A Deep Learning-Based Method for Automatic Detection of Epileptic Seizure in a Dataset With Both Generalized and Focal Seizure Types

A. Einizade[1], M. Mozafari[1], S. Hajipour Sardouie[1], S. Nasiri[2,3] and G. Clifford[3,4]

1. Sharif University of Technology, Tehran, Iran
2. Harvard Medical School, Massachusetts, USA
3. Emory School of Medicine, Georgia, USA
4. Georgia Institute of Technology, Georgia USA

aref.einizade@yahoo.com, mohsen.mozafari21@yahoo.com, hajipour@sharif.edu, snasiri@mgh.harvard.edu, gari@gatech.edu

Abstract—Epilepsy is the second most popular neurological disorder affecting 65 million people around the world. Seizures are classified into two kinds; focal and generalized ictal activities, reflecting the spread of seizure activity on the brain. Focal seizures start and affect specific regions of the brain, whereas the generalized propagates throughout the brain. Current approaches to developing an automatic seizure detection algorithm do not consider the types of seizures. However, to detect the focal seizures, the locations of onset of seizure must be identified by an expert through inspection of the electroencephalogram (EEG), which is an expensive and time-consuming procedure. Moreover, most proposed methods are patient-specific and cannot be generalized on an unseen patient, limiting the clinical usage of previous studies. This work presents a generalizable seizure detection algorithm by considering different seizure types. After pre-processing data and rejecting artifacts, a deep neural network is used to extract robust representations across seizures and a population. The proposed method includes deep recurrent and convolutional neural networks to capture spatial and temporal information simultaneously. Experiments on the TUH EEG seizure dataset, which contains both generalized and focal seizures, show that the proposed method increases the accuracy over state-of-the-art from 80.72% to 82%, precision from 67.55% to 71.69%, and sensitivity from 80% to 85%.

I. Introduction

Epilepsy is the second popular neurological disorder, where neurons produce abnormal signals and cause seizures, affecting 65 million people around the world [1]. Seizures are classified into two kinds; focal and generalized ictal activities, reflecting the spread of seizure activity on the brain. Focal seizures start and affect specific areas of the brain, whereas the generalized propagates throughout the brain [2]. The ground truth for seizure detection remains the multi-lead electroencephalogram (EEG) and manual labeling by an expert, which is a costly and time-consuming task. Over the last decade, various methods for seizure detection from EEG signals have been proposed. The performance of a seizure detection algorithm has been affected by different factors such as the density and energy of artifacts and noises in the EEG signals, and the type of seizure.

Wang et al. [3] used the public dataset of Bonn University [4] for seizure detection. First, they reduced the artifacts using the wavelet threshold method. Then, they extracted multi-domain features consisting of time, frequency, time-frequency, and information theory features. The extracted feature space was reduced using Principal Component Analysis (PCA) [5] and Analysis of Variance (ANOVA) [6]. In this study, intracranial electrodes in the Bonn dataset [4] were used which is not practical in real-life scenarios. Pathak et al. [7] investigated several linear features and found that the Line Length feature [8] has an important role in the classification of seizure and non-seizure epochs on the Freiburg dataset [9]. Bolagh et al. [10] proposed subject-selection and subject clustering to select relevant individuals based on the similarity between the EEG patterns of different individuals. They evaluated their method on the CHB-MIT dataset [11] in the cross-subject scenario. Mozafari et al. [8] proposed a method consisting of clustering and classification to detect seizure epochs and evaluated on the Temple University Hospital (TUH) dataset [12]. Multi-domain features such as Line Length were used, and the number of features was reduced using the Fisher feature reduction method [13]. To remove EEG artifacts, any activities that do not originate from the brain, Independent Component Analysis (ICA) and Canonical Correlation Analysis (CCA) were used [14].

Most of the mentioned studies used hand-engineered features, which need strong a priori knowledge of different types of seizures. To address this issue, Deep Learning (DL) has recently become popular for its ability to auto-discover features. Vidyarante et al. [15], used a Deep Recurrent Neural Network to learn both spatial and temporal features of the raw EEG data (CHB-MIT dataset [11]). Birjandtalab et al. [16] extracted frequency domain features by calculating normalized power spectrum density of EEG signals in different frequency bands. A multilayer perceptron is trained for each patient to classify seizure and non-seizure events. Cao et al [17] trained a Convolutional Neural Network (CNN) using the Short-Time Fourier Transform (STFT) of the EEG signals (CHB-MIT dataset [11]) to detect epileptic seizures. The results of this method showed that CNN could significantly improve the accuracy. Talha Avcu et al. [18] designed a CNN for seizure detection called SeizNet, where their proposed network uses a small set of channels to detect epileptic seizures. The SeizNet outperformed the state-of-the-art methods for seizure detection when only two EEG channels were used.

