

# BrainNet: Epileptic Wave Detection from SEEG with Hierarchical Graph Diffusion Learning

Junru Chen\*  
Zhejiang University  
jrchen\_cali@zju.edu.cn

Yang Yang\*†  
Zhejiang University  
yangya@zju.edu.cn

Tao Yu  
Zhejiang University  
yutaoyt@zju.edu.cn

Yingying Fan  
Zhejiang University  
yingyingfan@outlook.com

Xiaolong Mo  
Neuroechos Medical  
Xiaolong.mo@neurox.cn

Carl Yang  
Emory University  
j.carlyang@emory.edu

## ABSTRACT

Epilepsy is one of the most serious neurological diseases, affecting 1-2% of the world's population. The diagnosis of epilepsy depends heavily on the recognition of *epileptic waves*, *i.e.*, disordered electrical brainwave activity in the patient's brain. Existing works have begun to employ machine learning models to detect epileptic waves via *cortical electroencephalogram* (EEG), which refers to brain data obtained from a noninvasive examination performed on the patient's scalp surface to record electrical activity in the brain. However, the recently developed *stereoelectrocorticography* (SEEG) method provides information in stereo that is more precise than conventional EEG, and has been broadly applied in clinical practice. Therefore, in this paper, we propose the first data-driven study to detect epileptic waves in a real-world SEEG dataset. While offering new opportunities, SEEG also poses several challenges. In clinical practice, epileptic wave activities are considered to propagate between different regions in the brain. These propagation paths, also known as the *epileptogenic network*, are deemed to be a key factor in the context of epilepsy surgery. However, the question of how to extract an exact epileptogenic network for each patient remains an open problem in the field of neuroscience. Moreover, the nature of epileptic waves and SEEG data inevitably leads to extremely imbalanced labels and severe noise. To address these challenges, we propose a novel model (*BrainNet*) that jointly learns the dynamic diffusion graphs and models the brain wave diffusion patterns. In addition, our model effectively aids in resisting label imbalance and severe noise by employing several self-supervised learning tasks and a hierarchical framework. By experimenting with the extensive real SEEG dataset obtained from multiple patients, we find that *BrainNet* outperforms several latest state-of-the-art baselines derived from time-series analysis.

\*Both authors contributed equally to this research.

† Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

KDD '22, August 14–18, 2022, Washington, DC, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9385-0/22/08...\$15.00

<https://doi.org/10.1145/3534678.3539178>

## CCS CONCEPTS

• Applied computing → Health care information systems.

## KEYWORDS

Epileptic wave detection; Graph diffusion; SEEG data

## ACM Reference Format:

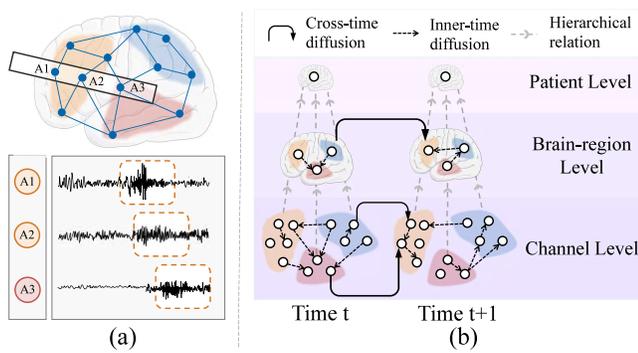
Junru Chen, Yang Yang, Tao Yu, Yingying Fan, Xiaolong Mo, and Carl Yang. 2022. BrainNet: Epileptic Wave Detection from SEEG with Hierarchical Graph Diffusion Learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, August 14–18, 2022, Washington, DC, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3534678.3539178>

## 1 INTRODUCTION

**Background.** *Epilepsy*, which is one of the most common serious neurological diseases, is characterized by abnormal neurophysiological activity, leading to epileptic seizures or abnormal behavior. Today, in 2021, there are more than 65 million epilepsy patients globally, approximately one-third of whom are medication-resistant [26]. In other words, medication will not be effective for these patients, and surgical removal of the area of the brain involved in seizures is considered the only effective treatment.

To measure the *seizure onset zone* (SOZ) or so-called *epileptogenic foci*, and guide epilepsy surgery, it is necessary to record the electrical activity of the patient's brain. There are two main types of such electrophysiological monitoring methods: EEG and SEEG. The former is noninvasive, while the latter is invasive (*i.e.*, requires electrodes to be inserted into the brain), and thus contains more stereo information. For example, when the SOZ is located in the deeper structures of the brain (such as the hippocampus or the insula), or when the *laterality of seizures* [21] is unknown, non-invasive testing will fail to pinpoint the exact seizure focus, in which case the SEEG approach is necessary.

**Problem.** To facilitate the development of epilepsy treatment, we collect a real-world SEEG dataset, which is made up of high-frequency and multi-channel SEEG signals obtained from multiple epilepsy patients (each patient has a 77GB record of 53 hours on average) in a specific first-class hospital. Based on the dataset, we further propose to automatically detect epileptic waves. Besides, the fact that epileptic wave activities propagate among different brain regions in clinical practice inspires us to further study the



**Figure 1: Illustration of epileptic wave detection.** (a) An electrode with three channels (A1, A2, and A3) is inserted into the brain to collect SEEG signals across two regions of a patient’s brain. (b) Our solution is to jointly learn the process by which epileptic seizures diffuse and detect epileptic waves, which are marked by yellow squares among the SEEG signals in (a), on multiple levels.

underlying *epileptogenic network* [37], which characterizes the relations among the brain regions involved in the production and propagation of epileptic activities. It is deemed to be a key factor in the context of epilepsy surgery, but how to extract an exact epileptogenic network for each patient is still an open problem in neuroscience [3].

By the example illustrated in Figure 1, we further introduce the details of our problem. The top part of Figure 1(a) shows a human brain into which an electrode with three channels has been inserted from the upper left corner. Notably, in real clinical diagnosis, doctors will insert multiple electrodes (each with multiple channels) into the suspected epileptogenic areas of the brain, which vary across different patients and may span multiple brain regions. After inserting the electrode, the doctor can collect and monitor the patient’s SEEG data; as the bottom part of Figure 1(a) shows, this can be regarded as a contiguous multi-channel time series.

Given the SEEG data, our target is to formulate an automatic data-driven method to pinpoint the time and location at which epileptic waves appear (marked by yellow squares in Figure 1(a)). It is worth mentioning that, in this work, we focus on epilepsy detection for individual patients. More specifically, given a particular patient, we aim to train a model based on his/her historical SEEG data, then utilize the model to identify epileptic waves in the current SEEG data, following existing works with similar settings on EEG data [22]. An alternative solution would involve a model to handle all patients together. However, individual differences in epileptogenic foci result in deep subdural electrodes being utilized with different numbers and locations from person to person, which further leads to dramatic variation in the collected signals across patients. Before properly handling the signals of individual patients, it is hard to directly study the possible shared patterns and generalization of models across patients. Therefore, we focus on patient-specific models and predictions in this work as the first study on SEEG-based epilepsy detection.

**Challenges.** Compared with existing works that study epileptic waves derived from the much simpler EEG data, there are several

unique challenges for us in fully leveraging the SEEG data, which are caused by the nature of the data and the lack of understanding regarding the diffusion mechanisms of brain waves.

