Received May 10, 2021, accepted May 23, 2021, date of publication May 28, 2021, date of current version June 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084635

Early Prediction of Epileptic Seizure Based on the BNLSTM-CASA Model

MENGNAN MA[1], YINLIN CHENG[1], YAO WANG[1], XINGYU LI[2], QIANXIANG MAO[1], ZHEN ZHANG[3], ZIYI CHEN[4], AND YI ZHOU[2,5]

[1]School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 510006, China
[2]Department of Medical Informatics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China
[3]Department of Software Engineering, School of Computer Science and Engineering, Huizhou University, Huizhou 516000, China
[4]Department of Neurology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China
[5]Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou 510080, China

Corresponding author: Yi Zhou (zhouyi@mail.sysu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0116900 and Grant 2016YFC0901602, in part by the National Natural Science Foundation of China (NSFC) under Grant 61876194, in part by the Key Research and Development Program of Guangdong under Grant 2018B010109006, in part by the Science and Technology Program of Guangzhou under Grant 202011020004, in part by the 111 Project under Grant B12003, and in part by the Science and Technology Innovation Special Project of Guangdong Province, China, under Grant 202011020004.

ABSTRACT Epilepsy is one of the world's most common neurological diseases. Reliable early prediction and warning of seizures can provide timely treatment for patients with epilepsy, and improve their quality of life. Compared with most hand-designed prediction methods, an automatic prediction model that can process the original electroencephalogram (EEG) signals directly and take into account the leads optimization problem is needed. In this paper, we proposed an end-to-end automatic seizure prediction model based on the Batch Normalization Long Short Term Memory networks (BNLSTM) and Channel and Spatial attention (CASA). Firstly, raw EEG signals without any preprocessing are used as the input to the system, which can reduce the computation amount. Secondly, BNLSTM and CASA retained the time and spatial information of the raw EEG data respectively. Channel attention (CA) achieved the automatic optimization of EEG full-lead data and improved the prediction accuracy. Spatial attention (SA) achieved the adaptive learning of feature parameters. Finally, a fully connected layer is applied to predict the seizures. The performance of the seizure prediction model we proposed is evaluated on the data of 14 patients with Area Under the Curve (AUC) of 0.986, accuracy (Acc) of 0.956, specificity (Spe) of 0.968, and sensitivity (Sen) of 0.942. In addition, the proposed method provided an accurate prediction for all 50 seizures of the other 5 patients in the generalization dataset. Experimental results show that the proposed model has a certain generalization performance, which can provide a reliable basis for early warning of epileptic seizures.

INDEX TERMS Epilepsy, electroencephalogram, seizure prediction, BNLSTM, CASA.

I. INTRODUCTION

Epilepsy is a chronic cerebral dysfunction syndrome. Nearly 65 million people worldwide suffered from epilepsy, accounting for about 1% of the world's population [1]. The seizures of the patients with epilepsy are transient, repetitive, and unpredictable. Patients spend most of their lives unsure of when and where epilepsy will occur [2]–[5]. Uncontrollability is a major problem with epilepsy. About a third of patients have drug-resistant epilepsy. Therefore, it is of great practical significance to design a reliable epileptic seizure prediction algorithm and administer drugs in the early warning period [6].

In the past few decades, it has become a research hotspot to detect different seizure periods based on the EEG of epilepsy patients [7]–13]. The brain activity of patients with epilepsy is different between the interictal and the ictal period. As the brain activity progresses from one state to another, the brain's electrical signals change dramatically. At present, the common used detection algorithm theories include nonlinear dynamics [14], machine learning [15], and deep learning [16], etc, and good research results have been achieved. However, far fewer researches have been conducted on seizure prediction because the high similarity makes it challenging to

The associate editor coordinating the review of this manuscript and approving it for publication was Taous Meriem Laleg-Kirati.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

distinguish between the pre-ictal state and the inter-ictal state. Considering the limitations of the detection algorithms, many researchers have focused on accurate epileptic prediction.

The difficulty of seizure prediction lies in the identification of pre-ictal and the location of epileptic seizure prediction points [17]. With the development of deep learning, neural network models are gradually favored by more researchers. Among them, recurrent neural networks (RNN) and convolutional neural networks (CNN) based on EEG are research hotspots in recent years [18]–[21]. The former focuses on the context of time series, while the latter focuses more on the EEG feature extraction. Zhang *et al.* [22] proposed a seizure prediction solution by using Common Spatial Pattern (CSP) and Convolutional Neural Network. The wavelet packet decomposition and common spatial pattern feature extractor were used to extract the time domain and frequency domain features of the reconstructed two-dimensional EEG. Finally, a shallow CNN was used to distinguish between inter-ictal and pre-ictal. The model was evaluated on 23 patients' from the MIT Scalp EEG dataset of Boston Children's Hospital with a sensitivity of 92.2%, and an error prediction rate of $0.12/h$. Truong *et al.* [23] realized the prediction of epileptic seizure based on Fourier transform (FFT) and convolutional neural networks. Truong used the short-time Fourier transform (STFT) on the 30s window of EEG to extract the frequency domain and time domain information, and sent the obtained time-frequency map as input to the neural network for model training. On the MIT Scalp Brain at Boston Children's Hospital, the sensitivity of 81.2% and the false alarm rate of $0.16/h$ were achieved. Convolutional neural networks often use the reconstructed EEG images in epileptic seizure prediction, but with the changes of the dimensions of EEG data, important information may be lost. The existence of the pooling layer will also lead to the loss of a lot of valuable information, as well as the neglect of whole and part relationships. Therefore, the research on the model based on the convolution neural networks needs to be improved.

Different from the convolutional neural network, the epileptic seizure prediction models based on the recurrent neural network can directly learn the original EEG data to ensure the maximum time domain information retention of EEG [24]–[26]. Tsiouris *et al.* [8] extracted the original EEG information based on the feature extraction methods and outputted prediction results through the Long Short-Term Memory (LSTM). The model was evaluated on the dataset of 23 patient from the MIT Scalp EEG dataset of Boston Children's Hospital. The sensitivity was 99.28%, the specificity was 99.28%, and the false alarm rate was $0.107FP/h$. Petrosian *et al.* [27] combined the wavelet packet decomposition and LSTM to realize the prediction of epileptic seizure. By inputting the original EEG into the recurrent neural network after the wavelet decomposition, the author finally proved that it is feasible to identify the pre-ictal within a few minutes on the dataset of two patients who were undergoing long-term electrophysiological monitoring for epilepsy. Because the network parameters of each loop layer are shared, problems such as gradient explosion or gradient disappearance will occur with the increase of RNN depth, which is also the key to the further optimization of the RNN model.

We want to emphasize the important role of EEG leads in the epileptic seizure prediction models. Many articles have focused on the analysis of single-lead or multi-lead EEG data. This analytical methods have limitations and cannot well reflect the interrelationship of each lead during the seizure period. The current research trend is to focus on the temporal and spatial information of EEG in order to comprehensively guide the analysis of EEG. Wei *et al.* [28] proposed a long-term recurrent convolutional network (LRCN) prediction model by using the full-lead data of the EEG. The team converted the EEG time series into a two-dimensional EEG, performed multi-channel fusion into a three-dimensional structure, used the CNN to learn the spatial features, and LSTM to retain the timing information. On the private dataset of 15 patients with epilepsy. (5 males and 10 females, aged 6–51 years), the deep seizure prediction model achieved an accuracy of 93.40%, prediction sensitivity of 91.88%, specificity of 86.13% in segment-based evaluations and a false prediction rate (FPR) of 0.04 F P/h. However, the limitations lie in the optimization of EEG leads and the determination of different lead weights. Attention Model (AM) can integrate the mutual relationship between EEG leads and automatically optimize the weights of different EEG leads to achieve model optimization.

