An Effective Dual Self-Attention Residual Network for Seizure Prediction

Xinwu Yang, Jiaqi Zhao, Qi Sun, Jianbo Lu, and Xu Ma

***Abstract*— As one of the most challenging data analysis tasks in chronic brain diseases, epileptic seizure prediction has attracted extensive attention from many researchers. Seizure prediction, can greatly improve patients' quality of life in many ways, such as preventing accidents and reducing harm that may occur during epileptic seizures. This work aims to develop a general method for predicting seizures in specific patients through exploring the time-frequency correlation of features obtained from multi-channel EEG signals. We convert the original EEG signals into spectrograms that represent time-frequency characteristics by applying short-time Fourier transform (STFT) to the EEG signals. For the first time, we propose a dual self-attention residual network (RDANet) that combines a spectrum attention module integrating local features with global features, with a channel attention module mining the interdependence between channel mappings to achieve better forecasting performance. Our proposed approach achieved a sensitivity of 89.33%, a specificity of 93.02%, an AUC of 91.26% and an accuracy of 92.07% on 13 patients from the public CHB-MIT scalp EEG dataset. Our experiments show that different EEG signal prediction segment lengths are an important factor affecting prediction performance. Our proposed method is competitive and achieves good robustness without patient-specific engineering.**

***Index Terms*— Dual self-attention, multi-channel EEG signals, residual network, seizure prediction.**

I. Introduction

ACCORDING to the International League Against Epilepsy (ILAE) report [1], epilepsy is defined as a group of neurological brain disorders due to excessive abnormal brain activities. Epileptic seizures may cause loss of consciousness or perception and disorders of mood or other cognitive functions, even an increased risk of premature mortality [2]. In addition, there are differences in the frequency of seizures between different epileptic patients, ranging from less than one seizure per year to several seizures per day. It has been counted that approximate 50 million people around the world have epilepsy and up to 2 million new patients suffer from epilepsy every year [3].

For decades, the treatment methods for epilepsy mainly include pharmacological and surgical treatments. However, roughly 25% of the epileptic patients cannot be completely controlled by above two methods [4]. Due to the frequent occurrence of seizures, epilepsy has a great impact on patients and their families psychologically and physically. Hence, being able to predict epileptic seizures is crucial for patients to prevent accidents and improve the quality of life.

Similar to other neurological disorders, epilepsy can be recorded and analyzed by electroencephalogram (EEG) which is considered as the most powerful diagnostic tool of epilepsy. EEG signals can be divided into two categories: scalp EEG (sEEG) signals [5] recorded directly by placing electrodes on the surface of patients scalp, and intracranial EEG (iEEG) [6] signals recorded by implanting the electrodes in the brain tissue during surgery. Due to the high risk of collecting signals from brain tissue and the need for lots of professional knowledge, current research work is mainly carried out by sEEG. As shown in Fig. 1, most epileptic seizure prediction studies suppose that there are four consecutive states of brain activity in epileptic EEG signals, including preictal state (the period just before the seizure), ictal state (the period of a seizure), postictal state (the period following a seizure) and interictal state (the remaining peried) [3], [4], [7]. The study of seizure prediction is concerned with identifying the preictal state. If this time is identified, patient can be given a warning, so that they can take action to neutralize an incoming seizure or limit the injuries from a seizure. Therefore, the aim of seizure prediction is to distinguish between the interictal stage and the preictal stage [6], [8]–[13].

A flowchart of a general seizures prediction system is shown in Fig. 2. The whole process includes data acquisition, EEG signal preprocessing, feature extraction, classification and evaluation results. Some of epileptic seizure prediction studies were based on extracting features from EEG signals and applying thresholds to discriminate between the preictal state and interictal state. Chu *et al*. [14] extracted the fourier coefficients of the six frequency bands on 16 patients from the EEG datasets and set a threshold for classification, obtaining a sensitivity of 86.67% and a false alarm rate (FPR)

Manuscript received March 9, 2021; revised July 11, 2021; accepted July 14, 2021. Date of publication August 9, 2021; date of current version August 20, 2021. This work was supported by the National Key Research and Development Program of China under Grant 2016YFC1000307 and Grant 2016YFC1000307-10. *(Corresponding authors: Jianbo Lu; Xu Ma.)*

Xinwu Yang and Jiaqi Zhao are with the College of Computer Science, Faculty of Information Technology, Beijing University of Technology, Chaoyang, Beijing 100124, China (e-mail: yangxinwu@bjut.edu.cn; 1144059352@qq.com).

Qi Sun is with the Human Genetics Resource Center, National Research Institute for Family Planning, Beijing 100081, China, and also with the Graduate School, Peking Union Medical College, Beijing 100730, China (e-mail: cmusunqi@163.com).

Jianbo Lu and Xu Ma are with the Human Genetics Resource Center, National Research Institute for Family Planning, Beijing 100081, China (e-mail: jblu@lsec.cc.ac.cn; genetic88@126.com).

Digital Object Identifier 10.1109/TNSRE.2021.3103210

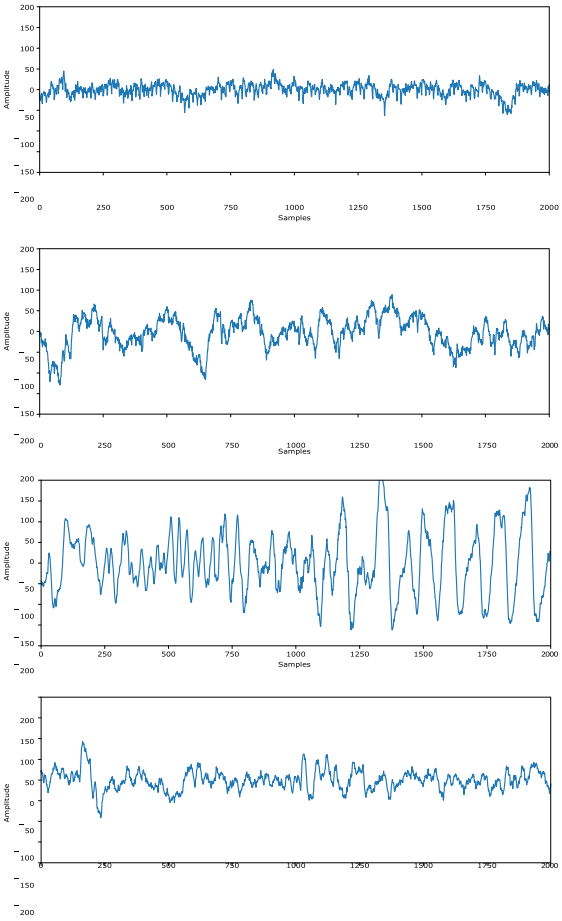


Fig. 1. Above of the figure represent the interictal, preictal, ictal, and postictal states of EEG recordings, respectively.