Important approaches and utilized datasets are listed in Table 1 and 2, respectively.

Table 1: Performance metrics of the mentioned studies

Related works	Accuracy	Sensitivity	Specificity	Precision	F-measure	Dataset
Vidyarante [15]	-	100	99.2	-		CHB-MIT
Birjandtalab [16]	-	96.27	-	94.21	95.3	21 patients
Yuzhen Cao [17]	90.13	96.5	93	-	-	CHB-MIT
Talha Avcu [18]	-	95.8	-	-	-	29 patients

Table 2: Public datasets used in the literature of the seizure detection

Name of dataset	# subjects	recording	Types of seizure
CHB-MIT [11]	22	21 channel scalp EEG	Not mentioned
Freiburg [9]	21	126 channels Intracranial electrodes	Focal
Bonn [4]	-	Intracranial electrodes, single channel	Focal
TUH [12]	ongoing	21 channel scalp EEG 20	Generalized and Focal

There are some drawbacks in the aforementioned papers, such as not considering time alterations in the patterns of ictal activities, using the information of the seizure onset region, using intracranial electrodes, and using a dataset that consists of only one kind of seizure. Therefore, the results are not guaranteed for clinical data, which has both kinds of seizures, to be the same as reported. In this study, we propose a seizure detection method based on DL that considers the alterations of the signal to have a better understanding of the ictal activity. Since we usually have no information about the patient in real clinical scenarios, the Leave-One-Subject-Out (LOSO) method is used to evaluate our method which is close to real-life applications. To evaluate the proposed method, the TUH dataset [12] is used which includes a pair of generalized and focal ictal activity.

The rest of this paper is organized as follows. Sec II presents the proposed framework for seizure detection, including a general description of the TUH dataset, the preprocessing steps, and the proposed CNN-LSTM network. The experiments, results, and discussion are presented in Sec III, and finally, Sec IV concludes the paper.

II. Methodology

A. Dataset

This study used the TUH dataset [12], which consists of both generalized and focal seizures. EEGs had been recorded in noisy clinical circumstances, which makes seizure detection a challenging task. Using an anti-epileptic drug or having a stroke changes EEG signals' dynamic in the normal and ictal states, therefore having a meaningful comparison across subjects is impossible. Hence, those patients who used an anti-epileptic drug or had a stroke were excluded from the dataset. Based on this criterion, 70 out of 107 subjects were used in this study, and only one recording from each subject was used. The sampling frequency was 250 or 256 Hz, and the length of each recording was between 10 to 60 minutes. In the aforementioned dataset, the onset and ending seconds of all ictal activities, and the montages which were under effect of ictal activities are annotated.

B. Preprocessing

To automatically reduce noise and artifact, the proposed framework used a method based on Independent Component Analysis (ICA) [19] and Canonical Correlation Analysis (CCA) [20] methods. After reducing artifacts, frequency bands related to line noise (59-61 Hz) were filtered using a notch filter. Frequencies less than 0.5 Hz were also filtered to reduce the sweat artifacts. After artifact reduction, signals were transformed into Temporal Central Parasagittal (TCP) montage with 22 channels [21]. Figure 1 and Figure 2 show an example of signals before and after artifact reduction in TCP montage. Figure 3 shows the location of employed channels in the dataset.

C. Deep Learning (DL) based network

This section presents the proposed hybrid DL model for seizure detection tasks. The model consists of two kinds of DL network architectures which has been shown in Figure 4. The proposed network takes the advantages of both CNN and RNN advantages, where CNN and RNN capture spatial and temporal information in EEGs, respectively.

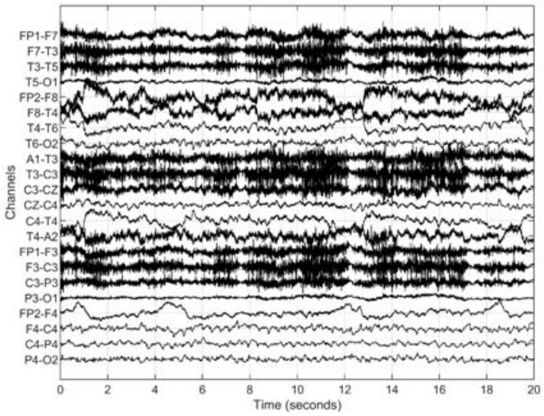

Figure 1: EEG signal contaminated with EMG artifacts. This artifact can be seen with high frequency (>30 Hz) activity in some channels such as FP1-F7 (patient ID 00000177).

