Seizure Prediction Using Directed Transfer Function and Convolution Neural Network on Intracranial EEG

Gang Wang*, *Member, IEEE*, Dong Wang, Changwang Du, Kuo Li, Junhao Zhang, Zhian Liu, Yi Tao, Maode Wang, Zehong Cao, *Member IEEE*, and Xiangguo Yan*

Abstract — **Automatic seizure prediction promotes the development of closed-loop treatment system on intractable epilepsy. In this study, by considering the specific information exchange between EEG channels from the perspective of whole brain activities, the convolution neural network (CNN) and the directed transfer function (DTF) were merged to present a novel method for patient-specific seizure prediction. Firstly, the intracranial electroencephalogram (iEEG) signals were segmented and the information flow features of iEEG signals were calculated by using the DTF algorithm. Then, these features were reconstructed as the channel-frequency maps according to channel pairs and the frequency of information flow. Finally, these maps were fed into the CNN model and the outputs were post-processed by the moving average approach to predict the epileptic seizures. By the evaluation of cross-validation method, the proposed algorithm achieved the averaged sensitivity of 90.8%, the averaged false prediction rate of 0.08 per hour. Compared to the random predictor and other existing algorithms tested on the Freiburg EEG dataset, our proposed method achieved better performance for seizure prediction in all patients. These results demonstrated that the proposed algorithm could provide an robust seizure prediction solution by using deep learning to capture the brain network changes of iEEG signals from epileptic patients.**

Index Terms — **Intracranial electroencephalogram (iEEG), seizure prediction, convolution neural networks, directed transfer function**

G. Wang, D. Wang, J. Zhang, Z. Liu, Y. Tao, and X. Yan are with the Key Laboratory of Biomedical Information Engineering of Ministry of Education, Institute of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China, and also with National Engineering Research Center for Healthcare Devices, Guangzhou 510500, China, and also with the Key Laboratory of Neuro-informatics and Rehabilitation Engineering of Ministry of Civil Affairs, Xi'an 710049, China. (e-mail: ggwang@xjtu.edu.cn and xgyan@xjtu.edu.cn)

C. Du, K. Li, and M. Wang are with the Department of Neurosurgery, First Affiliated Hospital, Xi'an Jiaotong University, Xi'an, 710061, China.

Z. Cao is with the School of Information and Communication Technology, University of Tasmania, Hobart, TAS, 7001, Australia.

I. INTRODUCTION

Epilepsy is a chronic nervous system disease caused by sudden abnormal discharges of neurons in the brain, which leads to dysfunction of the brain in a short period [1]. It is estimated that about 60 million people worldwide suffer from epilepsy. Approximately two-thirds of epileptic patients can be controlled or even cured through the first part of specific drug treatment [2]. For some patients with refractory epilepsy, they can be treated by surgical resection. However, the scopes of application of the surgical treatment are still also limited. Only after accurate source localization of the epileptogenic zone, the operation can be performed. Additionally, it should be guaranteed that the extent of epileptogenic foci should be concentrative; otherwise the excision may cause significant functional deficits. About 25% of patients have no suitable treatments to control their seizure symptoms. For patients who cannot be surgically removed, intervention such as medication or vagus nerve simulation, is needed for prolonged periods to suppress seizures. Specially, if an alarm is given before seizure and then human intervention is carried out, it can greatly improve the quality of life of epileptic patients and is helpful for treatment [3]. Therefore, the study of epileptic seizure prediction is becoming increasingly important.

Electroencephalogram (EEG) is an important portable device to measure the electrical activity of brain cortex and allows to explore various information related to brain functions [4]. Therefore, EEG signals are of great value in the diagnosis of brain diseases, and especially, EEG has been widely used in the diagnosis and treatment of epilepsy [5-7]. In most cases, the epileptic seizures cannot be detected in the short term, so it is necessary to record EEG signals continuously over a long period of time. The long-term EEG monitoring can effectively provide information about the electrical activity of brain and the number of seizure, which is helpful for the diagnosis and prediction of epilepsy. Due to the advances in high time resolution and effective brain function representation of EEG signals, it has been applied for the prediction of epileptic seizures [8]. There are two types of the state-of-the-art (SOTA) methods to predict epileptic seizures using EEG signals. The first type of SOTA methods is to extract features from EEG signals. If the features cross a certain or dynamic threshold, there will be an alert that epilepsy seizure is about to occur. In

terms of second type of SOTA methods, EEG is artificially divided into interictal and preictal periods and features are extracted, and then machine learning is used to identify preictal and interictal periods. It is possible to predict the arrival of epileptic seizures when preictal EEG segments are identified. In recent decades, there are many kinds of algorithms for epileptic seizure prediction. These algorithms mainly included time domain, frequency domain, time-frequency domain and nonlinear methods, as the most of classical signal processing method, the frequency spectral analysis [9-11] had achieved good results. Previous investigations also demonstrated that the EEG signals are nonlinear, non-stationary random processes, so Lyapunov exponent, correlation dimension [12, 13], and approximate entropy [14] had been introduced into the feature analysis of epileptic seizure prediction. In addition to the single-channel EEG analysis, recent investigations started to focus on the interrelationship between multi-channel EEG signals and the information exchange among various parts of brain [15-17]. However, current studies still neglect further investigations for improving seizure prediction performance that will benefit the treatment of epilepsy patients, via providing an accurate early alert for seizure attacks.

Deep learning is one category of machine learning technique which can classify the samples through multiple layers in the hierarchical architectures of neural networks. Deep learning has already proven its capability and has outperformed humans in audio and image recognition tasks. For example, Convolutional Neural Network (CNN) is a fundamental and broadly used deep neural network. There are several characteristics of CNN, such as local connections, shared weights, pooling etc. These features can reduce the complexity of the network and the number of training parameters, and they can also make the model create some degrees of invariance to shift, distortion and scale and have strong robustness and fault tolerance. So it is easy to train and optimize its network structure. Based on these predominant characteristics, it has been shown to outperform the standard fully connected neural networks in a variety of signal and information processing tasks.

At present, deep learning methods have been applied to the field of clinical medicine. In addition to medical images, they have begun to be used for a variety of physiological signals, including EEG [17, 18], ECG [19], and EMG [20]. Recently, some CNN-based methods have also significantly improved the accuracy of seizure prediction. Mirowski et al. extracted the phase synchronization feature that combines the time as a pattern to the feed to the CNNs, and achieved the sensitivity of 71% without false prediction on 15 patients using out-of-sample test method [21]. Truong et al. used a technique of epileptic seizure prediction based on CNN, which uses the extracted short-time Fourier transform (STFT) time-frequency map as the inputs of CNN to identify preictal and interictal phases [22]. Khan et al. proposed an epileptic prediction method based on wavelet transform and convolution neural network using scalp EEG [23]. Ozcan et al. achieved the seizure prediction by giving spectral band power, statistical moment, and Hjorth parameters as inputs to a multi-frame 3D CNN model [24]. Zhang et al. presented a novel solution on epilepsy seizure prediction using common spatial pattern (CSP) and CNN to improve overall accuracy while reducing the training time [25].