Fig. 2. A flowchart for predicting seizures.

of 0.367/h. Ibrahim *et al*. [9] proposed a statistical time–domain approach which depends on estimating probability density functions (PDFs) for the signals. Then they preseted probability thresholds for EEG channel selection and seizure prediction. However, the method of setting the threshold didn't take the complexity of EEG signals into account, as well as reducing flexibility of seizures prediction. Tradictional machine learning algorithms have been widely used on epileptic seizure prediction to distinguish between preictal and interictal periods. Rasekhi *et al*. [10] extracted 22 univariate features, including statistics and spectral moments, entropy, Hjorth parameters, and Lyapunov exponents, achieving the sensitivities of 73.9% and 73.5% on the iEEG dataset by using SVM [15] and multilayer perceptrons. Vipin *et al.* [16] proposed a seizure classification method based on weighted multiscale Renyi permutation entropy (WMRPE) and rhythms obtained with FourierBessel series expansion (FBSE) of EEG signals. Rishi *et al.* [17] proposed an automatic classification method for epileptic EEG signals based on iterative filtering (IF) of EEG signals. Features were extracted from an intrinsic mode function (IMF) obtained by IF decomposition and an envelope function (AE) obtained by a discrete energy separation algorithm, and epilepsy signals were classified according to their p values. This method was evaluated on the Bonn University data set, using 10-fold cross-validation to achieve an ACC of 99.5% in a two-way classification (AE) and an ACC of 98% in a three-way classification (AB-CD-E). Abhijit *et al.* [18] proposed a method using an empirical wavelet transform to obtain the combined transient value and frequency of the signal at an adaptive frequency scale to identify epileptic seizure and seizure-free signals. Six classifiers with 10-fold cross-validation were used for epilepsy classification on the CHB-MIT data set, achieving sensitivity of 97.91%, specificity of 99.57%, and accuracy of 99.41%. With the development of deep learning in the field of image recognition, text classification and speech recognition, Some seizure prediction methods using deep learning have also emerged. Khan *et al*. [11] proposed to use the wavelet transform of the original EEG signal as the input of the convolutional neural networks and tested it on the CHB-MIT datasets, reaching a sensitivity of 87.8% and a false positives rate of 0.147/h. Truong *et al*. [12] first used short-time Fourier transform to convert EEG data into spectrograms, then used convolutional neural networks to extract and classify features. Evaluated on the Freiburg, CHB-MIT, and American Epilepsy Society Seizure Prediction Challenge datasets respectively, this work obtained sensitivities of 81.4%, 81.2%, and 75%, as well as 0.06/h, 0.16/h, and 0.21/h false positive rate. Ozcan *et al*. [13] proposed a multi-frame 3DCNN model to generally evaluate the spatio-temporal dependency of training data. They extracted the time–domain and frequency-domain features of the EEG signal, such as spectral band power, statistical moment and Hjorth parameters, converting them into a series of multi-color images according to the topology of the EEG channel, then classified them by a multi-frame 3DCNN model. Their method provided a sensitivity of 85.71%, a false positive rate of 0.096/h and an AUC of 88.60%.

In fact, the majority of methods still failed to provide acceptable performance for some patients in different datasets utilizing EEG data. It seems to be caused by two main reasons. On the one head, there is a lack of uniform labeled data. It is tough to distinguish the preictal and ictal periods by eyes alone, because the boundary of four periods is difficult to define. On the another head, the time, characteristics and dynamics of various epileptic states vary greatly among different patients, thus the typical characteristics of seizures in some patients may not be applicable to other patients. At present, there is no general method to achieve high predictive performance for each patient instead of being specially trained for specific patient. As a result, most studies that achieve high prediction performance have adopted patient-specific methods [14], [19]–[24].

In this paper, we propose a patient-specific seizure prediction method by developing deep learning-based models to improve the performance of epileptic seizure prediction. The time-frequency characteristics of EEG signals used in our method are very important for seizures prediction. Some studies applied convolutional neural networks to epileptic seizure prediction, and confirmed that convolutional neural networks are an effective method

for EEG classification [11], [12]. However, because of the complexity and diversity of EEG signals and the simple structure of convolutional neural networks, many studies have obtained low seizure prediction performance. In the present work, we used a residual network to improve the performance of epileptic seizure prediction, and for the first time proposed a dual self-attention residual network (RDANet) to predict epileptic seizures. This RDANet adaptively integrates the local features of EEG signals and the global features through the spectral attention module and channel attention module, and strengthens the correlation between multi-channel EEG signals. Furthermore, we used leave- one-out cross-validation to evaluate the prediction results to ensure they were representative of real conditions. In general, compared with existing seizure prediction algorithms, numerical experiments demonstrated that the proposed algorithm is effective.

II. Materials and Methods

A. Dataset

The CHB-MIT scalp EEG dataset, collected by Boston Children's Hospital, contains scalp electroencephalogram (sEEG) data of 22 pediatric subjects with 844 hours of successive recording and is publicly available [25] and [26]. Using bipolar montage technology of the international 10-20 system, EEG signals are collected from 22 electrodes at a sampling rate of 256 Hz. Litt *et al.* [27] proved that complicated epileptic discharges are common 7 hours before seizures, while activities similar to seizures are about 2 hours before the real onset. At the same time, the accumulated energy augments during the 50 minutes before the seizure. In this study, we define the preictal state as the 30 minutes EEG signal before the onset of the epileptic seizure, and define the interictal state as the period between 4 hours after the end of a seizure and 4 hours before the beginning of the next seizure. In order to allow therapeutic intervention, it is necessary to allow a short time window before the onset of the seizure [28]. From a clinical point of view, it is best to have a long enough intervention to allow effective therapeutic intervention or preventive measures [12]. In this study, an intervention period is set to 5 minutes and removed from the training data. Furthermore, considering that seizures can happen very close to each other, we are interested in predicting the leading crisis, which is about less than 30 min away from the next one [29]. Based on Truong *et al.* [12], we only consider patients with less than 10 seizures per day, because it is not very critical to perform the task for patients having a seizure every 2 h on average. Based on above definitions and limitations, Table I shows that we have selected 13 patients with 64 seizures and 268.6 hours interictal data from this dataset.

B. Pre-Processing

Since epileptic seizures are rare events, the number of records in the interictal period is higher than the number of records in the preictal period. Most machine learning algorithms presume that the data of different categories are evenly distributed. If the amount of data is unbalanced, we will

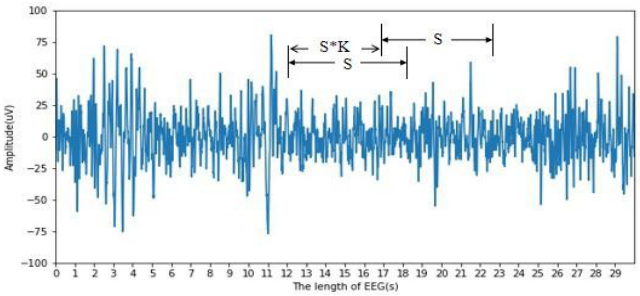


Fig. 3. EEG signal segmentation process.

TABLE I
The Number of Seizures and Interical Data Duration of 13 Patients in the CHB-MIT Dataset

Patient id	No. of seizures	Iterictal hours(/h)
Pt 1	7	14
Pt 2	3	23
Pt 3	6	22
Pt 5	5	14
Pt 9	4	46.7
Pt 10	6	26
Pt 13	5	14
Pt 14	5	5
Pt 18	6	24
Pt 19	3	25
Pt 20	5	20
Pt 21	4	22
Pt 23	5	12.9
totlal	64	268.6

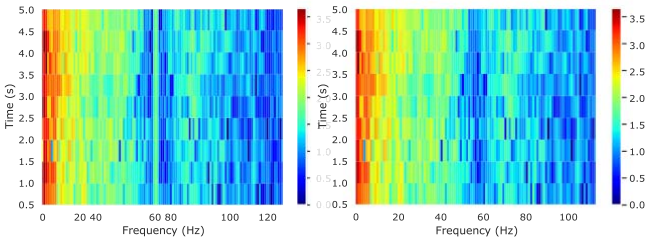
obtained a classifier biased towards the larger number of categories [30]. In order to solve the problem of data imbalance, we use overlapping sampling to obtain more preictal data during the training phase. As shown in Fig. 3. Specifically, We define the length of the pre-ictal signal as M, and the length of the inter-ictal signal as N, and calculate the length ratio of the two types of data K (formula 1). Set the sampling window S to 5 seconds. In order to obtain the same amount of the two types of data, the pre-seizure data are collected with a moving step of $S \times K$ during the training phase, and the interictal data is collected with a moving step of S. The EEG signal is segmented to obtain the preictal and interictal signal fragments of a and b respectively (see formula 2 and 3).