- 1 Capturing the true *epileptogenic network* [3]. The propagation of epileptic seizures cannot be observed directly and does not follow any known routines. Indeed, an epileptic seizure will not diffuse in a manner consistent with the anatomical brain structure [7], while its propagation paths may dramatically change over time [3]. Therefore, quantifying the dynamic diffusion graph (*i.e.*, epileptogenic network) structure is challenging.
- 2 Handling imbalanced labels. SEEG often generates extremely large amounts of data. This is partly due to its high frequency of data acquisition (mostly ranging from 256Hz to 1024Hz). Moreover, patients are monitored with electrodes for an average of 11 days, and sometimes for up to 33 days [25]. However, the epileptic seizure process lasts only for tens of seconds among days of records. Consequently, the low epileptic wave rate leads to the imbalanced label issue.
- 3 Handling data noise. Due to inherent problems like flaws in electrode artifacts, mechanical noises, and the interference of epileptic interval waves, SEEG data is severely affected by noise.

**Solution.** To address the above challenges, in this paper, we propose a novel epileptic wave detection model, referred to as *BrainNet*.

To handle the first challenge, we aim to find out the underlying epileptogenic network tracking the process of epileptic wave diffusion across time. To this end, *BrainNet* adopts graph neural networks along with a structure learning algorithm in order to both learn and quantify the epileptic wave diffusion process. More specifically, as the duration of epileptic waves may be longer or shorter, *BrainNet* learns two types of diffusion processes. Longer epileptic waves naturally result in diffusion from one time segment to the next, which is called *cross-time diffusion*, denoted by solid black lines in Figure 1(b). Meanwhile, within the same time segments of each channel, there also exists a diffusion process that occurs as the electric signals spread quickly. The dotted black lines in Figure 1(b) represent the *inner-time diffusion* described above.

For the second challenge, the *BrainNet* adopts a self-supervised learning approach to overcome extremely imbalanced labels. Specifically, we propose *bidirectional contrastive predictive coding* (BCPC) to pre-train the representation of every segment in each channel. Compared with existing self-supervised algorithms, BCPC makes *BrainNet* more capable of extracting bidirectional information by taking full advantage of the sufficient (unlabeled) SEEG data.

Finally, to make the model more robust to noise, we propose auxiliary learning tasks with a hierarchical framework. Inspired by the diagnosis process used by doctors, which usually considers information from different levels simultaneously—including channels (micro-level), brain regions (meso-level), and patients (macro-level)—to make more appropriate and accurate diagnoses, we propose to make *BrainNet* further predict whether a particular brain region or patient will be epileptic at a specific time. More specifically, a sample in the brain-region/patient level is said to be normal if none of its corresponding channels/brain-regions are epileptic. By adopting this approach, the input sensitivity of lower levels will be weakened at higher levels. Intuitively, *BrainNet* aggregates accurate information at higher levels, and in turn feeds

it back to the lower levels, reducing the likelihood of inaccurate information being accumulated there.

To sum up, by utilizing all technical designs discussed above, *BrainNet* possesses the ability to capture the dynamic diffusion process and enhance the accuracy of epileptic wave detection tasks. The main contributions can be summarized as follows:

- We are the first to formulate and study the epileptic wave detection problem using an automatic end-to-end data-driven method for SEEG data.
- We propose *BrainNet* to jointly learn the dynamic diffusion graphs and model the brain wave diffusion patterns thereon to achieve accurate epileptic wave detection under conditions of imbalanced labels and severe noise.
- We conduct extensive experiments on a large-scale real-world SEEG dataset involving multiple patients. Our results validate the effectiveness of our model on epileptic wave detection, while case studies show its superiority to capture the diffusion process.

## 2 PROBLEM DEFINITION

We take Figure 1 as an example to illustrate our studied problem. In Figure 1(a), for an individual patient, we have the set of three brain regions<sup>1</sup>  $B$ , represented by three colors. All the electrode contacts (*i.e.* blue solid nodes) consist of a set of channels  $C$ . Each channel  $c \in C$  belongs to a unique brain region  $b(c) \in B$ , where  $b(\cdot)$  is a function that maps a channel to its corresponding brain region, which is derived from domain knowledge. For example, channels  $A1$  and  $A2$  both belong to the same orange brain region, while  $A3$  is located in another red region. After locating the contactors, the machine begins to record the signals of every channel so as to collect the SEEG data—shown at the bottom of Figure 1(a)—as a multivariate time series  $T = \{X_i\}_{i=1}^{|T|} \in \mathbb{R}^{|T| \times |C|}$ , where  $X_i = \{x_{i,c}\}_{c \in C} \in \mathbb{R}^{|C|}$  represents a vector of the channel signals belonging to time point  $i$ . Moreover,  $Y = \{Y_i\}_{i=1}^{|T|}$  denotes multivariate labels, where  $Y_i = \{y_{i,c}\}_{c \in C}$  are the labels of every channel at the  $i$ -th time point. Here,  $y_{i,c} \in \{0, 1\}$  indicates whether an epileptic seizure is occurring (*i.e.*  $y_{i,c} = 1$ ) or not (*i.e.*  $y_{i,c} = 0$ ). The positive sample ratio in practice is pretty low (around 0.003).

In line with existing works on time-series analysis [2, 24, 31], we use a sliding window with length  $k$  and stride  $l$  to divide the raw time-series  $T$  into smaller segments  $S = \{S_t\}_{t=1}^{|S|}$  where  $|S| = \lfloor (|T| - k) / l \rfloor + 1$  is the number of segments. The annotation  $S_t = \{s_{t,c}\}_{c \in C} \in \mathbb{R}^{|C| \times k}$  represents the segment with  $|C|$  channels and  $s_{t,c} = \{x_{l*(t-1)+i,c}\}_{i=1}^k \in \mathbb{R}^k$  is the  $t$ -th time segment of one channel  $c$ . Using the same division strategy, we can also divide the labels  $Y$  into small segments  $Y^S = \{Y_t^S\}_{t=1}^{|S|}$ . If one segment includes a seizure point, we consider it to be a seizure segment; *i.e.*,  $y_{t,c}^S = \max_{i=1}^k \{y_{l*(t-1)+i,c}\}$ . By means of basic time-series data segmentation, we formally define the studied problem as follows:

**DEFINITION 1 (EPILEPTIC WAVE DETECTION).** *Given the historic segment set  $\hat{S}$  and corresponding label set  $Y^{\hat{S}}$  of an individual patient,*

*we aim to predict the labels  $Y^S$  of the future segment set  $S$ :*

$$P(y_{t,c}^S | S, \hat{S}, Y^{\hat{S}}), \quad t = 1, \dots, |S|, \quad c \in C,$$

*where the target is the probability of an epileptic state occurring in the  $t$ -th segment of channel  $c$  in the future segment set.*

We emphasize here that our problem is defined on an individual patient. A similar problem has been defined in [33]; however, their results were all based on EEG data, which are much simpler than SEEG data due to limited tracking of deep brain activities. To the best of our knowledge, we are the first to formally propose and analyze the epileptic detection task on SEEG data. We summarize important notations in Appendix A.1.

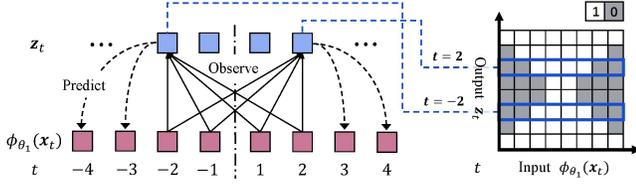
## 3 PROPOSED MODEL

### 3.1 General Description

In real-world scenarios, doctors diagnose epilepsy patients step by step: whether the patient suffers from seizures, which brain regions are suspected epileptic areas, and the particular location in the brain that is directly causing the seizures (which might be surgically removed). Inspired by this, we propose a novel framework *BrainNet*, which employs a hierarchical structure to jointly model epileptic waves and their diffusion process in three different levels, ranging from high to low: the patient, the brain-region, and the channel levels (with each channel corresponding to a particular location in the brain). While the problem under study focuses on the channel level, we find that higher levels provide a macroscopic view, which is helpful in capturing epileptic waves more precisely and improves the model's robustness to handle severe noise in SEEG data. Figure 1(b) illustrates the hierarchical structure of *BrainNet*.

