

A Robust Deep Learning Approach for Automatic Seizure Detection

Xinghua Yao, Xiaojin Li, Qiang Ye, Yan Huang, Qiang Cheng*, and Guoqiang Zhang*

Abstract—Detecting epileptic seizure through analysis of the electroencephalography (EEG) signal becomes a standard method for the diagnosis of epilepsy. In a manual way, monitoring of long term EEG is tedious and error prone. Therefore, a reliable automatic seizure detection method is desirable. A critical challenge to automatic seizure detection is that seizure morphologies exhibit considerable variabilities. In order to capture essential seizure patterns, this paper leverages an attention mechanism and a bidirectional long short-term memory (BiLSTM) model to exploit both spatially and temporally discriminating features and account for seizure variabilities. The attention mechanism is to capture spatial features more effectively according to the contributions of brain areas to seizures. The BiLSTM model is to extract more discriminating temporal features in the forward and the backward directions. By accounting for both spatial and temporal variations of seizures, the proposed method is more robust across subjects. The testing results over the noisy real data of CHB-MIT show that the proposed method outperforms the current state-of-the-art methods. In both mixing-patients and cross-patient experiments, the average sensitivity and specificity are both higher while their corresponding standard deviations are lower than the methods in comparison.

Index Terms—bidirectional LSTM, attention, seizure detection, time split, deep learning.

I. INTRODUCTION

EPILEPSY is a central nervous system disorder, in which brain activity becomes abnormal, causing seizures or periods of unusual behaviors, sensations, and sometimes loss of awareness. More than 50 million people in the world suffer from epilepsy [1]. An important technique to diagnose epilepsy is an electroencephalography (EEG). An EEG records the electrical activities of the brain, and may reveal patterns of normal or abnormal brain electrical activities. In current clinical practices, EEG readings are mostly analyzed by trained neurologists to identify characteristic patterns of the disease, such as seizures and pre-ictal spikes. This manual way of analyzing is laborious and error prone, for it generally takes several hours for a trained professional to analyze one-day of recordings from one patient [2]–[6]. These limitations have motivated researchers to develop automated approaches to detect seizures.

The study of automatic seizure detection has been extensively explored. A critical challenge is that seizure mor-

phologies exhibit considerable inter-patient and intra-patient variabilities. Different machine learning methods and computational technologies have been applied to address this challenge. There are many studies for constructing patient-specific detectors capable of detecting seizure onsets [6]–[15]. In early studies, hand-crafted features are usually used as characteristics of seizure manifestations in EEG. More recent studies focus on designing deep learning models for seizure detection [4] [13] [16] [17]. There are some components shared by most of these studies. For example, signal processing techniques are used to filter the data; certain modules need to be pre-trained; multiple channels are utilized to extract spatial features, and temporal features are extracted by the sliding windows. However, to the best of our knowledge, the data over channels are processed in the same way; i.e. the channels are not differentiated. About extracting temporal features, most studies only work in the forward direction. In fact, for seizure detection, the EEG signals can potentially provide some additional information in the backward direction.

Different brain areas are likely to have different contributions to the seizure. The characteristics of EEG data for epilepsy at different brain areas are different. The features of EEG signals at a time point are correlated with the past data and with the future data. Besides, though EEG signals are in general dynamic and non-linear, during a sufficiently small time period, the signal may be considered to be stationary. Based on the above three observations and inspired by the architecture in [18], we design a new approach by using bidirectional long short-term memory (BiLSTM) model integrated with an attention mechanism. Firstly, we introduce an attention mechanism over EEG channels. Different weights are automatically assigned to the signal channels at different brain areas according to how much they would affect the seizures. Secondly, the bidirectional long short-term memory technique is adopted to extract temporal features of EEG signals in both the forward and the backward directions. Thirdly, the output sequence of the BiLSTM module in the whole architecture is split into patches according to the time steps. Each patch only contains the data in one time step. All the patches are separately processed to extract features. With these three new ideas, we develop a novel approach to seizure detection in EEG signals. By using the proposed approach, mixing-patients and cross-patient experiments are conducted. In the mixing-patients experiments, we obtain the average sensitivity and specificity of 86.6% and 86.0%, respectively, and the corresponding standard deviations of 0.0258 and 0.0349 respectively. For the cross-patient experiments, the average sensitivity and specificity of 83.72% and 84.06%

* Corresponding authors.

Xinghua Yao, Xiaojin Li, Yan Huang, Qiang Cheng, and Guoqiang Zhang are with the Institute of Biomedical Informatics, University of Kentucky, Lexington, Kentucky 40536-0082, USA (e-mail: xhyaosues@aliyun.com; lxj9173@gmail.com; yanhuang.uky@uky.edu; Qiang.Cheng@uky.edu; gqat-case@gmail.com).

Qiang Ye is with the Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506-0027, USA (e-mail: qiang.ye@uky.edu).

are respectively achieved, and the standard deviations being 0.1349 and 0.1379, respectively. These results exceed the current state-of-the-art performance on the noisy real data of CHB-MIT. The extensive experimental results show that the performance of the proposed new approach is promising and has strong stability, in that its performance has much smaller variations compared to existing methods.

In brief, the main novelties of our paper include the following: (1) An attention mechanism is utilized to account for spatial variations of seizures for the first time; (2) The bidirectional long short-term memory model is adopted to extract temporal features of seizures; (3) Extensive experimental results on the noisy real EEG data of CHB-MIT demonstrate that, using the new approach, more robust seizure patterns can be captured, and the inter-patient seizure variations can be overcome better than current state-of-the-art deep learning approaches.

The rest of this paper is organized as follows. Section II describes some related research work on automatic seizure detection. Section III presents our designed model BiLSTM with attention. In Section IV, mixing-patients and cross-patient experimental results are provided to evaluate the proposed model. Section V discusses the model BiLSTM with attention. Finally, Section VI concludes this paper with descriptions of some future work.

II. RELATED WORK

Seizure detection is to detect whether a data record contains seizure or not. For this task, there are extensive studies, and many researchers have used machine learning methods to do it [4] [7] [8] [10]–[14] [16] [17] [19]–[23]. Seizure detection is often viewed as a classification problem, in which data records need be classified to be seizure versus non-seizure records. According to sources of subjects data records used in training and testing sets, different automatic approaches are to be grouped into three types: patient-specific approach, mixing-patients approach, and cross-patient approach.

A. Patient-Specific Approach

The patient-specific approach detects data records of one subject based on training data of the same subject. That is, all the data are from one subject only. So, in this approach there is no variations caused by different subjects between the testing data and the training data. On the one hand, the testing data and training data have more similarities than the other two types of approaches. On the other side, data records of one subject are limited. Usually, the quantity of data records is small so that classical signal processing and machine learning methods have been successfully used while deep learning models can hardly help us achieve satisfactory detecting results.