Frequency-domain granger causality analysis describes the connections between channels in the frequency dimension and the brain effective connectivity can be achieved by this causality analysis to localize the epileptic foci and analyze the epileptic brain networks [26, 27]. Although the graph theory index of brain networks constructed by the directed transfer function (DTF) had been used to predict the seizure onsets [28], this method would lose some detailed information of brain connectivity. Hence, by considering the specific information exchange between EEG channels or brain regions from the perspective of whole brain activities, a novel algorithm based on DTF and CNN was proposed for epileptic seizure prediction by combining the frequency-domain causality analysis of brain connectivity and deep learning.

II. METHODS

A. Patients and iEEG Data

The Freiburg EEG dataset (http://epilepsy.uni-freiburg.de/) was used to test the proposed seizure prediction algorithm in this study. The EEG dataset contains iEEG recordings of 21 patients suffering from medically intractable focal epilepsy. The data were recorded during invasive pre-surgical epilepsy monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany. For each patient, the iEEG data includes at least 50 minutes of preictal data near epileptic seizures and at least 24 hours of interictal data. For 13 patients, there are over 24 hours of continuous interictal recordings. For the remaining patients, interictal iEEG consisted of certain segments of separate recordings. Each patient has 6 channels of iEEG recordings from grid, strip, or depth electrodes. Near the seizure focus (marked as electrodes 1, 2, and 3) are three intrafocal electrodes and the other three are extrafocal electrodes distal to the focus (marked as electrodes 4, 5, and 6). According to the clinical manifestation and iEEG recordings, the epileptic seizure onset time-points were determined by certified epileptologists. The sampling rate is 256 Hz (for patient 12, the sampling rate is 512 Hz). In order to avoid an under-fitting or over-fitting problem in the seizure prediction model, the training dataset should keep in a large size. Therefore, the 19 patients with at least 3 seizures were selected from this dataset, and the last 2 cases with fewer seizures were excluded. There are 82 seizures in total, and the length of entire interictal periods was 459.1 hours.

B. Seizure Prediction Algorithm

The block diagram of the proposed method is shown in Fig. 1. Firstly, the information flow features of iEEG signals were obtained by using DTF algorithm in a sliding window. Secondly, these features were reconstructed as the channel-frequency feature maps, which were inputs to a CNN model. Finally, the outputs of CNN were post-processed to achieve epileptic seizure prediction.

1) Pre-processing of iEEG Signals

As the presence of power frequency interference in iEEG
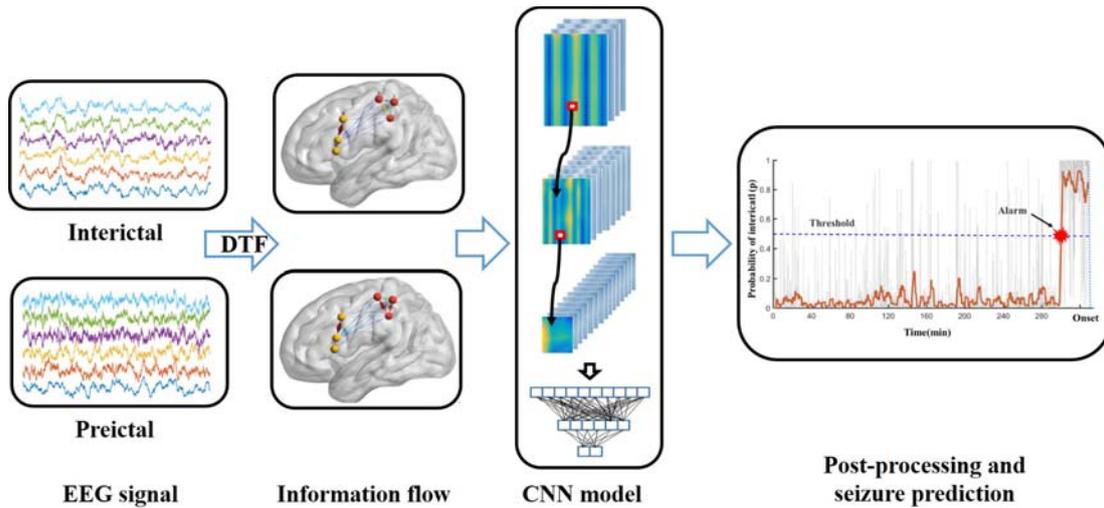
Fig. 1. The block diagram of the proposed seizure prediction method.

recordings might have an impact on the DTF algorithm, the 50 Hz notch filter was set to remove the effect of power line. In this study, we chose 30 minutes preceding a seizure onset as the preictal period.

2) Feature Extraction

The iEEG signals were segmented according to 10 s of analysis window without overlap. The first step of feature extraction was to normalize the iEEG segments by subtracting the mean value and dividing by the variance. Then, the multivariate autoregressive (MVAR) model was established and the features were extracted by the DTF algorithm using the coefficients of MVAR model. These features reflected the intensity and direction of information flow between any two channels of iEEG signals.

The N channels of iEEG signals at time t can be defined as a vector:

$$X(t) = \left[X_1(t), X_2(t), \cdots X_N(t) \right]^T \quad (1)$$

where N is channel number, T is the transpose of matrix and $X_n(t)(n=1,2,\cdots,N)$ is the nth channel of iEEG signals. For each segment of iEEG signals, a MVAR model with p order can be built as follows:

$$X(t) = \sum_{r=1}^{p} A_r X(t-r) + E(t) \quad (2)$$

where p is the order of MVAR model, A_r is a $N \times N$ matrix of coefficients and $E(t)$ is the estimation error which is an uncorrelated white noise sequenced with zero mean. The order p can be determined by Schwarz's Bayesian Criterion (SBC) [29]. The optimal order was calculated for each sample, and the maximum p was 5 in all MVAR models. The coefficient matrix A_r can be estimated by ARFIT algorithm [30]. Then, the Fourier transform of coefficient matrix A_r can be calculated by:

$$A(f) = \sum_{r=1}^{p} A_r e^{-2\pi i\, rf} \quad (3)$$

The transfer matrix is defined as:

$$H(f) = \left(I - A(f) \right)^{-1} \quad (4)$$

Then, $H(f)$ can be used to calculate the information flow between any two channels of iEEG signals:

$$\gamma^2_{ij}(f) = \frac{\left| H_{ij}(f) \right|^2}{\sum_{j=1}^{N} | H_{ij}(f) |^2} = \frac{\left| H_{ij}(f) \right|^2}{h_i^T(f) h_i(f)} \quad (5)$$

where $H_{ij}(f)$ is the element in the i th row of column j of transfer matrix $H(f)$, $h_i(f)$ is the column i of transfer matrix $H(f)$, $\gamma^2_{ij}(f)$ denotes the intensity and direction of information flow from the jth channel to the ith channel at frequency f. Because the dataset used in this study consisted of 6-channel iEEG signals, 36 channel pairs of information flow characteristics were calculated from these signals. These features extracted by DTF algorithm are reorganized to form the channel-frequency feature maps where the abscissa axis represents frequency and the vertical axis indicates channel pairs, respectively. The relationship between the order number

TABLE I
THE RELATIONSHIP BETWEEN THE ORDER NUMBER OF CHANNEL PAIRS AND THE CORRESPONDING INFORMATION FLOW.