In this paper, we considered the full-lead EEG data and the relationship between the different EEG leads. In order to realize the prediction model of epileptic seizure, we proposed to use BNLSTM [29] to process and predict EEG time series. On the basis of the BNLSTM architecture, we added the CASA module to optimize the model. Among CASA [30], channel attention (CA) [31] can perform secondary weighting on the weights of different EEG leads in the channel dimension and integrate the correlation between the leads. Spatial attention (SA) [32] can extract the features of EEG data at the convolution level, so as to preserve the spatiotemporal information of EEG as much as possible. Therefore, a seizure prediction model based on the combination of BNLSTM and CASA was established in this paper.

The main contributions of our work are the following aspects: (1) We propose a novel seizure prediction model based on the BNLSTM and CASA architecture, which can directly process the original EEG of all leads and retain as much temporal and spatial information of EEG signals as possible. The first part is the BNLSTM module which automatically extract the features of EEG signals, and outputs the feature matrix. Then the CASA module and the fully connected layer are constructed to predict the oncoming seizure. (2) We use the CASA module to realize the automatic optimization of the weight between the full leads and improved prediction accuracy. Channel attention enables automatic optimization of the weight between full leads.

Spatial attention realizes the adaptive learning of feature parameters and improves prediction accuracy. (3) The performance of the seizure prediction model we proposed is evaluated on the 14 patients' data. The Area Under the Curve (AUC) is 0.986, Accuracy (Acc) is 0.956, Specific (Spe) is 0.968, and Sensibility (Sen) is 0.942. In addition, the proposed method provides an accurate prediction for all 50 seizures of other 5 patients in the generalization dataset. The experimental results outperform most state-of-the-art seizure prediction methods in recent literature.

The rest of this article is organized as follows: Section 2 describes the details of the data we used. Section 3 describes the details of our proposed method, including the overall prediction model framework and the BNLSTM and CASA modules. Section 4 describes the details of our model training and the design of the experiment. Section 5 describes the results of our experiment. Finally, Section 6 is the conclusion of the whole article.

II. DATA

A. DATA RESOURCES

In our experiment, we used two datasets. All the EEG data of the epilepsy patients came from the Neurology EEG Center of the First Affiliated Hospital of Sun Yat-sen University and the First Affiliated Hospital of Xinjiang Medical University from 2013 to 2016. In order to reduce the interference of physical activity on EEG signals, epileptic patients were placed in a resting state. The scalp electrodes were placed according to the international 10–20 system. The synchronized recording of 22-lead EEG signals was performed by the bipolar lead method. The sampling frequency was 500 Hz.

The first dataset is shown in table 1. For the first dataset, we randomly select 70% as the training set and 30% as the test set. The training set is used to train the model, and the test set is used to test the overall performance of our model. The second dataset, as shown in Table 2, is a generalization dataset that was collected separately to verify the generalization performance of our prediction model. Table 1 shows the training dataset, consists of 14 patients, including 5 males and 9 females, aged 6 to 51 years. The total duration of the available EEG recordings is approximately 360h. The time nodes of the onset and end of the seizures were manually annotated by clinical experts after visual inspection. A total of 154 seizures were recorded, with an average of 11 seizures per person. The duration of the ictal period was 9394 seconds (approximately 156 minutes).

Table 2 shows the generalization performance verification dataset (the generalization dataset) consists of 5 patients, including 2 males and 3 females, aged 15 to 39 years. The total duration of available EEG recordings is approximately 120h, with a total of 50 seizures, and an average of 10 seizures per person.

B. DATA PREPROCESSING

The method we proposed is based on the raw data of all leads in the EEG. In the data preprocessing step, we did not perform manual feature extraction. Since there were some unavoidable human errors or accidents in data collection, we cleaned our data to eliminate the influence of incorrect data before performing the EEG analysis.

TABLE 1. Characteristics of the collected training data.

ID	Gender	Age	Consciousness	Monitoring time	Number	Seizure time
1	F	36	wake⟶sleep	24h	14	654s
2	F	22	wake⟶sleep	48h	12	274s
3	F	40	wake⟶sleep	24h	6	302s
4	M	6	wake⟶sleep	24h	21	453s
5	F	16	wake⟶sleep	24h	7	329s
6	F	16	wake⟶sleep	24h	8	254s
7	F	28	wake⟶sleep	24h	5	400s
8	F	31	wake⟶sleep	24h	9	423s
9	M	51	wake⟶sleep	24h	30	1064s
10	M	20	wake⟶sleep	24h	19	4072s
11	M	46	wake⟶sleep	24h	6	208s
12	F	15	wake⟶sleep	24h	8	137s
13	F	28	wake⟶sleep	24h	5	824s
14	M	39	wake⟶sleep	24h	4	895s

PS:F:female, M:male Number: the number of seizures.

TABLE 2. Characteristics of the collected test data.

ID	Gender	Age	Consciousness	Monitoring time	Number	Seizure time
15	M	19	wake⟶sleep	24h	13	981s
16	F	16	wake⟶sleep	24h	14	405s
17	F	39	wake⟶sleep	24h	13	503s
18	M	32	wake⟶sleep	24h	5	106s
19	F	15	wake⟶sleep	24h	5	342s

PS:F:Female, M:Male Number: the number of seizures.

1) DEFINITION OF SEIZURE PERIOD

As shown in Figure 1, we focus on three stages of patients with epilepsy: interictal period, pre-ictal period and ictal period. For the ictal period, the clinical experts have marked the start and end points of the ictal. In order to predict the seizure of epilepsy, accurate pre-ictal recognition is the key to the seizure prediction model. There is no fixed time for the pre-ictal period, it is basically set artificially. Some studies chose a fixed pre-ictal duration of 20 minutes to 90 minutes [33]. In this study, EEG signal within half an hour before the ictal of each patient were uniformly selected as the pre-ictal EEG data screening based on the data situation. EEG data from one hour before the ictal and one hour before the next ictal of each patient were selected as the interictal data [34].

