attention to model the dependency between EEG channels and improved the classification performance.

Note that the above baseline methods conduct experiments using 8, 13 and 19 patients from CHB-MIT dataset in their original paper, respectively. For a fair and comprehensive comparison, all these methods are retested under the same environment and use 19 patients from CHB-MIT dataset. From Table III, the proposed CLEP-STS-Net yields an average AUC of 0.918, while other baseline methods only get average AUC of 0.856, 0.886 and 0.857 respectively, which shows the robust classification ability of our proposed method. Especially, AUC values from patient 1, 23 are greater than 0.99, indicating that our method is capable of distinguishing preictal EEG signals from interictal ones. Moreover, our seizure predicting system warns total 96 out of 99 seizures, which also outperforms all other methods. Meanwhile, the average FPR/h of our method is 0.072/h, which is lower than the baseline methods. Additionally, the p-values of our predicting system are less than 0.05 for all patients, which proves the robustness of the proposed CLEP-STS-Net. The comparisons on Xuanwu dataset are given in the Table IV.

D. Influence of the Contrastive Pretraining Strategy

In order to measure how the contrastive pretraining strategy contributes to the model, we remove the CLEP and retest our STS-Net on each patient. From Table V, our CLEP-STS-Net achieves a higher S_n of 2.4% on CHB-MIT comparing to the STS-Net. Moreover, with the CLEP, the FPR/h values gets 0.055 and 0.06 lower in two datasets. These increases show that the proposed contrastive learning can effectively benefits the seizure predicting model. In addition, Fig. 5(a) and Fig. 5(b) shows the comparing results on two datasets. We can see that AUC of all patients are boosted with the

TABLE III
THE OVERALL COMPARISON OF THE PERFORMANCE ON CHB-MIT DATABASE

Patient	DCNN+Bi-LSTM [7]				STFT+CNN [12]				CE-stSENet [35]				**Our CLEP-STS-Net**			
ID	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	0.994	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	0.419	66.67	0.000	<0.001	0.799	66.67	0.000	<0.001	0.836	100.0	0.000	<0.001
3	0.838	100.0	0.000	<0.001	0.940	83.33	0.000	<0.001	0.892	83.33	0.070	<0.001	0.962	100.0	0.140	<0.001
5	0.845	100.0	0.000	<0.001	0.850	80.00	1.505	<0.001	0.824	100.0	0.052	<0.001	0.925	100.0	0.000	<0.001
6	0.824	85.7	5.616	<0.001	0.946	85.71	2.150	<0.001	0.832	100.0	0.274	<0.001	0.899	100.0	0.000	<0.001
7	0.712	33.3	0.000	0.004	0.679	33.33	0.000	0.006	0.743	66.67	0.000	0.006	0.653	66.7	0.000	0.005
8	0.930	80.00	0.000	0.001	0.932	80.00	0.000	0.003	0.927	100.0	0.000	0.003	0.962	100.0	0.000	0.001
9	0.919	100.0	0.000	0.002	0.869	25.00	0.000	0.001	0.712	50.00	0.000	<0.001	0.934	100.0	0.047	<0.001
10	0.929	100.0	0.000	<0.001	0.889	83.33	10.19	<0.001	0.961	83.33	0.000	<0.001	0.988	100.0	0.000	<0.001
11	0.938	100.0	0.060	<0.001	0.988	100.0	0.005	<0.001	0.897	66.67	0.000	<0.001	0.969	100.0	0.272	<0.001
13	0.847	80.00	1.084	<0.001	0.995	100.0	0.000	<0.001	0.943	100.0	0.000	<0.001	0.996	100.0	0.346	<0.001
14	0.707	83.3	0.105	<0.001	0.832	83.33	2.440	<0.001	0.708	66.67	0.210	<0.001	0.797	83.3	0.000	<0.001
16	0.926	100.0	0.000	<0.001	0.984	100.0	1.427	<0.001	0.916	100.0	0.000	<0.001	0.988	100.0	0.000	<0.001
17	0.892	100.0	0.059	0.001	0.905	100.0	5.652	0.001	0.882	100.0	0.059	0.001	0.910	100.0	0.476	0.001
18	0.833	100.0	0.000	<0.001	0.807	66.67	0.000	0.001	0.817	83.33	0.036	<0.001	0.945	100.0	0.000	<0.001
20	0.994	100.0	0.000	<0.001	0.989	75.00	0.000	<0.001	0.982	100.0	0.000	<0.001	0.963	87.5	0.091	0.013
21	0.765	75.0	3.438	<0.001	0.933	100.0	0.116	<0.001	0.778	100.0	2.181	<0.001	0.854	100.0	0.000	<0.001
22	0.916	100.0	0.884	0.015	0.877	100.0	0.008	<0.001	0.679	66.67	4.133	0.016	0.858	100.0	0.000	0.015
23	0.943	85.7	0.000	<0.001	0.999	100.0	0.000	<0.001	0.998	100.0	0.000	<0.001	0.999	100.0	0.000	<0.001
Aver	0.865	90.7	0.592	-	0.886	82.2	1.237	-	0.857	86.0	0.369	-	**0.918**	**96.7**	**0.072**	-

