

## Research Article

# A Novel MEGNet for Classification of High-Frequency Oscillations in Magnetoencephalography of Epileptic Patients

Jun Liu,<sup>1,2</sup> Siqi Sun ,<sup>2</sup> Yang Liu,<sup>2</sup> Jiayang Guo ,<sup>3</sup> Hailong Li,<sup>4</sup> Yuan Gao,<sup>5</sup> Jintao Sun,<sup>5</sup> and Jing Xiang<sup>6</sup>

<sup>1</sup>Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China

<sup>2</sup>School of Computer Science & Engineering, Wuhan Institute of Technology, Wuhan 430205, China

<sup>3</sup>Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH, USA

<sup>4</sup>Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

<sup>5</sup>Department of Neurosurgery, Nanjing Brain Hospital, Nanjing, China

<sup>6</sup>Department of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Correspondence should be addressed to Siqi Sun; 21807010006@wit.edu.cn and Jiayang Guo; guojy@mail.uc.edu

Received 17 April 2020; Revised 7 August 2020; Accepted 19 August 2020; Published 4 September 2020

Academic Editor: Zhile Yang

Copyright © 2020 Jun Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Epilepsy is a neurological disease, and the location of a lesion before neurosurgery or invasive intracranial electroencephalography (iEEG) surgery using intracranial electrodes is often very challenging. The high-frequency oscillation (HFOs) mode in MEG signal can now be used to detect lesions. Due to the time-consuming and error-prone operation of HFOs detection, an automatic HFOs detector with high accuracy is very necessary in modern medicine. Therefore, an optimized capsule neural network was used, and a MEG (magnetoencephalograph) HFOs detector based on MEGNet was proposed to facilitate the clinical detection of HFOs. To the best of our knowledge, this is the first time that a neural network has been used to detect HFOs in MEG. After optimized configuration, the accuracy, precision, recall, and F1-score of the proposed detector reached 94%, 95%, 94%, and 94%, which were better than other classical machine learning models. In addition, we used the k-fold cross-validation scheme to test the performance consistency of the model. The distribution of various performance indicators shows that our model is robust.

## 1. Introduction

Epilepsy is a spectrum of neurological disorders, caused by the abnormal firing of neurons in the brain with sudden and recurrent characteristics. It has tremendous adverse impacts to the epileptic patients. Many previous researches explored the pathogenesis of epilepsy from the cellular level to the molecular level and the gene level [1]. Neurosurgery is often required to gain seizure freedom [2]. A successful epileptic surgery highly depends on accurate localization of the origin of epileptic foci, the areas of brain cortex generating the epileptic seizures and understanding postoperative changes in epilepsy network. Unfortunately, localization of epileptic foci is usually very challenging. Invasive surgical intracranial electroencephalography (iEEG) with intracranial electrode placement has been used before neurologic surgery.

Magnetoencephalography (MEG) has been utilized to locate epileptic foci through magnetic field signals using spike signals. MEG has higher temporal and spatial resolution than electroencephalography (EEG) and can be used as input signal to establish brain-computer interface system [3, 4]. However, only 80% of epileptic patients show spikes during MEG recordings, and approximately 50% of epileptic surgeries failed when the brain areas that generate spikes were resected. Recent studies suggest that localized high frequency oscillations (HFOs) detected in MEG recordings are closely linked to the epileptic seizures areas [5, 6] and HFO-generating regions can be used to identify seizure onset zone [7]. Increasing evidence indicate that pathological HFOs are significantly related to the seizure area. The literatures suggest that HFOs reflect the epileptogenic capacity of underlying tissues because HFOs is more frequent

after the reduction of antiepileptic drugs. Although HFOs are primarily recorded on intracranial electroencephalograms, the new study suggests that it is possible to identify HFOs on scalp MEG or EEG [8]. Furthermore, HFOs include ripples (80–250 Hz) and fast ripples (FRs) (250–500 Hz). Recent evidences indicated that FRs are more useful than ripples for localization of epileptogenic zones, particularly true in the case of multiple epileptic foci [9–11]. Thus, it is desirable to identify ripples and FRs in the presurgical evaluation. Nevertheless, compared to spikes, ripples and FRs have short duration and low amplitude, making the visual identification by human experts of HFOs very time consuming, labor-intensive, subjective, and error prone, especially for the large volume of MEG signals data.

Prior researches have applied machine learning to automatic HFO identification in epilepsy studies. Chaibi et al. proposed an automatic algorithm for detection and classification of HFOs, combining smoothed Hilbert Huang Transform (HHT) and root mean square (RMS) feature. Performance evaluation in terms of sensitivity and false discovery rate (FDR) were, respectively, 90.72% and 8.23% [12]. In order to specifically minimize false positive rates and improve the specificity of HFOs detection, they also developed another approach, combining tunable Q-factor wavelet transform (TQWT), morphological component analysis (MCA), and complex Monet wavelet (CMW), improving sensitivity and specificity to 96.77% and 85.00%, respectively [13]. Another study used decision tree analysis for HFOs detection. The results demonstrated that the decision tree approach yielded low false detection (FDR = 8.62 %), but with a sensitivity of 66.96% [14]. Raj et al. used Fishers Linear Discriminant Analysis (FLDA) and logistic regression for classification of HFOs. The accuracy, sensitivity, and specificity of their method were 76.1%, 85.0%, and 66.6%, respectively [15]. Recently, deep learning methods have been applied to HFOs classification. Wan et al. proposed a stacked sparse autoencoder-based HFOs (SMO) detector to distinguish HFOs signals from normal biological signals [16]. In another study, Ting Wana developed Fuzzy entropy (FuzzyEn) and Fuzzy neural network (FNN) for automatic HFOs detection [17]. These studies demonstrated the superior capability of deep learning models on learning latent patterns of biological signal for HFO identification. For CapsuleNet, some studies link it with video content and propose a 3D capsule network for motion detection. And experiments on the UCF-Sports, J-HMDB, and UCF-101 datasets have obtained good results [18]. The literature [19] shows a framework based on CapsuleNet for extracting spectral and spatial features to improve the classification of hyperspectral images. It is proved that the framework can optimize feature extraction and classification. This provides a certain theoretical basis for the implementation of related experiments in this paper.