Shoeb and Guttag propose a method to construct a patient-specific detector for seizure detection in [7]. The method leverages filters to extract spectral features over each channel, and then concatenates the feature vectors according to a fixed time length. The obtained feature vectors are input to the support vector machine (SVM) to train. The method is validated through patient-specific experiments. And a sensitivity

of 96% is achieved. The sensitivity result is often used as a benchmark for patient-specific seizure detection on the data set CHB-MIT. The authors observe that the identity of channels could help differentiate between the seizure and the non-seizure activity. However, the proposed method does not adopt different processing ways for the data on different channels.

Amin and Kamboh in [8] design an algorithm RUSBoost to process imbalanced seizure/non-seizure data, and use RUSBoost and the decision tree learner to conduct patient-specific experiments over the data set CHB-MIT. The method is fast in training and has good performance in the patient-specific experiments. Although data on multiple channels are analyzed in the method, they are not also distinguished.

Fan and Chou in [9] utilize a complex network model to represent EEG signals, and integrate it with spectral graph theory to extract spatial-temporal synchronization patterns for detecting seizure onsets in real-time. The method is tested on 23 patients from CHB-MIT Scalp EEG database. The resulting patient-specific sensitivity surpasses the benchmark.

In [6], Zandi et al. propose a wavelet-based algorithm for real-time detection of epileptic seizures using scalp EEG. In this algorithm, the EEG from each channel is decomposed by wavelet packet transform, and a patient-specific measure is developed by using wavelet coefficients to separate the seizure and non-seizure states. Utilizing the measure, a combined seizure index is derived for each epoch of every EEG channel. Through inspecting the combined seizure index, proper channel alarms are generated. The method is not completely automated.

Hunyadi et al. in [10] present a patient-specific seizure detection algorithm, which uses a nuclear norm regularization to convey spatial distribution information of ictal patterns. The algorithm extracts features from each channel, and then stacks them to analyze as one entity.

Esbroeck et al. in [11] utilize a multi-task learning framework to detect patient-specific seizure onset in the presence of intra-patient variability in seizure morphology. They consider distinguishing the windows of each seizure from non-seizure data as a separate task, and treat the individual-seizure discrimination as another task. Some testing results over the data set CHB-MIT show that the performance of the method over most cases has improvements compared with using the standard SVM.

Truong et al. [12] present a automatic seizure detection method over intracranial electroencephalography (iEEG) data. First, supervised classifiers are used to select those channels which contribute the most to a seizure. Features in both frequency and time domains, including spectral power and correlations between channel pairs, are extracted. Then, Random Forest is adopted for classification. This method has the state-of-the-art computational efficiency while maintaining the accuracy. In the method, selecting channels with the most contributions to a seizure is to reduce the number of channels so that the computational efficiency could be improved. And in [14], Truong et al. focus on the applicability of seizure detection method to hardware implementation, and proposes an integer convolutional neural network.

Vidyaratne et al. [13] propose a deep recurrent architecture

by combining Cellular Neural Network and Bidirectional Recurrent Neural Network. The bidirectional recurrent neural network is deployed into each cell in the cellular neural network, and it is utilized to extract temporal features in the forward and the backward directions. Each cell interacts with its neighbor cells to extract local spatial-temporal features. The computed results in the cellular neural network are output into a multi-layered perceptron. In the perceptron, samples are classified based on a trained threshold. In order to satisfy the input requirements of cellular neural network, the authors propose a mapping which organizes EEG signals into a 2D grid arrangement. Patient-specific experiments are conducted over the EEG data of five patients from the data set CHB-MIT. The obtained sensitivities are all 100% for the five patients. However, the raw EEG data are preprocessed using a bandpass filter between 3Hz and 30Hz in order to extract seizure activity data.

B. Mixing-Patients Approach

Mixing-patients seizure detection has no subject requirements for testing data and training data. The training data and the testing data are partitioned from a pool of the segmented signals from all patients. In this approach, more samples could be obtained to support utilizing deep learning models.

In [19], Fergus et al. present a method for seizure detection based on traditional machine learning techniques, and obtain 88% in Sensitivity and 88% in Specificity over the data set CHB-MIT. The method mainly consists of four steps, which are data filtering, feature extraction, feature selection and training classifiers. In the mixing-patients experiments, EEG signals in CHB-MIT are sliced for one seizure segment per seizure with 60 seconds per segment, and non-seizure segments that are as many as seizure segments and are randomly selected. Finally, the experiment data consist of 171 seizure segments and 171 non-seizure segments. In the slicing way, for one seizure with a duration less than 60 seconds, all the corresponding seizure data are included in one segment; for one seizure with a duration long than 60 seconds, only the first 60 seconds of seizure data are used in one segment. The average segment contains 40s ictal data. Note that this kind of splitting is different from ours. Our splicing way does not require that all the ictal data in one seizure are included in one segment. Compared with the slicing way in [19], most of our segments contain less ictal data. Additionally, after slicing EEG signals [19] uses a bandpass filter and second order butterworth filters to extract the EEG data in the bandwidth 0.5-30Hz.

Golmohammadi et al. [16] explore seizure-detection performances of two neural networks over the data source of TUH EEG Corpus. Their experiment results show that the convolutional long short-term memory (LSTM) network is better than the convolutional GRU network. And also the impacts of initialization methods and regularization methods over the performance are experimented. The two models do not utilize attention mechanism.

Hussein et al. in [18] designs a deep neural network for seizure detection by using LSTM as a main module, which is

called LSTM approach. The LSTM approach extracts temporal features by using LSTM. Some experiments are conducted on the EEG data set provided by University of Bonn. The testing results mostly reach 100%. In [17], Acharya et al. present a 13-layers deep neural network for seizure detection by using convolutional neural network (CNN), which is called CNN approach. Over the Bonn EEG data set, the obtained average sensitivity and specificity are 95% and 90% respectively. Because each record in the Bonn EEG data set is the data from only one channel. For the experiments in [18] and [17], the LSTM approach and the CNN approach extract seizure features from the data on one channel to conduct detection.

C. Cross-Patient Approach

The cross-patient approach requires that the testing data and the training data could not be from the same subject. It uses data from other subjects for training a model that is further used to detect seizures in the testing data of a new subject. The seizure detection model in this approach should overcome the subject differences, and should extract robust features shared by different subjects. The detection task is the most difficult among the three types of approaches. The cross-patient approach generally obtains more samples than the patient-specific approach.