TABLE IV
THE OVERALL COMPARISON OF THE PERFORMANCE ON OUR XUANWU DATABASE

Patient	DCNN+Bi-LSTM [7]				STFT+CNN [12]				CE-stSENet [35]				**Our CLEP-STS-Net**			
ID	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	0.869	75.00	0.000	0.029	0.819	75.0	1.780	0.003	0.998	75.0	0.000	0.001
2	0.982	100.0	0.000	0.013	0.786	100.0	2.275	0.021	0.817	66.67	0.000	0.003	0.997	100.0	0.000	0.019
3	0.984	100.0	0.329	0.015	0.990	100.0	0.000	0.096	0.984	100.0	0.000	0.359	0.987	100.0	0.000	0.014
4	0.996	100.0	0.000	0.004	0.950	75.0	2.179	0.012	0.940	100.0	0.733	0.004	0.989	100.0	0.434	0.003
5	0.995	100.0	0.000	0.018	0.980	100.0	0.164	0.062	0.999	100.0	0.000	0.019	0.999	100.0	0.000	0.018
Aver	0.924	90.0	0.264	-	0.915	90.0	0.924	-	0.912	88.3	0.503	-	**0.994**	**95.0**	**0.087**	**-**

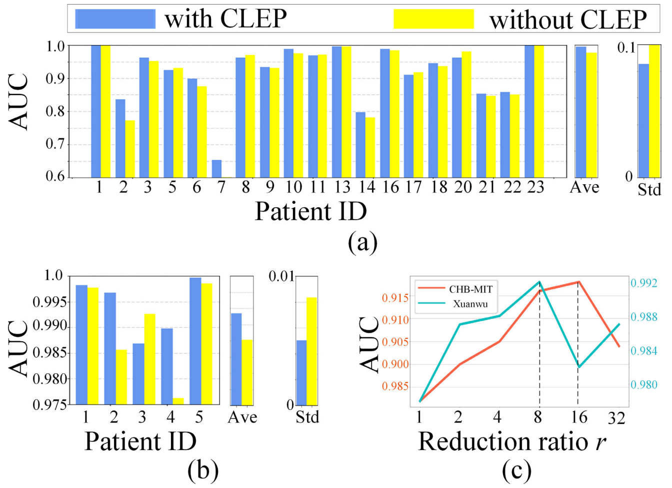

Fig. 5. (a) Comparison with and without the CLEP on CHB-MIT; (b) Comparison with and without the CLEP on our Xuanwu dataset; (c) Comparison between the different reduction ratio r.

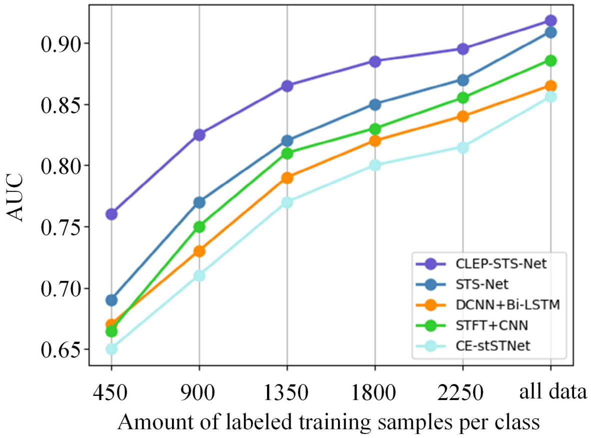

Fig. 6. Performance comparison with baseline methods under different amount of training samples per class in the fine-tuning step on CHB-MIT dataset. “all data” means the model is trained using all available data.