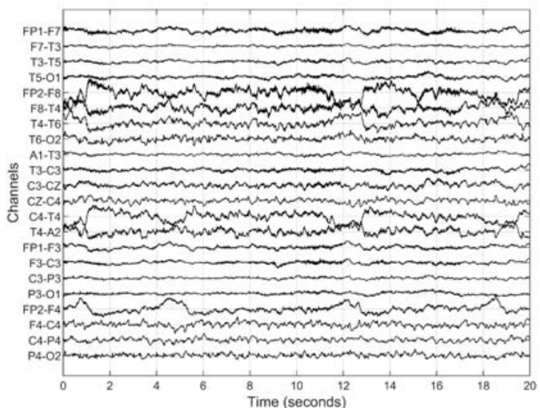

Figure 2: EEG signal of Figure 1. The EMG activities at channels such as FP1-F7 are reduced using BSS method (patient ID 00000177).

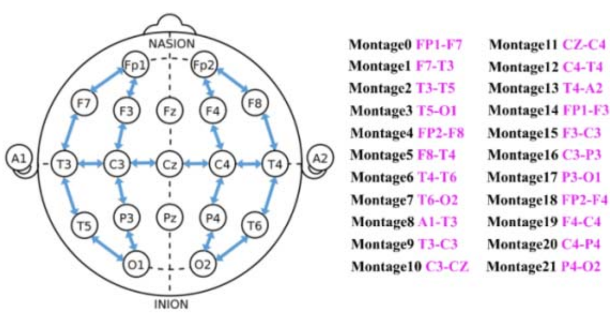

Figure 3: TCP or double banana montage which is used in clinical seizure detection.

It is known that the Long Short-Term Memory (LSTM) unit is used to learn the temporal dependencies in the sequential sequences, which is expressed as follows [22]:

$$\{T_i\}_{i=1}^{n} \rightarrow Y \tag{1}$$

where Y is the label of an epoch (e.g. seizure, non-seizure), T_i denotes the ith segment of the raw multi-channel EEG signals in the corresponding trial and n is the number of the segments in the entire trial. The final label of the entire trial in our model is determined by the average of the outputs of all steps (sub-segments). Due to the non-stationary nature of EEGs, the label of an epoch is computed by taking averages over segments (time-steps of the LSTM unit). These explanations can be formalized as follows:.

$$Y = argmax_j \left[\frac{1}{n}\sum_{i=1}^{n} y_i\right] \tag{2}$$

where $y_i \in R^{1\times 2}$ denotes the output of the $softmax$ layer corresponding to the i-th LSTM time-step. If the first element of the average vector (i.e. $\frac{1}{n}\sum_{i=1}^{n} y_i$) is larger than the other, the final label for the entire initial trial is 0 (non-seizure epoch); otherwise, it is 1.

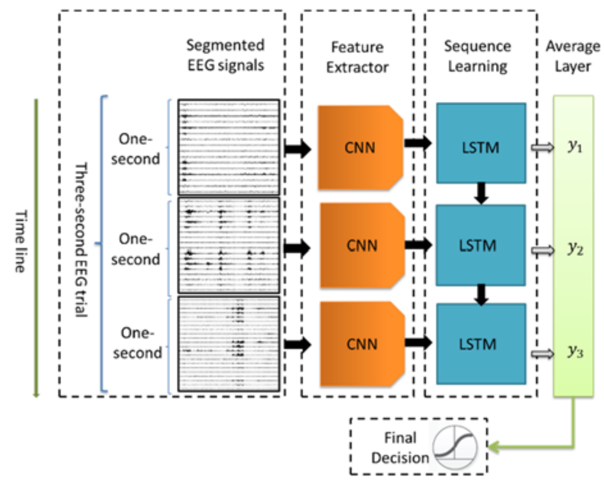

Figure 4: Framework of proposed method: The sequential data consists of three sub segments of the three-second EEG trial. Each time step data is EEG signals of the one-second segment. The CNN extracts features from the preprocessed EEG signals, and the following LSTM network learns the long-term dependencies in the sequential data. An epoch's label is obtained by taking the average on outputs of the Softmax layers of all time-steps.