In this study, we proposed a multiclass MEGNet model to identify ripples and FRs from MEG signals. MEGNet, closely mimicking the biological neural organization, is a specialized artificial neural networks model that improves the learning hierarchical relationships. We hypothesized that MEGNet is able to achieve improved biological signal

classification performance than peer deep neural networks (DNN) as well as traditional machine learning models. In addition, dimension reduction approaches were investigated to couple with MEGNet so as to achieve desirable classification performance of ripple, FRs, and normal control (NC) signals. If our hypothesis is valid, this work may facilitate the presurgical evaluation of epileptic patients.

## 2. Materials and Methods

*2.1. MEG Data and Gold Standard Dataset.* In this study, MEG data was acquired under approval from an Institutional Review Board. We obtained MEG data from 20 clinical epileptic patients (age: 6–60 years, mean age 32; 10 female and 10 male), who were affected by partial seizures arising from one part of the brain. Full details of MEG data acquisition can be found in our prior study [17]. Briefly, MEG recordings were performed using a 306-channel, whole-head MEG system (VectorView, Elekta Neuromag, Helsinki, Finland) in a magnetically shielded room. As one part of presurgical evaluation, sleep deprivation and reduction of antiepileptic drugs were used to increase the chance to capture ripples and FRs during MEG recordings. The sampling rate of MEG data was set to 2,400 Hz, and approximately 60 minutes of MEG data were recorded for each patient.

For identifying MEG system noise, the noise floor which was calculated with MEG data acquired without subject (empty room) was applied in the MEG systems. The noise level was about 3–5 fT/Hz. The empty room measurements were also used to compute noise covariance matrix for localizing epileptic activities (i.e., ripples and FRs). A three-dimensional coordinate frame relative to the subjects head was derived from these positions. The system allowed head localization to an accuracy of 1 millimeters (mm).

The changes in head location before and after acquisition were required to be less than 5 mm for the study to be accepted. For identifying the system and environmental noise, we routinely recorded one background MEG dataset without patients just before the experiment.

MEG data were preliminarily analyzed at a sensor level with MEG Processor [20]. The ripples were visually identified in waveform with a band-pass filter of 80–250 Hz, while FRs were analyzed with band-pass filter of 80–500 Hz. For the model evaluation purpose, the clinical epileptologists selected ripples and FRs signal segments based on intracranial recordings iEEG for these patients. These ripples and FRs coincided with slower spikes in more than 80% of patients [21]. By comparing the MEG sources and the brain areas generating ripples and FRs, the clinical epileptologists marked ripples and FRs. The duration of each signal segment which contains a series of 1000 signal time points is 500 milliseconds. A total of 150 signal segments (50 NC samples, 50 ripples, and 50 FRs) were composed as a gold standard data set for model evaluation.

*2.2. Overview of MEGNet Detector.* As shown in Figure 1, overall MEG data pipeline is composed of four steps: (1) signal segmentation; (2) signal dimension reduction