CLEP involved, which shows the effectiveness of the CLEP in learning general representations for the seizure prediction. Especially, benefitting from the contrastive learning, a maximum AUC increasement of 0.01 is reached on the patient 2 from CHB-MIT, and this indicates that the learned representations are not only invariant to subjects but also generalizable to different epileptic patterns. Besides, the decrease in standard deviation (Std) among all patients indicate that the CLEP can combat the heterogeneity between different patients.

In addition, we illustrate that our CLEP facilitates the training process of seizure prediction. In Fig. 6, we compare the performance of our CLEP-STS-Net with STS-Net and three baseline methods using different amount of the fine-tuning data. While it is intuitive to expect model with less data to underperform model with more data, we instead find that our CLEP-STS-Net fine-tuned with 450 samples per class almost matches the performance of the STS-Net fine-tuned with 900 samples per class, and outperforms all other methods with 900 samples per class. This is likely due to the important pretrained representations generated from the CLEP.

TABLE V
ABLATION STUDIES ON TWO DATASETS

Dataset	Methods	AUC	S_n(%)	FPR/h
CHB-MIT	without Spectral Pyramid	0.912	93.5	0.470
	without Temporal Pyramid	0.907	93.9	0.272
	without CLEP	0.909	94.3	0.127
	without sdGCN	0.883	95.4	0.078
	without TAL	0.914	96.7	0.116
	Our CLEP-STS-Net	**0.918**	**96.7**	**0.072**
Xuanwu	without Spectral Pyramid	0.980	83.3	0.491
	without Temporal Pyramid	0.993	93.8	0.688
	without CLEP	0.990	95.0	0.147
	without sdGCN	0.979	95.0	0.246
	without TAL	0.989	95.0	0.098
	Our CLEP-STS-Net	**0.994**	**95.0**	**0.087**

The contrastive pretraining strategy allows for EEG features to be general which facilitates the fine-tuning step. By contrast, normal supervised learning must train from scratch which do not utilize any pretrained knowledge [44]. Besides, from Fig. 6, we can observe that, benefited by the contrastive pretraining, our CLEP-STS-Net fine-tuned with 2250 samples per class matches the performance of the STS-Net trained with all data, which reduces the amount of data required for training to achieve the same seizure predicting performance. Specifically, for CHB-MIT dataset, each patient has an average of 3800 samples for training in this paper. Therefore, an average amount of 2250 samples are required on each patient for the fine-tuning of CLEP-STS-Net, which is able to match the performance of model without the contrastive pretraining, and outperforms three baseline methods.

E. Impact of the Pyramid Convolution Net

In this subsection, we evaluate the effectiveness of the pyramid convolution net by comparing our CLEP-STS-Net with two simplified models: (1) model without spectral pyramid net; (2) model without temporal pyramid net. The seizure predicting performances on both datasets are given in Table V with average AUC, Sn and FPR/h. First, with the waveConv layers extracting multi-level spectral features, our CLEP-STS-Net gains higher AUC than the model without spectral pyramid on two datasets, and gains increases of 0.006 and 0.014, respectively. Also, the S_n shows increases of 3.2% and 11.7% and the FPR/h declines by 0.398 and 0.404, which further demonstrate that the waveConv layers contribute to the seizure predicting abilities. Moreover, comparing to the model without temporal pyramid, which only captures spectral features from EEG, our CLEP-STS-Net yields higher AUC of 0.011 and 0.001, higher S_n of 2.8% and 1.2%, and lower FPR/h of 0.2 and 0.601. These improvements show that extracting only spectral representations may omit important temporal feature responses, which also proves the efficiency of Temporal Pyramid in extracting multi-scale temporal features. In summary, we can learn that the proposed CLEP-STS-Net outperforms the two simplified models on two datasets, which intuitively demonstrates the effectiveness of the spectral pyramid net and temporal pyramid net.