The two components of the hybrid proposed model are defined as follows:

1) *Convolutional Neural Networks (CNN)*

Deep neural networks are more powerful tools for learning general and extracting robust features from the raw data [23, 24]. The various applications of these networks include speech recognition, natural language, etc. CNNs are generally made up of one or more convolutional layers. Each of these layers consists of three processing parts: the convolution stage, the activation function stage, and the pooling stage.

- The convolution stage slides the convolution filters onto the two-dimensional input, and then several feature maps are obtained from this step.
- The activation function is a nonlinear transformation, e.g. Sigmoid, ReLU, and ELU, which applies to the extracted feature from the previous step. The activation function helps the network to better map features and more accurately predicts the main labels.
- The last part of transformation is called pooling (e.g., average pooling, max pooling, etc.) that helps the network to resist possible small changes, such as translation. Also, this part dramatically reduces the dimensions of the feature maps processed from the previous step. In this work, the CNN part of our network is inspired by EEGNET [23]. EEGNET is a well-known CNN architecture that has shown its effectiveness in classifying EEG signals.

Table 3 shows the characteristics of the CNN network in our proposed model. Note that before connecting to the LSTM unit, a flattening operation is needed to vectorize the final feature maps.

Table 3: The CNN architecture in the proposed network (C= # channels)

Layers	# filters	size	Activation	Options
Conv2D	16	(1,64)	Linear	Mode=same
BatchNorm				
Conv2D	16	(C,1)	Linear	Mode=valid, max norm=1
BatchNorm				
Activation			ELU	
AveragePool2D		(1,4)		
Dropout				P = 0.5
Conv2D	16	(1,16)	Linear	Mode=same
BatchNorm				
Activation			ELU	
AveragePool2D		(1,8)		
Dropout				P = 0.5

2) *Recurrent Neural Networks (RNN)*

After extracting feature maps from time segments of the entire trial using a CNN network, the learnable relationship of the sequential data is used to predict the whole trial label. The reuse of weights at each stage is the critical difference between RNN and CNN. RNNs can define which parts of the time sequence play an essential role in the input data. However, one of the main drawbacks of RNNs is the explosion or vanishing of the gradient in the case of having long-term learning tasks. [25]. To address this problem, certain types of these units, such as GRU and LSTM, were designed [26, 27]. These gated RNNs have been effective in various applications such as emotion recognition [27], speech recognition [28], machine translation [29], and image captioning [30].

The RNN part of our model is inspired by [31] .In [31], first, the Scalograms of the windowed segments are calculated across the temporal axis which forms a three-way tensor data for each segmented window. Then these tensors are averaged over the temporal axis and form 2D images. Finally, a CNN processes these images. An averaging function is applied to the following RNN's outputs to obtain the final label for the whole initial epoch. In this work, an LSTM unit is used, where by unfolding this unit, we can understand how information is processed over time. Utilizing the structure of gates, it can be seen that in this type of RNNs, the dependence of information flow over time is learned using concepts such as forget [32]. The formulation of LSTM is presented as follows [32]:

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

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

$$\overline{C_t} = tanh(W_c.[h_{t-1}, x_t] + b_c) \tag{5}$$

$$C_t = C_{t-1} \odot f_t + \overline{C_t} \odot i_t \tag{6}$$

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

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

where at each time step t, the hidden state h_t of the last LSTM layer is taken as the sequence encoding, and h_{t-1} denotes for the hidden state of the previous time step. Note that $\odot$ is the Hadamard product [33], and x_t is an input vector. Also, f_t, i_t, C_t and o_t are outputs of Forget, Input, Update and Output gates, respectively. The weight learnable parameters of the LSTM unit are W_f, W_c, W_i, W_o, b_f, b_i, b_c and b_o. σ and $tanh$ denote for the Sigmoid and Hyperbolic tangent functions, respectively.

In this work, each trial consists of three-second recorded epileptic EEG signals. In order to take advantage of LSTM, we divided these trials into three one-second sub-trials without any overlapping. Figure 5 shows the framework of the proposed method.

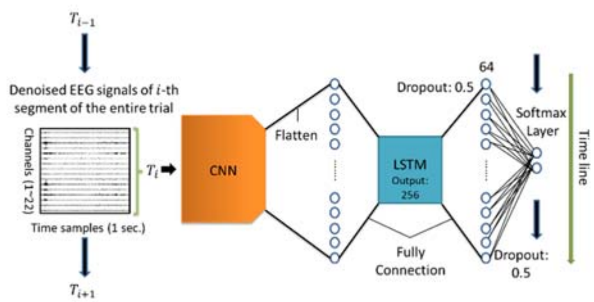

Figure 5: General structure of our model for data processing in one time-step.

