Automated EEG-Based Epileptic Seizure Detection Using Deep Neural Networks

J. Birjandtalab, M. Heydarzadeh, M. Nourani
Quality of Life Technology Laboratory
The University of Texas at Dallas
Richardson, TX 75080, USA
Email: {birjandtalab, mehhey. nourani}@utdallas.edu

***Abstract*—Millions of people around the world suffer from epilepsy. It is very important to provide a method to efficiently monitor the seizures and alert the caregivers to help patients. It is proven that EEG signals are the best markers for diagnosis of the epileptic seizures. In this paper, we used the frequency domain features (normalized in-band power spectral density) to extract information from EEG signals. We applied a deep learning technique based on multilayer perceptrons to improve the accuracy of seizure detection. The results indicate that our nonlinear technique is able to efficiently and automatically detect seizure and non-seizure episodes with an F-measure accuracy of around 95%.**

***Index Terms*—Seizure detection; EEG signals; Feature extraction; Multi Layer Perceptron; Deep Neural Networks.**

I. Introduction

A. Background

Epilepsy is a common neurological disorder all around the world. This can leads to severe damages including falls and sudden unexpected death in epilepsy. The risk of death will be reduced if seizures are managed and medical help is provided when seizures happen. Approximately 30% of epileptic patients do not respond to medication. It is very critical to extract new knowledge from their physiological signals in order to improve quality of their lives [1].

An Electroencephalography (EEG) signal is able to discover any neurons misfiring or excessive neural activity which can be a sign of a neurological disorder. Epileptic seizure monitoring require extensive amount of long-term collected EEG signals. It can be used as an accurate evidence of the seizures especially when seizures happen in an invisible manner [2].

Several works have employed EEG signals to automatically detect epileptic seizure events. Recently, EPLIPSIAE project had offered a database for long-term epileptic seizure studies [3]. Basically, EEG datasets are highly unbalanced since there is a few number of seizure activities recorded among large amount of non-seizure data. Some studies restrict the non-seizure dataset to only a few minutes. Authors in [4] presented a semi-automated patient-dependent unsupervised technique. However, the size of analyzed non-seizure data is relatively small (e.g. 60 seconds before and after seizures as non-seizure data).

Using the notion of deep neural network for EEG signal analysis is quite challenging. Authors in [5] used a deep belief network on high dimensional EEG signals for seizure detection. Authors in [6] compares raw data and feature selection with deep learning based approaches in order to monitor the seizure.

B. Main Contribution

In this work, we applied a different state-of-the-art approach namely deep neural networks on the power spectral density of the EEG signals. The proposed approach proved high accuracy and it is robust against different variabilities in the power spectral density of EEG signals. Deep neural network provides a rich learning capabilities which can capture differences among seizure and non-seizure events. Since the feature space of EEG signals has a non-linear structure, highly non-linear approaches like DNN provides significantly larger set of functions to learn the seizure patterns. Moreover, it is very challenging to mathematically model complex patterns of seizure from multi-channel EEG signals. In other words, DNN makes it possible to learn those complex patterns without access to any clear mathematical modeling.

II. Methodology

The overall view of our proposed model is presented in Figure 1. Inputs of this model are multi-channel (18-23) scalp EEG signals. A frequency domain feature extraction technique is applied to transform time domain EEG signals to meaningful features. Finally, a patient-specific deep neural network based classification technique detects seizure and non-seizure events.

A. Feature Extraction

EEG signals are usually categorized to uniquely well-known bandwidths for clinical and research studies. Table I represents all normal and abnormal functions affecting each frequency band [7]. Particularly, seizure events have been observed in majority of EEG bandwidths. Frequency domain analysis for EEG-based epileptic seizure detection has been used and reported in the literature [8][9][10]. We calculate the power spectrum density of the EEG signal per window per channel and consider the results as spectral features to find the classify each window as seizure or nonseizure event. We use ten second windows with no overlap. We apply a Fourier transform to each window and group the result into five frequency bands represented in Table I, i.e. $w = \{\delta, \theta, \alpha, \beta, \gamma\}$.