F. The Influence of the Triple Attention Layer

Our next attempt is to adopt the triple attention fusion net which helps to alleviate the differences between the spectral features and temporal features generated from pyramid convolution. In this section, we exploit the advantage of using TAL in the training of seizure predicting model. We first evaluate the performance of our CLEP-STS-Net with or without TAL, and show the average AUC, S_n and FPR/h on two datasets in Table V. Concretely, our CLEP-STS-Net achieves a higher AUC of 0.004 and 0.005 comparing to the model without TAL. Moreover, the FPR/h values gets 0.044 and 0.011 lower in two datasets with TAL. These increases show that the proposed triple attention fusion method is able to fuse the temporal-spectral features under the preictal transition, which combats the heterogeneity between different patients and effectively benefits the seizure predicting model.

G. The Performance of the Spatio Dynamic Graph Convolution Network

In order to embed spatial epileptic activities into the temporal-spectral responses, the sdGCN is adopted to dynamically model the relationships between electrodes and aggregate spatial information. We further compare the classification performance between our CLEP-STS-Net and a simplifier one without sdGCN, and Table V shows the results. We can see that with sdGCN, AUC and S_n are 0.035 and 1.3% higher on CHB-MIT, and AUC also gets improvements of 0.015 on Xuanwu dataset. Moreover, the FPR/h declines by 0.06 and 0.159 on two datasets. These improvements show that the proposed sdGCN algorithm actually contributes to the seizure predicting, and this is probably due to its learnable transformation which turns the static electrode position into dynamic spatial graph. Next, the reduction ratio r has influence on the sparsity of the dynamic adjacent matrix A_d, which allows us to vary the capacity and computational cost of the sdGCN. The best hyperparameters are subjective and vary for each dataset, since the amount of data and the number of patients are different. To investigate this influence, we set r with the range from 1 to 32 to evaluate the classification performance. From Fig. 5(c), we can observe that with $r = 8$, it reached the best performance on the Xuanwu datasets; with $r = 16$, it achieved the highest AUC for CHB-MIT dataset.

We further investigate how the dynamic adjacent matrix A_d changes during training process, and Fig. 7 shows the transitions from the original adjacent matrix A to the final A_d in the training process. We can observe that the zero value in A are replaced by the learned values in the A_d, and this indicates that certain spatial correlations are built between the corresponding electrode pairs. Moreover, the original A is transformed from a symmetric matrix into an asymmetric directed matrix. From the perspective of the causal interaction, the direction of the information flow between brain regions can reveal more information about brain interactions [45]. Therefore, this directed graph provides more precise information than simpler undirected graphs [46], which makes the sdGCN to learn more diverse information from the electrode position embeddings. Fig. 7 also presents the visualization of A_d on scalp topologies, which clearly shows how the electrode correlation is built during the training of the sdGCN. In addition, we can see that our sdGCN learns a patient-specific A_d. If we use the static position embedding method which all patients share the same adjacent matrix, the het-

TABLE VI
THE COMPARISON OF MODEL COMPACTNESS

Methods	S_n(%)	FPR/h	Parameter ($\times 10^6$)	Inference time (μs)
DCNN+Bi-LSTM [7]	90.7	0.592	0.348	842.6
STFT+CNN [12]	82.2	1.237	0.115	948.0
CE-stSENet [35]	86.0	0.369	0.290	182.2
Our CLEP-STS-Net	**96.7**	**0.072**	**0.285**	**689.1**

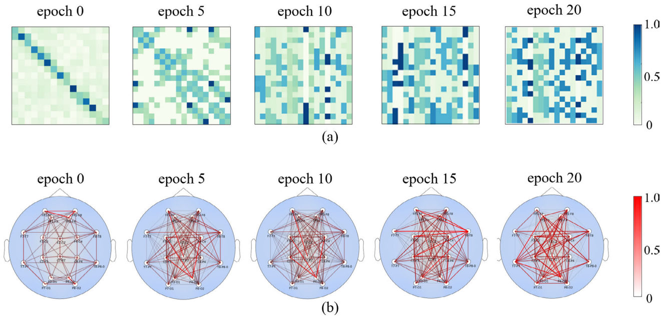

Fig. 7. Transitions of the dynamic adjacent matrix A_d during the training process of the patient 8 from CHB-MIT. (a) the values of A_d; (b) the 2D visualization of A_d on scalp topologies.