Orthogonal to the three levels, *BrainNet* consists of three major components: *pre-training* for input SEEG data, *graph diffusion* over time and a *prediction* module. We begin from the channel level to explain how these components work together. Given the massive SEEG data, with low epileptic wave rate, we first pre-train the representation of every segment to fully capture and leverage the patterns of normal brain waves by performing a self-supervised learning task (Section 3.2). To capture the dependency between each segment and the propagation of brain waves ignored by the pre-training component, we design the *graph diffusion* component. Specifically, with the pre-trained representation taken as input, the *graph diffusion* component incorporates a graph neural network along with a structure learning algorithm to explore how brain waves diffuse among channels (or brain regions in the higher level), and encodes these diffusion patterns into the segment representation (Section 3.3). The obtained representation then can be directly piped to a classifier for *prediction*. Similar procedures are followed at the brain-region and patient levels, where we use an aggregation strategy to obtain the brain-region/patient level representations from the channel/brain-region level (Section 3.4). In summary, each component of the model is designed to capture every segment's own characteristics and its property of dynamic diffusion on the three different levels. In the remainder of this section, we introduce the details of each component.

<sup>1</sup>We define brain regions according to the automated anatomical labeling (AAL) [35], which is a digital atlas of the human brain and defines 116 different regions in total.



**Figure 2: Schematic diagram of BCPC: the left represents the pretext task; the right is the corresponding mask matrix.**

### 3.2 Model Pre-training

Within the massive amount of SEEG records obtained from a patient lasting for days, there are often only seconds of epileptic waves. To handle this issue and fully leverage the massive (unlabeled) data, we propose a self-supervised learning (SSL) algorithm, Bidirectional Contrastive Predictive Coding (BCPC), for pre-training the representation of SEEG data.

For a self-supervised algorithm, the key to learning discriminative representations is building an effective *pretext task* [18]. By following Contrastive Predictive Coding (CPC) [27], the classical SSL model for time series, we set the pretext task as to predict the  $P$  local features obtained from the encoder on the head and tail of the segment, with global contextual features acquired from the autoregression model based on the local features from the middle part. Different from CPC, we believe a bidirectional model will be more effective by enabling the contextual features to have the semantic information from both directions in the time dimension, which has also been illustrated in existing works such as BERT [11] and ELMo [28]. Therefore, we design our pre-training model to predict the pretexts in both directions in a skip-gram fashion.

Specifically, we use a pre-defined multi-layer CNN  $\phi_{\theta_1}$  to embed the raw SEEG signals as *local features*  $\phi_{\theta_1}(x_t)$ . In order to ensure the model’s ability to extract bidirectional information, we adopt a Transformer [36]  $\psi_{\theta_2}$  with a designed mask matrix as the autoregression model to obtain the *global contextual features*  $z$ . As Figure 2 shows, we divide a sequence into two equal half sub-sequences: the left-hand one is used to encode reverse direction information, while the right-hand one is for the forward direction. We set the center of the sequence as 0 in the time coordinates, and assign positive (negative) subscripts to elements in the forward (backward) sequence. With this definition, the  $t$ -th row in the mask matrix indicates the observable local features when constructing the global contextual features  $z_t$ , i.e.,  $z_t = \psi_{\theta_2}(\phi_{\theta_1}(x_t), \dots, \phi_{\theta_1}(x_{-t}))$ . After that, based on the InfoNCE loss, we let  $z_t$  predict the unobserved local features that is  $p$ -step away from the global context, i.e.,  $\phi_{\theta_1}(x_{t+\text{sgn}(t) \cdot p})$ . Formally, we define our contrastive loss at time  $t$  as

$$\mathcal{L}_t = -\frac{1}{P} \sum_{p=1}^P \log \left[ \frac{\text{Score}_p(\phi_{\theta_1}(x_{t+\text{sgn}(t) \cdot p}), z_t)}{\sum_{\mathbf{n} \in \mathcal{N}_t} \text{Score}_p(\phi_{\theta_1}(\mathbf{n}), z_t)} \right], \quad (1)$$

where  $\mathcal{N}_t$  denotes the random noise subset of the negative samples plus one positive sample, while  $\text{sgn}(\cdot)$  is the signed function. A bilinear classifier is utilized for prediction.

### 3.3 Graph Diffusion Learning

Although we have obtained the representation of every segment, interactions between channels and the underlying diffusion patterns have been ignored. We further propose a *graph diffusion* component to explicitly model the diffusion process in the human brain. More specifically, we adopt the approach of alternating between graph structure learning and brain wave diffusion to achieve the target, in an end-to-end data-driven fashion.

**Graph structure learning.** Overall, the challenge for modeling graph diffusion is that we do not know the underlying correlation and diffusion paths among the channels. We therefore propose to learn a graph structure where all the channels are the nodes. The key to tackling this problem is how to quantify the impact from one channel to another. Inspired by the phenomenon that in the brain’s electrical activities, there universally exist *traveling waves* [6], which keep some characteristics such as shape and frequency during the propagation across different channels, we aim to propose our structure learning algorithm based on it.

Under the assumption that similar channels are more likely to carry the same *traveling waves* that propagate over time, we use the *cosine* function as our metric to measure the correlation between pairs of channels. We leave the exploration of other similarity functions that might be more appropriate for time series data as future work. Considering the asymmetric time delays of the traveling waves, we need to distinguish the direction of correlation. Therefore, we use a pair of source and target learnable weight parameters  $\mathbf{W}_1$  and  $\mathbf{W}_2$  to identify important features before the similarity computation. Formally, we obtain the score matrix  $\mathcal{A}_0$  between every node pair from the cartesian product of source and target node sets  $v_1 \times v_2$  with node features  $\mathbf{h}_1$  and  $\mathbf{h}_2$  as follows:

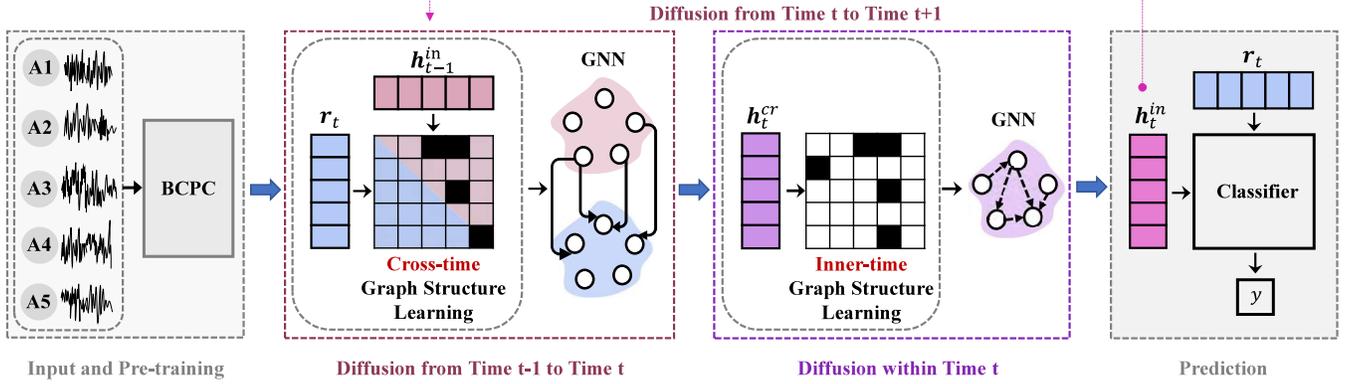
$$\mathcal{A}_0(i, j) = \cos(\mathbf{W}_1 \odot \mathbf{h}_1(i), \mathbf{W}_2 \odot \mathbf{h}_2(j)), \quad (2)$$

where  $i \in v_1, j \in v_2$  and  $\odot$  denotes the Hadamard product. Notice that,  $\mathcal{A}_0$  is very likely to represent a dense graph. To maintain the sparsity of graph structure that works of neuroscience suggest [1, 38], and remove insignificant and spurious connections caused by low frequency fluctuation or physiological noises, we filter unnecessary edges by using a threshold-based filter function  $F_\theta(\cdot)$  with a tunable hyper-parameter  $\theta$  as follows:

$$F_\theta(x) = \begin{cases} x, & x \geq \theta; \\ 0, & x < \theta. \end{cases} \quad (3)$$

Finally, the graph structure  $\mathcal{A}(i, j) = F_\theta(\mathcal{A}_0(i, j))$  is obtained.