In [4], Thodoroff et al. design a recurrent convolutional neural network to capture spectral, spatial and temporal patterns of seizures. The EEG signals are firstly transformed into images by using such techniques, including Polar Projection, cubic interpolation and Fast Fourier transform. The image-based representation of EEG signals is to exploit the spatial locality in seizures. Created images are fed into the convolution neural network. The output vectors of the convolution neural network are organized to be sequences in chronological order. Then, the sequences are input into the bidirectional recurrent neural network to make classification. Both patient-specific experiments and cross-patient experiments are conducted. The patient-specific experiment results are similar to the results in [7]. And the cross-patient testing sensitivity is 85% on average. In the two kinds of experiments, the convolution neural network is pre-trained alone. And the transfer learning technology is utilized to overcome the problem of small amount of data in the patient-specific experiments. As we see, the recurrent convolutional neural network in [4] need work with some other techniques to efficiently detect seizure.

III. METHODS

A. Model Design

Electroencephalogram signal is an important modality for the diagnosis of epilepsy. EEG signal data is generally collected through placing electrodes on the scalp. Each electrode records brain activities in its located brain area. As different brain areas play different roles in the seizure procedure, the data collected at different brain areas record different characteristics of seizures. [7] observes, over some channels the collected data in seizure are quite different from non-seizure activity. To exploit the signal difference between different

brain areas, we will adopt an attention mechanism to give different weights to the data over different channels.

Brain activities are continuous. EEG signals could be viewed to be continuous records of brain activities when ignoring the sampling effects. The brain activity at a time point could be correlated with some past signal data, and could also be analyzed from some future signal data. This kind of analysis in two directions could help extract more discriminating features of seizures. To exploit correlations from both directions, we employ the BiLSTM model for analyzing the EEG sequence data.

EEG signal is dynamic and non-linear. Because of the dynamic nature, some statistic characteristics of EEG signal change with the time. In a sufficiently small time duration, the EEG signal segments have similar statistical temporal and spectral features [18] [24]. After bidirectionally processing, the sequence is split into time-step patches. Each patch only contains data in a time step. The patches are further extracted features through full connection operations separately and concurrently.

The raw EEG signal is split into data segments according to a fixed time span. Our task is to detect whether a data segment contains seizure or not. The split data segments are automatically weighted through an attention mechanism. That is to say, for each segment, data over different channels are multiplied with different weights. The weights are achieved through a fully connected module and a non-linear function in training procedure. After adding weights, the data segments are passed into bidirectional LSTM module. The BiLSTM module extracts features in the forward and the backward directions. Next, another time-step feature extraction operation is executed. For the output sequence of BiLSTM, the data at each time step are separately input into a full connection module to be processed. Then, the extracted features are averaged over all the time steps in order to achieve global features of a segment. Finally, the labels of data segments are computed through a fully connected module with the Softmax function.

B. Model Architecture and Algorithm

Our model architecture consists of five modules, including attention layer, BiLSTM module, time-distributed fully-connected layer, pooling layer and fully-connected layer with Softmax. The designed architecture is presented in Fig. 1.

1) *Attention Layer*: The attention layer, described in Fig. 2, is to generate attention weights for each channel and then executes an element-wise multiplication. The original data are input into a fully connected module with a nonlinear activation function. The outputs of the fully connected module are averaged over all the time steps. Then, the obtained averages are copied to be shared over all the time steps. So, an attention weight matrix is achieved. Finally, the attention matrix is element-wisely multiplied with the original inputs. The attention layer is computed using the following equations:

$$X = f_{\text{reshape}_1}(X_{\text{input}}) \quad (1)$$

$$Y_{FC-\text{atten}} = \sigma(X * W_{FC-\text{atten}} + B_{FC-\text{atten}}) \quad (2)$$

$$Y_1 = f_{\text{reshape}_2}(Y_{FC-\text{atten}}) \quad (3)$$

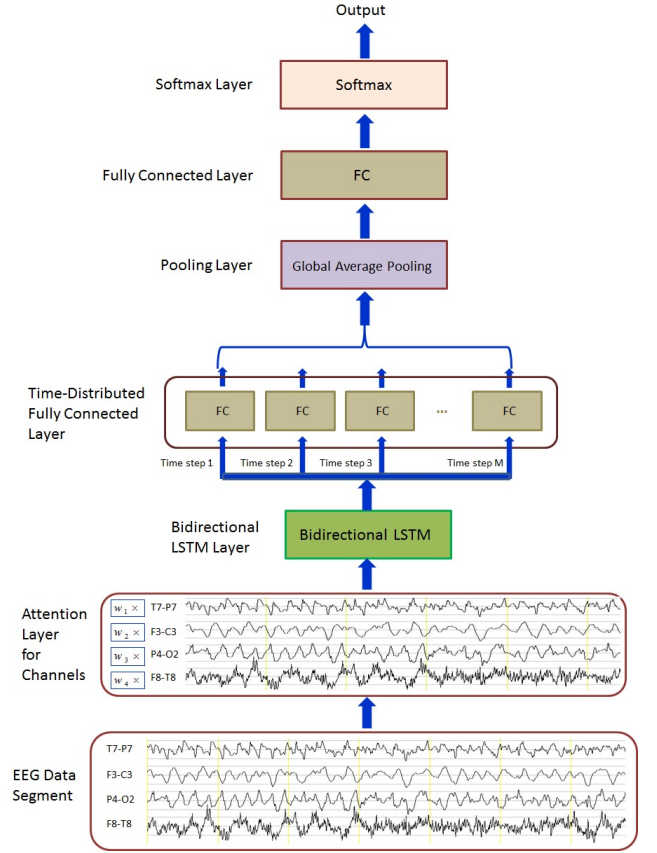


Fig. 1: Architecture of the proposed model

$$Y_2 = f_{\text{average}}(Y_1) \quad (4)$$

$$Y_3 = f_{\text{copy}}(Y_2) \quad (5)$$

$$Y_{\text{atten}} = X_{\text{input}} \odot Y_3 \quad (6)$$

Here, X_{input} denotes an input tensor of size (n_S, n_T, n_C) . Symbols n_S , n_T , n_C respectively represent the number of samples, the number of time steps, and the number of signal channels. X is a matrix of size (n_{ST}, n_C) , $n_{ST} = n_S * n_T$, $W_{FC-\text{atten}}$ a weight matrix of size (n_C, n_C) , a bias matrix $B_{FC-\text{atten}}$ of size (n_{ST}, n_C) , and $Y_{FC-\text{atten}}$ with size (n_{ST}, n_C) . A symbol $\sigma(\cdot)$ represents a non-linear function, like $\text{softmax}(\cdot)$ and $\text{sigmoid}(\cdot)$. Y_1 is a matrix of size (n_S, n_T, n_C) , Y_2 of size (n_S, n_C) , Y_3 of size (n_S, n_T, n_C) , and Y_{atten} with size (n_S, n_T, n_C) . Functions $f_{\text{reshape}_1}(\cdot)$ and $f_{\text{reshape}_2}(\cdot)$ are to reshape a matrix, $f_{\text{average}}(\cdot)$ is a function of computing averages along with the second axis of matrix, and $f_{\text{copy}}(\cdot)$ is an copying operation to share the averages over all the time steps. The symbol \odot means an element-wise multiplication between matrices.