The power spectrum density (PSD) of a specific channel c with sampling frequency of f_s is defined as:

$$P^c(f) = \frac{1}{f_s N} | \sum_{n=0}^{N-1} x_n^c e^{-j2\pi fn} |^2 \qquad \frac{-f_s}{2} < f < \frac{f_s}{2} \quad (1)$$

where x_n^c denotes the time domain data of channel c with N samples. P_w^c is defined to compute the PSD of channel c signal in frequency band $w = [w1, w2]$ as follows:

978-1-5090-4881-6/17 $31.00 © 2017 IEEE
DOI 10.1109/ICHI.2017.55

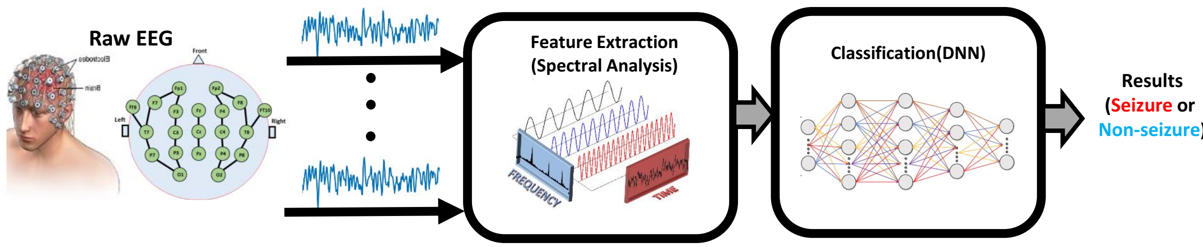

Fig. 1. The overall view of our proposed model.

TABLE I
COMPARISON OF EEG SIGNALS BANDWIDTHS AND FREQUENCIES [7].

Bandwidth	Frequency	Normal Functions	Abnormal Functions
Delta (δ)	0.1-4 Hz	artifacts, sleep, hyperventilation	structural lesion, encephalopathy, seizures
Theta (θ)	4-8 Hz	drowsiness, idling	encephalopathy
Alpha (α)	8-12 Hz	closing the eyes, inhabitation	coma, seizures
Beta (β)	12-30 Hz	effect of medication, drowsiness	drug overdose, seizures
Gamma (γ)	30-70 Hz	voluntary motor movement, learning and memory	seizures

$$P_w^c = \frac{\sum_{f=w1}^{f=w2} P^c(f)}{\sum_{f=0}^{f=\frac{f_s}{2}} P^c(f)} \tag{2}$$

where $w1$ and $w2$ are starting and end points of the frequency band. Note that for each frequency band, we consider power in band over its corresponding frequency range (i.e. $f_s = 256$ Hz). In order to have a fair comparison among all windows, we normalize the power in band value of each window by dividing it by the total spectral power of the window. Five normalized in-band power spectral density (NIPSD) features calculated for each of $M = 21$ channels generate a total of 105 features. Figure 2 illustrates an example of the extracted features of EEG signal in frequency domain for one window.

B. Deep Neural Network

A deep neural network (DNN) is a conventional multilayer perceptron (with more that three layers including input and output layers) [11][12]. Figure 3 represents an example of a DNN. Considering a DNN with L hidden layers, vector $\boldsymbol{v}^l$ shows the output vector of the l^{th} layer, where the input and output layers are represented by layers 0 and $L+1$, respectively. Then, $\boldsymbol{v}^l$ vector is calculated as follows:

$$\boldsymbol{v}^l = f(\boldsymbol{z}^l) = f(\boldsymbol{W}^l \boldsymbol{v}^{l-1} + \boldsymbol{b}^l) \qquad 0 < l < L \tag{3}$$

where $\boldsymbol{z}^{N_l} \in \mathbb{R}^{N_l}$ is the activation vector of the l^{th} layer with N_l neurons, $\boldsymbol{W}^l \in \mathbb{R}^{N_l \times N_{l-1}}$ is the weight matrix and $\boldsymbol{b}^l \in \mathbb{R}^{N_l}$ is the bias vector. Note that in Equation 3, $\boldsymbol{v}^0$ is the input feature vector, $\boldsymbol{x}$ and $f(\cdot) : \mathbb{R}^{N_l \times 1} \to \mathbb{R}^{N_l \times 1}$ is the activation function. In this work, we used sigmoid transformation as activation function as follows:

$$f(\boldsymbol{z}) = \frac{1}{(1 + e^{-z})} \tag{4}$$

The activation function needs to be selected regarding to type of classification. For example, in problems with more that two classes, the value of i^{th} output neuron shows the posterior probability of class $i \in \{1, \cdots, C\}$ as $P(c_i|\boldsymbol{x})$. We choose the class c_i only when it provide the maximum probability for an input vector $\boldsymbol{x}$. In order to consider the output neuron's values as probability, they need to meet $0 \leq \boldsymbol{v}^L \leq 1$ and $\sum_{i=1}^{C} v_i^L = 1$. This type pf normalization requirement can be met by a Softmax function as follows:

$$\boldsymbol{v}^L = P(c_i|\boldsymbol{x}) = softmax_i(\boldsymbol{z}^L) = \frac{e^{z_i^L}}{\sum_{i=1}^{C} e^{z_i^L}} \tag{5}$$

where z_i^L is the element with i^{th} index in the activation of vector $\boldsymbol{z}^L$.

C. Training & Regularization

In training phase the weight matrices $\boldsymbol{W}$, and bias vectors $\boldsymbol{b}$, are calculated per layer. Assume, M sample pairs sample as training set $\mathbb{S} = \{(\boldsymbol{x}_k, \boldsymbol{y}_k)\}$ where the vector $\boldsymbol{y}_k$ is posterior probability vector related to the k^{th} input vector $\boldsymbol{x}_k$. In this case, hard class labels are considered for $\boldsymbol{y}_k$ in training phase which means $\boldsymbol{y}_k$ has only one element equal to 1 which belongs to corresponding class of the $\boldsymbol{x}_k$. The training process is obtained by optimizing a cost function as follows:

$$J(\boldsymbol{W}, \boldsymbol{b}; \mathbb{S}) = \frac{1}{M} \sum_{k=1}^{M} J_{CE}(\boldsymbol{W}, \boldsymbol{b}; \boldsymbol{x}_k, \boldsymbol{y}_k) + \lambda |\boldsymbol{W}|_F^2 \tag{6}$$

where $J_{CE}(\boldsymbol{W}, \boldsymbol{b}; \boldsymbol{x}_k, \boldsymbol{y}_k) = -\sum_{i=1}^{C} y_i log(v_i^L)$ represents the cross entropy of corresponding posterior probability and estimated posterior from deep neural network. The parameter J shows a regularization factor to avoid over-fitting issue. Parameter $|\boldsymbol{W}_F^2|$ shows the Frobenius norm of matrix $\boldsymbol{W}$ and λ is a scalar value. In Training phase the difference between the posterior probability of training dataset and one estimated by deep neural network is optimized. This optimization is widely performed by using the notion of back propagation which is an algorithm work based on the gradient decent algorithm.

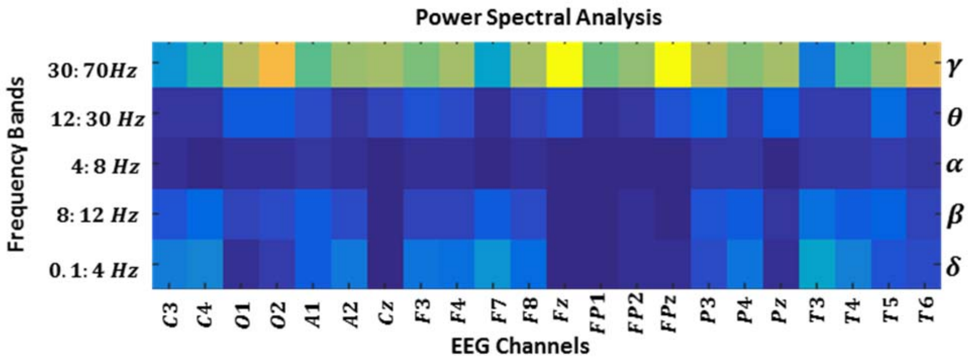

Fig. 2. An example of extracted features of EEG signal in frequency domain from one 10 seconds window from patient ID # 1.