**Brain wave diffusion.** The constructed graph  $\mathcal{A}$  represents the relative correlation between channels. The greater the edge weight, the more possible diffusion will occur. We aim to trace the diffusion along the constructed graph to enhance the representation of the traveling waves. During seizures, more rapid and significant propagation of spike-and-wave discharges will appear [29], which implies more distinguishable representations after propagation. We therefore adopt GNNs to model the brain wave diffusion process due to their natural message-passing ability on a graph. For simplicity and clear performance attribution, we use the standard GCN model in this work and take the one-layer directed GCN [39] as an example to describe the process. Specifically, given the graph



**Figure 3: Architecture of *BrainNet* at the channel level.** The first part represents the process of obtaining segment representation  $r_t$  using the pre-trained BCPC. The second part describes the cross-time diffusion process from  $t-1$  to  $t$ , in which the cross-time diffusion graph is first generated by the cross-time graph structure learning sub-module, and then a GNN is utilized on the learned graph. The third part corresponds to the inner-time diffusion process, which is based on a mechanism similar to the cross-time process. The outputs of the part will be utilized in the cross-time diffusion process of the next time step. The fourth part takes the input of the iteratively diffused representation  $h_t^{in}$  and combines it with the original representation  $r_t$  to predict whether an epileptic seizure occurs at time  $t$ .

$\mathcal{G} = ((v_1, v_2), \mathcal{A}, (h_1, h_2))$ , we obtain the representation of the target node  $j \in v_2$  after diffusion as follows:

$$h'_2(j) = \sigma \left( \frac{h_2(j) + \sum_{i \in v_1} \mathcal{A}(i, j) \cdot h_1(i)}{1 + \sum_{i \in v_1} \mathcal{A}(i, j)} \Theta \right), \quad (4)$$

where  $\Theta$  is the learnable linear transformation matrix for directed GCN and  $\sigma(\cdot)$  denotes the ReLU activation function.

**Putting it all together.** Owing to the epileptic waves lasting longer or shorter, *BrainNet* learns two types of diffusion processes. Concretely, *cross-time diffusion* naturally models the propagation of longer epileptic waves between two consecutive time segments. Meanwhile, fast signal spreading within the same time segments of each channel are captured by *inner-time diffusion*. Based on the structure learning and brain wave diffusion process, we formulate (2)-(4) as a function  $\mathcal{F}_{W_1, W_2, \Theta}(v_1, v_2, h_1, h_2, \theta)$ . We will then use the notation to introduce the graph diffusion component involving the two types of diffusion.

Graph diffusion component executes the two diffusion steps in the order of cross-time followed by inner-time. Given the representations obtained from the  $(t-1)$ -th inner-time diffusion process  $h_{t-1}^{in}$ , we derive the cross-time diffusion at the  $t$ -th time segment:

$$h_t^{cr} = \mathcal{F}_{W_1^{cr}, W_2^{cr}, \Theta^{cr}}(C_{t-1}, C_t, h_{t-1}^{in}, r_t, \theta^{cr}), \quad (5)$$

where both the source and target node sets consist of all channels, *i.e.*,  $C_t = C$  for  $t = 1, \dots, |S|$ . Here we use the subscript only to emphasize the time index. Essentially, the cross-time diffusion models the impact from the representation of last segments ( $h_{t-1}^{in}$ ) to that of current segments ( $r_t$ ), which is obtained from Section 3.2.

Following the cross-time diffusion at the  $t$ -th time segment, with a slight difference in the formula, the inner-time process traces the correlations among the representation  $h_t^{cr}$  of segments within the same time:

$$h_t^{in} = \mathcal{F}_{W_1^{in}, W_2^{in}, \Theta^{in}}(C_t, C_t, h_t^{cr}, h_t^{cr}, \theta^{in}). \quad (6)$$

The output representation  $h_t^{in}$  obtained after inner-time diffusion will then be involved in the next cross-time diffusion and, eventually, to the prediction stage. It is worth mentioning that the first cross-time diffusion has no historical representations of previous time segments. To handle this, we define a virtual node set  $v_0$  and virtual representation  $h_0$  to construct an empty graph and diffuse on it.

As shown above, Eq.(5) and Eq.(6) separately models the two diffusion patterns, and the two steps are alternately connected. Until now, we have acquired the representations after the graph diffusion component for further classification. Moreover, we also process the segments in the reverse time direction independently to obtain the reverse representations after inner-time diffusion  $h_{-t}^{in}$  at the  $t$ -th time segment.

### 3.4 Hierarchical Predictions

Given the representations obtained from the graph diffusion component, we concatenate  $h_t^{in}$ ,  $h_{-t}^{in}$  and  $r_t$  together so as to obtain the predicted probability of seizures  $\hat{y}_t$  through a discriminator  $D$  that implemented by a two-layer MLP. The objective function for epileptic wave detection of the channel is then defined as the binary cross-entropy:

$$\mathcal{L}_{ch} = - \sum_{t=1}^{|S|} \sum_{c \in C} \left[ y_{t,c}^S \log \hat{y}_{t,c} + (1 - y_{t,c}^S) \log(1 - \hat{y}_{t,c}) \right]. \quad (7)$$

However, the diffusion process for the channel level alone has limited horizons and lacks a more macroscopic perception. As Section 3.1 inspires, the synthesized information of the channel, brain-region, and patient level will be considered simultaneously to facilitate more accurate diagnoses. We describe the hierarchical task design combining these three levels in more details below.

**Hierarchical label construction.** The labels are originally situated at the channel level. Following the reverse logic order of doctor's diagnosis, we mark a brain region as epileptic if at least

one of the channels in it is epileptic. Similarly, the patient is regarded as being in a seizure state if at least one brain region is abnormal. More formally, given the  $t$ -th segment labels with all channels  $y_t^S$  (see Section 2), we first divide the channels into different brain regions through the mapping  $b(\cdot)$ . Then for each brain region  $\mathbf{b} \in \mathbf{B}$ , we assign the segment label as follows:

$$y_{t,\mathbf{b}}^{\text{brain}} = \max_{c:b(c)=\mathbf{b}} y_{t,c}^S. \quad (8)$$

In a similar fashion, we can also construct the patient-level’s segment labels based on labels of brain regions:

$$y_t^{\text{patient}} = \max_{\mathbf{b} \in \mathbf{B}} y_{t,\mathbf{b}}^{\text{brain}}. \quad (9)$$

**High-level representation learning.** We aggregate representations from the lower levels to obtain representations in the higher levels. While there are many available choices for permutation-invariant pooling, we adopt max pooling to aggregate the representations, because this approach is more aligned with the hierarchical label construction process and we need the features that contribute most to the epileptic state in the lower levels. Formally, taking the brain region representations  $\mathbf{r}^{\text{brain}}$  as an example, we use the element-wise max-pooling method as follows:

$$\mathbf{r}_{t,\mathbf{b}}^{\text{brain}}(i) = \max_{c:b(c)=\mathbf{b}} \mathbf{r}_{t,c}(i), \quad i = 1, \dots, d, \quad (10)$$

where  $\mathbf{b} \in \mathbf{B}$  and  $d$  denotes the dimension of the representations.