2) *BiLSTM Module*: The BiLSTM module processes the input sequence separately according to the forward order and the backward order, and synthesize the forward outputs and the backward outputs [25] [26]. Its main procedure is illustrated in Fig. 3. In either forward order or backward order, the sequence is computed in the same way as LSTM, in which the computation can be described by using Eqs. (7)–(12) according to [27] and [28]. The synthesizing operations can be concatenation or summation.

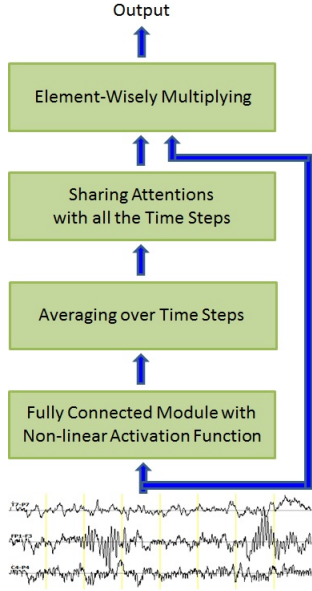


Fig. 2: Attention layer

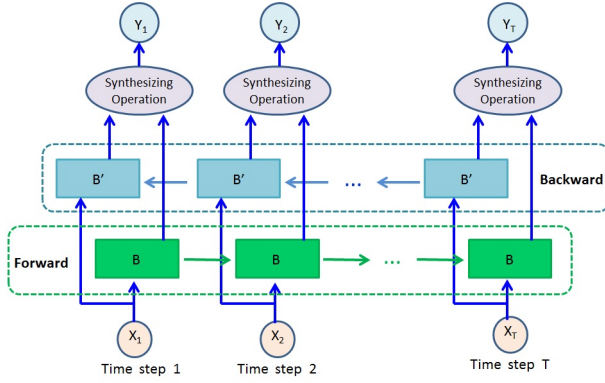


Fig. 3: BiLSTM module

$$\text{Block input } \tilde{\mathbf{c}}^t = g(\mathbf{x}^t * \mathbf{W}_c + \mathbf{y}^{t-1} * \mathbf{R}_c + \mathbf{b}_c) \quad (7)$$

$$\text{Input gate } \mathbf{i}^t = \sigma(\mathbf{x}^t * \mathbf{W}_i + \mathbf{y}^{t-1} * \mathbf{R}_i + \mathbf{b}_i) \quad (8)$$

$$\text{Forget gate } \mathbf{f}^t = \sigma(\mathbf{x}^t * \mathbf{W}_f + \mathbf{y}^{t-1} * \mathbf{R}_f + \mathbf{b}_f) \quad (9)$$

$$\text{Output gate } \mathbf{o}^t = \sigma(\mathbf{x}^t * \mathbf{W}_o + \mathbf{y}^{t-1} * \mathbf{R}_o + \mathbf{b}_o) \quad (10)$$

$$\text{Cell } \mathbf{c}^t = \mathbf{c}^{t-1} \odot \mathbf{f}^t + \tilde{\mathbf{c}}^t \odot \mathbf{i}^t \quad (11)$$

$$\text{Block output } \mathbf{y}^t = h(\mathbf{c}^t) \odot \mathbf{o}^t \quad (12)$$

Here, \mathbf{x}^t is an input vector at the time step t , and \mathbf{y}^t an output vector at the time step t . Input weights matrices \mathbf{W}_c , \mathbf{W}_i , \mathbf{W}_f and \mathbf{W}_o are with shape (n_C, n_{units}) . Recurrent weights matrices \mathbf{R}_c , \mathbf{R}_i , \mathbf{R}_f , and \mathbf{R}_o are of size (n_{units}, n_{units}) . Bias weights \mathbf{b}_c , \mathbf{b}_i , \mathbf{b}_f , and \mathbf{b}_o are of size $(1, n_{units})$. $g(\cdot)$, $\sigma(\cdot)$, and $h(\cdot)$ are non-linear activation functions. The symbol \odot means element-wise multiplication.

Based on Eqs. (7)–(12), a forward output sequence \vec{Y}_{seq} can be obtained corresponding to an input sequence $\mathbf{x}_1 \mathbf{x}_2 \cdots \mathbf{x}_{n_T}$, and a backward output sequence \overleftarrow{Y}_{seq} corresponds to the inverse sequence $\mathbf{x}_{n_T} \cdots \mathbf{x}_2 \mathbf{x}_1$. We use $\vec{Y}_{seq}(t)$ to denote the t -th item in the sequence \vec{Y}_{seq} , i.e. the forward output at the time step t . And $\overleftarrow{Y}_{seq}(t)$ means the backward output at the time step t . The two output sequences \vec{Y}_{seq} and

\overleftarrow{Y}_{seq} are then synthesized as follows:

$$Y_{seq-BiLSTM}(t) = \Phi(\vec{Y}_{seq}(t), \overleftarrow{Y}_{seq}(n_T + 1 - t)). \quad (13)$$

Here, $t = 1, \dots, n_T$. $\Phi(\cdot)$ means an operation, which can be chosen to be concatenation or summation. $Y_{seq-BiLSTM}$ represents the synthesized sequence of the forwarding output sequence and the backward output sequence, and $Y_{seq-BiLSTM}(t)$ is the t -th item in the sequence $Y_{seq-BiLSTM}$, i.e. the output of BiLSTM module at the time step t .

3) *Time-Distributed Fully-Connected Layer*: The time-distributed fully-connected layer is to further extract features at each time point. It executes fully-connected operations separately and simultaneously for inputs at each time step. And the fully-connected operations adopt linear functions as activation functions. Time-distributed layer could help improve executing efficiency when processing signal data with high sampling frequency. At each time step, the computation procedure is described as follows:

$$Y_{seq-TimeDistr}(t) = Y_{seq-BiLSTM}(t) * \mathbf{W}_t + \mathbf{B}_t. \quad (14)$$

Here, $t = 1, 2, \dots, n_T$. Matrix $Y_{seq-BiLSTM}(t)$ of size (n_S, n_{units}) , is the t -th item of the output sequence of BiLSTM module. \mathbf{W}_t denotes a weight matrix of size $(n_{units}, n_{features})$, \mathbf{B}_t a bias matrix of size $(n_S, n_{features})$, and $Y_{seq-TimeDistr}(t)$ a matrix of size $(n_S, n_{features})$. All the time-step components $\{Y_{seq-TimeDistr}(t), t = 1, \dots, n_T\}$ form a sequence $Y_{seq-TimeDistr}$, and further compose a matrix $Y_{TimeDistr}$ of size $(n_S, n_T, n_{features})$ as the output of the time-distributed fully-connected layer.

4) *Pooling Layer*: The pooling layer in our architecture executes the average pooling operation in order to extract global features of each sample. The operation computes the mean value of the time-step data for each sample.