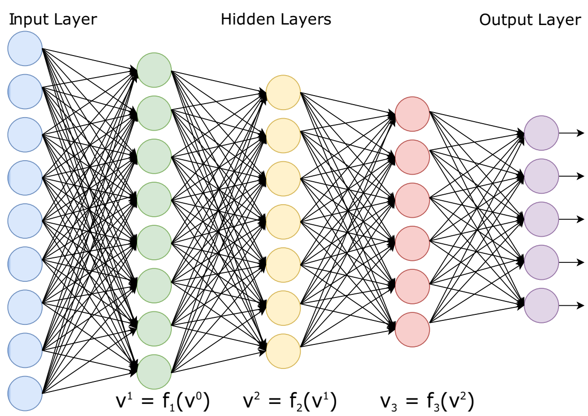

Fig. 3. An example of a multilayer perceptron neural network structure.

TABLE II
CLINICAL STUDY INFORMATION.

ID.	Age	Gender	Channels	Seizures	Seizure (MM:SS)
1	11	F	23	7	7:10
2	11	M	23	3	2:50
3	14	F	23	7	5:40
4	22	M	23	4	4:00
5	7	F	23	5	9:00
6	1.5	F	21	10	2:00
7	14.5	F	21	3	5:10
8	3.5	M	23	5	15:10
9	10	F	21	4	4:00
10	3	M	21	7	6:50
11	12	F	23	3	13:20
12	2	F	23	27	14:50
13	3	F	18	12	8:10
14	9	M	23	8	2:30
15	16	F	22	20	27:20
16	7	F	23	8	1:20
17	12	F	23	3	4:40
18	18	F	23	6	4:50
19	19	F	23	3	3:40
20	6	F	23	8	3:30
21	13	F	23	4	3:10
22	9	F	23	3	3:10
23	6	F	21	7	6:40

D. Classification

We parsed the EEG readings into sample windows with 10 seconds duration. We selected 10 seconds windows experimentally because the shortest seizures are happened about 10 seconds in this dataset. We found spectral features mentioned in Section II-A per sample window. Then, we performed a multilayer perceptrons in order to classify seizure and non-seizure events. The multilayer perceptrons provides significantly larger set of functions to learn the seizure patterns. Note that EEG signal pattern associated with seizure is different among epileptic patients. So, we personalized each DDN (patient-specific method) by training the classification model of each patient. This leads a higher average accuracy for seizure detection. Then, we performed a 10-fold cross-validation for evaluating our proposed method. Because the current dataset is grossly unbalanced we only chose 8 hours of known nonseizure data for each patient.

III. EXPERIMENTAL RESULT

A. EEG Dataset

23 epileptic patient's EEG data are investigated from the open access scalp EEG signals recorded at the Childrens Hospital Boston [13][14]. The EEG signals are collected in sequential montage which means each signal in corresponding to the difference of two neighbor electrodes. The sampling frequency is 256 Hz with 16-bit resolution. There is an annotation file in dataset which indicates beginnings and end of all the seizures which is considered as the ground truth in this study. Table II shows the clinical study information for this dataset.

B. Discussion

We tested the multilayer perceptron neural network using Weka software tool in this paper [15]. We compared the multilayer perceptron for the different number of hidden layers. Figure 4 shows the sensitivity, precision, and F-measure for one, two, and three hidden layers. Sensitivity shows the number of correctly detected seizure instances divided by all the instances detected as the seizure. Precision is the number of correctly detected seizures divided by all the

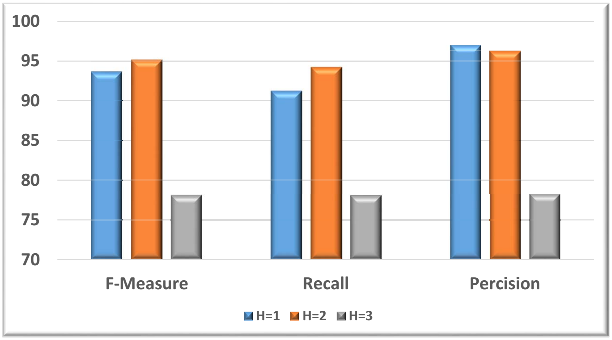

Fig. 4. Comparing results in [%] for different number of hidden layers

TABLE III
ACCURACY RESULTS IN [%].