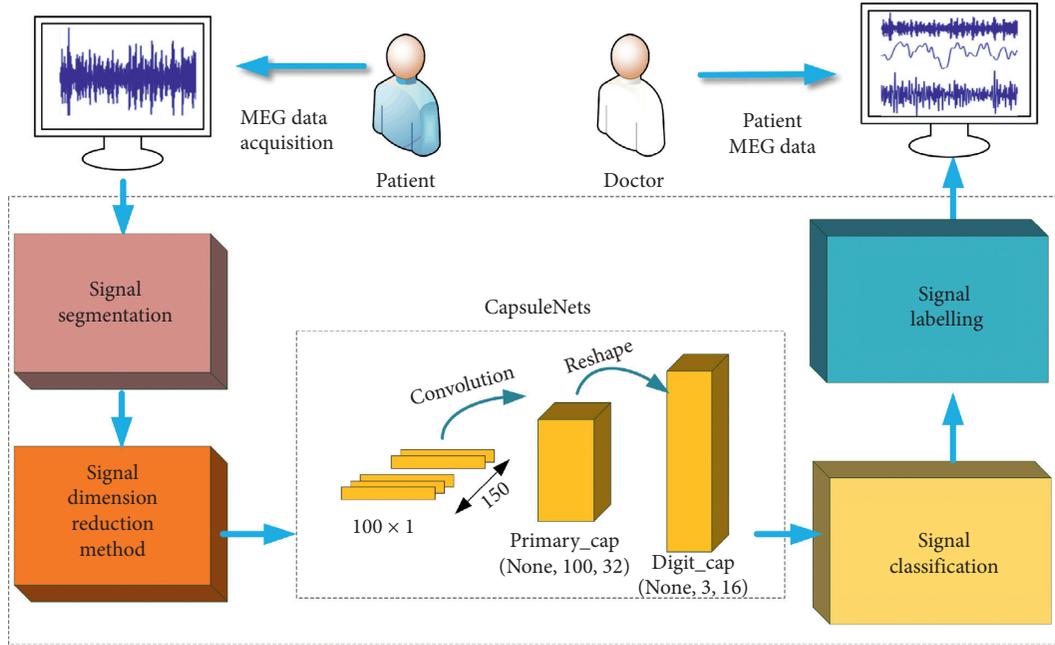


FIGURE 1: Overview of MEGNet detector working for HFOs detection in clinical practice.

method; (3) signal classification; and (4) signal labelling. With this pipeline, MEG data of an epileptic patient can be automatically analyzed and presented to a neurology clinician. Signal segmentation and labelling are simple functions of MEG processing software (e.g., MEG processor). In this work, we detailed the signal dimension reduction and signal classification steps.

**2.3. Signal Dimension Reduction.** Since the gold standard set is a set of 150 samples (50 Normal, 50 Ripple, and 50 FR signals), sample size is less than the dimension of feature 1000 signal points. This may lead to the overfitting of machine learning models. Therefore, we first reduced the dimension of the signal segment. Reducing or eliminating statistical redundancy between the components of high-dimensional vector data enables a lower-dimensional representation without significant loss of information. In this work, we investigated three dimension reduction methods, including principal component analysis (PCA), Kernel Principal Component Analysis (KPCA), and Local Linear Embedding (LLE):

- (1) PCA is a multivariate analysis technique in which dependent variables are determined by the values of several independent variables. Its goal is to extract the important information and represent it as a set of new orthogonal variables called principal components and to display the pattern of similarity of the observations and the variables as points [22, 23].

Assume that all  $m$   $n$ -dimensional data  $(x^{(1)}, x^{(2)}, \dots, x^{(m)})$  have been centralized, that is,  $\sum_{i=1}^m x^{(i)} = 0$ . After projection transformation, the

new coordinate system is  $\{\omega_1, \omega_2, \dots, \omega_n\}$ , where  $\omega$  is the orthonormal basis, that is,  $\|\omega\|_2 = 1, \omega_i^T \omega_j = 0$ .

Reduce the data from  $n$ -dimensional to  $n'$ -dimensional, set the new coordinate system to  $\{\omega_1, \omega_2, \dots, \omega_{n'}\}$ , and project the sample point  $x^{(i)}$  as  $z^{(i)} = (z_1^{(i)}, z_2^{(i)}, \dots, z_{n'}^{(i)})^T$  in the  $n'$ -dimensional coordinate system, where  $z_j^{(i)} = \omega_j^T x^{(i)}$  is the coordinate of the  $j$ th dimension of  $x^{(i)}$  in the  $n'$ -dimensional coordinate system.

For any sample  $x^{(i)}$  is projected as  $W^T x^{(i)}$  in the new coordinate system, and the projection variance is  $x^{(i)T} W W^T x^{(i)}$  in the new coordinate system. To maximize the sum of projection variances of all sample points, that is,  $\arg \max_W \text{tr}(W^T X X^T W)$  s.t.  $W^T W = I$

Using Lagrangian function, we can obtain the following:

$$J(W) = \text{tr}(W^T X X^T W + \lambda(W^T W - I)). \quad (1)$$

Take the derivative of  $W$  and we obtain the following:

$$X X^T W + \lambda W = 0. \quad (2)$$

After finishing the previous formula, we can obtain the following:

$$X X^T W = (-\lambda)W. \quad (3)$$

It can be known from the above formula that  $W$  is a matrix composed of  $n'$  eigenvectors of  $X X^T$ , and  $-\lambda$

is a matrix composed of several eigenvalues of  $XX^T$ , the eigenvalues have the main diagonal, and the remaining positions are 0. When the data is reduced from  $n$ -dimensional to  $n'$ -dimensional, it is necessary to find the largest  $n'$  eigenvalues and corresponding eigenvectors. The matrix  $W$  of these  $n'$  eigenvector matrices is the desired matrix. For the original formula  $z^{(i)} = W^T x^{(i)}$ , the original data set can be dimensionally reduced to the  $n'$ -dimensional data set with the minimum projection distance to complete the dimensional space transformation.

- (2) KPCA is a new method for performing a nonlinear form of Principal Component Analysis. By using integral operator kernel functions, one can efficiently compute principal components in high-dimensional feature spaces, related to input space by some nonlinear map [24].

In the above PCA algorithm reasoning, it is assumed that a linear hyperplane can be used to project the data. But sometimes the data is not linear, and the kernel function idea is used here. The data set can be mapped from  $n$ -dimensionality to linearly separable high-dimensionality  $N$ , and then reduced to  $n'$ -dimensionality, where  $N > n > n'$ . It is assumed here that the data in the high-dimensional space is generated from the data in the  $n$ -dimensional space by mapping  $\varphi$ . Then, the characteristic decomposition of  $n$ -dimensional space is  $\sum_{i=1}^m x^{(i)} x^{(i)T} W = \lambda W$

The mapping is as follows:

$$\sum_{i=1}^m \varphi(x^{(i)}) \varphi(x^{(i)})^T W = \lambda W. \quad (4)$$

Perform the eigenvalue decomposition of the covariance matrix through high-dimensional space, and the following steps are the same as PCA. The mapping  $\varphi$  does not need to be calculated explicitly, but is done through a kernel function when it needs to be calculated. Linear kernel function is used in this study.