OC: IF	OC: IF	OC: IF	OC: IF	OC: IF	OC: IF
1: 1->1	7: 2->1	13: 3->1	19: 4->1	25: 5->1	31: 6->1
2: 1->2	8: 2->2	14: 3->2	20: 4->2	26: 5->2	32: 6->2
3: 1->3	9: 2->3	15: 3->3	21: 4->3	27: 5->3	33: 6->3
4: 1->4	10: 2->4	16: 3->4	22: 4->4	28: 5->4	34: 6->4
5: 1->5	11: 2->5	17: 3->5	23: 4->5	29: 5->5	35: 6->5
6: 1->6	12: 2->6	18: 3->6	24: 4->6	30: 5->6	36: 6->6

Order number of channel pairs: OC;
Corresponding information flow: IF;
The $i \to j$ represents the information flow from the ith channel to the jth channel.
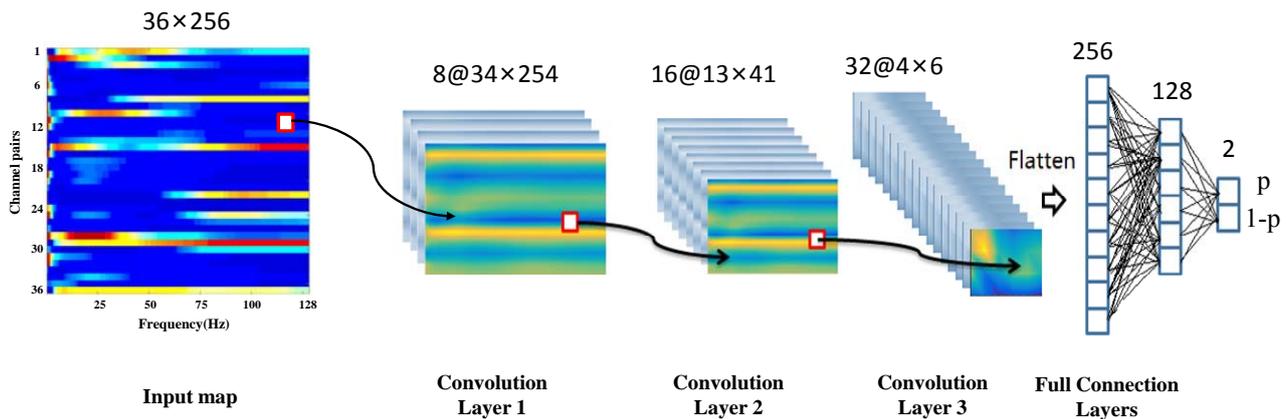
Fig. 2. The architecture of Convolutional Neural Network.

of channel pairs and the corresponding information flow is shown in Table I.

3) Convolution Neural Network

In recent years, CNN has made a great breakthrough in the field of neuro-computing. It has many advantages, such as automatic feature extraction, feature selection, and weight sharing, etc. According to AlexNet network, six-layer CNN was used for feature selection and classification in this study [31]. The block architecture of this CNN model is shown in Fig. 2. The first convolutional layer filtered the channel-frequency feature maps with 8 kernels of size 3 x 3 with a stride of 1 x 1. The second convolutional layer takes as inputs the outputs of the first convolutional layer and filtered it with 16 kernels of size 5 x 5 with a stride of 1 x 3. The third convolutional layer had 32 kernels of size 3 x 3 with a stride of 1 x 3 connected to the outputs of the second convolutional layer. The first and second convolutional layers were followed by max-pooling layers. The pooling size was 2 x 2 and the stride was 2 x 2. The ReLU nonlinearity was regarded as the activation function applied to the outputs of every convolutional layer. The remaining three layers were fully-connected and used to identify preictal and interictal recordings. The outputs of the last fully-connected layer were fed to a softmax activation function which produced the probability of preictal (p) and interictal ($1-p$) samples. In order to prevent the over-fitting problem of CNN model, the dropout layers were placed in the first two fully-connected layers with the dropout rate of 0.5. The batch size was assigned as 100 and the categorical cross entropy was selected as loss function. The optimizer used in this network was the Adaptive Moment Estimation (Adam) with the learning rate of 0.001, the beta1 of 10.9, and the beta2 of 0.999 [32]. Finally, the CNN was implemented using Keras 2.2.4 framework upon Tensorflow 1.12.0 backend in Python 3.5 environment.

4) Training and Testing Strategy of CNN

In order to reliably evaluate the performance of the proposed method, the seizure prediction results were calculated using cross-validation method. For a given patient with N seizures, the iEEG signals were divided into N mutually exclusive folds which contained approximately equal time length of interictal recordings and 30 minutes of paired preictal recordings. In

these N folds, one fold was left aside as test data to test the performance of CNN model and the remaining ($N-1$) folds were used as training data to train and validate the CNN model. To prevent the over-fitting problem, 20% of training dataset was choose for model validation and the remaining 80% dataset was used for training the CNN model. After each training epoch, the error rate of validation dataset was calculated to check if the CNN model was over-fitting on the training dataset. The training process should be stopped when the maximal iteration times were achieved or the error rate of validation dataset increased continuously in four training epochs. In this study, the training process could be terminated after about 50 times of iteration. Once the training stage of CNN model was completed, the interictal and preictal recordings in the fold left aside were tested using the trained CNN model. This process was then repeated N times until every fold was tested and the cross-validation results were obtained by averaging N individual evaluation measures (sensibility and false prediction rate).

Fig. 3 illustrated the training and testing strategy of CNN model for patient 15. This patient had 24 hours of interictal recordings and 4 epileptic seizures. So the 24 hours of interictal recordings were divided into 4 folds (each fold had 6 hours). One fold included interictal and paired preictal data was reserved for testing and the other three folds were used for training. The imbalanced ratio of interictal periods to preictal
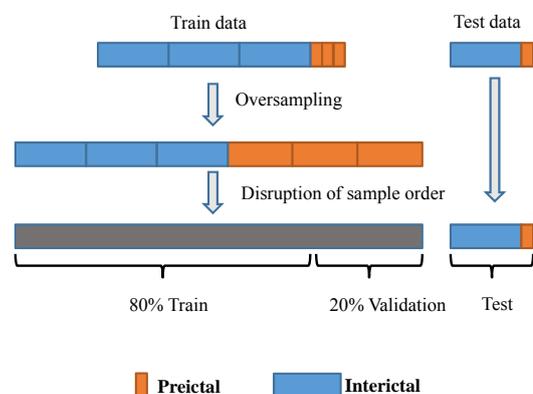
Fig. 3. The training and testing strategy of CNN for patient 15.

periods was about 12:1. The CNN learning of imbalanced data would lead to the seizure prediction algorithm with low sensitivity. To solve this problem, the random oversampling technique was used to replicate the preictal samples according to the same number as the interictal samples.