2) DATA SEGMENTATION

The previous section introduced the three categories of EEG. In each category, the data was selected according to the criteria for division, and the time window for extraction was 5000 points, that is, 10 seconds. For the ictal data, as shown in Figure 2, we performed a 500-point slid extraction (that is, 1 second). In consideration of category imbalance data, the interictal and pre-ictal EEG data were extracted by non-overlapping time windows. Based on the data segmentation, three categories of EEG signal segments were obtained. Even though we have taken the class imbalance into account in data segmentation, the data of the ictal was still in the

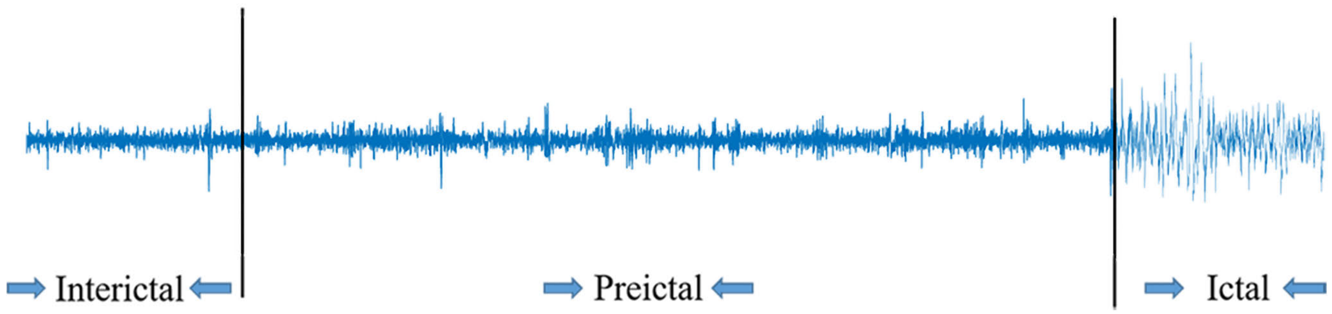


FIGURE 1. Definition of the different periods of epileptic seizure.

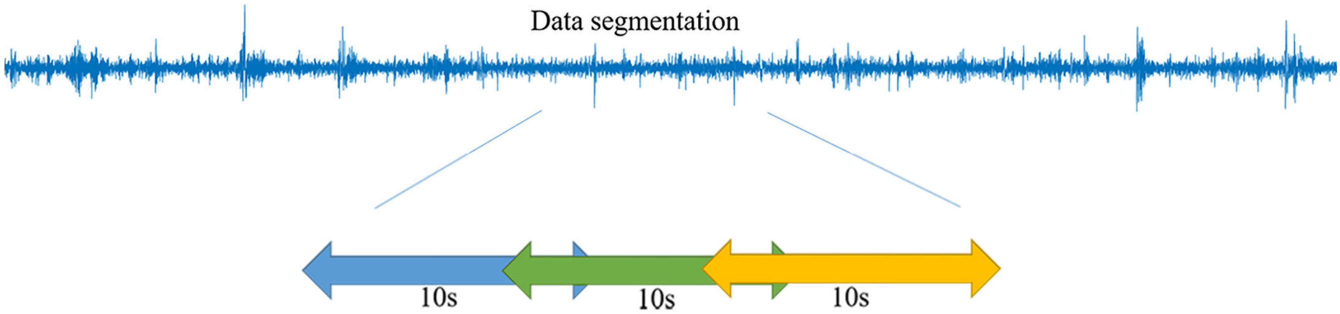


FIGURE 2. Data segmentation with 1s sliding window.

class imbalance state compared with the interictal period and the pre-ictal period. We randomly screened the EEG data segments of the interictal and pre-ictal phases with the ratio of1:1:1. In the end 23,536 EEG segments were obtained with a size of 5000×22.

III. METHODS

After pre-processing the data, this section mainly introduces our epileptic seizure prediction model. First, we will introduce the overall framework structure of our proposed model, including the whole process from data input, model training, and output. In addition, we also introduce some of our understanding of BNLSTM and Attention.

A. THE FRAMEWORK OF THE PREDICTION MODEL

In this study, a seizure prediction model based on the combination of BNLSTM and CASA was proposed. Firstly, the full-lead EEG data was pre-processed to obtain three kinds of corresponding EEG segments. Then, the three kinds of EEG segments were labeled with corresponding tags as the input of BNLSTM. The BNLSTM module can learn the relationship between time series from all-lead EEG segments, automatically extract the features of EEG signals, and output the feature matrix. In addition, the feature matrix output by BNLSTM is used as the input of the CASA module. In the CASA module, Channel Attention can automatically optimize the weight parameters of all leads, and Spatial Attention can perform adaptive learning of features. Finally, the fully connected layer and the softmax function outputted the results and determine whether the current EEG segment belonged to the pre-ictal period. The seizure prediction model in our paper is shown in figure 3.

In this study, we combined the BNLSTM module with the Attention module to establish a spatio-temporal adaptive lead-optimized epileptic seizure prediction model. The traditional recurrent neural network took the input data of the time series through the hidden layer to get the feature extraction, and then directly outputted it to the fully connected layer. In view of the fact that the EEG-lead optimization problem was not present in the traditional prediction model, we connected the CASA module after the BNLSTM module, which can be adaptively optimized and learn the feature matrix from both the channel and spatial dimensions. First, the EEG data at t time step passed through the BNLSTM module to output a feature matrix $F \in \mathbb{R}^{C*S}$, where C is the channel, S is the processed one-dimensional matrix. Then the feature matrix was inputted to the Attention module. In order to make the network pay more attention to meaningful information, Channel Attention used the interdependence between the channels to perform the quadratic weighting on the 22-lead weight parameters, so as to achieve the automatic optimization of the weight parameters of all leads. By inputting the optimized feature matrix into spatial attention, adaptive feature learning can be carried out for the parameter features of spatial dimension. Finally, after the fully connected layer, we used "softmax" to determine the period. In this process, the BNLSTM module retained the time information of the EEG signal, while the CASA module optimized the spatial

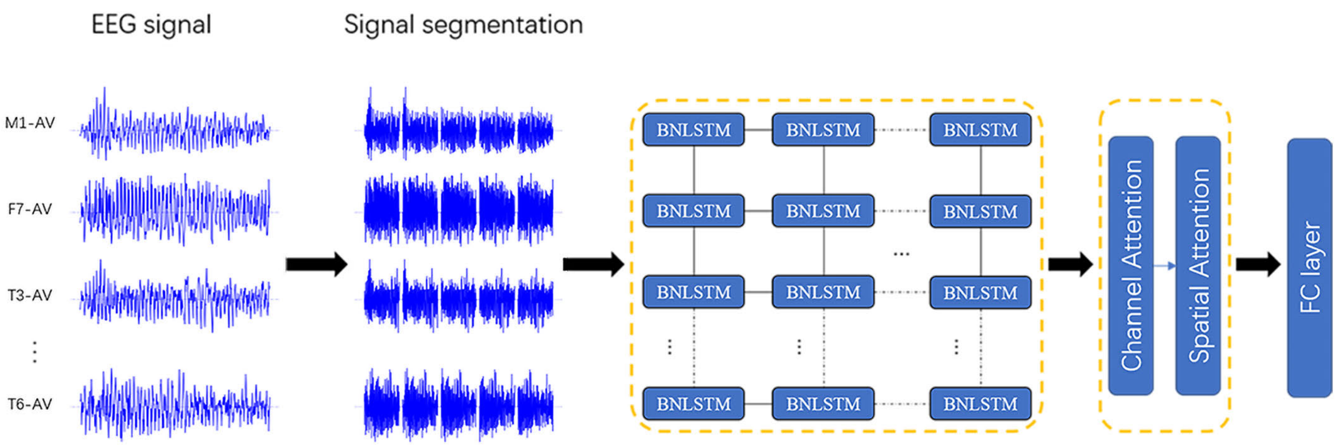


FIGURE 3. The BNLSTM–CASA epileptic seizure prediction model framework.

information of the EEG signal, and then made an epileptic seizure prediction alert based on the temporal and spatial information.