After the pooling operation, the graph diffusion component is implemented on the high-level representations to obtain the respective objectives, *i.e.*,  $\mathcal{L}_{\text{br}}$  and  $\mathcal{L}_{\text{pa}}$ , each of which is defined as the binary cross-entropy similar to Eq.(7). Considering that the completely independent parameters cannot guarantee a consistent optimization direction, we therefore let the three levels share the same parameter sets in the graph diffusion component and the discriminator  $D$  in order to align their representation spaces.

Finally, we jointly optimize the three tasks in different levels. With the guidance of labeled data, we optimize the proposed model via back propagation and learn the representation spaces for epileptic wave detection. Through the hierarchical task framework, *BrainNet* is expected to aggregate accurate information at higher levels, and in turn feed it back to the lower levels, resisting the data noise and improving the model performance.

## 4 EXPERIMENTS

### 4.1 Dataset

**Data collection.** The SEEG dataset used in our experiment is provided by a first-class hospital. The dataset statistics is summarized in Table 4 (Appendix A.2). More specifically, for a patient suffering from epilepsy, 4 to 10 invasive electrodes with 52 to 126 channels are used for recording 256Hz to 1024Hz SEEG signals; these figures vary from patient to patient. Notably, as SEEG signals are collected with high-frequency across multiple channels, our dataset is massive. In total, we have collected 526 hours of SEEG signals with 769GB. Although each prediction is patient-specific, to validate the generalizability and stability of our model, we repeat

all experiments on multiple patients. Based on our dataset, two professional neurosurgeons helped us to label the epileptic waves. We regard all time points within an epileptic wave as positive samples and the remainder as negative ones. The positive sample ratio of a single patient in our dataset is around 0.003 on average, which is extremely imbalanced. The dataset will be released after further cleansing and scrutinization.

**Preprocessing.** In order to conduct the self-supervised learning task (Section 3.2), for each patient, we randomly sample 10000 normal segments with a window length of 1 second. Ninety percent of the sampled segments are then used for training, with the remainder used for validation. As for the epileptic wave detection task, for each patient, we first obtain 13300 segments to train our model (85% for training and 15% for validation). For the testing, to explore the performance of different models under datasets with different positive sample ratios, we sample three test sets, each of which includes 1140, 9690 and 95190 segments for each patient, with positive-negative sample ratio of 1:5, 1:50 and 1:500 respectively.

### 4.2 Experimental Setup

First, we train a model for each patient independently and evaluate the performance. Then we repeat the experiments on all patients and obtain the average results. Owing to different numbers of recording channels in different brain regions for different patients, generalization between patients is difficult, and we will leave this issue as future work. Therefore, we only conduct experiments on a single patient. Considering the advantage conferred by the hierarchical task framework, we validate the effectiveness of our proposed *BrainNet* at both the channel and patient levels. As for the baselines, to the best of our knowledge, no existing model can handle these two levels of tasks at the same time. Therefore, for each task, we adopt task-specific baselines. Specifically, for the *channel-level epileptic wave detection task*, we compare *BrainNet* with several univariate time series classification models including TSF [10], STSF [4], MiniRocket [9], WEASEL [32], LSTM-FCN [19] and TS-TCC [12]. For each baseline, we train an independent model for every single channel and obtain the average results on all channels for one patient. As for the *patient-level epileptic wave detection task*, we use the following multivariate time series classification models as baselines: EEGNet [23], TapNet [40], MLSTM-FCN [20] and NS [15]. We provide details of these baselines, the evaluation metrics and the hyperparameter analysis in the Appendix. These baselines are designed to deal with raw time series data rather than representation space, so we do not use representations pre-trained by BCPC as input to these models.

### 4.3 Experimental Results

The average performance over all patients of different methods are presented in Table 1. Overall, *BrainNet* outperforms all baselines on every evaluation indicator in both channel and patient levels.

**Results in channel-level.** In the channel-level task, *BrainNet* improves 12.31%, 36.66%, 142.03% in terms of  $F_2$  on the test datasets with positive-negative sample ratio of 1:5, 1:50 and 1:500 respectively. In particular, as the labels become more and more imbalanced, the performance of baselines drops rapidly, while *BrainNet* keeps a relatively much better performance than baselines. The increasing

**Table 1: The average performance of epileptic wave detection tasks at the channel and patient levels. Test datasets with different positive-negative sample ratios are used for evaluation.**

Models		Ratio		1 : 5					1 : 50					1 : 500				
		Pre.	Rec.	$F_1$	$F_2$	AUC	Pre.	Rec.	$F_1$	$F_2$	AUC	Pre.	Rec.	$F_1$	$F_2$	AUC		
Channel Level	TSF	51.22	45.33	44.02	44.16	83.47	16.99	46.62	21.92	29.71	85.00	2.42	36.52	3.96	7.52	79.66		
	STSF	53.51	57.65	51.79	54.26	86.75	18.21	56.98	24.69	34.48	87.80	2.61	49.73	4.63	9.22	84.07		
	MiniRocket	55.68	56.84	53.55	54.84	87.07	21.82	55.30	28.12	36.85	87.47	4.02	48.93	6.83	12.42	84.10		
	WEASEL	50.38	43.92	43.84	43.24	80.93	17.91	43.61	22.44	29.11	82.21	2.79	36.83	4.84	9.05	77.50		
	LSTM-FCN	36.32	33.85	30.35	30.99	73.41	10.73	35.64	13.91	19.65	75.68	1.68	30.12	2.79	5.38	70.88		
	TS-TCC	56.20	53.48	50.81	51.58	85.57	21.60	52.59	26.87	34.79	86.33	3.93	46.63	6.40	11.41	82.49		
<i>BrainNet</i>		<b>68.19</b>	<b>62.71</b>	<b>61.90</b>	<b>61.59</b>	<b>95.22</b>	<b>42.74</b>	<b>58.21</b>	<b>45.29</b>	<b>50.36</b>	<b>94.80</b>	<b>16.87</b>	<b>51.69</b>	<b>21.88</b>	<b>30.06</b>	<b>92.19</b>		
Patient Level	EEGNet	55.24	64.62	57.37	60.74	84.81	13.20	64.02	21.07	33.56	83.32	1.38	54.73	2.66	6.04	78.25		
	TapNet	71.48	57.93	63.37	59.88	91.73	30.55	61.89	38.33	47.56	90.15	OOM						
	MLSTM-FCN	68.68	70.96	66.77	68.70	89.60	28.00	73.32	34.38	46.14	90.69	4.39	62.25	7.43	13.47	85.50		
	NS	57.68	54.99	55.80	55.19	80.76	14.70	54.13	22.52	33.70	80.17	1.67	44.88	3.20	7.06	73.24		
<i>BrainNet</i>		<b>79.61</b>	<b>79.08</b>	<b>76.69</b>	<b>76.87</b>	<b>94.92</b>	<b>40.31</b>	<b>80.53</b>	<b>49.37</b>	<b>60.30</b>	<b>93.86</b>	<b>12.36</b>	<b>72.58</b>	<b>18.57</b>	<b>28.85</b>	<b>91.93</b>		

ratio of performance improvement implies our model has the ability to handle more imbalanced data, which is more aligned with the practical clinical scenarios. Compared with univariate baselines, *BrainNet* takes the advantage of learning how the epileptic waves diffuse across channels. For example, two channels A1 and A2 have a low correlation in the normal state, *i.e.*, the diffusion of the normal brain waves between them is very weak. However, during the seizures, their correlation increases significantly, which is reflected in the larger edge weight in the diffusion graph learned by our model. When *BrainNet* determines that A1 contains seizures, it can further infer the epileptic status of A2 with more certainty through the diffusion process along with the learned graph structure. This makes the prediction result of A2 more reliable compared with the situation only considering A2 alone.