5) Post-processing

For each testing sample, the CNN model would produce two probability values. One was the probability of identifying the samples as preictal recordings (p) and the other one was the probability of recognizing the samples as interictal recordings (1-p). The probability values p should be further processed to perform seizure prediction. To remove the fluctuation of probability values, the moving average method across 20 points was used to smooth the curves of probability values. The post-processing process of probability values produced by the CNN model is shown in Fig. 4. Once the smoothed probability p exceeded the threshold values, a seizure alarm would be raised.
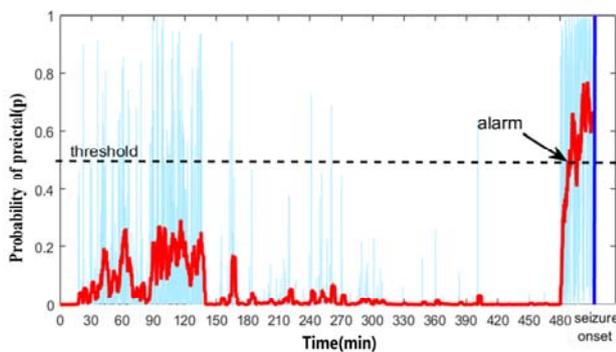
Fig. 4. Post-processing process of output probability produced by the CNN model. The pale blue line represents the probability values of identifying the samples as preictal recordings. The red line is the smoothed probability curve. The blue vertical line denotes the time-point of seizure onset. The horizontal dotted line is the alarm threshold of seizure prediction.

III. RESULTS

In this study, the prediction results of the proposed algorithm were evaluated by using cross-validation method. In order to further compare the seizures predicted by the proposed algorithm with those recognized by the certified epileptologists, the seizure prediction characteristic method was used and the relevant performance parameters are defined by the following indices [1]. The seizure prediction horizon (SPH) refers to the time period between seizure alarm and the occurrence of epileptic seizure. The seizure occurrence period (SOP) is followed by the SPH and defined as the time interval during which the seizure onset is expected to appear. In this study, the SPH and SOP were assigned as 5 minutes and 30 minutes respectively. If the seizure onset not appears during the SPH but occurs during the SOP, the seizure alarm can be recognized as a correct seizure prediction. True positive (TP) denotes the number of correct seizure predictions. False negative (FN) represents the number of missed seizure events. If no seizure occurs during SOP, the seizure alarm could be identified as a wrong seizure prediction. For a realistic seizure predictor, the

wrong prediction cannot be avoided. False prediction (FP) refers to the number of wrong seizure predictions. Then, we took into account two assessment criteria to evaluate the effectiveness of the proposed method quantitatively as follow:

1) The sensitivity (SEN) is defined as the ratio of the number of correct seizure predictions to all registered seizures.

$$SEN = \frac{TP}{TP + FN} \times 100\% \tag{6}$$

2) The false prediction rate (FPR) indicates the number of false prediction per unit time.

$$FPR = \frac{FP}{time\ period} \tag{7}$$

The proposed algorithm was examined on 19 patients with 459.1 hours of interictal recordings and 82 seizures in the Freiburg EEG dataset. When post-processing the outputs of CNN model, it is very important to choose an appropriate the threshold of seizure prediction for epilepsy patients. If the smoothed probability output exceeded this threshold, an epileptic attack alarm would be raised. In fact, different thresholds would result in different seizure prediction results, including SEN and FPR. As shown in Fig. 4, if the threshold was small, there would be high SEN and FPR. If the threshold became large, the SEN and FPR decreased at the same time. Nevertheless, a perfect seizure prediction algorithm should achieve as high SEN and low FPR as possible. In order to choose an appropriate threshold of seizure prediction, the assessment criterion SF compromising between SEN and FPR can be defined as [33]:

$$SF = \sqrt{\frac{SEN^2 + (1 - FPR)^2}{2}} \times 100\% \tag{8}$$

where SEN and FPR are the corresponding sensitivity and false prediction rate, respectively. The SEN ranged from 0 to 1 and reached the maximum value of 100% when the SEN was 1 and the FPR was 0. If the FPR was greater than 1, then it was considered equal to 1. Fig. 5 illustrates the assessment criterion SF of seizure prediction results for different thresholds. When the threshold of seizure prediction was 0.5, the SF index consisting of related SEN and FPR achieved the highest value
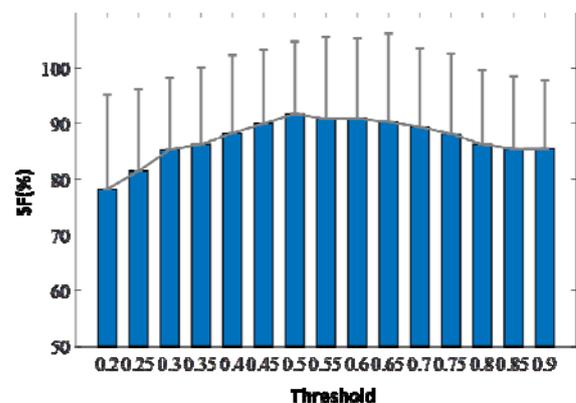
Fig. 5. Comparison of seizure prediction performance under different alarm thresholds. Histograms represent the average SF values of all patients. Bars: standard deviations.

of 91.7%. Hence, the alarm threshold of seizure prediction was set to 0.5 for all patients in this study.

TABLE II
SEIZURE PREDICTION PERFORMANCE ACHIEVED BY THE PROPOSED METHOD FOR ALL 19 PATIENTS.

Patients	No. of seizure	Interictal (hours)	SEN (%)	FPR (/h)	p-value
1	4	24	100	0	0.0000
2	3	24	100	0	0.0000
3	5	24	100	0	0.0000
4	4	23	100	0	0.0000
5	5	24	60	0.375	0.0380
6	3	24	100	0	0.0000
7	3	24.6	100	0	0.0000
9	5	23.9	100	0	0.0000
10	5	24.4	80	0	0.0000
11	4	24	75	0	0.0000
12	4	24.8	100	0	0.0000
14	4	23.8	75	0	0.0000
15	4	24	100	0.042	0.0000
16	5	24	100	0.333	0.0001
17	5	24	100	0	0.0000
18	5	24.8	100	0.040	0.0000
19	4	24.3	75	0.534	0.0426
20	5	25.6	60	0.195	0.0070
21	5	23.9	100	0	0.0000
Total	82	459.1	90.8	0.08	/

The results of epileptic seizure prediction using the proposed method on all patients are summarized in Table II. For all 19 patients, the averaged sensitivity of seizure prediction was 90.8% and the averaged false prediction rate was 0.08 per hour. The proposed method achieved the sensitivity of 100% for seizure prediction in 13 patients of 19 and the sensitivity of 100% without false prediction in 10 patients of 19. In order to further investigate the effectiveness of the proposed method, we compared this one with a random predictor based on Poisson processes. The probability to raise a seizure alarm in a period of duration equal to SOP can be calculated by [34]:

$$prob \approx 1 - e^{-FPR \bullet SOP} . \tag{9}$$

The probability values of predicting at least k events correctly from K seizures by an unspecific random predictor can be given by:

$$p(k, K, prob) = \sum_{j \geq k} \binom{K}{j} prob^j (1 - prob)^{K-j} . \tag{10}$$

Then, the probability of the random predictor producing the same SEN and FPR as our algorithm would be calculated to verify how superior the performance of this method is to chance. In this study, the significance level p was set to 0.05. As shown in Table II, the proposed method achieved significantly better seizure prediction results than a random predictor in all 19 patients ($p < 0.05$). These results demonstrated that our proposed method was able to perform the epileptic seizure prediction effectively.