5) *Fully Connected Layer and Softmax Layer*: Fully connected layer executes the fully connected operation to extract further features and to reduce the last dimension of input matrix into number of classes. And it utilizes an linear function as its activation function. Based on outputs of the fully-connected layer, Softmax layer computes probabilities that each sample belongs to a classification. In the following, we will use Eq. (15) and Eq. (16) to present the computations in the fully-connected layer and in the Softmax Layer.

$$Y_{FC} = Y_{AvePool} * \mathbf{W}_{FC} + \mathbf{B}_{FC} \quad (15)$$

$$Y_{Softmax} = \text{softmax}(Y_{FC}) \quad (16)$$

Here, $Y_{AvePool}$ is a matrix of size $(n_S, n_{features})$, which is an output of the pooling layer. \mathbf{W}_{FC} and \mathbf{B}_{FC} denotes respectively weights matrix of size $(n_{features}, n_{classes})$ and bias matrix of size $(n_S, n_{classes})$. Function $\text{softmax}(\cdot)$ is to compute the predicted probabilities of samples belonging to some classes. Y_{FC} and $Y_{Softmax}$ are respectively the outputs of the fully connected layer and the Softmax layer.

The pseudo-codes of the proposed seizure detection approach of BiLSTM with attention are presented in Algorithm 1.

Algorithm 1 Seizure Detection over EEG Data using BiLSTM with Attention

Input: X_{input} , the matrix of EEG data segments

Output: Y_{output} , the matrix of predicted label information

- 1: Initialize matrices $W_{FC-atten}$, $B_{FC-atten}$, W_c , W_i , W_f , W_o , R_c , R_i , R_f , R_o , b_c , b_i , b_f , b_o , W_{FC} , B_{FC} , W_t , B_t , $t = 1, 2, \dots, n_T$
 - 2: Compute the output matrix Y_{atten} using Eqs. (1)–(6)
 - 3: Split Y_{atten} into n_T components $\{x_1, x_2, \dots, x_{n_T}\}$ according to time steps, and compose a sequence $x_1 x_2 \dots x_{n_T}$ in chronological order
 - 4: Compute a forward output sequence \vec{Y}_{seq} for the sequence $x_1 x_2 \dots x_{n_T}$ based on Eqs. (7)–(12)
 - 5: Compute a backward output sequence \overleftarrow{Y}_{seq} for the inverse sequence $x_{n_T} \dots x_2 x_1$ based on Eqs. (7)–(12)
 - 6: Synthesize sequences \vec{Y}_{seq} and \overleftarrow{Y}_{seq} by using Eq. (13), and achieve a sequence $Y_{seq-BiLSTM}$
 - 7: Compute a sequence $Y_{seq-TimeDistr}$ by using Eq. (14), and then compose a matrix $Y_{TimeDistr}$ according to time steps
 - 8: Compute matrix $Y_{AvePool}$ by averaging the values over time steps for each sample in $Y_{TimeDistr}$
 - 9: Compute matrix $Y_{Softmax}$ according to Eqs. (15) and (16)
 - 10: Compute the column position of the maximal element in each row of $Y_{Softmax}$, and achieve Y_{output}
 - 11: Return Y_{output}
-

IV. EVALUATION

In this section, we evaluate the approach of BiLSTM with attention by using the noisy real pediatric scalp EEG data set CHB-MIT. Our evaluation uses two standard metrics, the sensitivity and the specificity.

A. Data

1) *CHB-MIT Dataset*: The data set CHB-MIT contains 686 EEG recordings from 23 subjects of different ages ranging from 1.5 years to 22 years. The recordings include 198 seizures. The used sampling frequency is 256 Hz. Each recording is a digital EEG signal. Most recordings are one hour long, and some are two hours long or four hours long. The EEG recordings are grouped into 24 cases and stored in EDF data files. Each EDF file corresponds to an EEG recording. In each case, the data recordings are from a single subject. Case chb21 was obtained 1.5 years after Case chb01 from the same subject. Each data file contains data over 23 or more channels. In some data files, the data over some channels were missing. And some data files, for example, chb12_27.edf, chb12_28.edf and chb12_29.edf, have different channel montages from other seizure files. In our experiments, we remove the above three EDF files.

2) *Data Segmentation*: In order to extract effective seizure features, 17 common channels are chosen. That means, only the data over the 17 common channels for each subject are analyzed to extract seizure/non-seizure features. The 17

common channels are respectively P4-O2, FP2-F4, P7-O1, C4-P4, F7-T7, C3-P3, FP1-F7, F8-T8, FZ-CZ, CZ-PZ, F3-C3, T7-P7, P8-O2, FP1-F3, F4-C4, FP2-F8, and P3-O1. According to a data segment length (i.e. 23 seconds), each data record in each case is split into data segments. When splitting, only the EEG data over the 17 common channels are collected into data segments. According to annotation files which mark the starting time and the ending time of each seizure, it could be determined whether a data segment contains a seizure or not. In our experiments, if a segment contains a seizure, it is considered as a seizure segment; otherwise, it is a non-seizure segment.

As a result of the splitting, 665 seizure segments and 152401 non-seizure segments are obtained. The 665 seizure data segments are taken as a part of our experiment data. And non-seizure segments for each experiment are randomly selected from the 152401 non-seizure segments. For evaluation over a balanced data, we take 665 non-seizure segments in each experiment.

B. Mixing-Patients Seizure Detection

The deep learning approach in [18] uses LSTM as a main module (shortly, LSTM approach) to detect seizures. The LSTM approach is evaluated through mixing-patients experiments over the EEG data set of Bonn University [29], showing state-of-the-art performance. We will compare our proposed approach with the LSTM approach. And also our approach will be compared with a convolutional neural network approach (for short, CNN approach) in [17]. The CNN approach demonstrates good performances on the Bonn dataset. Because the EEG data in Bonn dataset are heavily processed and contain no artifacts, and its size is small, we choose to use the noisy real dataset CHB-MIT to conduct mixing-patients experiments.

The LSTM approach [18] and the CNN approach [17] do not provide all the source codes. Thus, we implement the two approaches according to their descriptions. The implemented LSTM approach and CNN approach are tested. Our obtained testing results reach to the reported performances in [18] and [17]. Then based on the two implementations, we conduct experiments on the data set CHB-MIT to compare with the approach of BiLSTM with attention.

In each one of the experiments, all the seizure segments are utilized as a part of experiment data, and non-seizure segments with the same quantity are randomly selected. The training set, validation set and testing set are obtained by randomly splitting the experiment data set according to the ratio 70:15:15. Based on the experimental feedbacks, we tune and determine parameters to attain the best performance for the three approaches, including the LSTM approach, the CNN approach, and BiLSTM with attention. And for each approach, ten experiments are carried out based on the correspondingly selected parameters. The experimental results using the LSTM approach, including Sensitivity, Specificity, F1 score, Accuracy, the average and the standard deviation (denoted by Stan. Dev. in tables), are given in Table I. And the results by using the CNN approach and BiLSTM with attention are presented in Tables II and III, respectively.