**Results in patient-level.** At the patient level, *BrainNet* also improves the performance by 50.95% in terms of  $F_2$  in average on the three test datasets. The results show the superiority of our hierarchical task design to make more accurate predictions through utilizing more refined information at lower levels. In more detail, *BrainNet* can offer some evidence of the specific local (channel- and brain-region-level) information for the global system (patient), which benefits the detection task in patient-level. For example, being aware of which particular brain region has epileptic waves increases the model’s confidence when inferring whether the patient is suffering from epilepsy.

#### 4.4 Ablation Study

In this section, we conduct ablation experiments to verify the effectiveness of each major component in our model. More specifically, we remove each of the following components from our model to see how it influences the performance respectively: pre-training (*BrainNet*-BCPC), inner-time diffusion step (*BrainNet*-Inner), cross-time diffusion step (*BrainNet*-Cross), hierarchical task framework (*BrainNet*-Multi) and graph diffusion component (*BrainNet*-graph).

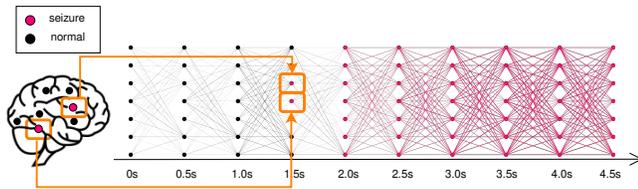
**Table 2: Results of ablation study.**

Models		Metrics				
		Pre.	Rec.	$F_1$	$F_2$	AUC
Channel Task	<i>BrainNet</i> -BCPC	3.86	33.84	5.26	8.42	80.12
	<i>BrainNet</i> -Graph	3.26	20.15	4.00	6.07	71.61
	<i>BrainNet</i> -Inner	11.43	38.58	13.76	17.99	91.14
	<i>BrainNet</i> -Cross	4.78	41.40	7.56	12.81	85.05
	<i>BrainNet</i> -Multi	11.80	44.19	14.70	19.68	88.81
<i>BrainNet</i>		<b>16.87</b>	<b>51.69</b>	<b>21.88</b>	<b>30.06</b>	<b>92.19</b>
Patient Task	<i>BrainNet</i> -BCPC	1.22	49.84	2.32	5.09	81.71
	<i>BrainNet</i> -Inner	5.27	62.24	9.20	17.08	88.12
	<i>BrainNet</i> -Cross	3.60	56.32	6.54	13.08	86.79
<i>BrainNet</i>		<b>12.36</b>	<b>72.58</b>	<b>18.57</b>	<b>28.85</b>	<b>91.93</b>

We report the evaluation results of the ablation experiments on the test dataset with 1:500 positive-negative sample ratio in Table 2. It can be observed that *BrainNet* achieves the best performance to all ablated model versions in all metrics, which demonstrates the effectiveness of each component in our model design. For *BrainNet*-BCPC, the striking drop in performance indicates the powerful representation ability of BCPC. Comparison with *BrainNet*-Graph, which obtains the representations from BCPC and feeds them into MLP directly, reveals that our model achieves superior performance (improves more than 390% in terms of  $F_2$ ). It suggests the significance of modeling the diffusion process.

#### 4.5 Case Study

We at last present a case study to illustrate how cross diffusion works. As shown in Figure 4, 10 contiguous time segments are presented with a corresponding learned cross diffusion graph structure. We can see that at time 1.5s epileptic waves appear in two brain regions for the first time, and then diffuse to other brain regions. The edge weights are lower in the normal state, while edges with



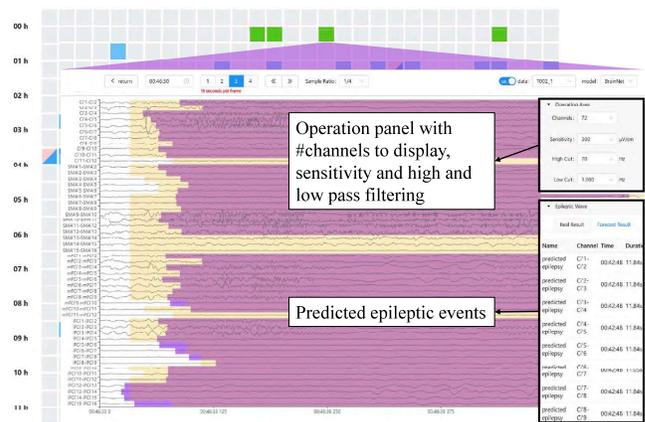
**Figure 4: A case of cross-time diffusion.** The horizontal direction is arranged in chronological order with 10 contiguous segments at intervals of 0.5s, while the 7 different brain regions in patient P-4 are expanded in the vertical direction. The black nodes denote the normal state of the corresponding brain regions predicted by *BrainNet*, and red ones indicate the predicted epileptic brain regions.

high weights are widely observed when epileptic waves appear in multiple brain regions. This phenomenon is consistent with some domain knowledge in the field of neuroscience. More specifically, dramatic changes of brain connectivity in SEEG can be tracked during seizures [3]. During normal status, most brain regions have relatively weak connections. When a seizure occurs, brain regions gradually form a resonance, which means that strong connections between brain regions will be observed. This phenomenon increases the credibility of our epilepsy diffusion graph. Moreover, if a seizure occurs in the two brain regions as first time predicted by our model, there is a higher probability that SOZ will be located in these two brain regions, which can assist the doctors with diagnosis in clinical practice.

#### 4.6 Online System

We have deployed *BrainNet* into an online system, where doctors could upload the SEEG data of patients. Then *BrainNet* will detect the epileptic waves from the data and present the results in the visualization panel. Therefore, doctors can review the predicted epileptic waves very quickly to obtain the basic seizure patterns of patients. This will save much time for doctors to develop further treatment plans.

Figure 5 shows the screenshot of the system interface. The top part of Figure 5 is the profile page of *patient overview*. We show the overview of a 12-hour patient file after being reviewed by doctors. Each square with different colors denotes a 1-minute data segment. The gray squares (□) denote that no epileptic waves exist, while the green ones (■), the blue ones (■) and the red ones (■) represent correct, wrong and missing predictions of our model respectively. Doctors can see the detailed SEEG data that may contain a seizure onset predicted by our model by clicking one square. The bottom part of Figure 5 shows the prediction results of epileptic wave detection of the clicked square. The top toolbar is used to change the presented time period. The data operation panel and epileptic wave events can be found on the right. In the center of the page, the purple part represents the real epileptic waves labels, if any, offered by doctors and the yellow part is our model’s prediction. As shown in the figure, the predictions of our model match the actual seizures well.



**Figure 5: Demonstration of the online system.**

## 5 RELATED WORK

Epilepsy is a disorder of the brain that can be detected using SEEG signals. Features of the SEEG are patient-specific in nature and vary largely from one person to another. Much work has been conducted over the past few decades in an attempt to design an automated system that can analyze and detect seizures and predict them before their occurrence, so that required measures can be taken to record them. Numerous works have been conducted by researchers to understand epilepsy and the characteristics of brain activity that accompanies an epilepsy attack so as to detect and predict the onset of the seizure. The earliest of these studies can be traced back to a 1982 work by *Gotman* [17], who developed patient non-specific detectors. Seizures consist of various kinds of wave-forms, such as spikes, sharp waves, sleep spindles, and periods [13]. Accurate patient-specific detectors were designed using SVM classifiers in [14]. Sensitivity and false detection rate were used as the standard for measuring performance by researchers in [34]. There are limitations in automatically detecting and predicting seizures [16].