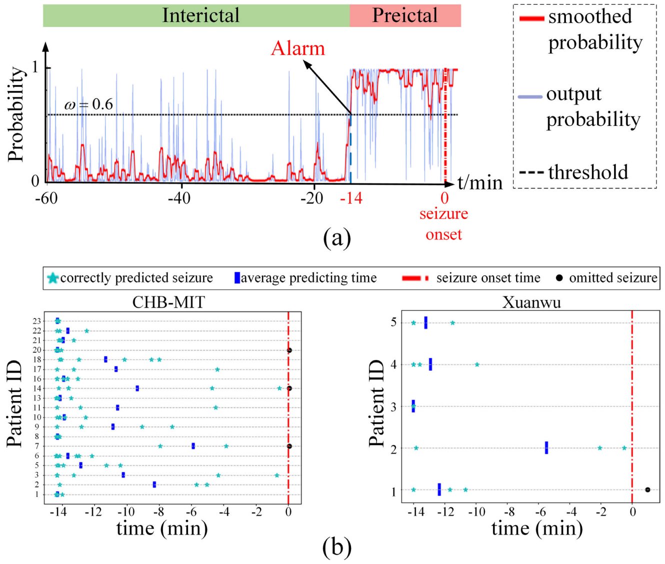

Fig. 8. (a) The seizure prediction results on one seizure from the typical patient 1 of CHB-MIT, where the results are shown one-hour before the seizure onset; (b) patient-specific seizure prediction time.

erogeneity of individuals is ignored and not precise enough for a patient-specific seizure prediction method. In summary, compared to the static position embedding method, our sdGCN learns the dynamic correlations among different EEG channels and embeds the spatial relationships into feature maps, which boosts the classification performance.

IV. DISCUSSIONS

A. Efficacy of Model Compactness

In order to evaluate the compactness of our CLEP-STS-Net, we compare the number of parameters with baseline methods in Table VI. The proposed CLEP-STS-Net involves 2.85×10^5 parameters which is less complex than the DCNN+BiLSTM and similar with the previous CE-stSENet. Although the network in STFT+CNN contains less parameters, it is not an end-to-end seizure predicting method and additional computation cost is spent in building the spectrums before the network. In addition, we evaluate the inference time which starts from the input of EEG sample and ends with output probability. We perform all experiments on Pytorch framework with Intel Core i7-4790 3.60GHz CPU and the NVIDIA V-100 GPU with 32GB. Although our proposed CLEP-STS-Net method takes longer time than the STFT+CNN methods, our inference time (689.1μs) is much less than the length of an input EEG sample (5s), which is suitable enough for real-time seizure prediction tasks.

B. Performance Comparison of Different Methods

Table VII lists the state-of-the-art methods in seizure prediction on CHB-MIT. It is difficult to decide which is the best approach since each method used a limited set of selected data according to the pre-defined preictal and interictal interval. For example, Truong et al. [12] and Yang et al. [17] combined the STFT with CNN and tested their approaches on 13 patients, which resulted in suboptimal performance compared with our proposed CLEP-STS-Net. This is probably due to the lack of temporal and spatial EEG pattern in their method. In comparison with Ozcan et al. [2] and Zhang et al. [6] which extract features from the frequency and time domain, our proposed method considers the spatio-temporal-spectral representation of EEG and yields 9.7%, 4.71% higher in *Sn* and 0.024, 0.048 lower in FPR/h. Also, compared with Gao et al. [13] which used a dilated CNN for spatial pattern extraction, our CLEP-STS-Net applies sdGCN to build patient-specific EEG graphs and embeds spatial information into the temporal-spectral feature maps, and gains a higher *Sn* of 3.41%. Moreover, compared with our LOOCV validation scheme which does not break the continuity of signals when testing, 10-fold CV in [11] and [14] shuffles EEG signals and ignores the continuous variation inside seizures. As results, our CLEP-STS-Net achieves more promising *Sn* and FPR/h against most of the recent studies.

C. Analysis of Seizure Predicting Time

To further evaluate the ability of predicting in time, the seizure prediction time of each patient is shown in Fig. 8. We use the filter length of 15 for CHB-MIT and 25 for Xuanwu, and threshold of 0.6 for both datasets, which are optimized in Section IV-H. Fig. 8(a) is an example of the prediction generated by our method on one seizure of the patient 1 from CHB-MIT. Fig. 8(b) shows the patient-specific seizure predicting time on two datasets. The proposed CLEP-STS-Net yields average prediction time of 12.62 min on CHB-MIT and 11.45 min on Xuanwu, respectively. Especially, for the patient 23 from CHB-MIT, our proposed method advances seizure onset with an average prediction time of 14.89 min, which is early enough for the patient to get prepared for the incoming seizure.