$$K = \frac{M}{N} \tag{1}$$

$$a = \frac{M - S}{S \times K} + 1 \tag{2}$$

$$b = \frac{N - S}{S} + 1 \tag{3}$$

The characteristics of time-frequency domain are very important in EEG data analysis, which usually are studied by means of the spectrograms that represents three parameters in a two-dimensional graph. The wavelet transform and short-time Fourier transform (STFT) are common methods for converting EEG signals into spectrograms [11], [12], [31]. We slide a 5-second,15-sencond and 30-second window and used STFT to transform the original EEG signal into a two-dimensional matrix with time and frequency as the axis (as shown in Fig. 3). The majority of EEG recordings in the CHB-MIT dataset are contaminated by 60 Hz power line noise, which



(a) (b)

Fig. 4. (a) Logarithmic spectrum before denoising, (b) Logarithmic spectrum after denoising.

can be effectively removed by excluding components in the frequency ranges of 57-63 Hz and 117-123 Hz. Meanwhile the DC component (0 Hz) is also excluded. The spectrogram of the 5-second EEG signal and the denoising spectrogram are shown in Fig. 4.

C. Model

Our proposed model is composed of a residual network (ResNet) [32] and a dual self-attention mechanism [33]. Taking the spectrum attention module in the upper part of Fig. 6 as an example, we use the spectrograms as the input to the network, and extract the potential time-frequency features through the ResNet. Then, the features are input into the spectrum attention module through the following three steps to generate new global spectral features. The first step is generating a spectrum attention matrix to describe the spatial relationship between any spectrograms. Second, we execute matrix multiplication on the attention matrix and the original features. Third, we perform the final global spectral feature by executing an element-wise sum between the result matrix of the last step and the original feature. The process of the channel attention module is similar to that of the spectrum attention module. Finally, we merge the features of the two modules and add them to the original features to better capture the characteristics of the EEG signal.

1) Residual Network: CNN [34] has been widely used in computer vision, natural language processing, etc. According to scientific studies, in order to get expression ability and fit the potential mapping relationship better, it is effective to deepen the number of network layers or widen the network structure. As the number of layers of the neural network deepens, there will be a problem of disappearing gradients, and the optimization of stochastic gradient descent becomes more difficult. Recently, He *et al.* [35] proposed a ResNet to combat the above problem during training very deep convolution networks, which is composed of several residual blocks. As shown in Fig. 5, the residual block adds an identity mapping to the network through shortcut connections, which neither adds additional parameters nor calculations, and solves model degradation problem to a certain extent. Each residual block is composed of two 3×3 convolution layers, batch normalization and ReLu. Besides, it has two paths added up, namely the residual path $F(x)$ and the identity map x. In this study, we used 4 residual blocks, which are connected with each other.

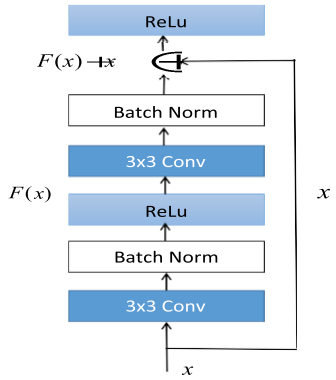


Fig. 5. A residual block.

2) Dual Self-Attention Mechanism: The attention mechanism originated from human visual perception. When human is perceiving an object, he usually first scans the global image and then focuses attention on a specific part to obtain more detailed information, as well as suppressing other useless information. With the further research of the attention mechanism, the self-attention mechanism proposed by the Google machine translation research team [36] has received extensive attention because it can learn the relationship between a certain position and other positions and capture the context dependence. Liu *et al.* [37] used self-attention generative adversarial network to complete the task of image completion. Bello *et al.* [38] applied self-attention mechanism to enhance image classification accuracy. For seizure prediction studies, we first propose to apply a dual self-attention mechanism to capture global information of EEG signals. Next, we will elaborate on the processes.

a) Spectrum attention module: As shown in Fig. 6, given a local feature $X \in R^{C\times H\times W}$, we first feed it into a convolution layer to produce two new feature maps Y and Z, respectively, where the dimensions of Y and Z are both $R^{C\times H\times W}$. Then we reshape them to $R^{C\times N}$, where $N = H^{\times}W$. Then we carry out a matrix multiplication between the transposition of Y and Z, and then get a weight matrix S with a dimension of by means of a softmax layer [33]:

$$S_{ji} = \frac{e^{Y_i \cdot Z_j}}{\sum_{i=1}^{N} e^{Y_i \cdot Z_j}} \tag{4}$$

where S_{ji} represents the influence of the i^{th} position on j^{th} the position. The more similar feature representations of two position lead to the greater relevance between them. At the same time, a new feature $T \in R^{C\times H\times W}$ representation will be generate by feeding X to the convolution layers and reshape it to $R^{C\times N}$. Then we perform a matrix multiplication of T and S, then reshape the result into $R^{C\times H\times W}$. Finally, we multiply it by a scaling parameter α and execute a element-wise sum operation with X to gain the final result $E \in R^{C\times H\times W}$ [33] (as shown in formula 5), where α is initialized to 0, allocated more weight little by little [39]. The final feature E, a sum of the weighted feature of the spectrum and the original feature, has a comprehensive contextual perspective and optionally collects global information based on the spectrum attention map.

$$E_j = \alpha \sum_{i=1}^{N} (S_{ji} * T_i) + X_j \tag{5}$$

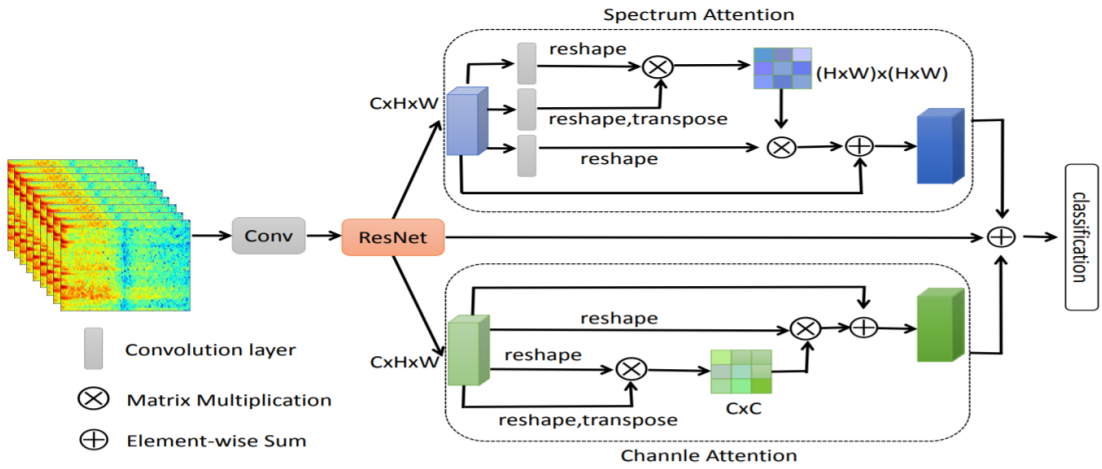


Fig. 6. The structure diagram of RDANet.