Using traditional machine learning methods to detect epileptic waves requires feature engineering, which in turn requires large amounts of domain knowledge and is also time-consuming. As an end-to-end method, deep learning is widely employed to perform this task. *Roy et al.* combine a one-dimensional convolutional layer and Gated Recurrent Unit (GRU) to perform epileptic wave detection [30]. *Hisham Daoud et al.* design a deep convolutional autoencoder architecture that pre-trains the model in an unsupervised manner. After pre-training, the trained encoder is connected to a Bidirectional Long Short-Term Memory (Bi-LSTM) Network for classification [5]. *Lawhern et al.* propose a model named EEG-Net, which is composed of two-dimensional convolution layers and pooling layers and was designed for EEG-based brain-computer interfaces; it can also be used to perform epileptic wave detection based on EEG [23].

## 6 CONCLUSION

In this paper, we proposed a novel model, *BrainNet*, to learn the diffusion graph via a hierarchical framework for epileptic wave

detection. Drawing on domain knowledge from neuroscience, to detect epileptic waves more accurately, we study the *epileptogenic network*, learn the underlying dynamic diffusion graph structures, and model the cross-time and inner-time diffusion patterns at the channel, brain region, and patient levels. Moreover, by conducting experiments on a large-scale real-world dataset, we demonstrate the effectiveness of our proposed model on detecting epileptic waves and modeling the diffusion process. Furthermore, we have deployed *BrainNet* into an online system to assist the diagnosis of epilepsy, where doctors could upload the SEEG data and obtain the epileptic waves identified by our model (see details in Section 4.6). In the future, motivated by our positive results on both the accuracy of epileptic wave detection and interpretability of learned epileptogenic graphs, it would be intriguing to further validate our framework across more epileptic patients and study its real-world clinical deployment.

**Acknowledgment.** This work is supported by NSFC (No.62176233) and the National Key Research and Development Project of China (No.2018AAA0101900) and the Fundamental Research Funds for the Central Universities.

## REFERENCES

- [1] Sophie Achard and Ed Bullmore. 2007. Efficiency and cost of economical brain functional networks. *PLoS computational biology* 3, 2 (2007), e17.
- [2] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. 2017. The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data mining and knowledge discovery* 31, 3 (2017), 606–660.
- [3] Fabrice Bartolomei, Stanislas Lagarde, Fabrice Wendling, Aileen Mcgonigal, Viktor Jirsa, Maxime Guye, and Christian Bénar. 2017. Defining epileptogenic networks: Contribution of SEEG and signal analysis. *Epilepsia* 58, 7 (2017), 1131–1147.
- [4] Nestor Cabello, Elham Naghizade, Jianzhong Qi, and Lars Kulik. 2020. Fast and Accurate Time Series Classification Through Supervised Interval Search. In *ICDM*. 948–953.
- [5] Hisham Daoud and Magdy A Bayoumi. 2019. Efficient epileptic seizure prediction based on deep learning. *IEEE T BIOMED CIRC S* 13, 5 (2019), 804–813.
- [6] Zachary W Davis, Lyle Muller, Julio Martinez-Trujillo, Terrence Sejnowski, and John H Reynolds. 2020. Spontaneous travelling cortical waves gate perception in behaving primates. *Nature* 587, 7834 (2020), 432–436.
- [7] Maarten De Vos, Anneleen Vergult, Lieven De Lathauwer, Wim De Clercq, Sabine Van Huffel, Patrick Dupont, Andre Palmi, and Wim Van Paesschen. 2007. Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *NeuroImage* 37, 3 (2007), 844–854.
- [8] Angus Dempster, François Petitjean, and Geoffrey I Webb. 2020. ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery* 34, 5 (2020), 1454–1495.
- [9] Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb. 2021. MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification. In *KDD*. 248–257.
- [10] Houtao Deng, George Runger, Eugene Tuv, and Martyanov Vladimir. 2013. A time series forest for classification and feature extraction. *Information Sciences* 239 (2013), 142–153.
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*. 4171–4186.
- [12] Emadelddeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwah, Xiaoli Li, and Cuntai Guan. 2021. Time-series representation learning via temporal and contextual contrasting. In *IJCAI*. 2352–2359.
- [13] Paul Fergus, A Hussain, David Hignett, Dhiya Al-Jumeily, Khaled Abdel-Aziz, and Hani Hamdan. 2016. A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics* 12, 1 (2016), 70–89.
- [14] Dimitrios I Fotiadis. 2015. *Handbook of research on trends in the diagnosis and treatment of chronic conditions*. IGI Global.
- [15] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. 2019. Unsupervised scalable representation learning for multivariate time series. In *NeurIPS*.
- [16] F Fürbass, P Ossenblok, M Hartmann, H Perko, AM Skupch, G Lindinger, L Elezi, E Patarai, AJ Colon, C Baumgartner, et al. 2015. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. *Clin Neurophysiol* 126, 6 (2015), 1124–1131.
- [17] Jean Gotman. 1982. Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and clinical Neurophysiology* 54, 5 (1982), 530–540.
- [18] Longlong Jing and Yingli Tian. 2021. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence* 43, 11 (2021), 4037–4058.
- [19] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2018. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* 6 (2018), 1662–1669.
- [20] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. 2019. Multivariate LSTM-FCNs for time series classification. *Neural Networks* 116 (2019), 237–245.
- [21] Simon Khuvis, Sean T Hwang, and Ashesh D Mehta. 2021. Intracranial EEG biomarkers for seizure lateralization in rapidly-bisynchronous epilepsy after laser corpus callosotomy. *Frontiers in Neurology* 12 (2021), 1757.
- [22] Isabell Kiral-Kornek, Subhrajit Roy, Ewan Nurse, Benjamin Mashford, Philippa Karoly, Thomas Carroll, Daniel Payne, Susmita Saha, Steven Baldassano, Terence O'Brien, et al. 2018. Epileptic seizure prediction using big data and deep learning: toward a mobile system. *EBioMedicine* 27 (2018), 103–111.
- [23] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *Journal of neural engineering* 15, 5 (2018), 056013.
- [24] Jason Lines, Luke M. Davis, Jon Hills, and Anthony Bagnall. 2012. A Shapelet Transform for Time Series Classification. In *KDD*. 289–297.
- [25] Jeffrey P Mullin, Michael Shriver, Soha Alomar, Imad Najm, Juan Bulacio, Patrick Chauvel, and Jorge Gonzalez-Martinez. 2016. Is SEEG safe? A systematic review and meta-analysis of stereo-electroencephalography-related complications. *Epilepsia* 57, 3 (2016), 386–401.
- [26] Min Ni, Bushra Afroze, Chao Xing, Chunxiao Pan, Yanqiu Shao, Ling Cai, Brandi L Cantarel, Jimin Pei, Nick V Grishin, Stacy Hewson, et al. 2021. A pathogenic UFSF2 variant in an autosomal recessive form of pediatric neurodevelopmental anomalies and epilepsy. *Genetics in Medicine* 23, 5 (2021), 900–908.
- [27] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation Learning with Contrastive Predictive Coding. *arXiv preprint arXiv:1807.03748* (2018).
- [28] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *NAACL*. 2227–2237.
- [29] Timothée Proix, Viktor K Jirsa, Fabrice Bartolomei, Maxime Guye, and Wilson Truccolo. 2018. Predicting the spatiotemporal diversity of seizure propagation and termination in human focal epilepsy. *Nature Communications* 9, 1 (2018), 1–15.
- [30] Subhrajit Roy, Isabell Kiral-Kornek, and Stefan Harrer. 2018. Deep learning enabled automatic abnormal EEG identification. In *EMBC*. 2756–2759.
- [31] Patrick Schäfer. 2015. The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery* 29, 6 (2015), 1505–1530.
- [32] Patrick Schäfer and Ulf Leser. 2017. Fast and accurate time series classification with weasel. In *CKM*. 637–646.
- [33] Alexandros T Tzallas, Markos G Tsipouras, and Dimitrios I Fotiadis. 2009. Epileptic Seizure Detection in EEGs Using Time-Frequency Analysis. *IEEE Trans Inf Technol Biomed* 13, 5 (2009), 703–710.
- [34] Alexandros T. Tzallas, Markos G. Tsipouras, Dimitrios G. Tsalikakis, Evaggelos C. Karvounis, Loukas Astrakas, Spiros Konitsiotis, and Margaret Tzaphlidou. 2012. Automated Epileptic Seizure Detection Methods: A Review Study. In *Epilepsy*. IntechOpen, Chapter 4.
- [35] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papanathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 1 (2002), 273–289.
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NeurIPS*.
- [37] Greg A Worrell, Andrew B Gardner, S Matt Stead, Sanqing Hu, Steve Goerres, Gregory J Cascino, Fredric B Meyer, Richard Marsh, and Brian Litt. 2008. High-frequency oscillations in human temporal lobe: simultaneous microwire and clinical macroelectrode recordings. *Brain* 131, 4 (2008), 928–937.
- [38] Renping Yu, Han Zhang, Le An, Xiaobo Chen, Zhihui Wei, and Dinggang Shen. 2017. Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification. *Human brain mapping* 38, 5 (2017), 2370–2383.
- [39] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. 2019. Graph Transformer Networks. In *NeurIPS*.
- [40] Xuchao Zhang, Yifeng Gao, Jessica Lin, and Chang-Tien Lu. 2020. TapNet: Multivariate Time Series Classification with Attentional Prototypical Network. In *AAAI*. 6845–6852.