TABLE I: Mixing-patients experimental results using the LSTM approach

Item	Sensitivity	Specificity	F1 Score	Accuracy
1	0.8500	0.8800	0.8629	0.8650
2	0.7700	0.8500	0.8021	0.8100
3	0.7900	0.8700	0.8229	0.8300
4	0.7100	0.9300	0.7978	0.8200
5	0.8200	0.8900	0.8497	0.8550
6	0.9100	0.7900	0.8585	0.8500
7	0.8600	0.8300	0.8473	0.8450
8	0.8600	0.8400	0.8515	0.8500
9	0.9400	0.7200	0.8468	0.8300
10	0.9300	0.8300	0.8857	0.8800
Average	0.8440	0.8430	0.8425	0.8435
Stan. Dev.	0.0696	0.0550	0.0259	0.0201

TABLE II: Mixing-patients experimental results using the CNN approach

Item	Sensitivity	Specificity	F1 Score	Accuracy
1	0.8400	0.8500	0.8442	0.8450
2	0.9200	0.7700	0.8558	0.8450
3	0.8000	0.8400	0.8163	0.8200
4	0.9000	0.6900	0.8145	0.7950
5	0.9200	0.8000	0.8679	0.8600
6	0.7900	0.8500	0.8144	0.8200
7	0.6300	0.9700	0.7590	0.8000
8	0.8500	0.8700	0.8586	0.8600
9	0.8700	0.7700	0.8286	0.8200
10	0.9600	0.6900	0.8458	0.8250
Average	0.8480	0.8100	0.8305	0.8290
Stan. Dev.	0.0891	0.0809	0.0301	0.0217

For the LSTM approach, the achieved average sensitivity and average specificity are respectively 84.4% and 84.3%. By using BiLSTM with attention, the obtained average sensitivity of 86.6% and specificity of 86.0% are better than the LSTM approach. For the F1 score, the approach of BiLSTM with attention also surpasses the LSTM approach. And the standard deviations by the BiLSTM with attention are less than the LSTM approach. It can be seen that our approach of BiLSTM with attention not only detects seizures better than the LSTM approach, but also more stably.

For the CNN approach, the obtained average sensitivity and average specificity are 84.8% and 81.0%, respectively. Our model outperforms the CNN approach in Sensitivity and Specificity. For the average accuracy and the average F1 score, our approach also outperforms the CNN approach. And the standard deviations of the sensitivity and specificity by our method are much smaller than the CNN approach. These experimental results show that, the proposed of model BiLSTM with attention has better performance in the seizure detection than the CNN approach.

C. Cross-Patient Task Detection

For cross-patient seizure detection, each experiment takes data of one subject as testing data, and other subjects data as training data and validation data according to the ratio 85:15. For each of the 23 subjects as a testing object, some experiments are carried out. Because the two cases chb01 and chb21 are records from the same subject. The two cases are utilized together either as testing data or as training-validation data. In each experiment, all the seizure data segments from

TABLE III: Mixing-patients experimental results using the BiLSTM with attention

Item	Sensitivity	Specificity	F1 Score	Accuracy
1	0.9000	0.8900	0.8955	0.8950
2	0.8200	0.8900	0.8497	0.8550
3	0.8500	0.8200	0.8374	0.8350
4	0.8700	0.9200	0.8923	0.8950
5	0.8300	0.8700	0.8469	0.8500
6	0.9000	0.8400	0.8738	0.8700
7	0.8600	0.8400	0.8515	0.8500
8	0.8700	0.8100	0.8447	0.8400
9	0.8900	0.8900	0.8900	0.8900
10	0.8700	0.8300	0.8529	0.8500
Average	0.8660	0.8600	0.8635	0.8630
Stan. Dev.	0.0258	0.0349	0.0210	0.0217

each patient are utilized, and non-seizure data segments are randomly selected to be as many as seizure segments. So, the data is balanced in each experiment.

Using each subject as a testing object, we obtain the sensitivity and the specificity, and all the results are listed in Table IV. Figs. 4 and 5 respectively depict the sensitivities and the specificities in the form of histogram. For the 23 subjects in CHB-MIT, the average sensitivity, specificity, and accuracy are 83.72%, 84.06%, and 83.89%, respectively. And the standard deviations of sensitivity and specificity are 0.1349 and 0.1379, respectively.

TABLE IV: Cross-patient experimental results using BiLSTM with attention

Patient ID (Cases)	Sensitivity	Specificity	F1 Score	Accuracy
P01 (chb01, chb21)	0.8974	0.7179	0.8235	0.8077
P02 (chb02)	0.8000	1.0000	0.8889	0.9000
P03 (chb03)	0.8846	0.9615	0.9200	0.9231
P04 (chb04)	0.9524	0.8095	0.8889	0.8810
P05 (chb05)	1.0000	0.4286	0.7778	0.7143
P06 (chb06)	0.8125	0.7500	0.7879	0.7813
P07 (chb07)	0.9412	0.8824	0.9143	0.9118
P08 (chb08)	0.9556	0.7333	0.8600	0.8444
P09 (chb09)	0.9375	0.6250	0.8108	0.7813
P10 (chb10)	0.9600	0.8800	0.9231	0.9200
P11 (chb11)	0.9730	0.8649	0.9231	0.9189
P12 (chb12)	0.5211	0.8451	0.6218	0.6831
P13 (chb13)	0.6000	0.8571	0.6885	0.7286
P14 (chb14)	0.6429	0.9286	0.7500	0.7857
P15 (chb15)	0.7379	0.9223	0.8128	0.8301
P16 (chb16)	0.6875	0.6250	0.6667	0.6563
P17 (chb17)	1.0000	0.8125	0.9143	0.9063
P18 (chb18)	0.9000	0.9000	0.9000	0.9000
P19 (chb19)	0.7857	1.0000	0.8800	0.8929
P20 (chb20)	0.7273	0.9545	0.8205	0.8409
P21 (chb22)	0.9167	0.9167	0.9167	0.9167
P22 (chb23)	0.9200	1.0000	0.9583	0.9600
P23 (chb24)	0.7027	0.9189	0.7879	0.8108
Average	0.8372	0.8406	0.8363	0.8389
Stan. Dev.	0.1349	0.1379	0.0888	0.0833

In [4], Thodoroff et al. utilize a recurrent convolutional neural network (recurrent CNN) and obtain an average sensitivity 85% in cross-patient experiments on the data set CHB-MIT. According to Figure 7.(a) and (c) in [4], for six cases chb06, chb12, chb13, chb14, chb15 and chb16, the obtained sensitivity results are not good, some even only about 20%. For other seventeen cases the sensitivity results are mostly 100%. The two cases chb01 and chb21 are tested separately