b) Channel attention module: Different channel features represent different semantics of EEG signals. The channel attention mechanism is used to mine the interdependence between channel mappings, so that different semantic representations are related to each other. First of all, we reshape X to $R^{C\times N}$, and then apply a matrix multiplication between X and the transpose of X. Finally, we apply a softmax layer to obtain the channel attention map $P \in R^{C\times C}$ [33] (as shown in formula 6), where P_{ji} measures the influence of the i^{th} channel on the j^{th} channel.

$$P_{ji} = \frac{e^{X_i \cdot X_j}}{\sum_{i=1}^{C} e^{X_i \cdot X_j}} \tag{6}$$

Next, we perform a matrix multiplication on P and X, then reshape the result to $R^{C\times H\times W}$. We multiply the result by a scaling parameter β and execute a element-wise sum operation with X to gain the final output $E \in R^{C\times H\times W}$ [33] (as shown in formula 7), where β is initialized to 0. In order to make full use of channel and spectrum context information, we integrate these two attention modules by an element-wise sum. After fusion with the original features, the final feature map can be obtained through average pooling.

$$E_j = \beta \sum_{i=1}^{C} (P_{ji} * X_i) + X_j \tag{7}$$

3) Training: In order to get results similar to the real conditions, we adopt a leave-one-out cross-validation method for each patient [40]. That is, if a patient contains n seizures and t hours of interictal recordings, the entire interictal recordings are divided into n parts and each part has approximately $\frac{t}{n}$ hours, which is randomly grouped with any preictal recordings. This round is done n times and in each time one of the preictal-interictal pairs is reserved for testing, while the remaining $n-1$ pairs are used in the training phase. Generally, some studies usually randomly divide 80% of the data as the training set, and the remaining 20% is used as the validation set to monitor overfitting [13]. However, this method is suitable for data that are independent in time, such as image classification. The EEG data is time-dependent, so we should select samples at a different time period from the training period to monitor whether the model has begun to overfit. In this study, we select 25% of the later samples from the preictal and interictal recordings in the training set as the validation set for monitoring, and the remaining 75% of the samples are used as the training set [12]. Although the number of iterations in the training phase increases the accuracy of training, there is still a problem of overfitting, which we used early-stop to solve it. In detail, when it is detected that the loss on the validation set has started to increase, we stop training and the network parameters are stored at the lowest validation loss, immediately.

TABLE II
THE PARAMETER TABLE OF RDANET STRUCTURE

Layer name	Output size	Network structure
Input	1×22×9×114	-
Conv1	64×1×7×55	22×3×5conv,64, stride1x1×2
Pooling	64×1×7×28	1×1×2 maxpooling, reshape
Resblock1	64×7×28	$\begin{pmatrix}3\times3\text{conv},64\\3\times3\text{conv},64\end{pmatrix}*2$
Resblock2	128×4×14	$\begin{pmatrix}3\times3\text{conv},128\\3\times3\text{conv},128\end{pmatrix}*2$
Res Block3	256×2×7	$\begin{pmatrix}3\times3\text{conv},256\\3\times3\text{conv},256\end{pmatrix}*2$
Resblock4	512×1×4	$\begin{pmatrix}3\times3\text{conv},512\\3\times3\text{conv},512\end{pmatrix}*2$
Daul self-attention	512×1×4	spectrum attention, channel attention
Average pooling	512×1	-
Classification	2×1	softmax

The parameters of the RDANet network are shown in Table II. The input of the model is $1\times22\times9\times114$, where 22 represents the number of the EEG signal channels and 9×114 represents the dimension of the spectrograms. Each convolutional layer is followed by a batch normalization, a dropout and a ReLu activation function. We firstly obtain a $64\times7\times28$ matrix by feeding the above feature map into a convolutional layer followed by reshape operation. Subsequently, we use four ResBlock layers to extract the deep features of the EEG signal. A dual self-attention layer fusing the global features is followed by a fully connected layers with a sigmoid activation function. We adopt a cross-entropy loss function as the cost function. The batch size is 32.

TABLE III
SEIZURE PREDICTION RESULTS OF CNN ON 13 PATIENTS IN THE CHB-MIT DATASET

Patient id	Sensitivity(%)	Specificity(%)	AUC(%)	Accuracy(%)
Pt 1	99.6 ± 00.2	99.9 ± 00.0	99.8 ± 00.1	99.9 ± 00.0
Pt 2	58.2 ± 21.0	81.7 ± 21.3	69.6 ± 00.8	80.3 ± 18.7
Pt 3	92.2 ± 03.9	95.1 ± 02.8	96.2 ± 01.2	94.8 ± 02.0
Pt 5	88.4 ± 01.6	87.9 ± 00.6	93.4 ± 01.4	88.0 ± 00.7
Pt 9	66.4 ± 07.3	91.4 ± 04.6	76.8 ± 06.9	90.3 ± 04.7
Pt 10	69.4 ± 00.8	85.0 ± 01.1	78.0 ± 03.7	83.4 ± 01.0
Pt 13	97.9 ± 01.9	96.2 ± 02.3	97.7 ± 00.8	96.4 ± 01.7
Pt 14	63.2 ± 07.6	84.6 ± 05.8	71.8 ± 03.3	77.4 ± 01.3
Pt 18	95.0 ± 01.3	94.0 ± 01.2	95.4 ± 00.8	94.1 ± 01.2
Pt 19	82.8 ± 02.6	98.4 ± 01.1	91.0 ± 01.0	97.5 ± 00.9
Pt 20	98.1 ± 00.0	98.2 ± 01.0	98.7 ± 00.8	98.2 ± 00.9
Pt 21	99.0 ± 00.4	86.4 ± 00.2	93.8 ± 00.1	87.5 ± 00.2
Pt 23	99.7 ± 00.0	98.9 ± 00.1	99.7 ± 00.1	99.1 ± 00.1

TABLE IV
SEIZURE PREDICTION RESULTS OF RESNET ON 13 PATIENTS IN THE CHB-MIT DATASET

Patient id	Sensitivity(%)	Specificity(%)	AUC(%)	Accuracy(%)
Pt 1	99.5 ± 00.0	99.9 ± 00.0	99.8 ± 00.0	99.8 ± 00.1
Pt 2	53.9 ± 04.7	91.2 ± 04.5	73.2 ± 02.5	88.9 ± 03.9
Pt 3	95.7 ± 00.7	95.6 ± 00.4	97.8 ± 00.0	95.6 ± 00.3
Pt 5	92.6 ± 01.1	89.6 ± 01.5	96.3 ± 01.2	90.0 ± 01.5
Pt 9	71.0 ± 02.1	92.1 ± 00.6	79.1 ± 01.7	91.2 ± 00.5
Pt 10	77.2 ± 16.8	82.6 ± 10.4	81.9 ± 03.4	82.1 ± 07.5
Pt 13	97.9 ± 02.0	94.7 ± 01.7	97.2 ± 01.2	95.1 ± 01.2
Pt 14	67.8 ± 13.7	70.4 ± 11.7	66.0 ± 01.0	69.5 ± 03.2
Pt 18	96.5 ± 01.0	96.8 ± 01.9	99.0 ± 00.5	96.8 ± 01.8
Pt 19	90.4 ± 03.3	97.1 ± 01.3	95.3 ± 01.0	96.7 ± 01.0
Pt 20	98.9 ± 00.3	99.7 ± 00.1	99.7 ± 00.0	99.6 ± 00.1
Pt 21	96.8 ± 01.5	87.3 ± 00.2	93.2 ± 01.0	88.1 ± 00.3
Pt 23	99.4 ± 00.2	98.3 ± 00.0	99.6 ± 00.0	98.5 ± 00.1