## A SUPPLEMENT

In the supplement, we provide the details not expanded in the text above, including the notations used in our model, dataset statistics, evaluation metrics of experiments, details of our baseline methods, and results of the hyperparameter analysis.

### A.1 Notations

The notations used in our model are summarized in Table 3.

**Table 3: Summary of important notations.**

Notation	Description
$B, C$	The set of brain regions and channels
$b(\cdot)$	The function mapping each channel to the corresponding brain region
$t, c, b$	The index of segments, channels and brain regions
$s_{t,c}$	The data of $t$ -th segment of channel $c$
$y_{t,c}^S$	The label of $t$ -th segment of channel $c$
$r_t$	The representation acquired by the pre-training BCPC and a linear transformation at the $t$ -th segment
$h_t^{\text{in}}$	The representation of the $t$ -th segment after inner-time diffusion process
$h_t^{\text{cr}}$	The representation of the $t$ -th segment after cross-time diffusion process

### A.2 Dataset Statistics

The SEEG data related to each patient is placed in Table 4, including recording time, sampling frequency, the number of electrodes inserted into the brain, the number of channels contained in the electrodes, the epileptic wave ratios and the total samples for each patient. During the experimental phase, the channels actually used by each patient removed those unrelated to seizures and bad channels that doctors told us.

### A.3 Evaluation Metrics

In our experiments, we use the following metrics to evaluate the models: precision, recall, two F-measures (*i.e.*  $F_1$  and  $F_2$ ) and AUC. Notice that we do not consider the accuracy metric because of the extremely imbalanced sample ratio. The F-measure is a metric defined as the weighted harmonic mean of precision and recall, with the following mathematical formulation:

$$F_\beta = \frac{(1 + \beta^2) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}.$$

In our experiments,  $F_2$  is more favored than  $F_1$ , as in clinical context, missing any epileptic waves is costly.

**Table 4: Statistics of SEEG data corresponding to each patient in our dataset. Notice that we detect epileptic waves on segment level and channel level, and obtain 55,961,080 samples in total (5,596,108 samples for each patient on average).**

Patients	Time (Hours)	Sample frequency	# Elec-trodes	# Chan-nels	Epileptic wave ratio	# Samples
P-1	72	1000Hz	10	126	0.003	8,113,760
P-2	21	1000Hz	4	52	0.004	3,460,280
P-3	114	1000Hz	10	126	0.001	9,664,920
P-4	56	256Hz	4	52	0.001	4,295,520
P-5	36	512Hz	5	63	0.002	3,818,240
P-6	6	512Hz	5	69	0.004	3,698,920
P-7	24	512Hz	7	89	0.009	6,801,240
P-8	36	1024Hz	4	52	0.001	1,789,800
P-9	24	512Hz	7	93	0.0002	4,176,200
P-10	137	1000Hz	8	110	0.0006	10,142,200

### A.4 Baselines

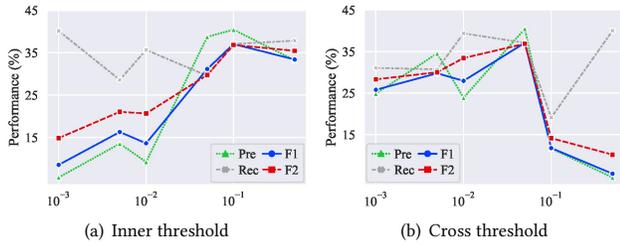
We give details of our baseline methods used in two levels of tasks.

**Channel-level epileptic wave detection.** We compare *BrainNet* with several univariate time series classification models. For each baseline, we train an independent model for every single channel and obtain the average results on all channels for one patient.

- Time Series Forest (TSF) [10]: This is a tree-ensemble method for time series classification. The features employed in TSF are basic time-series features, including the mean, standard variance, and slope of a segment.
- Supervised Time Series Forest (STSF) [4]: This is an interval-based tree model that adopts a top-down approach to search for relevant subseries in three different time series representations before training any tree classifier.
- MiniRocket [9]: This approach reformulates Rocket [8] into a fast version maintaining essentially the same accuracy.
- WEASEL [32]: This approach transforms time series into feature vectors using a sliding window approach. The vectors are then analyzed through a machine learning classifier.
- LSTM-FCN [19]: This is used for univariate time series classification consisting of an LSTM layer and CNN layer.
- TS-TCC [12]: This is a contrastive method that learns time series representation through two different views generated by weak and strong augmentations of the original segments.

**Patient-level epileptic wave detection.** We use the following multivariate time series classification models as baselines:

- EEGNet [23]: This is a model proposed in the field of neuroscience. Specifically, it is a compact convolutional neural network for EEG-based brain-computer interfaces.
- TapNet [40]: An attentional prototype network, which takes the strengths of both traditional and deep learning based approaches to perform multivariate time series classification.
- MLSTM-FCN [20]: This is a deep learning framework consisting of an LSTM layer and stacked CNN layer, along with a Squeeze-and-Excitation block for multivariate time series classification.



**Figure 6: The hyperparameter analysis of the thresholds in the inner-time diffusion (a) and the cross-time diffusion (b).**

- NS [15]: This approach uses an unsupervised method employing time-based negative sampling to learn embeddings. SVM is then applied to perform the final classification.

Considering the trade-off between efficiency and effectiveness, we run TapNet 200 epochs with learning rate 10 times larger than

the default. For NS, we run 100 epochs with default learning rate. All other baselines are initialized with the same parameters suggested in their respective papers.

## A.5 Hyperparameter Analysis

We analyze the influence of the hyperparameters—the two thresholds in the graph diffusion component—on the performance (Figure 6). Taking the patient P-6 on the test dataset with 1 : 50 sample ratio as an example, whose best inner-time and cross-time thresholds are 0.1 and 0.05 respectively, the results show that the performance of *BrainNet* reaches the best results near the two optimal thresholds and degrades on both sides of the optimal values. This can be interpreted as more noisy correlations will be introduced along with the more dense diffusion graph structure if we set a relatively low threshold. On the contrary, a too high threshold will also harm the performance because some representative diffusion patterns might be ignored.