IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 June 2023, accepted 16 June 2023, date of publication 20 June 2023, date of current version 27 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3287927

RESEARCH ARTICLE

A Multi-Channel Feature Fusion CNN-Bi-LSTM Epilepsy EEG Classification and Prediction Model Based on Attention Mechanism

YAHONG MA[1], ZHENTAO HUANG[1], JIANYUN SU[2], HANGYU SHI[2], DONG WANG[3], SHANSHAN JIA[3], AND WEISU LI[1]

[1]School of Electronic Information, Xijing University, Xi'an 710123, China
[2]Neurosurgery Department, Affiliate Children's Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710003, China
[3]Department of Neurology, Affiliate Children's Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710003, China

Corresponding authors: Yahong Ma (yahongma@sina.com) and Jianyun Su (doctorsu121817@163.com)

This work was supported in part by the General Projects of Shaanxi Science and Technology Plan under Grant 2023-JC-YB-504 and Grant 2017SF-292; in part by the Shaanxi Province Innovation Capacity Support Program under Grant 2018KJXX-095; in part by the Xijing University Special Talent Research Fund under Grant XJ17T03; and in part by the Research Project of Affiliate Children's Hospital of Xi'an Jiaotong University in 2020, under Grant 2020C13.

ABSTRACT Epilepsy is the unstable state caused by excessive discharge of brain cells. In more than 30 percent of epilepsy cases, seizures cannot be controlled with medication or surgery. Refractory epilepsy seriously affects the health of patients and brings great economic burden to families. Therefore, this requires an effective seizure classification and prediction method to reduce risk in epilepsy patients. Researchers proposed machine learning or deep learning methods to predict seizures. However, automatic screening of electrode channels and improvement of predictive accuracy remain a challenge. A multi-channel feature fusion model CNN-Bi-LSTM. This method only requires simple preprocessing. CNN is responsible for extracting spatial features, Bi-LSTM is responsible for extracting temporal features, and finally, two channel weights are allocated through the attention mechanism to filter out the results of the more weighted electrode channel output classification. The performance of the model is tested on the CHB-MIT dataset, and the output is divided into three categories, including normal, pre-seizure and mid-seizure. The ten-fold cross-validation average accuracy is 94.83%, the precision is 94.84%, the recall is 94.84%, the F1-score is 94.83%, and the MCC is 92.26% across CHB-MIT EEG. The ten-fold cross-validation average accuracy of UCI data set is 77.62%, the precision is 77.66%, the recall is 77.62%, the F1-score is 77.60%, and the MCC is 72.03%. The results showed that this method is superior to existing methods and can predict the EEG signals of epilepsy in advance. This work will be extended to design a removable epilepsy predictive device for real-time use.

INDEX TERMS Convolutional neural network (CNN), electroencephalogram (EEG), bi-directional long short-term memory (Bi-LSTM), attention mechanism, epilepsy.

I. INTRODUCTION

According to the World Health Organization (WHO), epilepsy is one of the most common neurological diseases characterized by seizures and the second most common neurological disease after stroke. It is an abnormal electrical discharge of neurons in the brain affecting the health of newborns and adults seriously [1]. Epilepsy sufferers are vulnerable to sudden and unexpected discharge of neurons during which they are unable to protect themselves and are prone to suffocation, injury even to death [2]. To date, the disease has been treated mainly through drugs and surgery. There is no cure or anticonvulsant therapy which can be completely effective for all types of epilepsy [3]. Therefore, early detection and timely treatment of epilepsy is of great significance.

At present, the most commonly used method to diagnose epilepsy is to record brain voltage fluctuations through

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kafiul Islam.

noninvasive electroencephalography (EEG) [4]. EEG can identify ongoing brain activity based on voltage fluctuations by placing multiple electrodes at different locations in the brain [5], [6]. The pathogenesis of epilepsy is mainly caused by abnormal discharge of neurons in the cerebral cortex, which can be transmitted to other neurons and muscles through nerve fibers, resulting in clinical symptoms such as muscle twitches. Pathological changes in brain activity causing seizures can be identified by EEG recording [7]. In addition, EEG recording has high time resolution, and the sampling rate ranges from hundreds to thousands of Hz, which can ensure timely capture of seizures. Moreover, EEG is non-invasive and economical, so EEG has been explored as an effective biomarker and diagnostic tool [8], [9], [10]. But neurologists have to examine EEG records for a long time before they can finally tell a seizure from a normal EEG, which takes a lot of time and effort. So it is important to automatically detect epileptic seizures from EEG.

In the past few years, a great deal of research has been done and a number of techniques have been developed to predict seizures given the magnitude of the problem. If seizures were identified and controlled in timely, 70% of epilepsy patients would be able to lead almost normal lives [11]. Traditional machine learning has been proposed for EEG processing [12], [13]. In Rincon et al.'s work, EEG was classified by using a linear classifier of generalized Gaussian distributive parameter estimates [14]. Tiwari et al. extracted features using filter bank common space pattern algorithms and then classified signals using extreme gradient elevation (XG Boost) [15], which performs well in low-dimensional data but has limitations in high-dimensional data. Chen et al. proposed chaos theory combining decision trees for seizure detection [16], which tends to ignore the relevance of properties in data. Tapani et al. proposed a new method for detecting neonatal epilepsy by extracting non-stationary periodic features in the time and frequency domains [17]. Moreover, deep learning epilepsy detection methods have advanced rapidly [18]. Hossain et al. used short-time Fourier transform to extract time and frequency domain information and used convolution neural networks (CNN) for feature extraction and classification [19]. Mandhouj B et al. Combined CNN with the spectral spectrum to detect and classify epilepsy [20]. Deep learning can avoid the limitations of manual design features compared to traditional machine learning, automatically generate effective features and enables better performance.