B. BNLSTM

BNLSTM [29] brings the benefits of batch standardization to the recurrent neural networks. Unlike the previous application of batch normalization to the input-to-hidden transformation of the recurrent neural networks, BNLSTM uses the batch standardization to the hidden-to-hidden transformation of the recurrent neural networks. This new neural network accelerates the training of neural networks by reducing internal covariate shifts, which makes the model have faster convergence speed and greater generalization ability. The implementation process of BNLSTM is as follows:

Firstly, the Long Short Term Memory networks (LSTM) can be expressed as:

$$\begin{pmatrix} \tilde{f}_t \\ \tilde{i}_t \\ \tilde{o}_t \\ \tilde{g}_t \end{pmatrix} = W_h h_{t-1} + W_x X_t + b \tag{1}$$

$$c_t = \sigma(\tilde{f}_t) \odot c_{t-1} + \sigma \tilde{i}_t \odot tanh(\tilde{g}_t) \tag{2}$$

$$h_t = \sigma(\tilde{o}_t) \odot tanh(c_t) \tag{3}$$

where $W_h \in \mathbb{R}^{d_h \times 4d_h}$, $W_x \in \mathbb{R}^{d_x \times 4d_h}$, $b \in \mathbb{R}^{4d_h}$ and the initial states $h_0 \in \mathbb{R}^{d_h}$, $c_0 \in \mathbb{R}^{d_h}$ are model parameters. σ is the logistic sigmoid function, and the $\odot$ operator denotes the Hadamard product. In our paper, d_h is 500, d_x is 2000.

Then, the Batch Normalization (BN) can be expressed as:

$$BN(h; \gamma, \beta) = \beta + \gamma \odot \frac{h - \mathbb{E}[\hat{h}]}{\sqrt{\hat{Var}[h] + \epsilon}} \tag{4}$$

where $h \in \mathbb{R}^d$ is the vector of activations to be normalized, $\gamma \in \mathbb{R}^d$, $\beta \in \mathbb{R}^d$ are model parameters that determine the mean and standard deviation of normalized activation, and $\epsilon \in \mathbb{R}$ is a regularization hyperparameter. The division should be understood to proceed elementwise.

Finally, the BNLSTM can be expressed as:

$$\begin{pmatrix} \tilde{f}_t \\ \tilde{i}_t \\ \tilde{o}_t \\ \tilde{g}_t \end{pmatrix} = BN\,(W_h h_{t-1}; \gamma_h, \beta_h) + BN\,(W_x X_t; \gamma_x, \beta_x) + b \tag{5}$$

$$c_t = \sigma(\tilde{f}_t) \odot c_{t-1} + \sigma \tilde{i}_t \odot tanh(\tilde{g}_t) \tag{6}$$

$$h_t = \sigma(\tilde{o}_t) \odot tanh(BN\,(c_t; \gamma_c, \beta_c) \tag{7}$$

BNLSTM uses batch standardization for both the input layer to hidden layer and the hidden layer to hidden layer transformations of the recurrent neural network, which makes the model have faster convergence and greater generalization capabilities. Formula 5 shows that BNLSTM normalizes the recurrent term $W_h h_{t-1}$ and the input term $W_x X_t$ separately.

C. ATTENTION MODULE

The attention module used in this study was an improved dual attention mechanism based on the Convolutional Block Attention Module (CBAM) proposed by S. Woo [30]. Different from the feature map $F \in \mathbb{R}^{C*H*W}$ of the three-dimensional middle layer of CBAM, where C is the channel, $H * W$ is the compressed image of the original picture, considering the change of the EEG time series dimension, we replace conv2d in the convolution module with conv1d, and the middle layer feature map of our CASA module was: $F \in \mathbb{R}^{C*S}$. Where Channel attention map was M_c, Spatial attention map was M_s. The entire process of the CASA module was as follows:

$$F' = M_c(F) \otimes F \tag{8}$$

$$F'' = M_s(F') \otimes F' \tag{9}$$

where $\otimes$ is element-wise multiplication, F'and F''are the intermediate variables of the feature matrix F.

1) CA

After the time sequence learning of the BNLSTM module, the preprocessed EEG data segments outputted

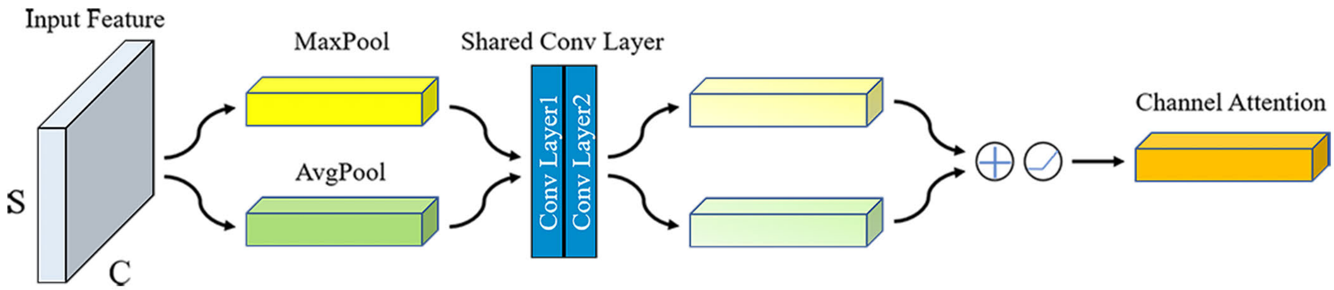


FIGURE 4. The schematic diagram of Channel attention.

TABLE 3. The parameters of Channel attention.

Hidden layer	Related parameters (in_kernel,out_kernel,kernel size,stride)
AdaptiveAvgPool1d	1
AdaptiveMaxPool1d	1
Conv1d+BN+ReLU	$22 \times 1 \times 1 \times 1$
Conv1d+Sigmoid	$22 \times 1 \times 1 \times 1$

TABLE 4. The parameters of Spatial attention.

Hidden layer	Related parameters (in_kernel,out_kernel,kernel size,stride)
AdaptiveAvgPool1d	1
AdaptiveMaxPool1d	1
Conv1d+Sigmoid	$2 \times 1 \times 7 \times 1 \times 3$

a $C * S$ feature matrix. $F \in \mathbb{R}^{C*S}$was inputted into Channel attention [31], [32] as the input matrix of the CASA module. Figure 4 is the schematic diagram of Channel attention. For the feature matrix outputted by BNLSTM, there are a total of 22 channel dimensions. The CA can compress the feature map feature matrix based on the channel dimension of the feature map, and compress the feature matrix into a weight coefficient. And the dimension of the weight coefficient matrix outputted by the module is consistent with the number of channels of the input feature map.

For efficient calculation efficiency, we used maximum pooling and average pooling to compress the feature matrix in spatial dimensions, and obtained two different spatial background descriptions. After that, the feature matrix map $\mathbb{R}^{C*1}$ of Channel attention was obtained by sharing the convolutional layer. Channel attention was used to assign different weights to each input channel C, and to extract more critical lead information from the EEG signal, and to realize the automatic optimization of all leads of the EEG signal. Table 3 shows the parameters of CA.