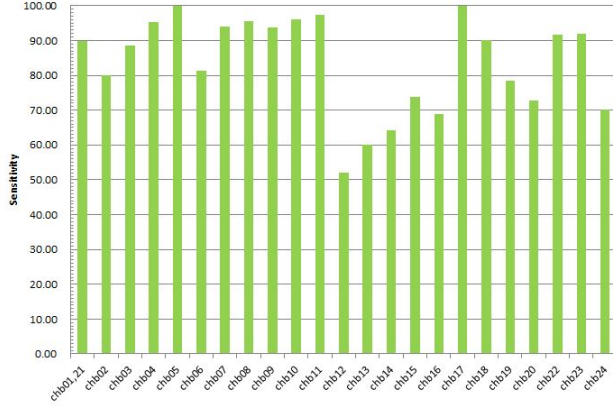


Fig. 4: Cross-patient sensitivity of BiLSTM with attention

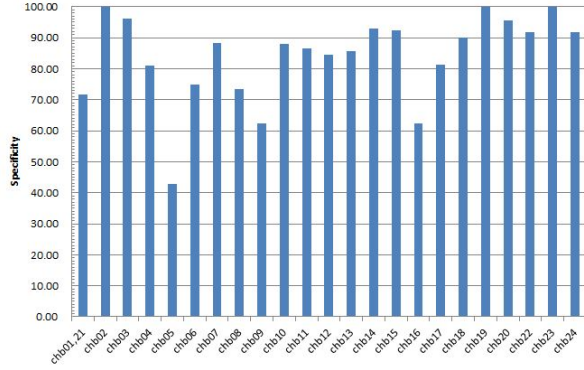


Fig. 5: Cross-patient specificity of BiLSTM with attention

for recurrent CNN. The proposed approach of BiLSTM with attention can obtain much better sensitivities for the above six cases, all being over 50%, although the sensitivities over the remaining cases do not reach 100%. Fig. 6 presents the sensitivity comparisons between the method of recurrent CNN and our approach of BiLSTM with attention for the above six cases. And Fig. 7 depicts the distribution of the sensitivities of 21 commonly-tested cases. The 21 cases do not contain chb01, chb21 and chb24. Over the commonly-tested cases, our standard deviations for sensitivity and specificity are 0.1374 and 0.1407, respectively. It can be seen that our sensitivity results are more concentrative, and in this sense, the proposed approach of BiLSTM with attention is more stable.

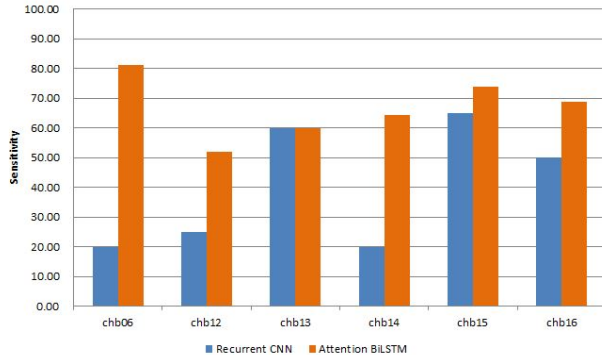


Fig. 6: Comparison of cross-patient sensitivity over 6 cases between attention BiLSTM and recurrent CNN

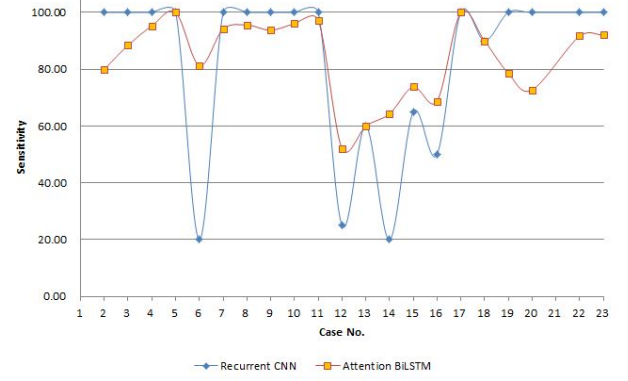


Fig. 7: Comparison of cross-patient sensitivity over 21 common cases between our approach and recurrent CNN

V. DISCUSSIONS

In this paper, we design a novel approach of BiLSTM with attention for seizure detection. The mixing-patients and cross-patient experiments are separately conducted over the pediatric data set CHB-MIT. The used data segments in each experiment contain 665 seizure segments and 665 non-seizure segments. For each segment, the time duration is 23 seconds. In the 24 cases, the total number of non-seizure segments generated in our slicing way is 152401. The 665 non-seizure segments for training, validation and testing in our each experiment are randomly selected from all the non-seizure segments. The selection is sparse enough. The randomness and sparsity of selection reduce temporal correlations among non-seizure data segments, and avoid resulting in overly optimistic specificity results [7]. Further, the two characteristics of the selecting way make the evaluation of the model BiLSTM with attention be reliable.

In the mixing-patients experiments, the obtained sensitivity and specificity by using the proposed approach of BiLSTM with attention are better than the LSTM approach in [18] and the CNN approach in [17]. The improvements in (Sensitivity, Specificity) over those two state-of-the-art approaches are 2.2%, 1.7% and 1.8%, 5%, respectively. And the standard deviations are much less than the two approaches in comparison. The CNN approach mainly utilizes a convolution neural network, leaky ReLU activation function and a max pooling layer. The LSTM approach mainly uses a LSTM network, a time-distributed layer and a global average pooling layer. By comparing the architectures of the three approaches, the better performances of the proposed approach of BiLSTM with attention are attributed to the attention mechanism over channels and the feature extraction in both forward and backward directions.

In the cross-patient experiments, the performances of BiLSTM with attention are more stable than the method of recurrent CNN proposed in [4]. The average sensitivity of our approach on 23 subjects is 1.3% less than that of recurrent CNN. When testing the method of recurrent CNN, the convolution neural network module is pre-trained separately in [4] before training the whole model. Our approach of BiLSTM with attention does not need to use pre-training,

and it directly processes raw data and extracts features. The REVEAL algorithm proposed in [30] achieves an average sensitivity of 61%. [5] uses the automatic seizure detection system EpiScan on the data set CHB-MIT and obtains an average sensitivity of 67%. The average sensitivity of our approach is much better than REVEAL and EpiScan.

For the setting of data segment length of 23 seconds in our experiments, it is the same as what is referred to as the signal time duration in the Bonn EEG data set [29]. We have not studied how to select the most proper segment length, which will be a future research topic. Regarding how to add attention weights, we have conducted some related experiments. In the case of using attention mechanism over time steps, the experimental results are not as good. Also another kind of attention mechanism over channels is tested, in which the channel attention weights at different time steps are different. Its testing results are not good. Finally, we choose to apply attention mechanism to channels and share the attention weights among time steps. Actually, different channels have different contributions to a seizure, and the contributions turn out to be much correlated to the locations of brain areas, rather than the time. In addition, some tests by using data on one channel have also been conducted. The experimental results by using multiple channel information are better. The results are in agreement with the observation in [7]; that is, over some channels, the data morphology in seizure state is similar to that in non-seizure state.