- (3) The LLE method is based on simple geometric intuitions: if a data set is sampled from a smooth manifold, then neighbors of each point remain nearby and are similarly collocated in the low-dimensional space [25, 26].

For LLE, the specific implementation steps are as follows: the first step is to calculate the  $k$  nearest neighbors of each sample point. Then, calculate the local reconstruction weight matrix  $W$  of the sample points, and define the reconstruction error:

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2. \quad (5)$$

The local covariance matrix  $C$  is as follows:

$$C_{j,k} = (x - \eta_j) \cdot (x - \eta_k), \quad (6)$$

where  $x$  represents a specific point, and its  $k$  neighbors are represented by  $\eta$ .

The objective function can be expressed as follows:

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2, \quad \sum_j w_j = 1. \quad (7)$$

Minimizing the above formula, we can obtain the following:

$$w_j = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}}. \quad (8)$$

Then, map all the sample points to the low-dimensional space, and the mapping conditions are as follows:

$$\min_Y \Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2. \quad (9)$$

The above formula can be transformed into

$$\Phi(Y) = \sum_{ij} M_{ij} (\vec{Y}_i \cdot \vec{Y}_j), \quad M = (I - W)^T (I - W). \quad (10)$$

The restrictions are as follows:

$$\sum_i \vec{Y}_i = \vec{0}, \quad \frac{1}{N} \sum_i \vec{Y}_i \vec{Y}_i^T = I. \quad (11)$$

Formulas (10) and (11) are centralization and unit covariance.

The final solution is the following:

$$MY = \lambda Y. \quad (12)$$

The eigen decomposition problem is to take the eigenvector corresponding to the smallest  $m$  nonzero eigenvalues with  $Y$  being  $M$ .

It is worth noting that various sampling factors can be optimized according to the sample size and segment dimension of the training data set, in order to obtain the best results. One hundred components were determined as the dimension after the final dimension reduction.

**2.4. MEGNet for Signal Classification.** Next, we described the signals classification using MEGNet. Our goal is to investigate the performance of MEGNet, a compact CapsuleNet architecture for MEG-based signals [27, 28]. The original CapsuleNet was based on image input and finally tag vector input. The architecture of MEGNet is illustrated in Figure 2. The data used in this paper was a time-spatial signal source based on the time axis. Thus, we designed one-dimensional convolution to extract the features of the data set. The developed MEGNet in this study used the convolutional kernel of 1X1 in the first layer.

As shown in Figure 2, the proposed MEGNet structure was fine-tuned in this paper. The purpose of adjusting parameters in the process of feature propagation is to obtain a more suitable data set for this experiment. After the

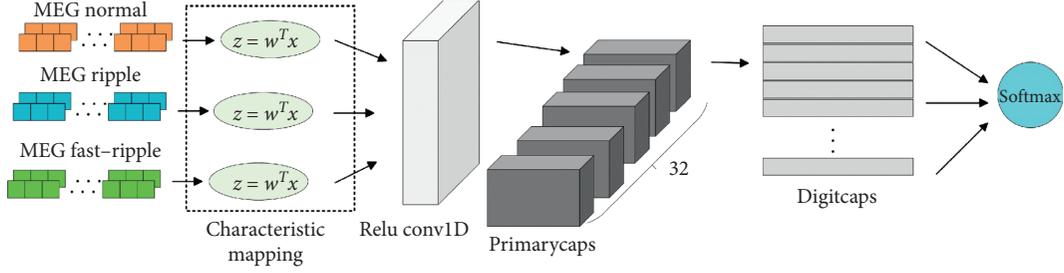


FIGURE 2: MEGNet network structure.

dimensionality reduction, the sample characteristics of the data set were reduced to 100 dimensions. Since this experiment is aimed at biological one-dimensional signals, the input vector in CapsuleNet is  $100 \times 1$ .

Similar to the convolutional neural network approach, MEGNet constructs layered image representations by transferring images across multiple layers of the network. The original CapsuleNet consists of two layers: the first layer is the main cap layer, with capturing low-level clues, and then a specialized secondary cap, which can predict the existence and attitude of objects in the corresponding images. The inputs and outputs of the model are vectors and routing algorithms are used. For the activation function, we used nonlinear compression and the length of the input vector is between 0 and 1. The MEGNet can be expressed by the following formula:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}, \quad (13)$$

where  $v_j$  is the vector output of capsule  $j$  and  $s_j$  is its input. And capsule  $s_j$  is the sum of the input  $\hat{u}_{jli}$ , and in a capsule, the definition of  $u_i$  is related to the matrix weight  $W$ :

$$s_j = \sum_i c_{ij} \hat{u}_{jli}, \hat{u}_{jli} = W_{ij} u_i, \quad (14)$$

where  $c_{ij}$  are coupling coefficients that are determined by the iterative dynamic routing process. The sum of coupling coefficient of capsule  $i$  and its upper capsule is 1, and the priori probability of  $b_{ij}$  of capsule  $i$  is initial:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}. \quad (15)$$

Log prior probabilities can be learned differently from other weights depending on the location and type of each capsule. In the routing algorithm, the routing process is: for capsule  $i$  of  $l$  layer,  $c_i$  can be obtained according to the above formula (3); then,  $s_j$  can be calculated for capsule  $j$  of  $l+1$  layer,  $v_j$  can be obtained through activation function; finally,  $b_{ij}$  can be obtained for capsule  $i$  of  $l$  layer and capsule  $j$  of  $l+1$  layer. During this process,  $b_{ij}$  is approaching to 0. Therefore,  $v_j$  is finally obtained to complete the routing and propagation process.