III. Results and Discussion

Due to the different types of seizure across the population, it is necessary to develop a model with high generalizability across subjects. Therefore, to evaluate the performance of the proposed model on an unseen patient, the Leave-One-Subject-Out (LOSO) cross validation is utilized. The proposed DL model was developed in Keras with Tensorflow backend [34]. For training, we used Adam optimizer with a learning rate of 0.001, binary cross entropy loss function and batch size of 64. Also, 70 iterations have been selected to execute without any validation split. Due to the imbalance issue in the seizure detection task, the weighting technique of two-class samples has been used. Therefore, the data is weighted in the final cost function according to the number of data in each class.

As a baseline, each segment (three second) of recorded are divided into three one-sec sub-epochs and a set of features for each sub-epoch is extracted. Then, the features of three sub-epochs were concatenated to have a feature vector with information about changes in time. In previous studies [8, 35], it is noted that the Line Length feature can discriminate seizure and non-seizure epochs.

It has been shown that the wavelet decomposition is a proper feature for seizure detection [3]. Therefore, this work uses the Line Length of raw signal, details coefficients for levels 2 to 5, and approximate of the discrete wavelet transform (DWT) decomposed by db4 mother-wavelet.. The db4 mother wavelet was chosen due to its morphological similarities with spike waves in seizure activities. Another group of features contains the normalized band power of Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), and Beta (12-30 Hz) band to 1 to 30 Hz. Linear Discriminant Analysis (LDA) classifier, which showed better performance compared to boosting methods and SVM, was used in the classification step. Table 4 shows the comparison of the proposed method with the baseline approach, and the method has been proposed in [8]. The seizure detection method proposed in [8] was evaluated on the same subset of the TUH dataset. The main parts of this method consist of clustering, classification, and voting on each cluster.

As shown in Table 4, the classification results show that statistical learning dependencies between time-steps of the entire trial, learned by the LSTM unit in our model, can significantly improve the learning quality. Non-stationary nature of the EEG signal through time can cause these changes in improvements. In the method of using the LDA classifier, features of the different segments of the entire trial are concatenated, regardless of the statistical dependencies, and the LDA considers these as independent features, while the relationship between the segments can yield different results. Our study also shows better performance compared to the method presented in [8] which is applied on the same subset of the TUH dataset. In fact, the proposed method automatically extracts features with no pre-defined restrictions. Besides, the proposed DL-based method is considering the changes in the pattern of the EEG signals in an epoch. Both of these specifications lead to a better performance than that of the other two methods as shown in Table 4.

Table 4: Classification results of our proposed method (over ten random runs) and the other two methods (LDA and [8]) for seizure detection task

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Mozafari et al. [8]	80.72	80.00	**81.08**	67.55
LDA on three one-sec epochs	71.12	67.10	71.44	67.98
Proposed Method	**82.00$\pm$0.63**	**85.01$\pm$0.84**	80.22$\pm$0.93	**71.69$\pm$1.01**

IV. Conclusion

This work proposed a retrospective EEG-based seizure detection algorithm that exhibited state of the art results. The dataset used in this paper included EEG signals recorded from epileptic patients with both kinds of seizures. The illustrated method was evaluated in a cross-subject scenario, which shows the proposed method's generalizability on an unseen patient. By combining BSS and DL-based techniques, our method can detect seizure periods in noisy EEG signals, without the need for hand-crafted features. Moreover, an LSTM is used to exploit its ability to process sequential data and learn time dependencies due to the non-stationary nature of EEG signals information. Results showed that the proposed method is more robust and accurate compared to previous method on the TUH dataset.

I. REFERENCES

[1] H. M. De Boer, M. Mula, and J. W. Sander, "The global burden and stigma of epilepsy," *Epilepsy & behavior,* vol. 12, no. 4, pp. 540-546, 2008.

[2] B. V. Maydell *et al.*, "Efficacy of the ketogenic diet in focal versus generalized seizures," *Pediatric neurology,* vol. 25, no. 3, pp. 208-212, 2001.

[3] L. Wang *et al.*, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy,* vol. 19, no. 6, p. 222, 2017.

[4] K. Lehnertz, "Epilepsy and nonlinear dynamics," *Journal of biological physics,* vol. 34, no. 3-4, pp. 253-266, 2008.