D. Visualization Interpretation of Our CLEP-STS-Net

The proposed CLEP-STS-Net can produce robust seizure prediction results using the epileptic EEG signals. However,

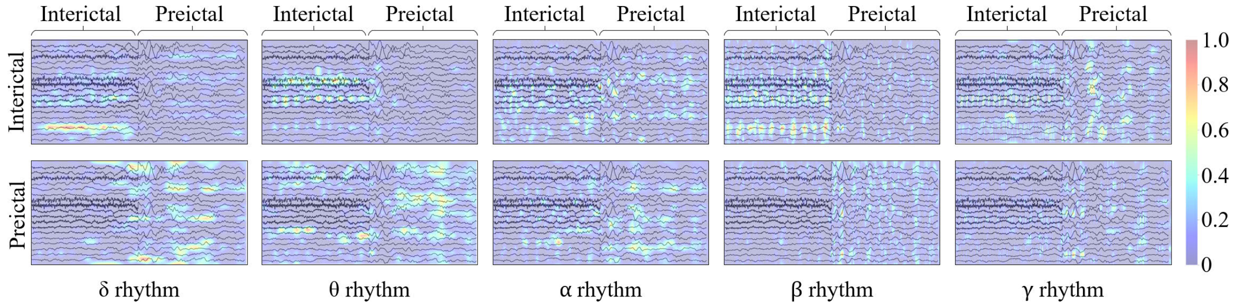

Fig. 9. Grad-CAM visualization results based on five outputs of the pyramid convolution net, where the EEG data is from the typical patient 1 of CHB-MIT, the brighter parts are the regions which the model pays more attention to the corresponding class, the top row is the results for interictal class, and the bottom row is the results for preictal class, respectively.

TABLE VII

Experimental Settings and the 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 intervels (min)	S_n(%)	FPR/h
Khan et al. [11]	2018	Wavelet transform+CNN	15	18	10-fold CV	10-10	87.80	0.147
Truong et al. [12]	2018	STFT+CNN	13	64	LOOCV	240-30	81.20	0.160
Ozcan et al. [2]	2019	Spectral power, statistical moments, Hjorth+3D CNN	16	77	LOOCV	60-60 120-60 240-60	86.80 87.01 85.71	0.292 0.186 0.096
Zhang et al. [6]	2020	CSP+CNN	23	156	LOOCV	NR-30	92.00	0.120
Yang et al. [17]	2021	STFT+RDANet	13	64	LOOCV	240-30	89.25	0.122
Gao et al. [13]	2022	Dilated CNN	16	85	LOOCV	60-30	93.30	0.007
Dissanayake et al. [14]	2022	Geometric Deep Learning	23	NR	10-fold CV	NR-60	95.94	NR
This work	**2023**	**Our CLEP-STS-Net**	**19**	**99**	**LOOCV**	**120-15**	**96.71**	**0.072**

it is hard for human to distinguish between interictal EEG and preictal EEG, so we wonder how the model is able to classify the EEG samples. Therefore, we adopt the Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize how our CLEP-STS-Net learns from the interictal and preictal EEG signals. We concatenate one interictal sample and one preictal sample as input and use the five outputs of the pyramid convolution net for the visualization, where each represents a certain rhythm. Fig. 9 shows the activation maps generated by Grad-CAM, and we can see that when the input EEG is labeled with interictal class, the activation maps are brighter in the interictal period. Also, when the input is labeled with preictal class, the model focuses mainly on the preictal period. The different regions of interests in temporal period indicates that our CLEP-STS-Net is sensitive to the temporal transitions of epileptic EEG. In addition, from Fig. 9, we can see that this difference is more distinct in δ and θ rhythms, where we can observe a clear transition of the f region of interests from interictal class to preictal class. This indicates that the preictal EEG tends to activate our CLEP-STS-Net especially in δ and θ rhythms, which is in line with the previous studies that the propagation of seizure causes a shifting from higher rhythm activities in a focal region to slower rhythms across widespread areas [15].