IV. DISCUSSION

A. Comparisons to the State-of-the-art Algorithms

At present, several algorithms have been reported to solve the epileptic seizure prediction. The performance of our proposed approach was compared with that of the state-of-the-art algorithms using the same Freiburg EEG dataset. Table III presents the detailed information of these relevant works in recent years. Compared to other existing algorithms tested on the Freiburg EEG dataset, our proposed method achieved high SEN, low FPR and the highest SF index.

The primary goal of epileptic seizure prediction is to find some characteristic differences between interictal and preictal recordings. In this study, the DTF algorithm was used to capture these differences of causal information flow between iEEG channels. Fig. 6 shows the comparison of DTF analysis between interictal and preictal recordings. It could be observed that there were distinct changes of the channel-frequency maps between interictal and preictal iEEG signals (Fig. 6(a) and Fig. 6(b)). For interictal recordings, the information flow mainly concentrated in high frequency bands. However, for preictal

TABLE III
COMPARISON TO RECENT SEIZURE PREDICTION ALGORITHMS ON FREIBURG EEG DATASET

Authors	No. of patients	Features	SPH	SOP	Classifier	No. of seizures	SEN (%)	FPR (/h)	SF (%)
Winterhalder et al., 2003 [1]	21	Phase coherence	10 minutes	30 minutes	Threshold crossing	80	60	0.15	73.57
Maiwald et al., 2004 [13]	21	Dynamical similarity index	2 minutes	30 minutes	Threshold crossing	87	42	0.15	68.28
Zheng et al., 2014 [14]	10	Mean phase coherence	10 minutes	30 minutes	Threshold crossing	50	70	0.15	77.86
Aarabi et al., 2014 [3]	21	Bayesian inversion of power spectral density	10 seconds	30 minutes	Rule-based decision	87	87.07	0.20	80.04
Aarabi et al., 2017 [5]	10	Univariate and bivariate features	10 seconds	30 minutes	Rule-based decision	28	86.7	0.13	87.05
Yuan et al., 2017 [8]	21	Diffusion distance	10 seconds	30 minutes	Bayesian linear discriminant analysis	87	85.11	0.08	88.62
Truong et al., 2018 [22]	13	STFT spectral images	5 minutes	30 minutes	CNN	59	81.4	0.06	87.93
The proposed method	19	DTF channel-frequency maps	5 minutes	30 minutes	CNN	82	90.8	0.08	91.40
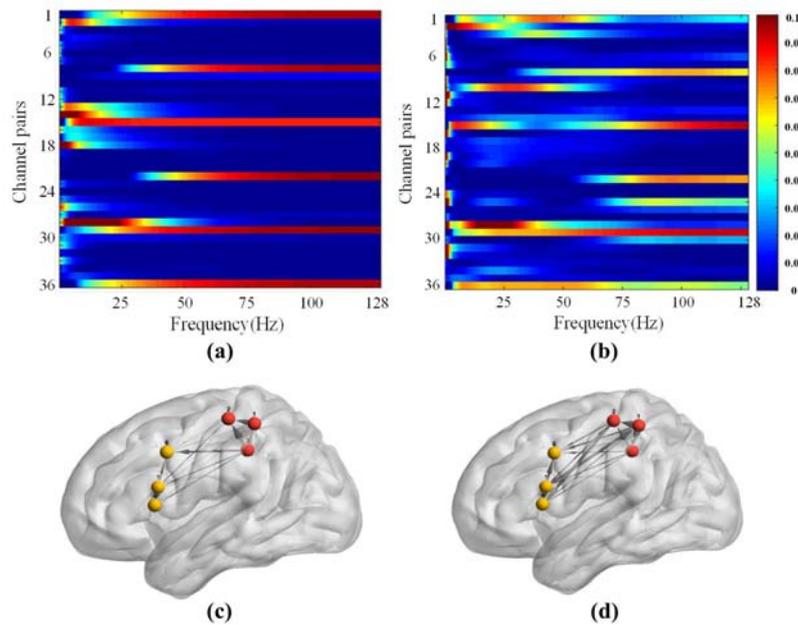
Fig. 6. Comparison of channel-frequency maps and related brain networks between interictal and preictal recordings in patient 11. The channel-frequency maps were reconstructed by the information flow characteristics using 10 s of interictal (a) and preictal (b) recordings. The brain networks of interictal (c) and preictal (d) data were connected by the information flow among six iEEG electrodes. The red nodes represent three intrafocal electrodes near the seizure focus, while the yellow nodes denote three extrafocal electrodes distal to the focus. The thickness of gray line indicates the strength of information flow of brain networks among iEEG electrodes.

recordings, the information flow distributed evenly in low and high frequency bands. In order to evaluate the total information flow between two channels of iEEG signals, all DTF values were summed up over the frequency bands of interests to obtain the integrated DTF. Then, the integrated DTF was visualized on the cerebral cortex by using BrainNet Matlab toolbox [35] to represent the brain networks of interictal and preictal recordings, as shown in Fig. 6(c) and Fig. 6(d). During the interictal phase, the brain connectivity between the seizure focus and the extrafocal region was relatively weak. However, from the interictal phase to the preictal phase, 22 of 36 connectivity between brain areas were increased prominently. These indicated that the seizure focus had largely exchanged the causal information flow with the extrafocal region during the preictal phase. In summary, the channel-frequency maps reflects the differences of information flow between interictal and preictal recordings in the aspect of frequency bands and channel locations. Since CNN captures these abnormal changes in the channel-frequency maps, the proposed method can effectively perform the epileptic seizure prediction.