It is worth mentioning that the conversion weight  $W_{ij}$  is not optimized by conventional back propagation, but by protocol routing algorithm. The main idea of the algorithm

is that the lower-level information capsule sends its input vector to the higher-level information capsule that is more consistent with the input, and then continuously modifies the information by routing protocols to achieve optimized performance. This method can establish a connection between the lower and higher levels of information [27].

The SoftMax layer with cross entropy is used loss function in our model:

$$L = - \sum_x gl(x) \log(pl(x)). \quad (16)$$

The length of the instantiation vector represents the probability that a capsule's entity exists. The  $L_k$  represents each digit capsule using the margin loss. And for the secondary capsule, its margin loss for class  $K$  is defined as follows:

$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda (1 - T_k) \max(0, \|v_k\| - m^-)^2, \quad (17)$$

where  $T_k$  are 1 if an entity of class  $k$  exists and  $m^+ = 0.9$  and  $m^- = 0.1$ .  $m^+$  represents the upper boundary, and when  $v_k > 0.9$ , the predicted value of class  $k$  entity is considered to be above 0.9.  $m^-$  represents the lower boundary, and when  $v_k < 0.1$ , the entity is not considered to exist. We set  $\lambda$  as 0.5 to obtain a more stable and reliable classifier. The final loss is based on the interval loss and the reconstruction loss.

**2.5. Experimental Design.** The main evaluation criteria used in this experiment are accuracy, precision, recall, and F1 score [29–31]. In each repetition of the experiment, we evaluated true positive (TP), false positive (FP), true negative (TN), false negative (FN), and true positive rate (TPR) for the classification by comparing the predicted labels and true labels. Based on the values of TP, TN, FP, and FN, the precision, recall, F1-score, and accuracy are estimated as follows:

$$\begin{aligned} \text{accuracy} &= \frac{TP + TN}{TP + FP + FN + TN}, \\ \text{recall} &= \frac{TP}{(TP + FN)}, \\ \text{precision} &= \frac{TP}{(TP + FP)}, \\ \text{F1-score} &= \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}. \end{aligned} \quad (18)$$

We compared the proposed MEGNet to multiple peer deep learning models, including CNN (convolutional neural network), DNN (deep neural network), and RNN (recurrent neural network).

CNN was originally used for image classification, which is to go through a series of convolutional layers, nonlinear layers, pooling layers, and fully connected layers to obtain the final output result. Now CNN is widely used for natural language segmentation and various classification problems [32, 33].

DNN is a fully connected neuron structure and previously used deep-level principles to solve the problem of local optimal solutions. However, it also brings various parameter problems, which experts and scholars have also improved when using DNN [34–36].

The core content of RNN is the use of long and short-term memory networks, which not only has the characteristics of backpropagation, but can propagate information in both directions in depth. RNN is widely used in speech recognition, natural language recognition, and other fields [37, 38].

The programming language used in this study is python 3, the development integration environment is anaconda 3, the programming framework is TensorFlow 2.0, keras2.1.0, and SKLearn library is used for the data analysis function libraries.

### 3. Results

*3.1. Optimization of Dimension Reduction for HFO Signal Classification.* In different methods of reducing dimension, the effect of capsule network designed in this paper is also different. The results of dimensionality reduction are also compared with those of nondimensionality reduction. The comparison results of different pretreatment methods are shown in Table 1.

Various preprocessing methods are used to remove noise and extract more useful features. For PCA, the most important parameter is the change of dimensionality. In this article, the original dimensionality is determined, and the number of dimensionality input to the model is the result of a large number of experiments. KPCA is based on PCA, setting various kernel functions to get the result. LLE selects the nearest neighbor number to obtain the linear relationship weight coefficient to obtain the distribution of sample points corresponding to the final dimension.

In the table above, different preprocessing methods based on the same classifier can result in different results. When it comes to raw data input directly, precision, and F1 score are only 0.41, and the recall rate is 0.40, accuracy is 0.41. After LLE pretreatment, the precision and recall rate, F1’s score and accuracy are 0.89, 0.89, 0.90, and 0.89. And then after PCA pretreatment, the precision and recall rate, F1-score, and accuracy are 0.95, 0.94, 0.94, and 0.94. From the above comparison results, it is very necessary for the model used in this paper to pretreat. And PCA is a better pretreatment method.

TABLE 1: Performance of HFOs classification with different dimension reduction methods.

Method	Accuracy	Recall	Precision	F1-score
Original	0.41	0.40	0.41	0.41
PCA	<b>0.94</b>	<b>0.94</b>	<b>0.95</b>	<b>0.94</b>
KPCA	0.86	0.87	0.86	0.86
LLE	0.89	0.90	0.89	0.89

The highest metrics were marked as bold.