V. Conclusion

In this paper, a novel epileptic seizure prediction system is built by using the proposed CLEP-STS-Net. Specifically, our STS-Net first extracts multi-scale temporal-spectral features under different rhythms through the pyramid convolution net. Meanwhile, an attention mechanism called TAL is adopted to construct inter-dimensional dependencies among feature maps and effectively fused the temporal-spectral features. Then, the proposed sdGCN is applied to dynamically construct the spatial correlations between EEG electrodes. Finally, the contrastive learning strategy CLEP learns the intrinsic epileptic patterns from source subjects and improves the generalization ability. Seizure prediction performance is evaluated on multiple patients with both scalp EEG and iEEG signals. Our proposed CLEP-STS-Net yields promising results in AUC, S_n and FPR/h, which outperform all the compared baseline methods. Moreover, we evaluate the effectiveness of the CLEP, pyramid convolution net, TAL and sdGCN through ablation studies. Additionally, the visualization investigation shows that our proposed method is able to extract spatio-temporal-spectral features related to different rhythms from epileptic EEG signals. Experimental results demonstrate that the proposed CLEP-STS-Net can predict the incoming seizures accurately and further facilitate epileptic patients' daily life.

References

[1] L. Xie, Z. Deng, P. Xu, K.-S. Choi, and S. Wang, "Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2200–2214, Jun. 2019.

[2] 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.

[3] G. Wang et al., "Seizure prediction using directed transfer function and convolution neural network on intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2711–2720, Dec. 2020.

[4] 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.

[5] P. Jin, D. Wu, X. Li, L. Ren, and Y. Wang, "Towards precision medicine in epilepsy surgery," *Ann. Transl. Med.*, vol. 4, no. 2, p. 24, Jan. 2016.

[6] 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.

[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] C. Li et al., "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 458–467, 2021.

[10] T. Liu, N. D. Truong, A. Nikpour, L. Zhou, and O. Kavehei, "Epileptic seizure classification with symmetric and hybrid bilinear models," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2844–2851, Oct. 2020.

[11] 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.

[12] 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.

[13] Y. Gao et al., "Pediatric seizure prediction in scalp EEG using a multiscale neural network with dilated convolutions," *IEEE J. Transl. Eng. Health Med.*, vol. 10, 2022, Art. no. 4900209.

[14] 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.

[15] 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.

[16] J.-S. Bang, M.-H. Lee, S. Fazli, C. Guan, and S.-W. Lee, "Spatio-spectral feature representation for motor imagery classification using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 3038–3049, Jul. 2022.

[17] 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.

[18] L. S. Vidyaratne, M. Alam, A. M. Glandon, A. Shabalina, C. Tennant, and K. M. Iftekharuddin, "Deep cellular recurrent network for efficient analysis of time-series data with spatial information," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6215–6225, Nov. 2022.

[19] Y. Zhao, C. Li, X. Liu, R. Qian, R. Song, and X. Chen, "Patient-specific seizure prediction via adder network and supervised contrastive learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1536–1547, 2022.

[20] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2496–2511, Jul. 2023, doi: 10.1109/TAFFC.2022.3164516.

[21] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, "Uncovering the structure of clinical EEG signals with self-supervised learning," *J. Neural Eng.*, vol. 18, no. 4, Mar. 2021, Art. no. 046020.

[22] T. Song et al., "Variational instance-adaptive graph for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 343–356, Jan. 2023.

[23] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2018, pp. 3–19.

[24] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.

[25] 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.

[26] D. Zhang, L. Yao, K. Chen, S. Wang, P. D. Haghighi, 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with C-LSTM," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016041.

[32] 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.

[33] K. P. Indiradevi, E. Elias, P. S. Sathidevi, S. D. Nayak, and K. Radhakrishnan, "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram," *Comput. Biol. Med.*, vol. 38, no. 7, pp. 805–816, Jul. 2008.

[34] L. Wang et al., "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, p. 222, May 2017.

[35] 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.

[36] 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.

[37] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.

[38] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[39] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[40] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[41] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, Jul. 2020.

[42] D. E. Snyder, J. Echauz, 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.

[43] 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.

[44] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 8748–8763.

[45] M. Lobier, F. Siebenhühner, S. Palva, and J. M. Palva, "Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions," *NeuroImage*, vol. 85, pp. 853–872, Jan. 2014.

[46] F. Hasanzadeh, M. Mohebbi, and R. Rostami, "Graph theory analysis of directed functional brain networks in major depressive disorder based on EEG signal," *J. Neural Eng.*, vol. 17, no. 2, Mar. 2020, Art. no. 026010.