TABLE V
SEIZURE PREDICTION RESULTS OF RDANET ON 13 PATIENTS IN THE CHB-MIT DATASET

Patient id	Sensitivity(%)	Specificity(%)	AUC(%)	Accuracy(%)
Pt 1	99.7 ± 00.4	99.6 ± 00.4	99.8 ± 00.3	99.7 ± 00.4
Pt 2	59.9 ± 02.5	91.6 ± 01.8	77.8 ± 00.5	89.7 ± 01.5
Pt 3	94.9 ± 01.3	96.9 ± 01.3	98.1 ± 00.2	96.6 ± 01.0
Pt 5	94.3 ± 00.6	92.7 ± 00.0	97.6 ± 00.2	92.9 ± 00.0
Pt 9	66.0 ± 02.3	92.1 ± 01.9	77.7 ± 02.3	91.0 ± 01.9
Pt 10	84.8 ± 03.4	76.9 ± 04.0	82.3 ± 00.3	77.7 ± 03.3
Pt 13	98.6 ± 01.5	97.0 ± 02.0	98.5 ± 01.0	97.3 ± 01.5
Pt 14	56.0 ± 00.1	82.6 ± 05.1	65.8 ± 03.7	73.7 ± 03.4
Pt 18	96.2 ± 00.9	96.0 ± 01.9	98.5 ± 01.0	96.0 ± 01.8
Pt 19	95.5 ± 02.3	96.1 ± 01.9	97.2 ± 01.9	96.0 ± 02.0
Pt 20	99.0 ± 00.2	99.7 ± 00.0	99.7 ± 00.0	99.7 ± 00.1
Pt 21	98.0 ± 01.1	87.4 ± 01.7	94.4 ± 00.2	88.3 ± 01.5
Pt 23	99.5 ± 00.1	98.3 ± 00.0	99.5 ± 00.1	98.5 ± 00.1

The dropout rate and the learning rate are set to 0.5 and 0.0005, respectively. Our new model is accomplished with Keras 2.2.2 of tensorflow 1.4.0 backend.

III. RESULTS

In this study, we use four parameters to evaluate the performance of the proposed model: sensitivity, specificity, accuracy and AUC which is a commonly used metric to evaluate the performance of classification tasks by calculating the area under the receiver operating characteristic curve (ROC).

With the aim of predicting segment 5s, the results of the CNN, ResNet, and RDANet models evaluated on 13 cases in the CHB-MIT dataset are shown in Tables III, IV and V, respectively. The experiments were executed twice and the average results with standard deviation are reported. In general, the numerical experiments show that the model performance varied from one patient to another. As the Table III shows, the results of Pt 2, Pt 9 and Pt 14 are lower than those of other patients. This is reasonable since Pt 2 includes only 3 seizures and there are few preictal recordings available for training, which makes it difficult for a simple CNN model to extract the characteristics of preictal data. Pt 9 had 46.7 hours of interictal recordings but only 4 seizures, which causes extremely unbalanced data and the poor classification performance. Similarly, Pt 14 has only a small number of interval recordings, which makes it difficult for the CNN model to achieve high classification performance.

The evaluation results of the ResNet model for 13 patients in the same dataset are shown in Table IV. Compared with Table III, the predictive results for Pt9 and Pt14 showed significant improvement. The sensitivity of Pt2 decreased a little, but the specificity improved at the same time, causing a corresponding increase in AUC and accuracy. It seems reasonable that the ResNet model can alleviate the problem of data imbalance through quick connections and deeper network architecture. Compared with the CNN model, it can also be demonstrated that the predictive performance of the ResNet model on other patients is obviously improved, except for Pt2.

Table V shows that the evaluation results of the RDANet model on 13 patients in the CHB-MIT dataset. Compared with CNN and ResNet models, the evaluation results of RDANet model has improved on many patients, except for Pt9 and Pt14. For example, the sensitivity and AUC of Pt1, Pt13, Pt20 and Pt23 are nearly 100%. It indicates that the dual self-attention module can enhance feature representations of EEG data, which results in improving predictive performance. However, the pre-period and inter-period data of pt9 showed a very large gap. Although resampling solved the problem of data imbalance, it was easy to overfit the previous samples, resulting in higher specificity than sensitivity. Pt 14 has only a small number of interval recordings, which made it difficult for the RDANet model to achieve high classification performance.

The AUC is a comprehensive parameter for evaluating classification tasks, and therefore we compared the ROCs of seizure forecasting performance for the three models testing different patients. The ROC curves obtained by evaluating the CNN, ResNet, and RDANet models on data from 13 patients of the CHB-MIT dataset are shown in Fig. 7. It can be seen that the ROC curves of the three models almost overlap for Pt 1 and Pt 23, and that the AUC values are almost the same. For Pt 1 and Pt 23, a simple CNN model can be used to obtain better results, and the AUC is close to 100%. This shows that when the original model has good classification performance, our proposed model does not show a great improvement because there is little room for improvement. In terms of the ROC curves of Pt 9 and Pt 18, it can be concluded that the AUC value of the ResNet model is larger than that of the RDANet model, indicating that ResNet had better classification performance than RDANet for these two