B. Impacts of Different Frequency Bands

In order to further discuss the impacts of different frequency bands on the seizure prediction results, the features extracted from different bands of iEEG recordings were used as the inputs of CNN model to perform the seizure prediction. After the DTF analysis of iEEG signals, the frequency resolution of 0.1 Hz was fitted to produce the channel-frequency maps. As the inputs of CNN, the size of these maps were 36 x 40 for delta (0-4 Hz), 36 x 40 for theta (4-8 Hz), 36 x 50 for alpha (8-13 Hz), 36 x 170 for beta(13-30 Hz), 36 x 400 for gamma1(30-70 Hz) and 36 x 580 for gamma2 (70-128 Hz), respectively. Fig. 7

illustrates the numerical differences of SF criterion assessing seizure prediction results among different frequency bands of iEEG signals. The average SEN of seizure prediction using six bands were 79.2%, 85.3%, 84.6% 81.3%, 79.8%, and 76.4%, respectively. The average FPR were 0.06, 0.09, 0.14, 0.12, 0.13, and 0.09 per hour, respectively. On the one hand, there were no significant differences of SF values among six frequency bands. On the other hand, the SF results of seizure prediction using full frequency band were significantly superior when compared to those using six bands of iEEG signals ($p < 0.05$). The epileptic seizures have been demonstrated to associate with not only low
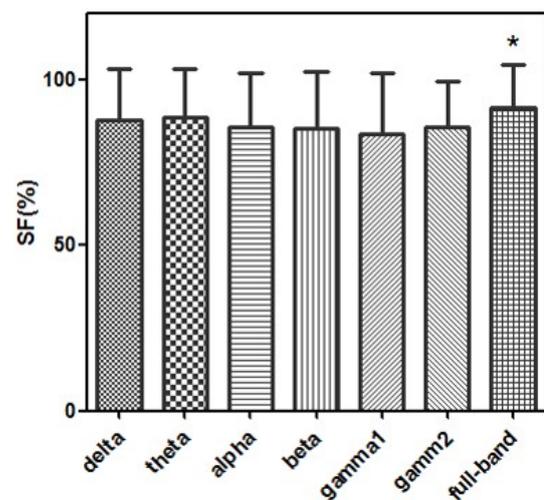
Fig. 7. Comparison of seizure prediction results for different frequency bands of iEEG signals. Histograms represent the average SF values. Bars: standard deviations.

frequency signals but also high frequency components [36-39]. Since each frequency band of iEEG signals provides different useful information for seizure prediction, the integration of these features can provide the CNN model with more information flow characteristics to discriminate interictal and preictal recordings. Hence, the proposed method can improve the seizure prediction results.

The feature selection method based on mutual information was used to explore the contribution of different channel-frequency features to epileptic seizure prediction [40]. The iEEG classes (interictal and preictal) could be regarded as a kind of random variables and another kind of random variables were the features related to specific frequency and channel pair in the channel-frequency maps. The mutual information between two kinds of random variables could be used to characterize the correlation between these variables. The higher the values of mutual information was, the greater the contribution of corresponding features to seizure prediction was. Fig. 8 shows the values of mutual information between information flow features and iEEG classes for patient 2 on channel-frequency maps. It could be observed from this figure that each frequency band would provide the useful features to distinguish between the preictal and interictal recordings. However, it was quite distinctive for the effect of different frequency and channel pairs on seizure prediction.
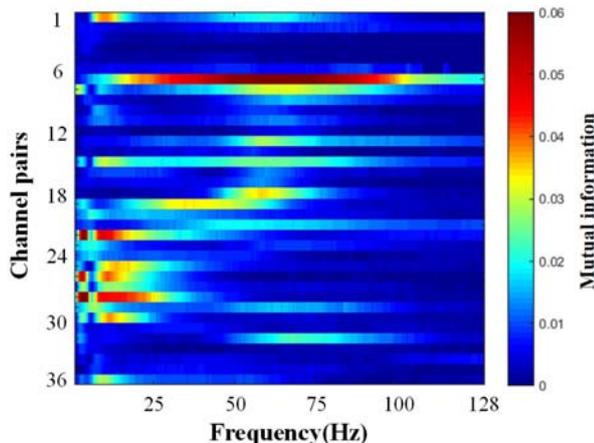
Fig. 8. The correlation analysis based on the mutual information between information flow features and iEEG classes (interictal and preictal) overlaid on the channel-frequency maps in patient 2.

To further investigate which frequency bands and channel pairs could contribute the useful features for seizure prediction, the correlation values were accumulated over six frequency bands according to each channel pair. Table IV summarizes the top two discriminative information flow features over 12 patients performed well in epileptic seizure prediction. Of these 12 patients, 13 features were the information flow from channel i to i which was related to the power spectrum of ith channel of iEEG signals. In fact, many previous studies used power spectrum features to predict epileptic seizures and were found to achieve good results [9-11]. The remaining 11 features were the information flow from channel j to i. This indicated that both spectrum-related features and information flow had a positive effect on seizure prediction. As far as the frequency bands were concerned, the feature number in the gamma frequency band (9/24) was the largest among those in all six bands. Hence, the gamma frequency bands greatly contributed to seizure prediction, which was in line with the close relationship between high frequency components of iEEG signals and epileptic seizures [41-43].

TABLE IV
TOP TWO DISCRIMINATIVE FEATURES RELATED TO SEIZURE PREDICTION IN TERMS OF FREQUENCY BANDS AND CHANNEL PAIRS.

Patients	First features	Second features
1	beta(5->5)	beta(6->6)
2	gamma1(2->1)	delta(5->4)
3	gamma2(2->2)	gamma1(2->2)
4	beta(4->4)	gamma1(4->4)
6	theta(6->5)	delta(6->5)
7	beta(4->4)	theta(4->4)
9	gamma2(4->4)	gamma2(3->4)
12	gamma2(4->5)	gamma2(5->5)
15	alpha(6->1)	theta(6->1)
17	theta(4->6)	alpha(4->6)
18	gamma2(3->3)	delta(5->5)
21	delta(1->1)	alpha(3->1)

C. Imbalance Learning for Seizure Prediction

For the iEEG signals from epileptic patients, interictal recordings are usually much longer than preictal recordings. In this study, the ratio of size of interictal class to preictal class is huge over 12:1. Such imbalanced data was unsuitable for the training of CNN model because of inadequate learning for minority class samples (preictal recordings). Since the CNN trained by imbalanced data was likely to ignore the minority class, the minority class samples could be identified as majority class samples (interictal recordings) [44]. Then, CNN learning will bring the low sensitivity of seizure prediction. In order to handle class-imbalance problems, previous studies have proposed some solutions, including sampling method, cost-sensitive learning, and so on [45]. The main idea of sampling method is to balance the proportion of majority class to minority class by altering the size of training dataset. The common approaches of changing the training dataset have random oversampling, random undersampling, and synthetic minority oversampling technique (SMOTE) [46]. Instead of creating class samples in the sampling strategies, the cost-sensitive learning attempts to balance distributions by considering the costs associated with misclassifying samples. For a binary classification of interictal and preictal recordings, there is usually no cost for correct classification. However, it is very important that the costs of identifying minority class samples as majority class samples should be larger than the costs of the contrary cases. In the process of CNN learning on the training dataset, misclassifying minority samples will result in more penalty costs than misclassifying majority samples. By minimizing the total costs on the training dataset, cost-sensitive learning can improve the recognition capability of minority

class samples.