Despite advances in automated epilepsy testing, the existing solutions still have limitations. In fact, EEG signals are nonlinear time series. The CNN-based method is good at extracting local features. It has good feature extraction performance for non-stationary and noisy signals [21], but there are some difficulties in capturing the global relevance of time series data. Long short-term memory (LSTM) is a variant of a recursive neural network that is often used to deal with nonlinear features of time series, such as EEG. It considers the long-term dependencies of time series but ignores local spatial information. CNN and LSTM are considered end-to-end models in most of the literature [22], [23], [24]. However, because not all brain regions produce abnormal discharges, EEG signals in different pathways have different manifestations. If an algorithm can be designed to screen for EEG channels that best respond to epilepsy in patients, it will not only reduce the computational and hardware costs of the model but also improve the efficiency of epilepsy prediction. Therefore, it is necessary to design an effective EEG channel selection algorithm before completing the seizure prediction task. At the same time, most researchers focus on the classification of epilepsy and normal EEG prediction, but there is little research on the prediction of epilepsy.

Aiming at the existing problems, a multi-channel feature fusion model CNN-Bi-LSTM is proposed in this paper for the automatic detection of epilepsy. The CNN module extracts local features, the Bi-LSTM module captures time-series information features of the multi-channel EEG, and the two modules are merged in parallel and stores these features into flatten layer. Then all the features are merged into a fully connected layer, and finally, the weight of the two-channel features is selected through the attention-mechanism module. The ability to predict fusion features is demonstrated by performing epilepsy detection tasks on CHB-MIT datasets. In summary, the main contributions and innovations of this paper can be summarized as follows.

Firstly, the proposed novel multi-channel feature fusion neural network provides a new way for researchers and doctors to classify EEG signals. Compared with the single-branch structure, the EEG classification accuracy is greatly improved due to the different feature extraction of the input signals.

Secondly, to make the most of the fusion feature information, an attention-mechanism module is added after the fusion of EEG features, weight each module's extracted feature, and filter out the information in important electrode channels, thus solving the problem of classification accuracy decreased due to information redundancy.

Thirdly, the epilepsy dataset is divided into three categories, including normal EEG, normal EEG seconds prior to onset of epilepsy, and seconds prior to onset of epilepsy. The result is very positive for the prediction and timely treatment of epilepsy.

The organizational structure of this article is shown below. The first section reviews previous work on epilepsy detection. The second section describes our proposed methods, including datasets, data preprocessing, experimental models and evaluation indicators. The third section focuses on the experimental results. The last part is the conclusion.

II. MATERIALS AND METHODS

A. DATASET

The database used in this study is the open-source EEG database from CHB-MIT [25], [26]. The records were collected from 22 epileptic children using scalp electrodes, with EEG data provided by the Massachusetts Institute of Technology. This study included 5 males, ages 3-22 and

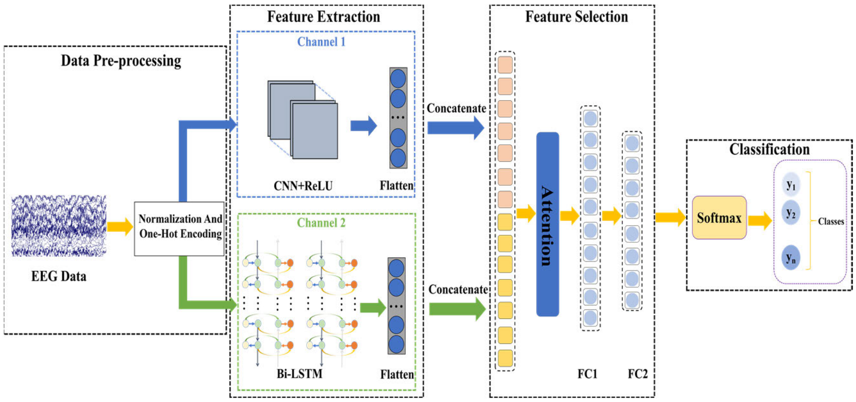

FIGURE 1. CNN-Bi-LSTM multi-channel fusion model based on attention mechanism.

17 females, ages 1.5-19. All participants were asked to stop treatment one week before data collection. The EEG signals were sampled at 256 Hz, 16-bit resolution. Most files contained 23 channels of EEG signals, in a few cases 24 or 26 channels. These records were made using the International 10-20 EEG Electrode Location and Naming System. The start and end time of seizures are marked based on expert judgment, and the number and duration of seizures varies from subject to subject. To ensure the reliability of the experiment and the balance of the sample, a dataset with the same 23 EEG channels was selected and divided into three categories. Neurologists can easily detect normal EEG and epileptic seizures from the original signals. The third category is the EEG signals a few milliseconds before the seizure, which has a positive impact on preventing seizures and taking measures to treat seizures. In this study, the EEG signals of each patient were collected twice, each time for 5 seconds, and data that did not meet the time requirement was also screened out. All these data sets were collected together, and more than 40,000 data were obtained.An open UCI epilepsy identification dataset are also used [27]. There was a 23.6 second record of brain activity in each file. After visual examination of the artifacts, such as muscle or eye movements, the segments were selected and cut from a continuous multichannel EEG signal. The corresponding time series was sampled at 4,097 data points. Each data point is an EEG recorded value at a different point in time. Therefore, there are $23 \times 500 = 11{,}500$ consecutive EEG samples, each containing 178 data points lasting 1 second (column), with the last column representing the label Y $\{1, 2, 3, 4, 5\}$. The raw dataset has been pre-processed by the UCI, which created the data in CSV file format. There are five status categories (a) Recording EEG signals when healthy subjects open their eyes. (b) Recording EEG signals when healthy subjects closed their eyes. (c) Recording healthy inter-hippocampal EEG signals in epilepsy patients. (d) Recording intermittent EEG signals at brain tumor sites in epilepsy patients. (e) Recording EEG signals of epilepsy activity in epilepsy patients.