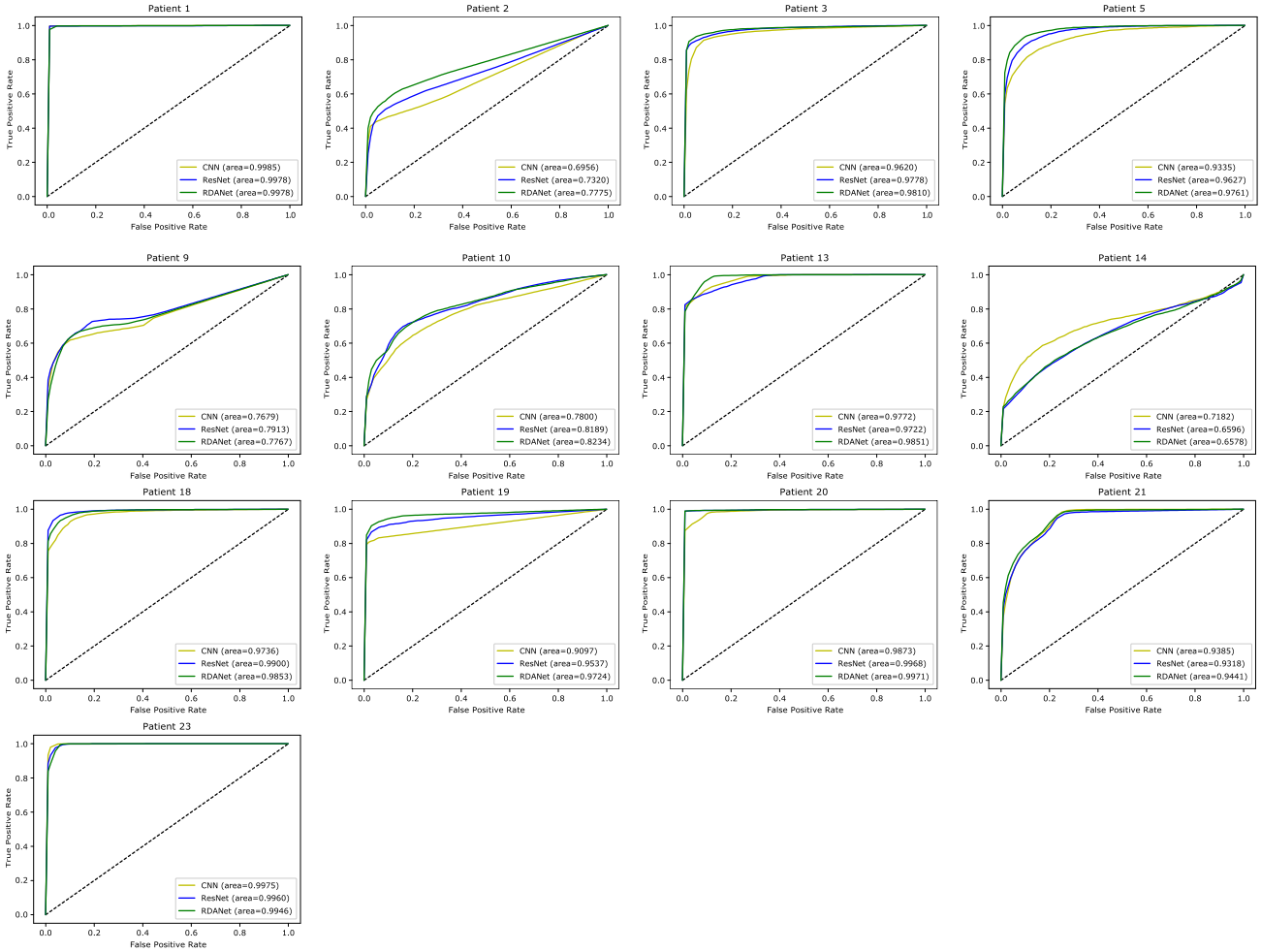


Fig. 7. Receiver operating characteristics (ROC) curves of seizure forecasting performance testing for 13 patients from the CHB-MIT dataset.

patients. According to the ROC curve of Pt 14, the AUC value of the CNN model was significantly higher than those of the other two models. The reason for this is that patient 14 has only 5 hours of interictal records, which easily results in overfitting and reduced performance if the ResNet and RDANet models are used to predict seizures. In general, although the AUC values achieved on some patients using the RDANet model were a little bit lower, our proposed model improved the overall prediction performance. From these figures, it can be clearly seen that the ratio of true positives was higher than that of false positives. Obviously, for most patients, the seizure prediction performance of the RDANet model performs best. This is because our proposed RDANet captures global features through a dual self-attention mechanism and mines the correlations of EEG signals in different channels, promoting the classification performance of preictal periods and interictal periods.

To test our model, the overall epileptic seizure prediction performance of the CNN, ResNet, and RDANet models can be represented by the weighted average of the above results (See Fig. 8). In general, the performance evaluations of the RDANet model were higher than those of the CNN and ResNet models. The CNN model consists of a four-layer convolutional neural network and two fully connected layers to classify EEG signals. The ResNet model is a network with

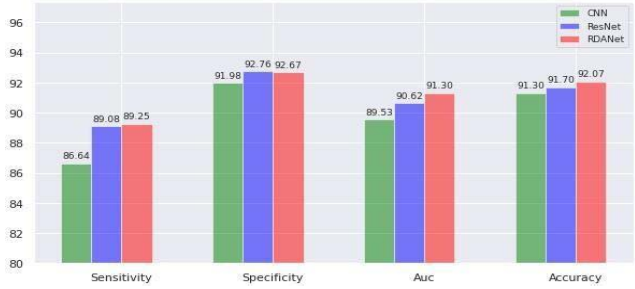


Fig. 8. The overall performance of epilepsy prediction of CNN, ResNet and RDANet models under the 5-second prediction segment.

only a four-layer ResBlock and an average pooling layer. Compared with the CNN model, the predictive results of the ResNet model were better than those of the CNN. This is because the ResNet model has a stronger expressive ability from the building of deeper networks and quick connections. By comparing the experimental results of the ResNet and RDANet models, it can be seen that after introduction of the dual self-attention module [28], the sensitivity of the RDANet model increased by 0.17%, the specificity increased by 0.09%, the AUC value increased by 0.68%, and the accuracy increased by 0.37%. Overall, the RDANet model offers more advantages for predicting seizures than the other two models.

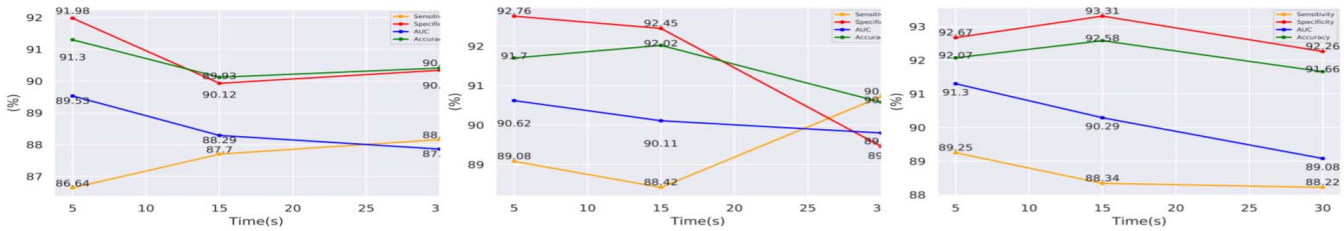
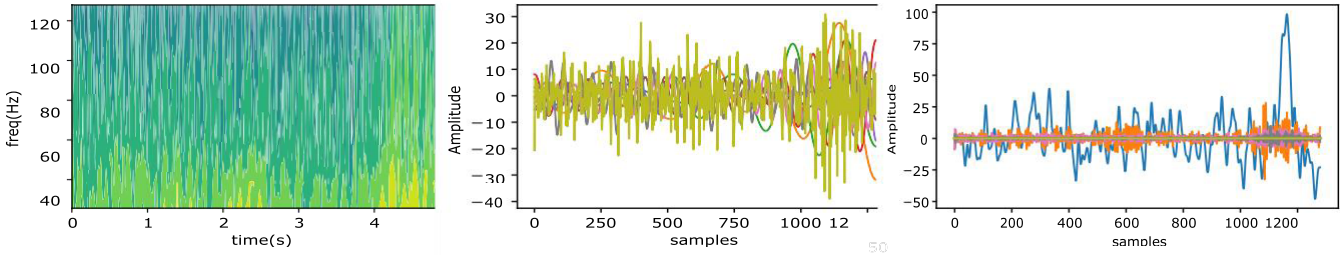


Fig. 9. Results of CNN, ResNet and RDANet on 5-second, 15-second and 30-second EEG signal segments.



(a) CWT (b) EWT (c) FBSE-EWT

Fig. 10. Spectrogram of 5-second EEG signal after CWT, EWT and FBSE-EWT.

Few studies have examined the influence of the EEG signal segment length on the prediction of epileptic seizures. We therefore explored the appropriate length of the EEG signal prediction segment to achieve the best prediction performance. In this study, we conducted repeated training on all models with different seizure prediction lengths of 15 seconds and 30 seconds. Fig. 9 is a comparison diagram of the CNN, ResNet, and RDANet models evaluated on EEG signal prediction segments of 5 seconds, 15 seconds, and 30 seconds. The experimental results show that although the sensitivity of some models increased, the specificity decreased accordingly. By comparing the comprehensive indicators of AUC and accuracy, it can be shown that when the prediction length increases, the prediction performance of the three models generally decreases.