TABLE 2: The performance of various models was compared.

Method	Accuracy	Recall	Precision	F1-score
1D-CNN	1	1	1	1
CNN	0.89	0.89	0.90	0.89
DNN	0.92	0.92	0.92	0.92
RNN	0.88	0.88	0.88	0.89
CapsuleNet	0.86	0.85	0.85	0.86
MEGNet	<b>0.94</b>	<b>0.94</b>	<b>0.95</b>	<b>0.94</b>

The highest metrics were marked as bold.

*3.2. Comparison of Performance on MEG Classification.* The MEGNet detector was compared with a number of deep learning models, including CNN, DNN, RNN, and CapsuleNet. The MEGNet proposed in this paper uses a capsule network, but the original structure is changed during the initial structure design. For the original image input changed to signal input, and the initial two layers are both one-dimensional convolutional layers, the two-layer capsule layer used before the final output is also the result of many experiments. After a large number of network parameter adjustments, a better model for processing MEG signals is finally obtained.

Table 2 shows the classification performance of the six methods repeated for more than 50 times. In our experiments, our MEGNet detector achieved 95% on precision, 94% on recall, 94% on F1 value, and 94% on accuracy. The accuracy, recall, precision, and F1-score of our MEGNet detector is slightly better than that of DNN, 2%, 2%, 3%, and 2%, respectively. In the test, the 1D-CNN (one-dimensional CNN) model performed worst among the six models and had overfitting. Compared with the original CapsuleNet, MEGNet has improved accuracy, recall, precision, and F1-score by 6%, 6%, 7%, and 5%, respectively.

*3.3. Comparison of Loss among Various Models.* In this paper, MEGNet has obtained good stability and robustness through various tests. Figure 3 shows the changes in the loss values of these six models. It can be seen that in the subsequent iterations of the 1-dimensional CNN, the loss value drops to 0, which indicates that overfitting has occurred. For CNN and DNN, both are also fast in convergence, but the loss value is relatively high after stable. For RNN and CapsuleNet, the loss value is relatively unstable. The MEGNet proposed in this paper has fast and stable convergence, and the final loss value is stable at about 0.04. This shows that the MEGNet detector is more stable and has better performance than other deep learning models.