B. DATA PRE-PROCESSING

In deep learning, the data was typically pre-normalized before entering the network, clipping the data to a certain extent to ensure that the data from different samples are of the same order of magnitude, which can speed up training and improve the generalization of the training model. To test the superiority of our model and promote the implantability of the brain-computer interface, raw EEG data was normalized only by dividing the data for each sampling point by the maximum value of the same electrode channel. In addition, since computer cannot understand non-digital data, data labels Y $\{1, 2, 3, 4, 5\}$ were converted to binary hot encoding with 0, 1 combination. After pre-processing, the training and test set can be divided and put into a deep learning model.

C. CNN-BI-LSTM MULTI-CHANNEL FUSION BASED ON ATTENTION MECHANISM

In this paper, a CNN-Bi-LSTM model based on the attention-mechanism is proposed for feature extraction, feature selection, and classification of epileptic EEG, as shown in Figure 1. The feature extraction of proposed model is composed of two channels. The normalized EEG signals with hot-encoding labels are put into two channels simultaneously. CNN with a ReLU activation function is used in channel 1 to extract the spatial feature of EEG data, which can achieve better accuracy with its excellent spatial feature extraction ability. However, EEG data are often significantly correlated in time dimensions. In order to make up for the shortage of CNN, channel 2 uses Bi-LSTM which is responsible for extracting the temporal features of the original data, because Bi-LSTM is good at processing data with sequential characteristics. Since the extracted feature size of each module is different, there is a Flatten layer at the end of each module that converts the dimensions of space-time features into one dimension. Then, the extracted spatial-temporal features are concatenated into a complete sequence. Considering the importance of different feature, the concatenated feature is put into the attention mechanism module and the weight of the each feature is automatically fitted by using the point

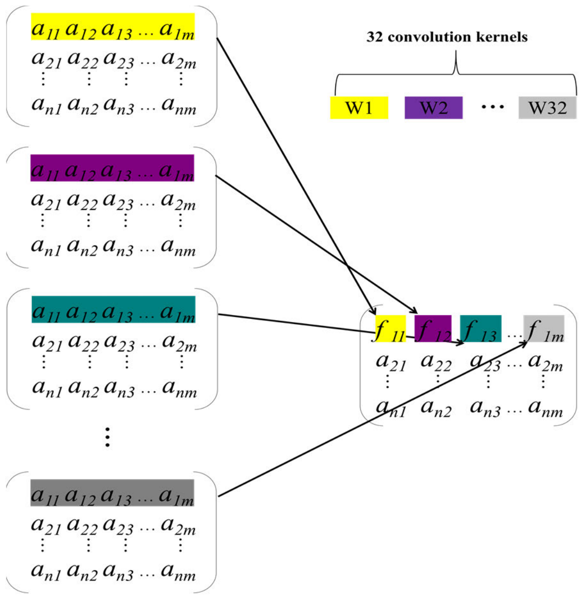

FIGURE 2. 1D CNN convolution process.

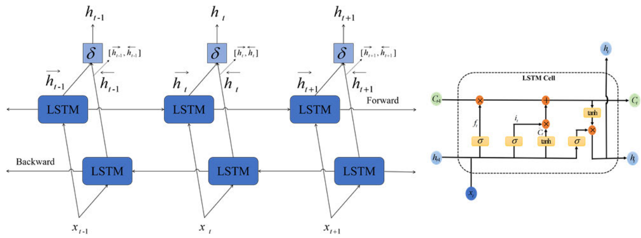

FIGURE 3. LSTM and Bi-LSTM structure.

product. Finally, through two fully connected layers, three classifications are obtained through Softmax function. Next, the main modules in this model are described.

1) CONVOLUTIONAL NEURAL NETWORK (CNN)

Because the EEG signal of each person is unique, classifying EEG information in different states is a challenging task. The deep learning model including complex neural networks has achieved excellent results in many applications, such as image classification, face recognition, and speech recognition. CNN can also be used to classify the status of EEG signals effectively. Unlike traditional machine learning algorithms, CNN does not need to design features manually. It simply uses the local sensory fields generated by convolutions nuclei to automatically learn abstract features from raw data to categorize, thus avoiding the loss of useful information. CNN can extract different levels (low, medium and high) of features from raw EEG data by using multiple convolution and merge operations. In contrast to classical frameworks, CNN typically has two separate steps, including feature learning and classification, which it can learn in one go through multiple layers of neural networks.

(1) Convolutional layer. The main purpose of the convolution layer is to extract features. Different types of features are extracted from input data using many convolution cores. The convolution process in this paper is shown in Figure 2.

Where, a is the EEG data, $W_1 \cdots\cdots W_{32}$ is the convolution kernel, m is the number of electrode channels, n is the maximum number of sampling points per channel and the output vector is f.

(2) Activation layer. The activation function is used to solve the problem of linear indivisibility. In the early days of neural network development, the Sigmoid function was used more. However, the Sigmoid function tends to cause gradient attenuation during back-propagation. Thus the ReLU activation function is used in channel 1, which is shown in Equation (1).

$$y = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{1}$$

When $x < 0$, the derivation equals to 0. When $x \geq 0$, the derivation equals to 1. It is possible to pass the gradient of y to x in its entirety without causing the gradient to disappear.

(3) Flatten layer. The purpose of the flatten layer is to compress high-dimensional data into vectors for classification by subsequent fully connected layers. The flatten layer follows convolutional layer and before the attention-mechanism layer and fully connected layer, which does not affect the size of the batch.

2) BI-DIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM)

Although convoluted neural networks show great advantages in feature extraction, they cannot retain the memory of previous time series patterns. A Circulating Neural Network (RNN) is a kind of neural network used to process sequence data, which preserves information through circulation. However, in some cases, we need more contextual information. As the distance increased, the RNN became unable to connect the information. These results in the RNN not being able to learn well about the long-term dependencies of time-series data. Compared with convolution neural networks, LSTM networks are more successful in processing time data. LSTM networks learn the long-term and short-term relevance of serial data through storage unit C, which has a self-connection to store the network's temporal state. LSTM and Bi-LSTM networks are shown in Figure 3.