IV. Discussion

A wide range of statistical techniques have been developed to analyze EEG signals according to their temporal and spatial resolution [41]. Many signal transformation techniques have been used for the interpretation of brain signals and detection of anomalies, such as the Fourier Transform, Short-Time Fourier Transform, and Wavelet-based Transform [31], [42]. The continuous wavelet transform [11], [43] and empirical wavelet transform [18] have been applied in the field of epileptic seizure prediction. The Fourier-Bessel series expansion (FBSE) [44]–[46] based empirical wavelet transform (EWT) [47] (FBSE-EWT) can effectively solve the problem of non-stationary signals, and was introduced for analyzing EEG signals which use empirical wavelets designed from FBSE for signal [48], [49]. Therefore, we compared four signal transformation methods in the RDANet.

We first use CWT, EWT and FBSE-EWT to convert the EEG signals respectively (see Fig. 10), and then use the RDANet network to automatically extract features and classify. The input of the model is $1 \times 22 \times 9 \times 1280$, where 22 represents the number of the EEG signal channels and 9×1280 represents the dimension after conversion. In order to improve computational efficiency and speed up training, we randomly downsample the above input results to obtain a new input of $1 \times 22 \times 9 \times 160$. The experimental results are shown in Table VI. Compared with STFT+RDANet, the CWT+RDANet model achieves a sensitivity of 91.03%, but the other three indicators are relatively low, especially the specificity. We can see that STFT and CWT perform well on interictal and pre-ictal signals, respectively. The four evaluation results of the EWT+RDANet model are very low. It may be that the signal characteristics in the time-frequency domain are more beneficial to distinguish different states of EEG signals. CWTrsp+RDANet and EWTrsp+RDANet indicate the model after downsampling, and their prediction results are poor. The reason may be that some important signals of the EEG signal are lost due to downsampling.

TABLE VI
The Result of the Signal Processed by CWT, EWT and FBSE-EWT Under the RDANet Model

Method	Sensitivity(%)	Specificity(%)	AUC(%)	Accuracy(%)
STFT	89.25	92.67	91.30	92.07
CWT	91.03	88.29	90.34	89.63
CWTrsp	82.05	94.95	81.49	76.87
EWT	88.33	82.23	87.67	84.05
EWTrsp	81.93	77.91	82.80	79.19
FBSE-EWT[1]	90.16	92.41	89.09	91.18

[1] we have only got the results of Pt 1, 2, 5, 10, 13, 14, 18, 19, 20, 21, 23, a total of 11 patients (except Pt 3 and Pt 9) in the FBSE-EWT method.

Table VII provides a comparison between the seizure prediction performance of our method and that of other methods. All methods were evaluated on the CHB-MIT scalp EEG dataset, which is a public dataset composed of long-term recordings. It is hard to decide which method was better, because each method was tested with limited data according to the different patients and different time definitions. Therefore, the generalizability of the proposed method without the need for patient-specific feature engineering is an important

TABLE VII
RESULTS OF RECENT STUDIES ON PREDICTING SEIZURES ON THE CHB-MIT SCALP EEG DATASET

Authors	Method	Sensitivity(%)	Specificity(%)	AUC(%)	Accuracy(%)
Khan *et al*. [11]	Wavelet coefficients+CNN	87.8	-	86.60	-
Truong *et al*. [12]	STFT spectral images+CNN	81.2	-	-	-
Tsiouris *et al*. [50]	Zero crossings, wavelet coefficients, PSD+LSTM	99.38	99.60	-	-
Truong *et al*. [51]	STFT spectral images+GAN	-	-	77.68	-
Ozcan *et al*. [13]	Spectral power, Statistical moments, Hjorth parameters+3DCNN	85.71	-	88.60	-
Sadeghzadeh *et al.* [52]	3 Features +Threshold	90.62	88.34	-	88.76
Detti P *et al*. [53]	EEG Synchronization+SVM/LightGBM/ThAlgo	100	95.97	-	-
Usman S M *et al*. [54]	STFT spectral images+CNN+SVM	92.7	90.8	-	-
This work	STFT spectral images+RDANet	89.25	92.67	91.30	92.07

indicator affecting seizure prediction performance. In a similar approach, Khan *et al*. [11] proposed a method using the wavelet transform of the original EEG signal as the input to the convolutional neural network, and evaluated their approach on 15 patients. Truong *et al*. [12] proposed a method that combined a short-time Fourier transform and convolutional neural network, and tested their method on 13 patients from the same dataset. Obviously, our proposed approach outperformed their methods.

To reveal the seizure prediction performance of a proposed model under conditions similar to the real situation, leave one-out cross-validation can be used for training. Detti *et al.* [53] proposed an approach based on finding synchronization patterns in the EEG that allowed them to distinguish preictal from interictal states in real time, and compared three classifiers: SVM, gradient boosting decision tree algorithms, and the threshold-based approach ThAlgo. There methods used five-fold cross-validation to perform evaluations on the CHB-MIT scalp EEG dataset. Because they did not use leave-one-out cross-validation, they correctly predicted all seizures using the algorithm ThAlgo, whereas LightGBM had a prediction rate of 98% and SVM a prediction rate 86.7%. Tsiouris *et al*. [50] proposed a method that combines the wavelet transform coefficients and power spectral density of the EEG signal with the LSTM to predict epileptic seizures, and obtained a high sensitivity of 99.84% and a false positive rate of 0.02/h. Their method also did not use leave-one-out cross-validation.

Truong *et al*. [51] used a Generative Adversarial Network (GAN) to predict epileptic seizures and obtained an AUC of 77.68%. Because their method used semi-supervised training, the amount of training data was insufficient, which led to a gap in the prediction performance of the supervised training. Ozcan *et al*[13] proposed a multi-frame 3DCNN model to predict seizures. Under conditions using the same number of patients and the same preictal periods, our proposed RDANet obtained a sensitivity of 88.63%, an FPR of 0.122/h, an AUC value of 89.91% and an accuracy of 89.78%. The sensitivity and AUC of our method were higher than those of Ozcan. Hoda *et al*. [52] proposed a real-time low computation epileptic seizure prediction method. They extracted level-3 features of early onset signals, and compared the third level information with predefined threshold levels to determine whether extracted features were associated with epilepsy. Except for sensitivity, the other two indicators were relatively lower. Usman *et al.* [54] proposed a simple seizure prediction system using convolutional neural networks to automatically extract features and perform SVM classification. Their method was evaluated on the CHB-MIT dataset, and resulted in average sensitivity and specificity of 92.7% and 90.8% respectively. However, they did not describe how the data were divided and what verification methods were used, and the details of the experiment were not very clear.

V. CONCLUSION

In this work, we propose a dual self-attention residual network (RDANet) for predicting epileptic seizures, which can integrate global features into local features through a self-attention mechanism. Specifically, the spectrum attention module and the channel attention module capture the global dependence on the spectrum and the interdependence on the channels, respectively, which improve the ability to express local features. In general, our proposed method is competitive with other latest methods and is generalizable because of no patient-specific engineering. However, the CHB-MIT dataset mainly consist of pediatric patients, our method will be comprehensive tested in more patients from different age groups under different clinical conditions to comfirm overall performance.