	Method	F-measure	Precision	Sensitivity
[13]	SVM	85.12	78.74	93.34
[16]	KNN (K=3)	61.50	48.98	90.00
This Work	DNN (H=2)	95.13	94.21	96.27

seizure instances. F-measure represents a harmonic mean of precision and sensitivity values. Note that the seizure dataset is extremely unbalanced because each patient experiences seizures for only short amount of time. So, the most part of the instances belongs to the non-seizure class. It means performance metrics (i.e. sensitivity, precision and F-measure), which are not affected by the number of correctly classified non-seizure class are considered more reliable metrics in this work.

Experimental results show that the multilayer perceptron with two hidden layers provides the highest accuracy. The rationale is that by increasing the number of hidden layers the model suffers more from the over-fitting problem. This leads to a higher value of the cross-validation error. The experimental results show that our proposed method efficiently classify seizure and non-seizure events in comparison with similar works. As shown in Table III, our proposed deep neural network outperformed SVM-based [13] and KNN-based [16] seizure detection approaches.

IV. CONCLUSION

Accuracy is a crucial measure of an epileptic seizure detection system. In this work, we extracted spectral domain features by calculating normalized in-band power spectrum density of EEG signals in different frequency bands. Next, we proposed multilayer perceptrons structure as a deep neural network. Then, we specifically trained a multilayer perceptrons for each patient to discriminate seizure and non-seizure events. The multilayer perceptrons provides significantly larger set of functions to learn the seizure patterns. The results support this fact that the proposed deep neural network provides better classification accuracy (F-measure more than 95%) and is able to effectively classify seizure and non-seizure data as separable groups.

REFERENCES

[1] E. Foundation, “www.epilepsy.com/learn/impact/mortality,” 2015.
[2] W. O. Tatum IV, *Handbook of EEG interpretation*. Demos Medical Publishing, 2014.
[3] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, “Epilepsiae–a european epilepsy database,” *Computer methods and programs in biomedicine*, vol. 106, no. 3, pp. 127–138, 2012.
[4] O. Smart and M. Chen, “Semi-automated patient-specific scalp eeg seizure detection with unsupervised machine learning,” in *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2015 IEEE Conference on*. IEEE, 2015, pp. 1–7.
[5] J. Turner, A. Page, T. Mohsenin, and T. Oates, “Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection,” in *2014 AAAI Spring Symposium Series*, 2014.
[6] A. Page, J. Turner, T. Mohsenin, and T. Oates, “Comparing raw data and feature extraction for seizure detection with deep learning methods.” in *FLAIRS Conference*, 2014.
[7] W. O. Tatum, “Ellen r. grass lecture: Extraordinary eeg,” *The Neurodiagnostic Journal*, vol. 54, no. 1, pp. 3–21, 2014.
[8] S. Khanmohammadi and C.-A. Chou, “Adaptive seizure onset detection framework using a hybrid pca-csp approach,” *IEEE Journal of Biomedical and Health Informatics*, 2017.
[9] P. Janwattanapong, M. Cabrerizo, H. Rajaei, A. Pinzon-Ardila, S. Gonzalez-Arias, and M. Adjouadi, “Epileptogenic brain connectivity patterns using scalp eeg,” in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*. IEEE, 2016, pp. 1161–1165.
[10] A. Ahmadi, V. Shalchyan, and M. R. Daliri, “A new method for epileptic seizure classification in eeg using adapted wavelet packets,” in *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2017*. IEEE, 2017, pp. 1–4.
[11] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, “Deep neural nets as a method for quantitative structure–activity relationships,” *Journal of chemical information and modeling*, vol. 55, no. 2, pp. 263–274, 2015.
[12] P. E. Rauber, S. G. Fadel, A. X. Falcao, and A. C. Telea, “Visualizing the hidden activity of artificial neural networks,” *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 101–110, 2017.
[13] A. H. Shoeb, “Application of machine learning to epileptic seizure onset detection and treatment,” Ph.D. dissertation, Massachusetts Institute of Technology, 2009.
[14] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, “Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. circulation [online]. 101 (23), pp. e215–e220,” 2000.
[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: an update,” *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
[16] J. Birjandtalab, M. B. Pouyan, D. Cogan, M. Nourani, and J. Harvey, “Automated seizure detection using limited-channel eeg and non-linear dimension reduction,” *Computers in Biology and Medicine*, vol. 82, pp. 49–58, 2017.