Bi-LSTM consists of two LSTM blocks that can better capture this information in both positive and negative directions to simultaneously process EEG signals in the opposite direction. Bi-LSTM not only has the ability to process context but also to process future context content to improve model accuracy. Therefore, we propose the use of Bi-LSTM for feature recognition in the local domain of EEG feature space. Bidirectional LSTM calculates the entire output h_t based on Equation (2).

$$h_t = \sigma(W_h \times [\overrightarrow{h_t}, \overleftarrow{h_t}] + b_h) \tag{2}$$

There are three main phases within the LSTM unit.

(1) Forget the stage. This phase consists mainly of selective forgetting of inputs from the previous node. Remember the important information and forget the unimportant. The layer reads the current input x and foreneuron information h, and the f_t determines the previous state of C_{t-1} and chooses to forget certain information.

(2) Choose the memory stage. This phase is mostly selective memory of input X_t. What is important is highlighted and what is not important is remembered less. This step consists of two layers. The sigmoid layer acts as the input gate layer, determining the value i we will update. *Tan*h layer creates a new candidate value vector $\tilde{C}_t$ to join the state.

(3) Output phase. This phase determines what information will be treated as an output of the current state. The C_t obtained in the previous stage is scaled down by the tan function. Similar to a normal RNN, the output y_t is ultimately obtained by h_t transformation.

The mathematical expression of the LSTM unit is defined as follows in the Equation (3-8):

$$f_t = \sigma(w_f \times [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(w_i \times [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(w_c \times [h_{t-1}, x_t] + b_c) \tag{5}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{6}$$

$$o_t = \sigma(w_o \times [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \times \tanh(C_t) \tag{8}$$

3) ATTENTION MECHANISM

The attention mechanism is a kind of resource allocation mechanism that simulates attention in the human brain. When the human brain processes things, it focuses on areas that need to be focused, reducing or even ignoring attention to other areas in order to get more detailed information that needs attention. As discussed in previous sections, ordinary RNN or LSTM structures use the time-dynamic properties of input data and map them to sequential output data. However, there are still correlations between the output produced by a given time step and the input sequence used to produce that output. Even if LSTM reduces the effects of long sequences of disappearing gradients and explosive gradients, it does not completely eliminate them. In addition, neural network architectures such as RNN, LSTM, or CNN may not be able to handle highly complex feature representations to produce accurate outputs. Attention mechanisms are an effective way to deal with long-term dependencies, especially for very long sequences. Attention mechanisms can be combined with neural network models such as CNN or LSTM to obtain important information.

This paper uses a dot-product attention mechanism that performs weighted summation of hidden layer vector expressions from CNN and Bi-LSTM outputs, in which weights represent the importance of the characteristics of each spatial and temporal point. Note that this mechanism can replace the original method of randomly assigning weights by assigning probabilities. Assuming an input of m eigenvector $h_{i,}$, $i = 1, 2, \ldots, k$. The model can get an environmental vector c_i based on h_i. These environmental vectors can be predicted together with the current hidden state. Environmental vector c_i can be calculated by weighted averages of previous states, as shown in Equation (9).

$$c_i = \sum_{i=1}^{k} a_i h_i \tag{9}$$

Since the weight of the added state is the attention weight a_i, in order to obtain a_i, we train a fully connected network whose input is the hidden vector of CNN and BI-LSTM output. The influence on the output is evaluated by calculating the score s_i of each hidden layer vector, as shown in Equation (10).

$$s_i = \tanh(w^T h_i + b_i) \tag{10}$$

where, s_i represents the degree of correlation between h_i and c_i. c_i is the output value of the i^{th} node, j is the total number of nodes that must always be counted, that is, the total number of nodes output. Then, Softmax function is used to normalize the score s_i to obtain the final weight factor a_i, as shown in Equation (11).

$$a_i = \text{softmax}(s_i) = \frac{e^{s_i}}{\sum j e^{s_j}} \tag{11}$$

After applying the attention mechanism to LSTM and CNN, we can pay attention to the features that have a great influence on the output variables and improve the accuracy of the method.

4) FULL CONNECTION LAYER

Convolution and pool sequences extract the most important features from the data. Finally, feature classification is needed to predict the actual types of input data. Therefore, apply one or more fully connected layers at the end of the network. The number of neurons in the last fully connected layer should be equal to the number of predicted output classes.

5) CLASSIFIER

In the output layer, we used the Softmax function. Softmax is as shown in Equation (12). Here, n represents the number of target classes and x_i represents the input value of i target class.

$$soft\max(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{12}$$

D. EVALUATION INDEXES

Since epilepsy prediction is a three-category issue, we used evaluation criteria commonly used in classification to measure the validity and robustness of our models from different perspectives, including accuracy, precision, recall, F1-score, and Matthews correlation coefficient (MCC). They are

TABLE 1. Parameters of the CNN-Bi-LSTM-Attention architecture.

Parameter	Value
epoch number	100
learning rate	0.001
batch size	1024
optimizer	Adam
loss function	categorical_crossentropy
convolution kernel	32
Bi-LSTM	16
FC1	32
FC2	16
classifier	Softmax

defined as formulas (13-17).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

$$precision = \frac{TP}{TP + FP} \tag{14}$$

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

$$F1\text{-}score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \tag{16}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{17}$$

Here TP and TN are symbols of the correct positive sample number and the correct negative sample number predicted by the model, respectively. FP and FN are symbols of the number of false positive and false negative samples predicted by the model, respectively. MCC is essentially a correlation between observed and predicted binary classifications. It returns a value between −1 and +1. A coefficient of +1 indicates perfect prediction, 0 indicates no better than random prediction, and −1 indicates a complete inconsistency between prediction and observation.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL SETUP

In this experiment, the number of training rounds is 100 and the batch is set to 1024. The Adam optimizer combines the advantages of both AdaGrad and RMS Prop optimization algorithms. The model's loss function is used using categorical cross entropy. In order to ensure the same distribution of data between the training set and the test set, the pre-training model data were all set to the same random seed, randomly scrambled, and transmitted to the network model. Multi-channel CNN-Bi-LSTM-Attention and other paired network models were implemented and modeled using Python 3.7 on GeForce RTX 2080Ti, all using the same parameter settings. Table 1 shows the CNN-Bi-LSTM-Attention model parameter Settings in this paper.