For this study, the channel attention module generated corresponding weight coefficients for each EEG lead characteristic channel by setting parameters to reflect the correlation between EEG lead channels. And CA used the final input weight matrix as the importance of the EEG lead channel, so as to realize the re-calibration of the EEG lead characteristics in the channel dimension and complete the lead optimization. The specific process can be described as:

$$M_c(F) = \sigma(Conv(AvgPool(F)) + Conv(MaxPool(F))) \quad (10)$$

$$\sigma(\cdot) = sigmoid(\cdot) \quad (11)$$

2) SA

Figure 5 is the schematic diagram of spatial attention [32]. The spatial attention module of this study used the convolutional layer to perform adaptive feature learning on each channel input $\mathbb{R}^{1*S}$, which makes the network pay more attention to the meaningful information and improve the accuracy of epileptic seizure prediction. Table 4 shows the parameters of SA. The specific process can be described as:

$$M_s(F) = \sigma(Conv(concat([AvgPool(F), MaxPool(F)]))) \quad (12)$$

$$\sigma(\cdot) = sigmoid(\cdot) \quad (13)$$

IV. TRAINING AND EXPERIMENTAL DESIGN

The method we proposed has certain advantages for unbalanced data. In addition, we introduced another solution in data segmentation. Taking into account the impact of the category imbalance data, the interictal and pre-ictal EEG data were extracted by non-overlapping time windows. Based on the data segmentation, we obtained three categories of EEG. Even though we have taken the class imbalance into account in data segmentation, the data of the ictal was still in a class imbalance state compared with the interictal period and the pre-ictal period. We randomly screened the EEG data segments of the interictal and pre-ictal phases with the ratio of1:1:1. This section mainly introduced our experimental design and evaluation indicators.

A. TRAINING AND TESTING METHOD

In the data section, data for training the model were obtained from 154 seizures in 14 patients. Figure 6 shows the schematic diagram of the experimental design. For all EEG segments in the training data, we randomly selected 70% as the training set and 30% as the test set. The training set is used for model training. For the EEG segments of the testing set, we reconstructed the form of streaming data, and

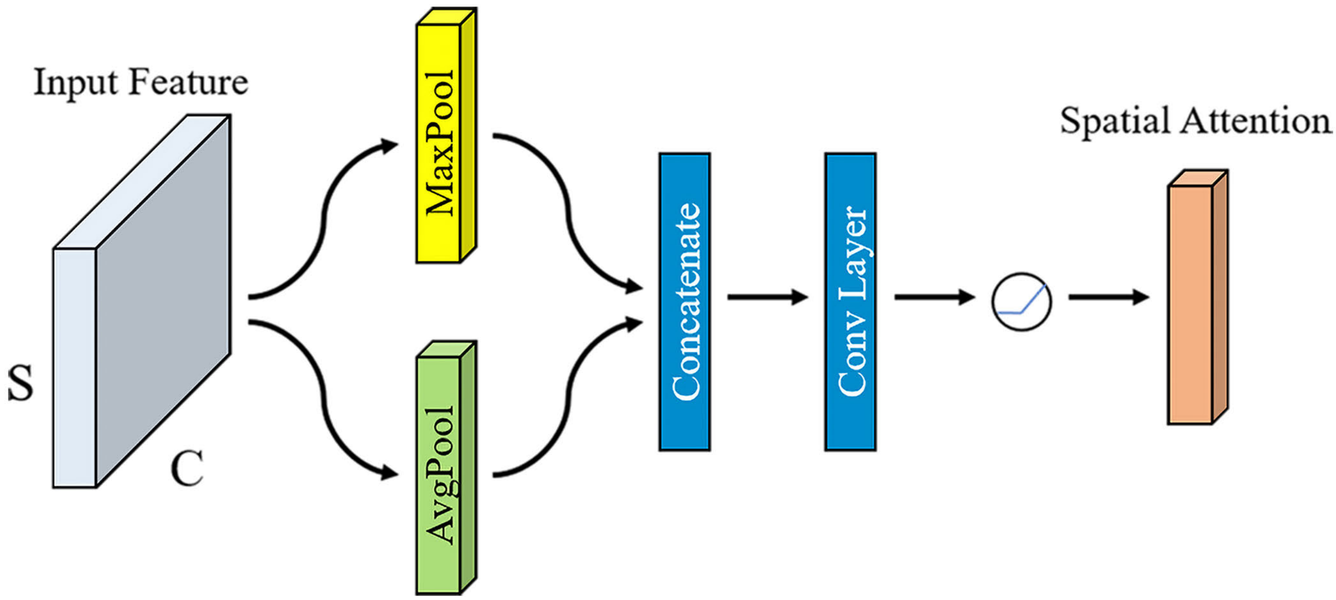


FIGURE 5. The schematic diagram of spatial attention.

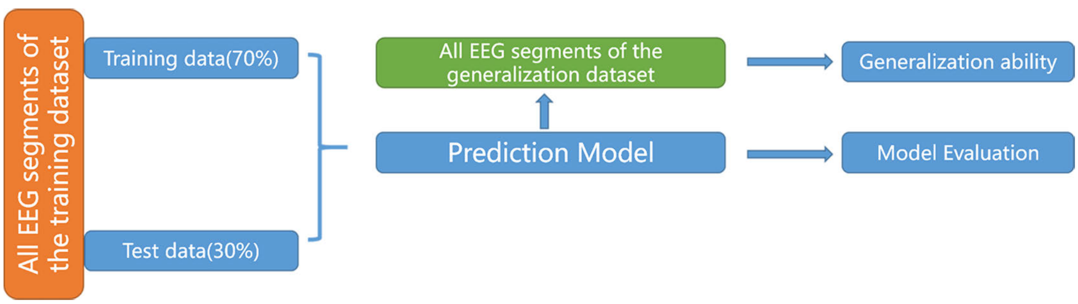


FIGURE 6. The schematic diagram of the experimental design.

take the test results as the performance output of the model. In addition, in order to verify the generalization performance of our model, we also selected the data of other five patients as the generalization dataset to test the generalization ability of the model.

In the model training, we used stochastic gradient descent (SGD) as the optimization function, and the momentum was selected as 0.9. SGD is often used to train various machine learning and deep learning models due to its fast learning rate and online updates. "Categorical_crossentropy" was the loss function of the prediction model which can be expressed as:

$$loss(y_{true}, y_{pre}) = -\sum_{i=1}^{n} y_{true}(x_i) log(y_{pre}(x_i)) \tag{14}$$

We used backpropagation with a batch size of 100 to train our seizure prediction model. The batch size was the number of signals used for each training update. The learning rate of our model was set to 0.1 and decayed every 10 batches with the decay coefficient of 0.95.