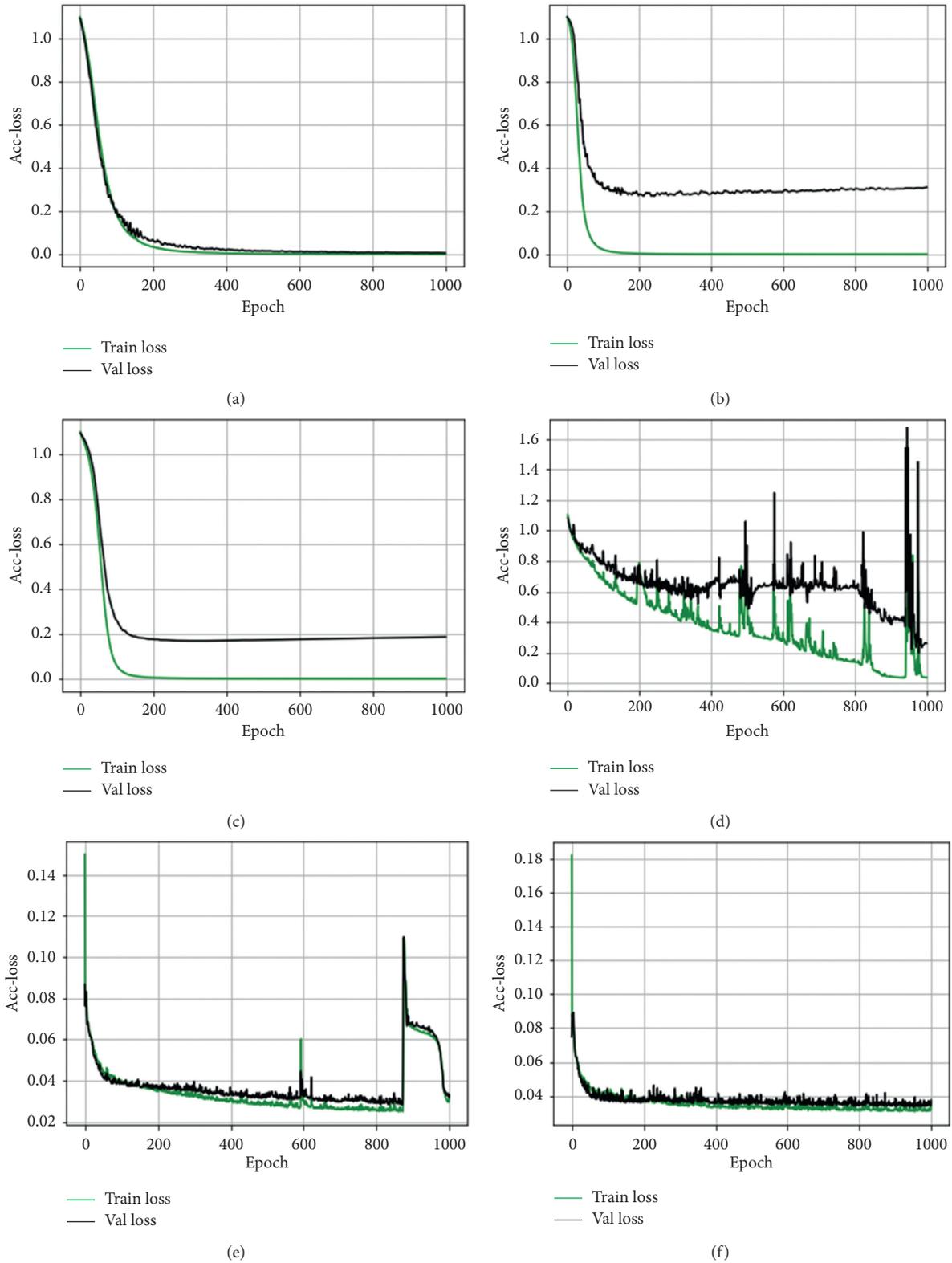


FIGURE 3: The changes in loss values of various models. (a) 1D-CNN. (b) CNN. (c) DNN. (d) RNN. (e) CapsuleNet. (f) MEGNet.

## 4. Discussion

This study proposed a multiclass HFOs detector by combining dimension reduction method and the advanced deep learning MEGNet model. The optimized detector achieved superior performance on classifying MEG HFOs signals into normal control, ripple, FRs. The detector was validated with the gold standard dataset. It may facilitate the clinical assessment of HFOs from preoperation evaluation of epileptic patients.

This study used dimension reduction method to decrease the feature size. Three dimension reduction methods (i.e., LLE, PCA, and KPCA) were compared. PCA had the best performance compared to other two methods. The linear discriminant analysis is not used due to the bed effects in theory and practice. We also compared CNN, DNN, RNN with the MEGNet. Our proposed HFOs detector built based on PCA and MEGNet achieved an accuracy up to 94%. Compared with other models, MEGNet firstly reduced MEG signal, takes advantage of dynamic routing process, layer by layer, features were extracted and reconstructed repeatedly, so as to obtain the optimized model. Therefore, MEGNet could repeatedly use useful feature information to obtain a higher accuracy after dimension reduction.

In the field of computer vision, processing classification, positioning, target detection, etc., the capsule network processing effect is better. Among them, CNN needs a large number of data sets for training to get better results, but the capsule network does not. CNN cannot handle ambiguity well and will lose a lot of information in the pooling layer. However, the capsule network has a better effect on local information feature processing and better shows the local hierarchical structure. Compared with CNN, at present, capsule network applications are not very extensive, and most of the time, the CNN framework is still adopted to solve problems. For the MEGNet proposed in this paper, it is based on the capsule network, but it uses one-dimensional convolution, which is better than two-dimensional convolution in processing serialized data.

This work has limitations. First, the gold standard training data set is small. We have only 150 samples in the gold standard dataset. More data would further improve the generalizability of the proposed CapsuleNet model. Second, the proposed model was tested on our collected gold standard data set. Additional external validation on dataset from other research groups would necessary to test our model. Third, single modality data (i.e., MEG) was utilized in this work. Multi-modality (i.e., concurrent EEG recording) may improve the identification of HFOs.

## 5. Conclusion

In summary, this paper proposed a new method that used the analysis of MEG during epileptic activity and develops a MEGNet detector to detect HFO in MEG signals. This paper first proposed a deep learning framework for HFO detection based on MEGNet detector and optimizes the detector. We have proved that based on this research, this method can accurately detect the area of epileptic activity with good

specificity. At the same time, we have other potential research directions. For example, the detector obtained in this article can finally be applied to EEG signals and further compared with existing EEG methods.

## Data Availability

The data in this paper cannot be made available publicly.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61906139, 61172150, and 61803286), Hubei Provincial Natural Science Foundation of China (Grant 2019CFB173), the Foundation of Hubei Provincial Key Laboratory of Intelligent Robot (HBIR 201802), Industry-University-Research Innovation Fund of Science and Technology Development Center, MOE (2019ITA03029) and the eleventh Graduate Innovation Fund of Wuhan Institute of Technology (CX2019240 and CX2019241).