B. CROSS VALIDATION

We use the average value of K-fold cross-validation as the model evaluation standard, and K-fold cross-validation is

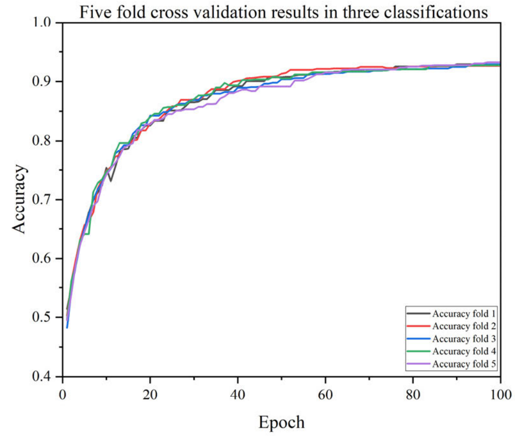

FIGURE 4. The accuracy of model based on five-fold cross-validation three classification task.

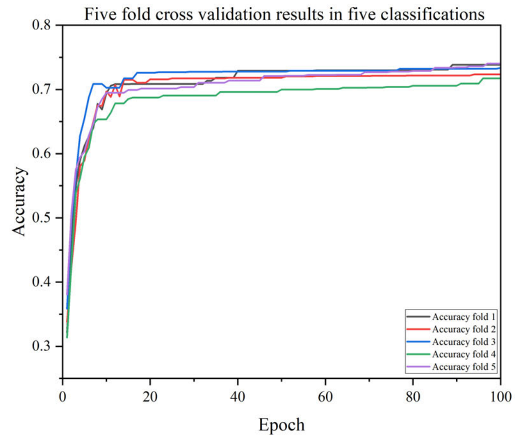

FIGURE 5. The accuracy of model based on five-fold cross-validation five classification task.

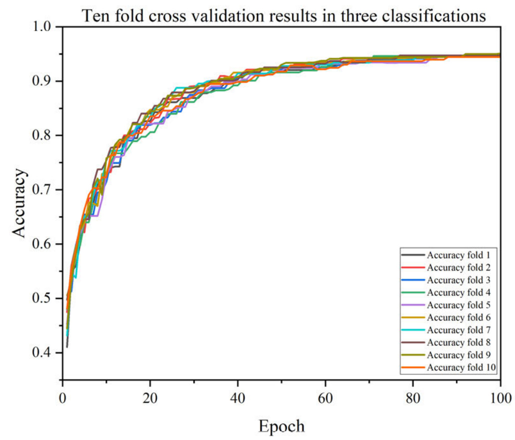

FIGURE 6. The accuracy of model based on ten-fold cross-validation three classification task.

often used in model training. K-fold cross-validation began by randomly dividing dataset into k mutually exclusive subsets of the same size, that is, each random selection of k-1 as a training set and the remaining 1 as a test set. When this round is complete, re-select k at random to train the data. After several rounds, the average value of the evaluation index is finally selected for evaluation. The model proposed

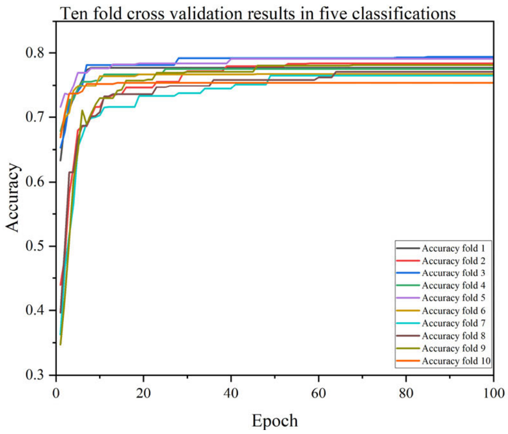

FIGURE 7. The accuracy of model based on ten-fold cross-validation five classification task.

in this paper was cross-validated with five and ten-fold cross-verification on two data sets, respectively, as shown in Figure 4 to Figure 7. The model proposed in this paper uses five-fold cross-validation, and the average accuracy of the two data sets is 92.90% and 73.06%, the average Precision is 92.96% and 73.23%, the average recall is 92.90% and 73.07%, the average F1-score is 92.90% and 73.12%, and the average MCC is 89.38% and 66.33%. Using ten-fold cross-validation, the average accuracy of the two data sets is 94.83% and 77.62%, the average Precision is 94.84% and 77.66%, the average recall is 94.84% and 77.62%, the average F1-score is 94.83% and 77.60%, and the average MCC is 92.26% and 72.03%. Since the model proposed in this paper performs well in 10-fold cross-validation, subsequent comparison models are all compared by 10-fold cross-validation.