B. EVALUATION INDICATORS

In order to evaluate the performance of our proposed model, four evaluation indicators were used: the per-class area under the ROC curve (AUC), Accuracy, Specificity, Sensitivity. AUC is an important curve to measure the prediction model. In this study, AUC was calculated to judge the performance of the seizure prediction model. The closer the AUC value is to 1, the better the performance of the seizure prediction model is. Sen indicated the sensitivity of the prediction model to the pre-ictal EEG data, which measured the classifier's ability to capture the pre-ictal. Spe represented the specific ability of the prediction model to correctly identify the pre-ictal EEG data as the pre-ictal phase, which measured the classifier's ability to recognize the non-pre-ictal. Acc represented the correct prediction ability of the prediction model. The specific formulas are as follows:

$$Sen = \frac{TP}{TP + FN} \tag{15}$$

$$Spe = \frac{TN}{TN + FP} \tag{16}$$

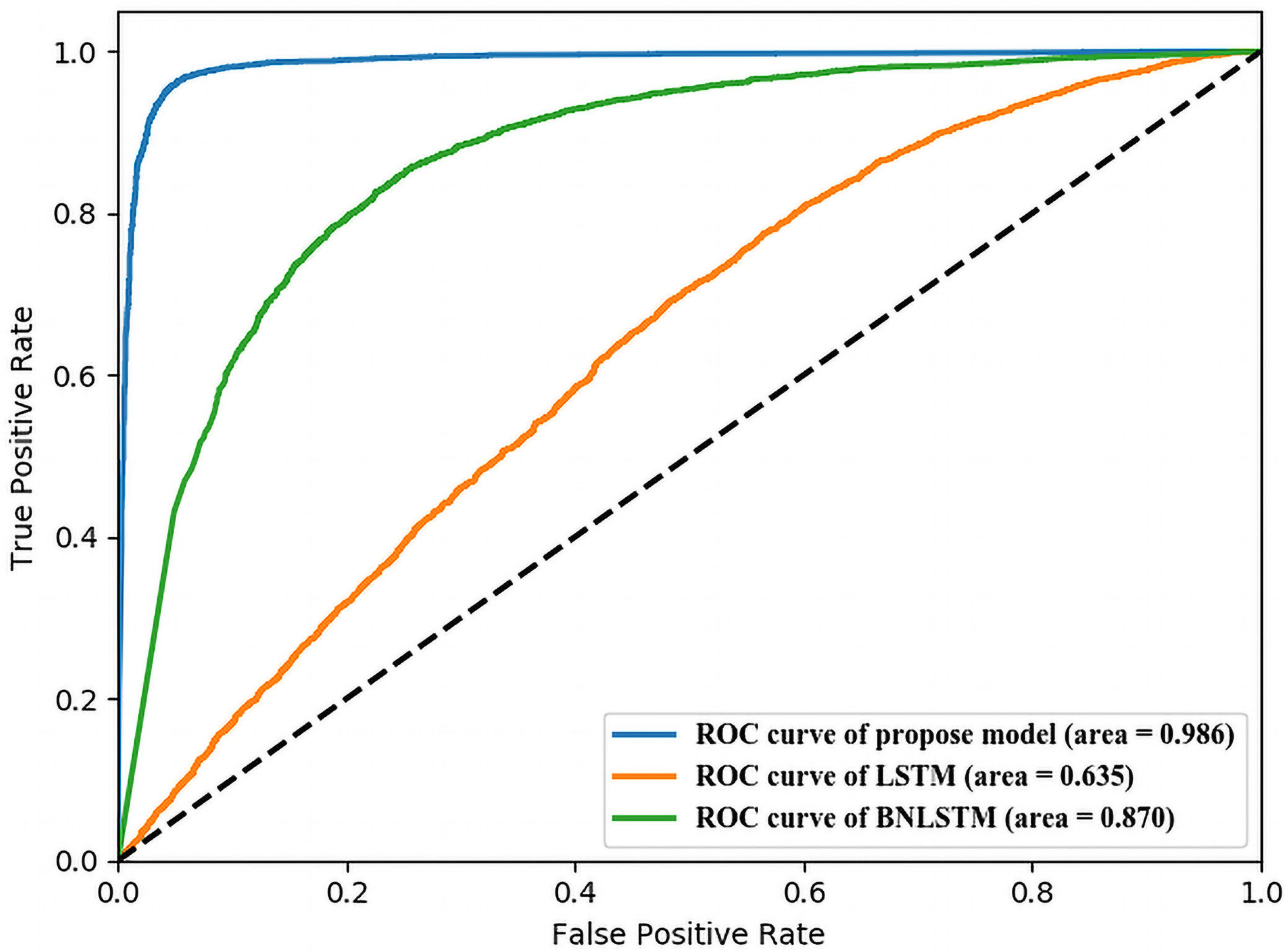


FIGURE 7. Schematic diagram of the area under the ROC curve.

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{17}$$

where TP is the number of correctly classified as positive cases, FN is the number of wrongly classified as positive cases, TN is the number of correctly classified as negative cases, and FP is the number of wrongly classified as negative cases.

V. RESULTS

A. COMPARISON OF THE BNLSTM-CASA MODEL WITH THE BASELINE METHODS

We used Pytorch (version=1.4) to build the automatic lead optimization seizure prediction model. The test results of the training dataset are shown in Table 5. In addition, we also chose LSTM and BNLSTM as the baseline methods. After the same data processing, the results are shown in table 5.

The epileptic seizure prediction model we proposed achieved AUC of 0.986, Acc of 0.956, Spe of 0.968, and Sen of 0.942. In order to demonstrate that our proposed method is superior to the baseline methods, all experimental treatments were consistent, including the neural network depth of each methods. Compared with the baseline methods, both LSTM and BNLSTM with 5 layers, our proposed algorithms achieved the best results. Most notably our AUC value reached 0.986. Figure 7 shows the area under the curve of the three methods, which shows the superior performance of our model. The comparison of different methods shows that the method of combining neural networks is superior to the method of the single type network. And this information is expected to be incorporated into clinical practice.

TABLE 5. The results of training dataset.

Methods	AUC	Acc	Spe	Sen
LSTM	0.635	0.616	0.763	0.617
BNLSTM	0.870	0.806	0.861	0.737
BNLSTM+CASA	0.986	0.956	0.968	0.942

In addition, in order to intuitively observe the features extraction capabilities of LSTM, BNLSTM, and BNLSTM-CASA modules for epileptic EEG data, we used a stochastic neighbor embedding (t-SNE) [35] algorithm to reduce the dimensionality of high-dimensional EEG data.

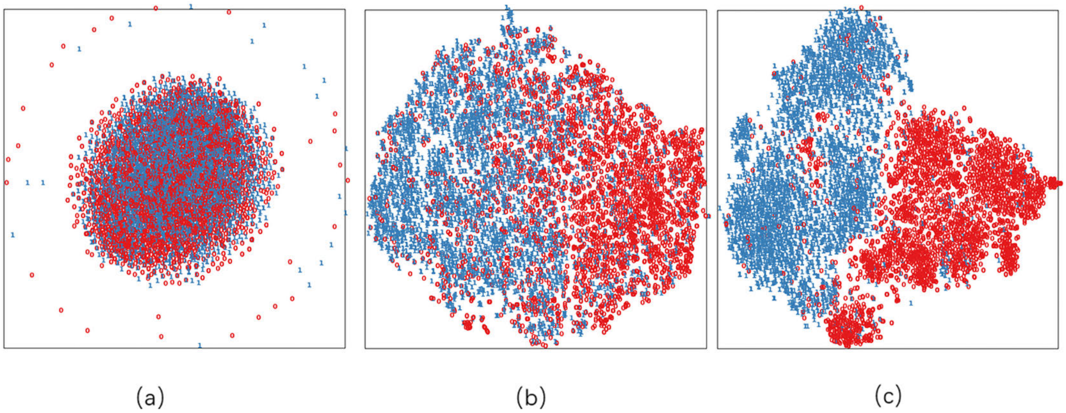


FIGURE 8. Feature distributions of Preictal and Nonpreictal using t-SNE algorithm: (a) Features extracted from LSTM, (b) Features extracted from BNLSTM, (c) Features extracted from BNLSTM-CASA.