[5] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems,* vol. 2, no. 1-3, pp. 37-52, 1987.

[6] H. Scheffe, *The analysis of variance*. John Wiley & Sons, 1999.

[7] A. Pathak, A. Ramesh, A. Mitra, and K. Majumdar, "Automatic seizure detection by modified line length and Mahalanobis distance function," *Biomedical Signal Processing and Control,* vol. 44, pp. 279-287, 2018.

[8] M. Mozafari and S. H. Sardouie, "Automatic epileptic seizure detection in a mixed generalized and focal seizure dataset," in *2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME)*, 2019: IEEE, pp. 172-176.

[9] U. o. Freiburg. "Seizure Prediction Project Freiburg." https://epilepsy.uni-freiburg.de/ (accessed 2020).

[10] S. N. G. Bolagh and G. Clifford, "Subject selection on a Riemannian manifold for unsupervised cross-subject seizure detection," *arXiv preprint arXiv:1712.00465,* 2017.

[11] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 975-982.

[12] A. Harati, S. Lopez, I. Obeid, J. Picone, M. Jacobson, and S. Tobochnik, "The TUH EEG CORPUS: A big data resource for automated EEG interpretation," in *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2014: IEEE, pp. 1-5.

[13] M. Mozafari, F. Firouzi, and B. Farahani, "Towards IoT-enabled Multimodal Mental Stress Monitoring," in *2020 International Conference on Omni-layer Intelligent Systems (COINS)*: IEEE, pp. 1-8.

[14] S. H. Sardouie, M. B. Shamsollahi, L. Albera, and I. Merlet, "Interictal EEG noise cancellation: GEVD and DSS based approaches versus ICA and DCCA based methods," *IRBM,* vol. 36, no. 1, pp. 20-32, 2015.

[15] L. Vidyaratne, A. Glandon, M. Alam, and K. M. Iftekharuddin, "Deep recurrent neural network for seizure detection," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016: IEEE, pp. 1202-1207.

[16] J. Birjandtalab, M. Heydarzadeh, and M. Nourani, "Automated EEG-based epileptic seizure detection using deep neural networks," in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 2017: IEEE, pp. 552-555.

[17] Y. Cao, Y. Guo, H. Yu, and X. Yu, "Epileptic seizure auto-detection using deep learning method," in *2017 4th International Conference on Systems and Informatics (ICSAI)*, 2017: IEEE, pp. 1076-1081.

[18] M. T. Avcu, Z. Zhang, and D. W. S. Chan, "Seizure detection using least eeg channels by deep convolutional neural network," in *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2019: IEEE, pp. 1120-1124.

[19] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[20] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE transactions on Biomedical Engineering,* vol. 53, no. 12, pp. 2583-2587, 2006.

[21] V. Shah, M. Golmohammadi, S. Ziyabari, E. Von Weltin, I. Obeid, and J. Picone, "Optimizing channel selection for seizure detection," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2017: IEEE, pp. 1-5.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[23] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 15, no. 5, p. 056013, 2018.

[24] L. J. Wong, W. C. Headley, S. Andrews, R. M. Gerdes, and A. J. Michaels, "Clustering learned cnn features from raw i/q data for emitter identification," in *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, 2018: IEEE, pp. 26-33.

[25] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 6, no. 02, pp. 107-116, 1998.

[26] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2016: IEEE, pp. 324-328.

[27] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Emotion,* vol. 8, no. 10, pp. 355-358, 2017.

[28] A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *2013 IEEE workshop on automatic speech recognition and understanding*, 2013: IEEE, pp. 273-278.

[29] Y. Cui, S. Wang, and J. Li, "LSTM neural reordering feature for statistical machine translation," *arXiv preprint arXiv:1512.00177,* 2015.

[30] C. Wang, H. Yang, C. Bartz, and C. Meinel, "Image captioning with deep bidirectional LSTMs," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 988-997.

[31] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*, 2016: IEEE, pp. 352-359.

[32] 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, 2016.

[33] R. A. Horn, "The hadamard product," in *Proc. Symp. Appl. Math*, 1990, vol. 40, pp. 87-169.

[34] F. Chollet, "Keras: Deep learning library for theano and tensorflow," *URL: https://keras.io/k,* vol. 7, no. 8, p. T1, 2015.

[35] L. Guo, D. Rivero, J. Dorado, J. R. Rabunal, and A. Pazos, "Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks," *Journal of neuroscience methods,* vol. 191, no. 1, pp. 101-109, 2010.