C. COMPARED WITH OTHER MODELS

To validate the taxonomic performance of our proposed multichannel CNN-BI-LSTM-Attention model for EEG detection, our model was compared with deep neural networks (DNN), convoluted neural networks (CNN), long and short term memory networks (LSTM) and bidirectional long short term memory networks (Bi-LSTM), as well as their combination models. To demonstrate the superiority of multichannel model predictions, we also compared CNN-BI-LSTM-Attention and CNN-BI-LSTM network models used End-To-End. And their combined models CNN-RNN, CNN-LSTM, CNN-Bi-LSTM, DSCNN-RNN, DSCNN-LSTM and DSCNN-Bi-LSTM. 1D Convolutional Auto-Encode(CAE) is composed of two convolutional layers replacing the fully connected layers, and the symbols of the input are down-sampled to provide a potential representation of smaller dimensions. 1D InceptionV1 is compared. 1D InceptionV1 is the replacement of InceptionV1 two-dimensional convolution nuclei with one-dimensional convolution nuclei. The results of the three-category experiment are shown in Table 2, where the three classification labels represent normal EEG data, pre-epileptic EEG data, and initial epilepsy EEG data. Of all the models compared, CNN-BILSTM-Attention

TABLE 2. Comparison between multi-channel CNN-BI-LSTM-Attention and other deep learning models in three classified tasks through ten-fold cross-validation method.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
DNN	55.09	55.21	55.12	54.50	33.05
CNN	59.64	59.89	59.67	59.40	39.74
RNN	56.84	55.35	55.41	55.03	35.49
GRU	61.05	59.52	59.53	59.32	41.72
LSTM	59.68	59.82	59.67	59.47	39.70
Bi-LSTM	64.62	65.05	64.63	64.49	47.20
CNN-Attention	82.34	82.88	82.32	82.36	73.73
Bi-LSTM-Attention	90.10	90.14	90.10	90.10	85.16
CNN-Bi-LSTM（End-To-End）	88.56	88.60	88.56	88.55	82.86
CNN-Bi-LSTM-Attention（End -To-End）	87.88	88.00	87.90	87.90	81.88
CNN-Bi-LSTM(Multi-Channel)	86.88	87.26	86.82	86.88	80.46
CNN-RNN	76.19	76.37	76.19	76.12	64.42
CNN-LSTM	85.54	85.75	85.55	85.55	78.40
DSCNN-RNN	74.56	74.92	74.54	74.50	62.02
DSCNN-LSTM	80.88	81.77	80.89	80.77	71.82
DSCNN-Bi-LSTM	86.25	8638	86.25	86.24	79.44
1D CAE	74.47	74.90	74.51	74.45	61.91
1DInceptionV1	65.62	65.89	65.57	65.32	48.69
Proposed method	**94.83**	**94.84**	**94.84**	**94.83**	**92.26**

(Multi-Channel) performed best, with 94.83% for Accuracy, 94.84% for Precision, 94.84% for Recall, 94.83% for F1-score, and 92.26% for MCC. Bi-LSTM-Attention was second only to CNN-Bi-LSTM-Attention (90.10%), Precision (90.14%), Recall (90.10%), F1-Score (90.10%) and, MCC (85.16%) on five evaluation metrics. The experimental results of the five categories are shown in Table 3. Among all the models, CNN-BILSTM-Attention (Multi-Channel) still has the best performance, with an accuracy is 77.62%, an accuracy rate is 77.66%, a recall is 77.62%, an F1-score is 77.60% and an MCC is 72.03%. The results show that the superiority of our method is proved by the parallel structure and fusion characteristics. When using the CNN-BI-LSTM-Attention parallel structure, both networks can simultaneously extract the characteristics of the input signal, making both temporal and spatial features rich in primitive features. Serial structures that use CNN features as input to Bi-LSTM can only extract features layer by layer, during which feature loss occurs, resulting in a decrease in final accuracy. The Bi-LSTM model performed best in a single model, suggesting that the Bi-LSTM model is well suited for classifying and predicting time-series features such as EEG.

D. COMPARED WITH TRADITIONAL MACHINE LEARNING MODELS

Traditional machine learning methods have been widely used in many computer fields. Traditional machine learning feature extraction relies on manual methods, which can be simple and effective for particularly simple tasks, and can

TABLE 3. Comparison between multi-channel CNN-BI-LSTM-Attention and other deep learning models in five classified tasks through ten-fold cross-validation method.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
DNN	45.68	45.63	45.71	44.19	32.86
CNN	63.40	63.74	63.38	63.18	54.42
RNN	44.08	41.24	44.16	38.82	33.15
GRU	62.00	60.81	60.95	58.28	53.78
LSTM	68.81	68.77	68.91	68.43	61.22
Bi-LSTM	69.62	69.92	69.69	69.15	62.34
CNN-Attention	74.24	74.66	74.25	74.11	67.95
Bi-LSTM-Attention	72.82	72.95	72.85	72.83	66.03
CNN-Bi-LSTM（End-To-End）	73.70	73.92	73.68	73.65	67.18
CNN-Bi-LSTM-Attention（End-To-End）	57.94	58.25	57.98	57.25	47.82
CNN-Bi-LSTM(Multi-Channel)	74.50	74.70	74.55	74.39	68.22
CNN-RNN	53.73	53.72	53.69	51.07	43.23
CNN-LSTM	74.93	75.11	74.86	74.81	68.72
DSCNN-RNN	58.89	58.80	59.02	58.13	48.98
DSCNN-LSTM	73.39	73.73	73.42	73.33	66.85
DSCNN-Bi-LSTM	73.23	73.45	73.25	73.13	66.62
1D CAE	73.57	74.56	73.43	73.08	67.37
1DInceptionV	65.73	65.97	65.74	65.63	57.26
Proposed method	**77.62**	**77.66**	**77.62**	**77.60**	**72.03**

TABLE 4. Comparison between multi-channel CNN-BI-LSTM-Attention and other traditional machine learning models in three classified tasks through ten-fold cross-validation method.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
Adaboost	51.83	51.08	51.86	50.70	28.19
Bayes	45.54	46.38	45.61	43.06	19.85
decision Tree	71.16	71.20	71.17	71.17	56.76
KNN	91.44	91.75	91.45	91.46	87.29
Random Forest	89.89	90.00	89.90	89.88	84.91
SVM	40.62	41.06	40.63	40.33	11.08
XGBoost	81.95	82.14	81.95	81.96	73.00
Proposed method	**94.83**	**94.84**	**94.84**	**94.83**	**92.26**