TABLE 6. Comparison with other methods.

study	method	dataset	AUC	Acc	Spe	Sen
Y. Kumar et.al [36]	ApEn+DWT+SVM	CHB-MIT	-	0.913	0.833	0.879
N.D. Truong et.al [23]	FFT+CNN	CHB-MIT	-	0.814	0.812	0.750
T.N. Alotaiby et.al [37]	CSP+LDA	CHB-MIT	-	-	0.61	0.89
Y. Zhang et.al [22]	CSP+CNN	CHB-MIT	-	-	-	0.922
T. Dissanayake et.al [38]	Feature extracted + CNN	CHB-MIT	-	0.915	-	-
S. Muhammad Usman et.al [39]	MFCC	CHB-MIT	-	-	0.908	0.927
H. Daoud et.al [40]	DCNN+RNN	CHB-MIT	-	0.996	-	-
Y. Xu et.al [41]	CNN	CHB-MIT	0.981	0.935	-	-
Y. Yang et.al [42]	PE+SVM	University of Freiburg	-	0.996	-	0.94
Proposed method	BNLSTM+CASA	CHB-MIT	0.961	0.914	0.195	0.962
Proposed method	BNLSTM+CASA	Private dataset	0.986	0.956	0.968	0.942

The dimensional data is mapped into a two-dimensional space for visualization. Figure 8 shows the features extracted from epileptic EEG through LSTM, BNLSTM and BNLSTM-CASA. Compared with (a), (b) shows that the BNLSTM nodule has a significant improvement in performance. The features in (c) in combination with the CASA model on the basis of BNLSTM show better separating ability through t-SNE visualization. This means that our model shows greater potential in processing epilepsy EEG data.

B. COMPARISION OF THE BNLSTM-CASA MODEL WITH THE STATE-OF-THE-ART METHODS

Table 6 briefly summarizes recent seizure prediction methods and their results. The method proposed by Kumar *et al.* [36] achieved an accuracy of 0.913, but it was not effective in terms of sensitivity and specificity. Compared with our proposed prediction model, the methods proposed by Truong *et al.* [23] and Alotaiby *et al.* [37] could not directly process the raw EEG data, and the prediction effect was not as good as our proposed method. The ''BNLSTM−CASA'' network can not only directly process the original epilepsy EEG without the feature extraction, but also automatically optimize the weight of all EEG-leads to improve the accuracy and reliability of epileptic seizure prediction.

In addition, for the fairness of comparison, we verified our method on the CHB-MIT dataset. The method of data preprocessing is the same as that of the first dataset we used. The result of the verification are shown in Table 6. Because of the different evaluation indicators, there will be some vacancies in the table. Our results achieved AUC of 0.961, Acc of 0.914, and Sen of 0.962. For the second data set we used in our experiment, his role is similar to the Boston Children's Hospital EEG dataset, and has verified the generalization performance of our model.

TABLE 7. Results on the generalized dataset.

ID	AUC	Acc	Spe	Sen
15	0.878	0.787	0.741	0.872
16	0.886	0.792	0.772	0.934
17	0.889	0.725	1.000	0.700
18	0.912	0.907	0.968	0.920
19	0.830	0.795	0.672	0.944

C. RESULTS OF THE GENERALIZATION DATASET

In order to verify the generalization performance of the ''BNLSTM−CASA'' prediction model, we also collected the long-term EEG data of the other five patients as the generalization dataset. The prediction results were shown in Table 7. For the important AUC of evaluating the model performance, all of 5 patients achieved 0.83 or more, and the AUC value of patient 18 was 0.912. The results shown in Table 7 demonstrate that the ''BNLSTM−CASA'' prediction model has a good generalization ability and could provide a reliable basis for early warning of epileptic seizures.

VI. CONCLUSION

In this study, a seizure prediction model based on BNLSTM and CASA was proposed. The model can directly process the raw EEG signals of all leads, and automatically optimize the lead weights to improve the accuracy and reliability of epileptic seizure prediction. In addition, to verify the generalization performance of the prediction model, we also performed verification on the generalization dataset except the training data. The experimental results show that the "BNLSTM+CASA" epileptic seizure prediction model has a certain generalization ability, which can provide a reliable basis for early warning of epileptic seizures. It is hoped that this research can promote the further development of epileptic seizure prediction system.

ACKNOWLEDGMENT

Mengnan Ma would like to thank Zhen Zhang for insightful discussion and language assistance and also he thank the anonymous reviewers for their careful review and valuable comments. *(Mengnan Ma and Yinlin Cheng are contributed equally to this work.)*

REFERENCES

[1] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "Towards accurate prediction of epileptic seizures: A review," *Biomed. Signal Process. Control*, vol. 34, pp. 144–157, Apr. 2017.

[2] K. G. van Leeuwen, H. Sun, M. Tabaeizadeh, A. F. Struck, M. J. A. M. van Putten, and M. B. Westover, "Detecting abnormal electroencephalograms using deep convolutional networks," *Clin. Neurophysiol.*, vol. 130, no. 1, pp. 77–84, Jan. 2019.

[3] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain, J. Neurol.*, vol. 130, no. 2, pp. 314–333, Feb. 2007.

[4] Z. C. Lipton, D. C. Kale, C. Elkan, and R. J. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, PR, USA, 2016, pp. 1–18.

[5] M. Bandarabadi, J. Rasekhi, C. A. Teixeira, M. R. Karami, and A. Dourado, "On the proper selection of preictal period for seizure prediction," *Epilepsy Behav.*, vol. 46, pp. 158–166, May 2015.

[6] A. R. Ozcan and S. Erturk, "Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 11, pp. 2284–2293, Nov. 2019.

[7] K. Das, D. Daschakladar, P. P. Roy, A. Chatterjee, and S. P. Saha, "Epileptic seizure prediction by the detection of seizure waveform from the preictal phase of EEG signal," *Biomed. Signal Process. Control*, vol. 57, pp. 1017–1020, Mar. 2020.

[8] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, Aug. 2018.

[9] X. Ma, S. Qiu, Y. Zhang, X. Lian, and H. He, "Predicting epileptic seizures from intracranial EEG using LSTM-based multi-task learning," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*. Cham, Switzerland: Springer, 2018.

[10] X. Liu, N. P. Issa, S. Rose, S. Wu, T. Sun, L. V. Towle, P. C. Warnke, and J. X. Tao, "The first-hour-of-the-day sleep EEG reliably identifies interictal epileptiform discharges during long-term video-EEG monitoring," *Seizure, Eur. J. Epilepsy*, vol. 63, pp. 48–51, Dec. 2018.

[11] F. Mormann and R. G. Andrzejak, "Seizure prediction: Making mileage on the long and winding road," *Brain, J. Neurol.*, vol. 139, no. 6, pp. 1625–1627, Jun. 2016.

[12] A. R. Johansen, J. Jin, T. Maszczyk, J. Dauwels, S. S. Cash, and M. B. Westover, "Epileptiform spike detection via convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 754–758.

[13] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka, S. Hosking, P. Lightfoot, V. Ruedebusch, W. D. Sheffield, D. Snyder, K. Leyde, and D. Himes, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study," *Lancet Neurol.*, vol. 12, no. 6, pp. 563–571, Jun. 2013.