REFERENCES

[1] R. S. Fisher *et al.*, "ILAE official report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.
[2] 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.
[3] W. H. Organization, "Neurological disorders: Public health challenges," *J. Policy Pract. Intellectual Disabilities*, vol. 5, no. 1, p. 75, 2010.
[4] A. Yadollahpour and M. Jalilifar, "Seizure prediction methods: A review of the current predicting techniques," *Biomed. Pharmacol. J.*, vol. 7, no. 1, pp. 153–162, 2015.
[5] T. C. Technologies. (2012). *10/20 System Positioning Manual*. [Online]. Available: https://www.trans-cranial.com/docs/1020posmanv10pdf.pdf
[6] M. Le Van Quyen *et al.*, "Anticipation of epileptic seizures from standard EEG recordings," *Lancet*, vol. 357, no. 9251, pp. 183–188, Jan. 2001.
[7] K. Rasheed *et al.*, "Machine learning for predict- ing epileptic seizures using eeg signals: A review," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 139–155, 2021.
[8] R. S. Delamont and M. C. Walker, "Pre-ictal autonomic changes," *Epilepsy Res.*, vol. 97, no. 3, pp. 267–272, Dec. 2011.
[9] F. Ibrahim *et al.*, "A statistical framework for EEG channel selection and seizure prediction on mobile," *Int. J. Speech Technol.*, vol. 22, no. 1, pp. 191–203, Mar. 2019.

[10] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, "Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods," *J. Neurosci. Methods*, vol. 217, nos. 1–2, pp. 9–16, Jul. 2013.

[11] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2109–2118, Sep. 2017.

[12] N. D. Truong *et al.*, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Netw.*, vol. 105, pp. 104–111, Sep. 2018.

[13] 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.

[14] H. Chu, C. K. Chung, W. Jeong, and K.-H. Cho, "Predicting epileptic seizures from scalp EEG based on attractor state analysis," *Comput. Methods Programs Biomed.*, vol. 143, pp. 75–87, May 2017.

[15] A. Mammone, M. Turchi, and N. Cristianini, "Support vector machines," *WIREs Comput. Statist.*, vol. 1, no. 3, pp. 283–289, Dec. 2009.

[16] V. Gupta and R. B. Pachori, "Epileptic seizure identification using entropy of FBSE based EEG rhythms," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101569.

[17] R. R. Sharma, P. Varshney, R. B. Pachori, and S. K. Vishvakarma, "Automated system for epileptic EEG detection using iterative filtering," *IEEE Sensors Lett.*, vol. 2, no. 4, pp. 1–4, Dec. 2018.

[18] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.

[19] A. S. Zandi, R. Tafreshi, M. Javidan, and G. A. Dumont, "Predicting epileptic seizures in scalp eeg based on a variational Bayesian Gaussian mixture model of zero-crossing intervals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1401–1413, May 2013.

[20] Z. Zhang and K. K. Parhi, "Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 3, pp. 693–706, Jun. 2016.

[21] 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.

[22] D. Cho, B. Min, J. Kim, and B. Lee, "EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1309–1318, Aug. 2017.

[23] A. Baldominos and C. Ramn-Lozano, "Optimizing eeg energy-based seizure detection using genetic algorithms," in *Proc. Congr. Evol. Comput.*, 2017, pp. 2338–2345.

[24] S. Ramgopal *et al.*, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy Behav.*, vol. 37, pp. 291–307, Aug. 2014.

[25] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, Sep. 2009.

[26] A. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[27] B. Litt *et al.*, "Epileptic seizures may begin hours in advance of clinical onset: A report of five patients," *Neuron*, vol. 30, no. 1, pp. 51–64, Apr. 2001.

[28] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H. U. Voss, A. Schulze-Bonhage, and J. Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Phys. D, nonlinear phenomena*, vol. 194, nos. 3–4, pp. 357–368, 2004.

[29] A. Affes, A. Mdhaffar, C. Triki, M. Jmaiel, and B. Freisleben, "A convolutional gated recurrent neural network for epileptic seizure prediction," in *Proc. Int. Conf. Smart Homes Health Telematics*. Cham, Switzerland: Springer, 2019, pp. 85–96.

[30] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," in *Proc. 4th Int. Conf. Natural Comput.*, vol. 4, 2008, pp. 192–201.

[31] K. Samiee, P. Kovács, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 541–552, Feb. 2015.

[32] Z. Wu, C. Shen, and A. Van Den Hengel, "Wider or deeper: Revisiting the ResNet model for visual recognition," *Pattern Recognit.*, vol. 90, no. 6, pp. 119–133, Jun. 2019.

[33] J. Fu *et al.*, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, vol. 32, no. 6, pp. 3146–3154.

[34] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clin. Neurophysiol.*, vol. 120, no. 11, pp. 1927–1940, 2009.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[36] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[37] X. Liu, K. Li, and K. Li, "Attentive semantic and perceptual faces completion using self-attention generative adversarial networks," *Neural Process. Lett.*, vol. 51, no. 1, pp. 211–229, Feb. 2020.

[38] I. Bello, B. Zoph, Q. Le, A. Vaswani, and J. Shlens, "Attention augmented convolutional networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3286–3295.

[39] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[40] D. R. Freestone, P. J. Karoly, and M. J. Cook, "A forward-looking review of seizure prediction," *Current Opinion Neurol.*, vol. 30, no. 2, pp. 167–173, 2017.

[41] C.-T. Lin, C.-S. Huang, W.-Y. Yang, A. K. Singh, C.-H. Chuang, and Y.-K. Wang, "Real-time eeg signal enhancement using canonical correlation analysis and Gaussian mixture clustering," *J. Healthcare Eng.*, vol. 2018, Jan. 2018, Art. no. 5081258.

[42] S. Garg and R. Narvey, "Denoising & feature extraction of eeg signal using wavelet transform," *Int. J. Eng. Sci. Technol.*, vol. 5, no. 6, pp. 1249–1253, 2013.

[43] R. Hussein, S. Lee, R. Ward, and M. J. McKeown, "Epileptic seizure prediction: A semi-dilated convolutional neural network architecture," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5436–5443.

[44] J. Schroeder, "Signal processing via Fourier-Bessel series expansion," *Digit. Signal Process.*, vol. 3, no. 2, pp. 112–124, 1993.

[45] R. B. Pachori and P. Sircar, "EEG signal analysis using FB expansion and second-order linear TVAR process," *Signal Process.*, vol. 88, no. 2, pp. 415–420, 2008.

[46] A. S. Hood, R. B. Pachori, V. K. Reddy, and P. Sircar, "Parametric representation of speech employing multi-component AFM signal model," *Int. J. Speech Technol.*, vol. 18, no. 3, pp. 287–303, 2015.

[47] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[48] A. Anuragi, D. S. Sisodia, and R. B. Pachori, "Automated alcoholism detection using Fourier-Bessel series expansion based empirical wavelet transform," *IEEE Sensors J.*, vol. 20, no. 9, pp. 4914–4924, May 2020.

[49] A. Bhattacharyya, L. Singh, and R. B. Pachori, "Fourier–Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals," *Digit. Signal Process.*, vol. 78, pp. 185–196, Jul. 2018.

[50] 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.

[51] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, D. Querlioz, and O. Kavehei, "Epileptic seizure forecasting with generative adversarial networks," *IEEE Access*, vol. 7, pp. 143999–144009, 2019.

[52] H. Sadeghzadeh, H. Hosseini-Nejad, and S. Salehi, "Real-time epileptic seizure prediction based on online monitoring of pre-ictal features," *Med. Biol. Eng. Comput.*, vol. 57, no. 11, pp. 2461–2469, Nov. 2019.

[53] P. Detti, G. Z. M. de Lara, R. Bruni, M. Pranzo, F. Sarnari, and G. Vatti, "A patient-specific approach for short-term epileptic seizures prediction through the analysis of EEG synchronization," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1494–1504, Jun. 2019.

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