be interpreted, but it is not universal. Feature extraction for deep learning is not manual, but machine generated. In addition to comparing the model to other deep learning models, we will also propose a comparison of the model with the more popular seven types of communication machine learning: Adaboost, Bayes, Decision Tree, KNN, Random Forest, Support Vector Machine and XGBoost. Similarly, we evaluated the model by means of a 10 fold cross-validation of the mean values, the results of which are shown in Table 4 and Table 5. In the tripartite task, KNN is second only to our proposed model, with Accuracy is 91.44%, Precision is 91.75%, Recall is 91.45%, F1-score is 91.46% and MCC is 87.29%. SVM had the worst performance with 40.62% for Accuracy, 41.06% for Precision, 40.63% for Recall, 40.33% for F1-score and 11.08% for MCC. In the five categories of tasks, Random Forest is second only to our proposed model, with an Accuracy is 70.34%, Precision is 69.99%, Recall is 70.37%, F1-score is 70.01% and MCC is 62.99%. SVM had the worst performance with 28.26% for Accuracy, 36.18% for Precision, 28.41% for Recall, 28.39% for F1-score, and 10.77% for MCC. Comparisons show that classification using machine learning classifiers is still significantly less effective than CNN-Bi-LSTM-Attention (Multi-Channel).

TABLE 5. Comparison between multi-channel CNN-BI-LSTM-Attention and other traditional machine learning models in five classified tasks through ten-fold cross-validation method.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
Adaboost	43.79	41.95	43.77	38.52	32.63
Bayes	43.72	44.17	43.77	41.98	30.75
Decision Tree	47.91	48.63	47.90	48.17	34.91
KNN	47.58	57.26	47.56	46.19	36.72
Random Forest	70.34	69.99	70.37	70.01	62.99
SVM	28.26	36.18	28.41	28.39	10.77
XGBoost	69.55	69.89	69.59	69.65	61.96
Proposed method	**77.62**	**77.66**	**77.62**	**77.60**	**72.03**

IV. CONCLUSION

Seizure prediction is useful for controlling seizures that are not treatable with medication or surgery. In this paper, we propose an effective method to predict epileptic seizures by using primitive EEG signals, which can alert epileptic patients to take necessary protective measures to avoid unnecessary life risks. The method requires only normalization to preprocess the raw data without additional pretreatment, which facilitates brain-computer interface transplantation. The data is then fed into a Multi-Channel CNN-Bi-LSTM-Attention network model. Our proposed model can automatically extract features from raw EEG signals. For the three classifications, the average cross-validation accuracy of this method was 94.83%. For five categories, the average cross-validation accuracy of this method is 77.62%. Compared with the End-To-End network of the same model, the performance of the model is significantly improved. The experimental results show that the feature can be extracted automatically from EEG signals by deep learning rather than by hand, and the weighted electrode channels can be selected automatically by the attention mechanism. However, the model proposed in this paper can only be applied to EEG signals. It is difficult to guarantee the stability of prediction with one kind of data input. In the future, we plan to use a multi-data fusion model. For example, combining signals such as Electrocardiograph(ECG) and Electromyography(EMG) with electroencephalogram (EEG) recordings could further improve methods for predicting seizures.

DATA AVAILABILITY STATEMENT

This study is an experimental analysis of a publicly available data set. The data can be found in this web page:

https://physionet.org/content/chbmit/1.0.0/ and https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition.

REFERENCES

[1] N. Murphy and T. Such-Neibar, "Cerebral palsy diagnosis and management: The state of the art," *Current Problems Pediatric Adolescent Health Care*, vol. 33, no. 5, pp. 146–169, May 2003.

[2] T. Yan, X. Bi, M. Zhang, W. Wang, Z. Yao, W. Yang, and J. Wu, "Age-related oscillatory theta modulation of multisensory integration in frontocentral regions," *NeuroReport*, vol. 27, no. 11, pp. 796–801, 2016.

[3] T. Yan, Y. Geng, J. Wu, and C. Li, "Interactions between multisensory inputs with voluntary spatial attention: An fMRI study," *NeuroReport*, vol. 26, no. 11, pp. 605–612, 2015.

[4] C. E. Elger and C. Hoppe, "Diagnostic challenges in epilepsy: Seizure under-reporting and seizure detection," *Lancet Neurol.*, vol. 17, no. 3, pp. 279–288, Mar. 2018.

[5] I. V. Petukhov, A. E. Glazyrin, A. V. Gorokhov, L. A. Steshina, and I. O. Tanryverdiev, "Being present in a real or virtual world: A EEG study," *Int. J. Med. Inform.*, vol. 136, Apr. 2020, Art. no. 103977.

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

[7] N. B. Finnerup, A. Fuglsang-Frederiksen, P. Røssel, and P. Jennum, "A computer-based information system for epilepsy and electroencephalography," *Int. J. Med. Informat.*, vol. 55, no. 2, pp. 127–134, Aug. 1999.

[8] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," 2015, *arXiv:1511.06448*.

[9] M. Sharma, P. V. Achuth, D. Deb, S. D. Puthankattil, and U. R. Acharya, "An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals," *Cognit. Syst. Res.*, vol. 52, pp. 508–520, Dec. 2018.

[10] S.-C. Liao, C.-T. Wu, H.-C. Huang, W.-T. Cheng, and Y.-H. Liu, "Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns," *Sensors*, vol. 17, no. 6, p. 1385, Jun. 2017.

[11] S. Yoo, K. Lim, H. Baek, S.-K. Jang, G.-Y. Hwang, H. Kim, and H. Hwang, "Developing a mobile epilepsy management application integrated with an electronic health record for effective seizure management," *Int. J. Med. Informat.*, vol. 134, Feb. 2020, Art. no. 104051.

[12] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., 2009.

[13] A. F. Gill, S. A. Fatima, M. U. Akram, S. G. Khawaja, and S. E. Awan, "Analysis of EEG signals for detection of epileptic seizure using hybrid feature set," in *Theory and Applications of Applied Electromagnetics: APPEIC 2014*. Cham, Switzerland: Springer, 2015, pp. 49–57.