[14] A. Nandy, M. A. Alahe, S. N. Uddin, S. Alam, A.-A. Nahid, and M. A. Awal, "Feature extraction and classification of EEG signals for seizure detection," in *Proc. Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST)*, Dhaka, Bangladesh, 2019, pp. 480–485.

[15] U. Seneviratne, P. Karoly, D. R. Freestone, M. J. Cook, and R. C. Boston, "Methods for the detection of seizure bursts in epilepsy," *Frontiers Neurol.*, vol. 10, p. 156, Feb. 2019.

[16] A. Subasi, J. Kevric, and M. A. Canbaz, "Epileptic seizure detection using hybrid machine learning methods," *Neural Comput. Appl.*, vol. 31, no. 1, pp. 317–325, Jan. 2019.

[17] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 804–813, Oct. 2019.

[18] A. M. Abdelhameed, H. G. Daoud, and M. Bayoumi, "Epileptic seizure detection using deep convolutional autoencoder," in *Proc. IEEE Int. Workshop Signal Process. Syst.*, Oct. 2018, pp. 223–228.

[19] U. R. Acharya, Y. Hagiwara, and H. Adeli, "Automated seizure prediction," *Epilepsy Behav.*, vol. 88, pp. 251–261, Nov. 2018.

[20] S. S. Talathi, "Deep recurrent neural networks for seizure detection and early seizure detection systems," 2017, *arXiv:1706.03283*. [Online]. Available: http://arxiv.org/abs/1706.03283

[21] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien, D. Grayden, M. Cook, D. Freestone, and S. Harrer, "Epileptic seizure prediction using big data and deep learning: Toward a mobile system," *EBioMedicine*, vol. 27, pp. 103–111, Jan. 2018.

[22] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 465–474, Feb. 2020.

[23] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw.*, vol. 105, pp. 104–111, Sep. 2018.

[24] M. Golmohammadi, S. Ziyabari, V. Shah, S. L. D. Diego, I. Obeid, and J. Picone, "Deep architectures for automated seizure detection in scalp EEGs," 2017, *arXiv:1712.09776*. [Online]. Available: http://arxiv.org/abs/1712.09776

[25] A. Ghosh, A. Sarkar, T. Das, and P. Basak, "Pre-ictal epileptic seizure prediction based on ECG signal analysis," in *Proc. 2nd Int. Conf. Converg. Technol.*, Mumbai, India, Apr. 2017, pp. 920–925.

[26] H. Namazi, V. V. Kulish, J. Hussaini, J. Hussaini, A. Delaviz, F. Delaviz, S. Habibi, and S. Ramezanpoor, "A signal processing based analysis and prediction of seizure onset in patients with epilepsy," *Oncotarget*, vol. 7, no. 1, pp. 342–350, Jan. 2016.

[27] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing*, vol. 30, nos. 1–4, pp. 201–218, Jan. 2000.

[28] X. Wei, L. Zhou, Z. Zhang, Z. Chen, and Y. Zhou, "Early prediction of epileptic seizures using a long-term recurrent convolutional network," *J. Neurosci. Methods*, vol. 327, pp. 188–195, Nov. 2019.

[29] T. Cooijmans, N. Ballas, C. Laurent, Ç. Gülçehre, and A. Courville, "Recurrent batch normalization," 2016, *arXiv:1603.09025*. [Online]. Available: http://arxiv.org/abs/1603.09025

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

[31] S. Sun, B. Zhao, X. Chen, M. Mateen, and J. Wen, "Channel attention networks for image translation," *IEEE Access*, vol. 7, pp. 95751–95761, 2019.

[32] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 6298–6306.

[33] U. B. Baloglu and Ö. Yildirim, "Convolutional long-short term memory networks model for long duration EEG signal classification," *J. Mech. Med. Biol.*, vol. 19, no. 1, 2019, Art. no. 1940005.

[34] X. Wei, L. Zhou, Z. Chen, L. Zhang, and Y. Zhou, "Automatic seizure detection using three-dimensional CNN based on multi-channel EEG," *BMC Med. Informat. Decis. Making*, vol. 18, no. S5, pp. 71–80, Dec. 2018.

[35] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[36] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," *Signal, Image Video Process.*, vol. 8, no. 7, pp. 1323–1334, Oct. 2014.

[37] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic seizure prediction using CSP and LDA for scalp EEG signals," *Comput. Intell. Neurosci.*, vol. 2017, Oct. 2017, Art. no. 1240323.

[38] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals," *IEEE Sensors J.*, vol. 21, no. 27, pp. 9277–9388, Apr. 2021.

[39] S. M. Usman, S. Khalid, and M. H. Aslam, "Epileptic seizures prediction using deep learning techniques," *IEEE Access*, vol. 8, pp. 39998–40007, 2020.

[40] H. Daoud and M. Bayoumi, "Deep learning based reliable early epileptic seizure predictor," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2018, pp. 1–4, doi: 10.1109/BIOCAS.2018.8584678.

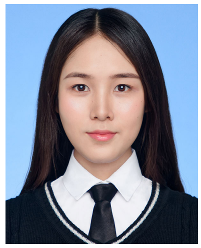
[41] Y. Xu, J. Yang, S. Zhao, H. Wu, and M. Sawan, "An end-to-end deep learning approach for epileptic seizure prediction," in *Proc. 2nd IEEE Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, Aug. 2020, pp. 266–270, doi: 10.1109/AICAS48895.2020.9073988.

[42] Y. Yang, M. Zhou, Y. Niu, C. Li, R. Cao, B. Wang, P. Yan, Y. Ma, and J. Xiang, "Epileptic seizure prediction based on permutation entropy," *Frontiers Comput. Neurosci.*, vol. 12, p. 55, Jul. 2018, doi: 10.3389/fncom.2018.00055.

MENGNAN MA received the bachelor's degree in biomedical engineering from Tianjin Medical University, Tianjin, China, in 2018. He is currently pursuing the master's degree with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include deep learning and seizure detection and prediction.

YINLIN CHENG received the bachelor's degree in biomedical engineering from the Beijing Institute of Technology, Beijing, China, in 2018. He is currently pursuing the master's degree with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include deep learning and computer vision.

YAO WANG received the bachelor's degree in biomedical engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2020. She is currently pursuing the master's degree with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. Her research interests include health big data and artificial intelligence.



XINGYU LI received the bachelor's degree in basic medical sciences from Sun Yat-sen University, Guangzhou, China, in 2021. Her research interest includes medical information systems.

QIANXIANG MAO received the bachelor's degree in biomedical engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2019. She is currently pursuing the master's degree with the School of Biomedical Engineering, Sun Yat-sen University, Guangzhou, China. Her research interests include deep learning and natural language processing.

ZHEN ZHANG received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China. He is currently an Instructor with Huizhou University. His research interest includes medical signal processing.

ZIYI CHEN received the Ph.D. degree from Sun Yat-Sen University, Guangzhou, China. She is currently a Professor and a Master Supervisor with First Affiliated Hospital, Sun Yat-sen University. Her research interests include mechanism of epilepsy drug tolerance, epilepsy with mental disorder and cognitive impairment, and the mathematical model of EEG.

YI ZHOU received the Ph.D. degree from Sun Yat-Sen University, Guangzhou, China. He is currently a Professor and a Doctoral Supervisor with the School of Zhongshan Medical, Sun Yat-sen University. He has published nearly 150 articles so far. His research interests include healthcare big data and medical artificial intelligence.