VI. CONCLUSIONS

This paper focuses on the problem of automatic seizure detection. Inspired by the architecture in [18], we analyze both spacial and temporal characteristics of seizures, and propose a novel deep learning-based approach by using the model of BiLSTM integrated with attention. The integration of an attention mechanism is to capture spatial features better, and the employment of the BiLSTM model is to extract more discriminating temporal features. The proposed approach is evaluated on the noisy real EEG data set of CHB-MIT. The evaluation is across different regions of the brain and across multiple subjects. In the mixing-patients experiments, we obtain sensitivity of 86.6% and specificity of 86.0%, which are better than the LSTM approach in [18] and the CNN approach in [17]. In the cross-patient experiments, the testing results are 83.72%-sensitivity and 84.06%-specificity on average. Comparing to the model recurrent CNN in [4], our model BiLSTM with attention is more stable.

In the model BiLSTM with attention, the pooling layer adopts a globally-averaging way to extract holistic features of data segments. The problem that whether such a way is the best or not for the seizure detection, will be explored in the future. And also we want to investigate whether the length of data segments has effects on the sensitivity and the specificity.

REFERENCES

- [1] I. Megiddo, A. Colson, D. Chisholm, T. Dua, A. Nandi, and R. Laxminarayan, "Health and economic benefits of public financing of epilepsy treatment in India: An agent-based simulation model," *Epilepsia*, vol. 57, no. 3, pp. 464-474, 2016.
- [2] J. Gotman, J. R. Ives, and P. Gloor, "Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings," *Electroencephalography and Clinical Neurophysiology*, vol. 46, no. 5, pp. 510-520, 1979.
- [3] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 54, no. 5, pp. 530-540, 1982.
- [4] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. 1st MLHC*, Los Angeles, CA, USA, 2016, pp. 178-190.
- [5] F. Furbass, P. Ossenblok, M. Hartmann, H. Perko, A. M. Skupch, G. Lindinger, L. Elezi, E. Patarai, A. J. Colon, C. Baumgartner, and T. Kluge, "Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units," *Clinical Neurophysiology*, vol. 126, no. 6, pp. 1124-1131, 2015.
- [6] A. S. Zandi, M. Javidan, G. A. Dumont, and R. Tafreshi, "Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1639-1651, 2010.
- [7] A. Shueb and J. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. 27th ICML*, Haifa, Israel, 2010, pp. 975-982.
- [8] S. Amin and A. M. Kamboh, "A robust approach towards epileptic seizure detection," in *Proc. IEEE 26th Int. Workshop MLSP*, Vietri sul Mare, Italy, 2016, pp. 1-6.
- [9] M. Fan and C. Chou, "Detecting abnormal pattern of epileptic seizures via temporal synchronization of EEG signals," *IEEE Transactions on Biomedical Engineering*, to be published.
- [10] B. Hunyadi, M. Signoretto, W. V. Paesschen, J. A. K. Suykens, S. V. Huffel, and M. D. Vos, "Incorporating structural information from the multichannel EEG improves patient-specific seizure detection," *Clinical Neurophysiology*, vol. 123, no. 12, pp. 2352-2361, 2012.
- [11] A. V. Esbroeck, L. Smith, Z. Syed, S. Singh, and Z. Karam, "Multitask seizure detection: Addressing intra-patient variation in seizure morphologies," *Machine Learning*, vol. 102, no. 3, pp. 309-321, 2016.
- [12] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, and J. Yang, "Supervised learning in automatic channel selection for epileptic seizure detection," *Expert Systems with Applications*, vol. 86, pp. 199-207, 2017.
- [13] L. Vidyaratne, A. Glandon, M. Alam, and K. M. Iftikharuddin, "Deep recurrent neural network for seizure detection," in *Proc. IJCNN*, Vancouver, BC, Canada, 2016, pp. 1202-1207.
- [14] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Integer convolutional neural network for seizure detection," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, to be published.
- [15] H. Qu and J. Gotman, "Improvement in seizure detection performance by automatic adaptation to the EEG of each patient," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 2, pp. 79-87, 1993.
- [16] M. Golmohammadi, S. Ziyabari, V. Shah, E. V. Weltin, C. Campbell, L. Obeid, and J. Picone, "Gated recurrent networks for seizure detection," in *Proc. SPMB*, Philadelphia, Pennsylvania, USA, 2017, pp. 1-5.
- [17] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, no. 1, pp. 270-278, 2018.
- [18] R. Hussein, H. Palangi, R. Ward, and Z. J. Wang, "Epileptic seizure detection: A deep learning approach," unpublished. Available: <https://arxiv.org/abs/1803.09848>
- [19] P. Fergus, A. Hussain, D. Hignett, D. A. Jumeily, K. A. Aziz, and H. Hamdan, "A machine learning system for automated whole-brain seizure detection," *Applied Computing and Informatics*, vol. 12, no. 1, pp. 70-89, 2016.
- [20] N. Nicalaou and J. Georgiou, "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines," *Expert Systems with Applications*, vol. 39, no. 1, pp. 202-209, 2012.
- [21] S. Nasehi and H. Pourghasem, "Patient specific epileptic seizure onset detection algorithm based on spectral features and IPSONN classifier," in *Proc. Int. Conf. CSNT*, Gwalior, India, 2013, pp. 186-190.
- [22] A. Kharbouch, A. Shueb, J. Guttag, and S. S. Cash, "An algorithm for seizure onset detection using intracranial EEG," *Epilepsy & Behavior*, vol. 22, no. 1, pp. S29-S35, 2011.
- [23] Y. Zheng, J. Zhu, Y. Qi, X. Zheng, and J. Zhang, "An automatic patient-specific seizure onset detection method using intracranial electroencephalography," *Neuromodulation*, vol. 18, no. 2, pp. 79-84, 2015.
- [24] H. Hassanpour and M. Shahiri, "Adaptive segmentation using wavelet transform," in *Proc. Int. Conf. Electrical Engineering*, Lahore, Pakistan, 2007, pp. 1-5.

- [25] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673-2681, 1997.
- [26] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602-610, 2005.
- [27] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. ICANN*, Edinburgh, UK, 1999, pp. 850-855.
- [28] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222-2232, 2017.
- [29] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, 061907, 2001.
- [30] S. B. Wilson, M. L. Scheuer, R. G. Emerson, and A. J. Gabor, "Seizure detection: Evaluation of the Reveal algorithm," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2280-2291, 2004.