## References

- [1] R. Kale, "Bringing epilepsy out of the shadows," *BMJ-British Medical Journal*, vol. 315, no. 7099, pp. 2-3, 1997.
- [2] A. J. Durnford, W. Rodgers, F. J. Kirkham et al., "Very good inter-rater reliability of Engel and ILAE epilepsy surgery outcome classifications in a series of 76 patients," *Seizure*, vol. 20, no. 10, pp. 809-812, 2011.
- [3] A. Bagić, "Look back to leap forward: the emerging new role of magnetoencephalography (MEG) in nonlesional epilepsy," *Clinical Neurophysiology*, vol. 127, no. 1, 2015.
- [4] J. S. Ebersole, "New applications of EEG/MEG in epilepsy evaluation," *Epilepsy Research. Supplement*, vol. 11, no. 11, pp. 227-237, 1996.
- [5] A. Bragin, J. Engel, C. L. Wilson, I. Fried, and G. Buzski, "High-frequency oscillations in human brain," *Hippocampus*, vol. 9, no. 2, pp. 137-142, 1999.
- [6] A. Bragin, J. Engel, C. L. Wilson, I. Fried, and G. W. Mathern, "Hippocampal and entorhinal cortex high-frequency oscillations (100-500 Hz) in human epileptic brain and in kainic acid-treated rats with chronic seizures," *Epilepsia*, vol. 40, no. 2, pp. 127-137, 1999.
- [7] A. L. Andrade-Valen, "Interictal high frequency oscillations (HFOs) in patients with focal epilepsy and normal MRI," *Clinical Neurophysiology*, vol. 123, no. 1, pp. 100-105, 2012.
- [8] J. Jacobs, R. Staba, E. Asano et al., "High-frequency oscillations (HFOs) in clinical epilepsy," *Progress in Neurobiology*, vol. 98, no. 3, pp. 302-315, 2012.
- [9] T. Akiyama, B. McCoy, C. Y. Go et al., "Focal resection of fast ripples on extraoperative intracranial EEG improves seizure outcome in pediatric epilepsy," *Epilepsia*, vol. 52, no. 10, pp. 1802-1811, 2011.
- [10] A. Bragin, I. Mody, C. L. Wilson, and J. Engel, "Local generation of fast ripples in epileptic brain," *The Journal of Neuroscience*, vol. 22, no. 5, pp. 2012-2021, 2002.

- [11] K. M. A. Van and V. K. N. E. Klink, "Residual fast ripples in the intraoperative corticogram predict epilepsy surgery outcome," *Neurology*, vol. 85, no. 2, pp. 120–128, 2015.
- [12] S. Chaibi, Z. Sakka, T. Lajnef, M. Samet, and A. Kachouri, "Automated detection and classification of high frequency oscillations (HFOs) in human intracerebral EEG," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 927–934, 2013.
- [13] S. Chaibi, T. Lajnef, Z. Sakka, M. Samet, and A. Kachouri, "A reliable approach to distinguish between transient with and without HFOs using TQWT and MCA," *Journal of Neuroscience Methods*, vol. 232, pp. 36–46, 2014.
- [14] S. Chaibi, "Detection of high frequency oscillations (hfos) in the 80-500 Hz range in epilepsy recordings using decision tree analysis," in *Proceedings of the International Image Processing, Applications and Systems Conference*, Genova, Italy, November 2015.
- [15] K. R. Raj, S. S. Rajagopalan, S. Bhardwaj et al., "Machine learning detects EEG microstate alterations in patients living with temporal lobe epilepsy," *Seizure*, vol. 61, pp. 8–13, 2018.
- [16] T. Wan, "Automated detection of high frequency oscillations based on Fuzzy entropy and Fuzzy neural network," in *Proceedings of the 2016 35th Chinese Control Conference (CCC)*, Chengdu, China, July 2016.
- [17] J. Guo, K. Yang, H. Liu et al., "A stacked sparse autoencoder-based detector for automatic identification of neuromagnetic high frequency oscillations in epilepsy," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2474–2482, 2018.
- [18] K. Duarte, Y. S. Rawat, and M. Shah, "Videocapsulenet: a simplified network for action detection," 2018, <https://arxiv.org/abs/1805.08162>.
- [19] B. K. P. A. Arun P, "Capsulenet-based spatial-spectral classifier for hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 99, pp. 1–17, 2019.
- [20] L. Tang, "Neuromagnetic high-frequency oscillations correlate with seizure severity in absence epilepsy," *Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology*, vol. 127, no. 2, pp. 1120–1129, 2015.
- [21] R. Kotecha, "Modeling the developmental patterns of auditory evoked magnetic fields in children," *PLoS One*, vol. 4, 2009.
- [22] P. Cunningham, "Dimension reduction," 2008.
- [23] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [24] B. Scholkopf, "Kernel principal component analysis," in *Proceedings of the 7th International Conference on Artificial Neural Networks*, Lausanne, Switzerland, January 1997.
- [25] D. L. Donoho and C. Grimes, "Hessian eigenmaps: locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.
- [26] J. Wang, *Locally Linear Embedding*, SpringerLink, New York, NY, USA, 2012.
- [27] A. Jiménez-Sánchez, S. Albarqouni, and D. Mateus, *Capsule Networks against Medical Imaging Data Challenges*, MICCAI, Granada, Spain, 2018.
- [28] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," 2017, <https://arxiv.org/abs/1710.09829>.
- [29] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Proceedings of the 27th European conference on Advances in Information Retrieval Research*, vol. 3408, no. 5, pp. 345–359, Santiago de Compostela, Spain, April 2005.
- [30] S. H. F. E. Radhakrishna Achanta, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009.
- [31] Y. J. Huang, R. Powers, and G. T. Montelione, "Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics," *Journal of the American Chemical Society*, vol. 127, no. 6, pp. 1665–1674, 2005.
- [32] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Doha, Qatar, October 2014.
- [33] H.-C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [34] A. Canziani, A. Paszke, and E. Cukurciello, "An analysis of deep neural network models for practical applications," *27th Signal Processing and Communications Applications Conference*, Sivas, Turkey, April 2016.
- [35] C. A. Dan, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, vol. 32, pp. 333–338, 2012.
- [36] D. Li, G. Hinton and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," in *IEEE International Conference on Acoustics*, May 2013.
- [37] K. Cho, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014, <https://arxiv.org/abs/1406.1078>.
- [38] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE International Conference on Acoustics*, May 2013.