To investigate the effects of different class-imbalance solutions on seizure prediction, we compared random oversampling, SMOTE, and cost-sensitive CNN learning. For random oversampling, the preictal samples were randomly duplicated according to the same number as the interictal samples. The SMOTE algorithm is a powerful tool and has widely been applied to a crowd of applications. This algorithm was used to produce the artificial samples based on the feature space similarities between existing preictal recordings. For cost-sensitive CNN learning, the costs of misclassifying interictal and preictal samples were designated according to the imbalanced ratio of interictal class to preictal class. The results of epileptic seizure prediction using different class-imbalance solutions are illustrated in Fig. 9. For random oversampling, the average SF was 91.7% with the SEN of 90.8% and the FPR of 0.08/h. The SMOTE achieved the average SF of 90.1%, the average SEN of 89.7% and the average FPR of 0.10/h. Additionally, the average SF, SEN, and FPR from cost-sensitive CNN learning were 86.9%, 77.6%, and 0.068/h, respectively. On one hand, there were no significant differences in SF values between the random oversampling and the SMOTE in terms of seizure prediction results. On the other hand, the random oversampling resulted in significantly superior SF to the cost-sensitive CNN learning ($p < 0.05$). As is well known, it is very important to choose the appropriate costs of misclassifying interictal and preictal samples for cost-sensitive learning. In this study, the cost of misclassifying interictal class was set to 1 and the cost of misclassifying preictal class was assigned as the ration of size of interictal class to preictal class. However, these misclassification costs may not be the optimal parameters for handling the imbalance learning of interictal and preictal recordings. If the optimal parameters can be used to perform seizure prediction, the performance of cost-sensitive CNN learning will be further enhanced. The grid search algorithm had been presented to find out the optimal misclassification costs when the cost-sensitive

Fig. 9. The results of epileptic seizure prediction using random oversampling, SMOTE, and cost-sensitive CNN learning. Histograms represent the average SF values related to different class-imbalance solutions. Bars: standard deviations.

SVM was applied to seizure prediction [9]. When the sample number of training dataset and the dimension of extracted feature vector is large, the computation complexity of parameter optimization will become high. If we perform the parameter optimization in deep learning algorithm, the computation complexity will be further increased dramatically. Hence, if deep learning was used to dealing with the class-imbalance problems in seizure prediction, the sampling method may be more convenient and feasible than the cost-sensitive learning.

D. Limitation and Future Works

The proposed method extracted the channel-frequency maps to perform the seizure prediction by using the DTF algorithm. The CNN model not only took advantage of frequency but also channel position information in the channel-frequency maps. However, the position in the channel-frequency maps was not the realistic location of iEEG channel. If the deep learning algorithm utilized the actual channel location, the seizure prediction results could be further improved. Recently, the graph convolution network (GCN) is proposed to especially analyze the network graph features and has been widely applied to recommendation systems, electronic transactions, computational geometry, brain signals, molecular structures, and so on [47]. Because the GCN is also suitable for addressing brain network graph [48], the GCN model may adequately explore the actual location information of iEEG channel to achieve better seizure prediction results than the CNN model. In this study, individual CNN classifier was trained according to each patient. In fact, if the seizure onset of a specific patient could be predicted accurately by the CNN classifier trained on other remaining patients, this method would be very useful for a clinical application system. In the future, the proposed method should be improved to effectively achieve the cross-patient seizure prediction.

V. CONCLUSION

In this study, a novel approach based on CNN and DTF algorithm was proposed to predict the epileptic seizures in long-term iEEG recordings. Our proposed method illustrated the possibility of using CNN to extract and recognize the features of DTF network and was performed on 19 epilepsy patients with a total 459.1 hours of iEEG signals. After the cross-validation measurement, the averaged SEN was 90.8% and the averaged FPR was 0.08 per hour. Of all 19 patients, 13 had 100% sensitivity for seizure prediction and 10 had 100% sensitivity without false prediction. Compared with the state-of-the-art algorithms using the same Freiburg EEG dataset, the proposed method achieved high SEN, low FPR and the highest SF index. Because the discriminative feature number in the gamma band was the largest among those in all frequency bands, the iEEG signals in the gamma band had a great impact on the seizure prediction results. In addition, since there was much imbalanced ratio of size of interictal class to preictal class, the class-imbalance solutions could greatly improve the seizure prediction results. These results indicated that the proposed method was suitable for performing the
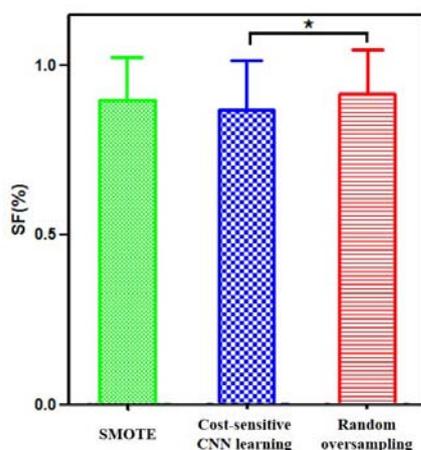
epileptic seizure prediction. This accurate and reliable seizure prediction method is likely to be applied to clinical practice and benefit the closed-loop treatment of epilepsy patients.

REFERENCES

[1] M. Winterhalder, T. Maiwald, H. U. Voss, R. Aschenbrenner-Scheibe, J. Timmer, and A. Schulze-Bonhage, "The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods," *Epilepsy Behav,* vol. 4, no. 3, pp. 318-25, Jun, 2003.

[2] H. Witte, L. D. Iasemidis, and B. Litt, "Special issue on epileptic seizure prediction," *IEEE Transactions on Biomedical Engineering,* vol. 50, no. 5, pp. 537-539, May, 2003.

[3] A. Aarabi, and B. He, "Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach," *Clinical Neurophysiology,* vol. 125, no. 5, pp. 930-940, May, 2014.

[4] G. Wang, and D. T. Ren, "Effect of Brain-to-Skull Conductivity Ratio on EEG Source Localization Accuracy," *Biomed Res Int,* 2013.

[5] A. Aarabi, and B. He, "Seizure prediction in patients with focal hippocampal epilepsy," *Clinical Neurophysiology,* vol. 128, no. 7, pp. 1299-1307, Jul, 2017.

[6] G. Wang, D. T. Ren, K. Li, D. Wang, M. D. Wang, and X. G. Yan, "EEG-Based Detection of Epileptic Seizures Through the Use of a Directed Transfer Function Method," *Ieee Access,* vol. 6, pp. 47189-47198, 2018.

[7] G. Wang, Z. J. Sun, R. Tao, K. Li, G. Bao, and X. G. Yan, "Epileptic Seizure Detection Based on Partial Directed Coherence Analysis," *IEEE J Biomed Health Inform,* vol. 20, no. 3, pp. 873-879, May, 2016.

[8] S. S. Yuan, W. D. Zhou, and L. Y. Chen, "Epileptic Seizure Prediction Using Diffusion Distance and Bayesian Linear Discriminate Analysis on Intracranial EEG," *International Journal of Neural Systems,* vol. 28, no. 1, Feb, 2018.

[9] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia,* vol. 52, no. 10, pp. 1761-1770, Oct, 2011.

[10] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clinical Neurophysiology,* vol. 126, no. 2, pp. 237-248, Feb, 2015.