[14] A. Quintero-Rincon, C. D'Giano, and M. Risk, "Epileptic seizure prediction using Pearson's product-moment correlation coefficient of a linear classifier from generalized Gaussian modeling," 2020, *arXiv:2006.01359*.

[15] A. Tiwari and A. Chaturvedi, "A multiclass EEG signal classification model using spatial feature extraction and XGBoost algorithm," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 4169–4175.

[16] C. Chen, J. Liu, and J. Syu, "Application of chaos theory and data mining to seizure detection of epilepsy," in *Proc. Conf. IPCSIT/Hong Kong*, vol. 25, 2012, pp. 23–28.

[17] K. T. Tapani, S. Vanhatalo, and N. J. Stevenson, "Time-varying EEG correlations improve automated neonatal seizure detection," *Int. J. Neural Syst.*, vol. 29, no. 4, May 2019, Art. no. 1850030.

[18] M. Golmohammadi, S. Ziyabari, V. Shah, I. Obeid, and J. Picone, "Deep architectures for spatio-temporal modeling: Automated seizure detection in scalp EEGs," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 745–750.

[19] M. S. Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, "Applying deep learning for epilepsy seizure detection and brain mapping visualization," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 1s, pp. 1–17, Jan. 2019.

[20] B. Mandhouj, M. A. Cherni, and M. Sayadi, "An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis," *Anal. Integr. Circuits Signal Process.*, vol. 108, no. 1, pp. 101–110, Jul. 2021.

[21] N. Lopac, F. Hržic, I. P. Vuksanovic, and J. Lerga, "Detection of non-stationary GW signals in high noise from Cohen's class of time–frequency representations using deep learning," *IEEE Access*, vol. 10, pp. 2408–2428, 2022.

[22] J. Zhou, M. Meng, Y. Gao, Y. Ma, and Q. Zhang, "Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 5600–5605.

[23] Z. Khademi, F. Ebrahimi, and H. M. Kordy, "A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105288.

[24] A. Echtioui, A. Mlaouah, W. Zouch, M. Ghorbel, C. Mhiri, and H. Hamam, "A novel convolutional neural network classification approach of motor-imagery EEG recording based on deep learning," *Appl. Sci.*, vol. 11, no. 21, p. 9948, Oct. 2021.

[25] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 975–982.

[26] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag, "Patient-specific seizure onset detection," *Epilepsy Behav.*, vol. 5, no. 4, pp. 483–498, Aug. 2004.

[27] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, Nov. 2001, Art. no. 061907.

YAHONG MA received the B.S. degree in biomedical engineering from Sichuan University, in 2005, and the M.S. degree in biomedical engineering and the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, in 2008 and 2013, respectively. From 2013 to 2016, she was a Postdoctoral Researcher of special medicine with Fourth Military Medical University. Since 2016, she has been an Assistant Professor with the Information Engineering Department, Xijing University. She is responsible for or participated in many national projects. She is the author of two books and more than 30 articles. Her research interests include intelligent information acquisition and processing, and bioelectromagnetic effect.

ZHENTAO HUANG received the Bachelor of Engineering degree from Yangtze Normal University, Chongqing, China, in 2021. He is currently pursuing the master's degree in electronic information with Xijing University, Xi'an, Shaanxi. His research interests include deep learning and quantum machine learning.

JIANYUN SU received the B.S. degree from the Clinical Medicine Department, Shaanxi University of Chinese Medicine, in 2012, and the M.S. degree from the Neurosurgery Department, Fourth Military Medical University, in 2015. He has been an Attending Physician with the Neurosurgery Department, Affiliate Children's Hospital, Xi'an Jiaotong University. He is responsible for one research project. He is the author of one book *Pediatric Neurosurgery*. His research interests include basic and clinical research of epilepsy, devoted to children's epilepsy networks, pathogenesis, and surgery treatment.

HANGYU SHI received the B.S. degree from the Clinical Medicine Department, Xi'an Medical University (renamed the Medical Science Center, Xi'an Jiaotong University), in 1997, and the M.S. degree from the Neurosurgery Department, Fourth Military Medical University, in 2007. He has been the Chief Physician with the Neurosurgery Department, Affiliate Children's Hospital, Xi'an Jiaotong University. He is responsible for several research projects. He is the author of one book *Pediatric Neurosurgery*. His research interests include basic and clinical research of neurosurgical diseases in children, such as the treatment of epilepsy, brain tumor, and congenital neural tube malformation.

DONG WANG received the bachelor's degree in medicine from Chongqing Medical University, in 1995. He has been engaged in pediatric neurology for 27 years since his graduation. He is currently the Director of the Neurology and Rehabilitation Department, Affiliate Children's Hospital, Xi'an Jiaotong University. He is a well-known specialist in pediatric refractory epilepsy, cerebral palsy, developmental disorders, central nervous system infection, immune-related neurological diseases, and pediatric neurological hereditary diseases in China. He is responsible for or participated in many scientific research projects. He has published many articles in international and national medical journals as the first or corresponding author. His research interests include the diagnosis and treatments of pediatric refractory epilepsy, immune-related neurological diseases, and pediatric neurological hereditary diseases.

SHANSHAN JIA received the B.S. degree in medicine from Southwest Medical University, in 2006, and the M.S. degree in pediatrics from Xi'an Jiaotong University, in 2016, where she is currently pursuing the Ph.D. degree in rehabilitation medicine and physiotherapy. Since 2006, she has been a Clinical Doctor with the Xi'an Children's Hospital. She is currently an Associate Chief Physician with the Department of Neurology. She passed the National Electroencephalography Advanced Level Examination organized by the National Examination Center and CAAE. She has published several professional articles, presided over a provincial research project, participated in a national natural science foundation project as a main project participant, and participated in the completion of a municipal project. Her research interests include epilepsy and electroencephalography. She received the Third Prize of the Xi'an University of Science and Technology.

WEISU LI received the B.Eng. degree from the Jining Medical College, in 2020. She is currently pursuing the degree with the Xijing College. Her research interest includes information data mining research.