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
yyingfan@outlook.com

Xiaolong Mo
Neuroechos Medical(Shenzhen) Co., Ltd.
xiaolong.mo@neurox.cn

Carl Yang
Emory Yang
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.

CCS CONCEPTS
• Applied computing → Health care information systems.

KEYWORDS
Epileptic wave detection; Graph diffusion; SEEG data

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
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, 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 electrode contact, the machine begins to record the signals of every channel so as to collect the SEEG data—shown at the bottom of the contactors, the machine begins to record the signals 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 \( S = \{ S_t \}_{t=1}^{T} \) where \( |S| = \lfloor (|T| - k)/l \rfloor + 1 \) is the number of segments. The annotation \( S_t = \{ s_{t,c}, \} \in \mathbb{R}^{(|C| \times k)} \) represents the segment with \( |C| \) channels and \( s_{t,c} = \{ x_{i,t-1} \}, i_c, k \in \mathbb{R}^{k} \) 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 \( Y^S = \{ Y^S_t \}_{t=1}^{S} \). If one segment includes a seizure point, we consider it to be a seizure segment; i.e., \( y^S_t = \max_{i_c} (y_{i,t-1} + l_i) \). 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 \( S \) and corresponding label set \( Y^S \) of an individual patient, we aim to predict the labels \( Y^S_t \) of the future segment set \( S \):

\[
P(y^S_t | S, Y^S), \quad t = 1, \ldots, |S|, \quad c \in C,
\]

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.
3.2 Model Pre-training

Within the massive amount of SEEG records obtained from a patient lasting for days, there are only often 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. 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 \( \phi_{\theta_{0}} \) to embed the raw SEEG signals as local features \( \phi_{\theta_{0}}(x_t) \). 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., \( z_t = \psi_{\theta_{2}}(\phi_{\theta_{0}}(x_t), \ldots, \phi_{\theta_{0}}(x_{t-1})) \). 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., \( \phi_{\theta_{0}}(x_{t+\text{sgn}(t-1)-p}) \).

Formally, we define our contrastive loss at time \( t \) as

\[
\mathcal{L}_t = -\frac{1}{P} \sum_{p=1}^{P} \log \frac{\text{Score}_p(\phi_{\theta_{0}}(x_{t+\text{sgn}(t-1)-p}), z_t)}{\sum_{n \in \mathcal{N}_t} \text{Score}_p(\phi_{\theta_{0}}(n), z_t)},
\]

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 \( W_{1} \) and \( W_{2} \) 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 \( \mathcal{V}_{1} \times \mathcal{V}_{2} \) with node features \( \mathbf{h}_{1} \) and \( \mathbf{h}_{2} \) as follows:

\[
\mathcal{A}_{0}(i, j) = \cos(W_{1} \odot \mathbf{h}_{1}(i), W_{2} \odot \mathbf{h}_{2}(j)),
\]

where \( i \in \mathcal{V}_{1}, j \in \mathcal{V}_{2} \) and \( \odot \) denotes the Hadamard product. 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:

\[
F_{\theta}(x) = \begin{cases} x, & x \geq \theta; \\ 0, & x < \theta. \end{cases}
\]

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...
where both the source and target node sets consist of all channels, i.e., $\mathcal{G} = ((v_1, v_2), \mathcal{A}, (h_1, h_2))$, we obtain the representation of the target node $j \in v_2$ after diffusion as follows:

$$h_j^2 = \sigma \left( \frac{h_j^2 + \sum_{i \in \mathcal{G}} \mathcal{A}(i, j) \cdot h_i^1}{1 + \sum_{i \in \mathcal{G}} \mathcal{A}(i, j)} \Theta \right),$$  \quad (4)

where $\Theta$ is the learnable linear transformation matrix for directed $\text{GCN}$ 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:

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

where both the source and target node sets consist of all channels, i.e., $C_t = C$ for $t = 1, \ldots, |S|$. 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 $h_{t}^{ct}$ of segments within the same time:

$$h_{t}^{in} = \mathcal{F}_{W_1^{in}, W_2^{in}, \Theta^{in}}(C_t, C_t, h_{t}^{ct}, r_t, \Theta^{in}).$$  \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 $h_{t-1}^{in}$ at the $t$-th time segment.

### 3.4 Hierarchical Predictions

Given the representations obtained from the graph diffusion component, we concatenate $h_{t}^{ct}$, $h_{t}^{in}$ 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:

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

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 $t = t - 1$ 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 $t$. 

---

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

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

---

**BrainNet:**

- **Input and Pre-training:**
  - Channels: A1, A2, A3, A4, A5
  - BCPC

- **Diffusion from Time t-1 to Time t**
  - Cross-time Graph Structure Learning
  - GNN

- **Diffusion within Time t**
  - Inner-time Graph Structure Learning
  - GNN

- **Prediction**
  - Classifier
  - Output: $y'$
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 \( y_{t}^{c} \) (see Section 2), we first divide the channels into different brain regions through the mapping \( b(\cdot) \). Then for each brain region \( b \in \mathcal{B} \), we assign the segment label as follows:

\[
y_{t}^{brain}_{b} = \max_{c:b(c)=b} y_{t}^{c}.
\]

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

\[
y_{t}^{patient} = \max_{b \in \mathcal{B}} y_{t}^{brain}_{b}.
\]

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 \( r^{\text{brain}} \) as an example, we use the element-wise max-pooling method as follows:

\[
r^{\text{brain}}_{t,b}(i) = \max_{c:b(c)=b} r_{t,c}(i), \quad i = 1, \ldots, d,
\]

where \( b \in \mathcal{B} \) 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, \( \mathcal{L}_{\text{br}} \) 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, \( \text{BrainNet} \) 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 \( 256\text{Hz} \) to \( 1024\text{Hz} \) 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 \( 769\text{GB} \). 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 scrutiny.

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 \( \text{BrainNet} \) 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 \( \text{BrainNet} \) 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, \( \text{BrainNet} \) outperforms all baselines on every evaluation indicator in both channel and patient levels.

Results in channel-level. In the channel-level task, \( \text{BrainNet} \) 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 \( \text{BrainNet} \) 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.

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

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.

4.4 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.

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

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.

4.5 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
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 4 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.

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 EEGNet, 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
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


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.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>B, C</td>
<td>The set of brain regions and channels</td>
</tr>
<tr>
<td>b(·)</td>
<td>The function mapping each channel to the corresponding brain region</td>
</tr>
<tr>
<td>t, c, b</td>
<td>The index of segments, channels and brain regions</td>
</tr>
<tr>
<td>s_t,c</td>
<td>The data of t-th segment of channel c</td>
</tr>
<tr>
<td>y_t,c</td>
<td>The label of t-th segment of channel c</td>
</tr>
<tr>
<td>r_t</td>
<td>The representation acquired by the pre-training BCPC and a linear transformation at the t-th segment</td>
</tr>
<tr>
<td>h_t^{in}</td>
<td>The representation of the t-th segment after inner-time diffusion process</td>
</tr>
<tr>
<td>h_t^{cr}</td>
<td>The representation of the t-th segment after cross-time diffusion process</td>
</tr>
</tbody>
</table>

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.

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).

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

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-Excitation block for multivariate time series classification.
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.