[11] Z. S. Zhang, and K. K. Parhi, "Low-Complexity Seizure Prediction From iEEG/sEEG Using Spectral Power and Ratios of Spectral Power," *Ieee Transactions on Biomedical Circuits and Systems,* vol. 10, no. 3, pp. 693-706, Jun, 2016.

[12] A. Aarabi, and B. He, "A rule-based seizure prediction method for focal neocortical epilepsy," *Clinical Neurophysiology,* vol. 123, no. 6, pp. 1111-1122, Jun, 2012.

[13] 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," *Physica D-Nonlinear Phenomena,* vol. 194, no. 3-4, pp. 357-368, Jul 15, 2004.

[14] Z. Zhang, Z. Y. Chen, Y. Zhou, S. H. Du, Y. Zhang, T. Mei, and X. H. Tian, "Construction of rules for seizure prediction based on approximate entropy," *Clinical Neurophysiology,* vol. 125, no. 10, pp. 1959-1966, Oct, 2014.

[15] Y. Zheng, G. Wang, K. Li, G. Bao, and J. Wang, "Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition," *Clinical Neurophysiology,* vol. 125, no. 6, pp. 1104-1111, Jun, 2014.

[16] P. van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, and D. Marinazzo, "Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization," *Progress in Neurobiology,* vol. 121, pp. 19-35, Oct, 2014.

[17] A. Supratak, H. Dong, C. Wu, and Y. K. Guo, "DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 25, no. 11, pp. 1998-2008, Nov, 2017.

[18] S. Sakhavi, C. T. Guan, and S. C. Yan, "Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks," *Ieee Transactions on Neural Networks and Learning Systems,* vol. 29, no. 11, pp. 5619-5629, Nov, 2018.

[19] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients," *Ieee Transactions on Systems Man Cybernetics-Systems,* vol. 48, no. 12, pp. 2095-2104, Dec, 2018.

[20] X. L. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin, "Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control Based on Convolutional Neural Network," *Front Neurosci,* vol. 11, Jul 11, 2017.

[21] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clinical Neurophysiology,* vol. 120, no. 11, pp. 1927-1940, Nov, 2009.

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

[23] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal Onset Seizure Prediction Using Convolutional Networks," *IEEE Transactions on Biomedical Engineering,* vol. 65, no. 9, pp. 2109-2118, Sep, 2018.

[24] A. R. Ozcan, and S. Erturk, "Seizure Prediction in Scalp EEG Using 3D Convolutional Neural Networks With an Image-Based Approach," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 27, no. 11, pp. 2284-2293, Nov, 2019.

[25] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network," *IEEE Journal of Biomedical and Health Informatics,* vol. 24, no. 2, pp. 465-474, Feb, 2020.

[26] M. Kamiński, and K. Blinowska, "A new method of the description of the information flow in the brain structures.," *Biol Cybern,* vol. 65, no. 3, pp. 203-10, 1991.

[27] L. A. Baccala, and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics,* vol. 84, no. 6, pp. 463-474, Jun, 2001.

[28] M. Hejazi, and A. Motie Nasrabadi, "Prediction of epilepsy seizure from multi-channel electroencephalogram by effective connectivity analysis using Granger causality and directed transfer function methods," *Cogn Neurodyn,* vol. 13, no. 5, pp. 461-473, Oct, 2019.

[29] G. Schwarz, "Estimating Dimension of a Model," *Annals of Statistics,* vol. 6, no. 2, pp. 461-464, 1978.

[30] A. Neumaier, and T. Schneider, "Estimation of Parameters and Eigenmodes of Multivariate Autoregressive Models," *ACM Transactions on Mathematical Software,* vol. 27, no. 1, pp. 27-57, Mar, 2001.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the Acm,* vol. 60, no. 6, pp. 84-90, Jun, 2017.

[32] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," in the 3rd International Conference for Learning Representations, San Deigo, USA, 2015, pp. 1-15.

[33] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger, "Epileptic seizures are preceded by a decrease in synchronization," *Epilepsy Res,* vol. 53, no. 3, pp. 173-185, Mar, 2003.

[34] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos,* vol. 16, no. 1, Mar, 2006.

[35] M. R. Xia, J. H. Wang, and Y. He, "BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics," *PLoS One,* vol. 8, no. 7, Jul 4, 2013.

[36] R. Hopfengartner, F. Kerling, V. Bauer, and H. Stefan, "An efficient, robust and fast method for the offline detection of epileptic seizures in long-term scalp EEG recordings," *Clinical Neurophysiology,* vol. 118, no. 11, pp. 2332-2343, Nov, 2007.

[37] B. Vanrumste, R. D. Jones, P. J. Bones, and G. J. Carroll, "Slow-wave activity arising from the same area as epileptiform activity in the EEG of paediatric patients with focal epilepsy," *Clinical Neurophysiology,* vol. 116, no. 1, pp. 9-17, Jan, 2005.

[38] M. Ayala, M. Cabrerizo, P. Jayakar, and M. Adjouadi, "Subdural EEG Classification Into Seizure and Nonseizure Files Using Neural Networks in the Gamma Frequency Band," *Journal of Clinical Neurophysiology,* vol. 28, no. 1, pp. 20-29, Feb, 2011.

[39] Y. F. Lu, G. A. Worrell, H. C. Zhang, L. Yang, B. Brinkmann, C. Nelson, and B. He, "Noninvasive Imaging of the High Frequency Brain Activity in Focal Epilepsy Patients," *Ieee Transactions on Biomedical Engineering,* vol. 61, no. 6, pp. 1660-1667, Jun, 2014.

[40] H. C. Peng, F. H. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 8, pp. 1226-1238, Aug, 2005.

[41] J. Jacobs, R. Zelmann, J. Jirsch, R. Chander, C. E. Chatillon, F. Dubeau, and J. Gotman, "High frequency oscillations (80-500 Hz) in the preictal period in patients with focal seizures," *Epilepsia,* vol. 50, no. 7, pp. 1780-1792, Jul, 2009.

[42] D. Wang, D. Ren, K. Li, Y. Feng, D. Ma, X. Yan, and G. Wang, "Epileptic Seizure Detection in Long-Term EEG Recordings by Using Wavelet-Based Directed Transfer Function," *IEEE Trans Biomed Eng,* vol. 65, no. 11, pp. 2591-2599, Nov, 2018.

[43] J. Jacobs, P. LeVan, R. Chander, J. Hall, F. Dubeau, and J. Gotman, "Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain," *Epilepsia,* vol. 49, no. 11, pp. 1893-1907, Nov, 2008.

[44] X. Y. Liu, J. X. Wu, and Z. H. Zhou, "Exploratory Undersampling for Class-Imbalance Learning," *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics,* vol. 39, no. 2, pp. 539-550, Apr, 2009.

[45] H. B. He, and E. A. Garcia, "Learning from Imbalanced Data," *Ieee Transactions on Knowledge and Data Engineering,* vol. 21, no. 9, pp. 1263-1284, Sep, 2009.

[46] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research,* vol. 16, pp. 321-357, 2002.

[47] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," in the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016, pp. 1-9.

[48] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *Neuroimage,* vol. 169, pp. 431-442, Apr 1, 2018.