Epileptic Seizure Detection and Anticipation using Deep Learning with Ordered Encoding of Spectrogram Features

Sameer Ranjan Sahu, Rama Krishna Sai Subrahmanyam Gorthi, Subrahmanyam Gorthi
Department of Electrical Engineering, *Indian Institute of Technology*, Tirupati, India
email: sahusr12@gmail.com, rkg@iittp.ac.in, s.gorthi@iittp.ac.in

***Abstract*—Electroencephalogram (EEG) signals of the brain play a vital role in the detection of epileptic seizures. This paper proposes a new spectrogram-based deep learning method for the detection and anticipation of epileptic seizures. Unlike the existing methods, the proposed method formulates the feature descriptor such that it retains the neighborhood order of spectrograms both in time and frequency, while significantly reducing the dimensionality of the feature descriptor. The spectrogram in each of the 18 EEG channels is constructed by dividing each EEG signal into 3 time-blocks and 19 frequency-blocks. The mean magnitude value of each of these blocks is computed and thereby compactly representing the input EEG signal by a 3D tensor of size $18 \times 19 \times 3$. This tensor descriptor is given as an input to the proposed convolution neural network for learning high-level features. Evaluations are performed on a publicly available EEG dataset of 23 patients and the results from the proposed method are compared with 9 other existing methods. Further, a five-class classification is performed using the proposed method for the anticipation of seizures. The proposed method is found to outperform the existing state-of-the-art methods both in detection and anticipation of epileptic seizures.**

***Index Terms*—Epileptic Seizures, Seizure Detection, Seizure Anticipation, Electroencephalogram (EEG) Signal, Multi-channel EEG, Spectrogram, Deep Learning, Convolutional Neural Network (CNN).**

I. Introduction

Epilepsy is a neurological disorder that can cause seizures due to unusual electrical activity in the brain. Seizures can lead to a disruption of daily life, causing physical symptoms such as attention lapse, hallucination, and convulsions. If the intensity of the seizure is high, it could be fatal for the patient. Electroencephalogram (EEG) signals are widely used to monitor, detect, and diagnose seizures. A manual inspection of EEG signals that are acquired over a long duration of time requires a lot of time and attention from the neurologist for data interpretation. Hence, an efficient automated seizure detection algorithm can save the neurologist's valuable time. Similarly, accurate anticipation of the preictal state (i.e., a state immediately before the occurrence of an actual seizure) also plays a vital role in the treatment process of seizures [1].

Seizure detection algorithms typically involve (i) prepossessing, (ii) feature extraction and (iii) classification steps [2]. Prepossessing is performed to remove unwanted noise like power line noise. Feature extraction step aims at effective representation for distinguishing different classes of EEG signals like seizure and non-seizure. The features are conventionally extracted in the time-domain, frequency-domain, or jointly in the time-frequency domain (e.g., spectrogram, scalogram). In the final step, the extracted features are fed to a classifier (e.g., support vector machine, random forest, nearest neighbor classifier) for the detection of seizures. With the recent developments in deep learning, more distinctive features than the conventional hand-crafted features are extracted along with performing a more accurate classification [3].

In the last one year, there are a few works that aim at the anticipation of epileptic seizures [2], [4]. For this purpose, the EEG signal is typically partitioned into 5 classes and is illustrated in Fig. 1. The EEG signal is broadly classified into (i) seizure, (ii) preictal, and (iii) interictal states. The seizure state is also referred to as "ictal" state. Notice that the one-hour duration before the occurrence of the seizure activity is referred to as "preictal" state. The preictal state is further sub-divided into three classes with each class having a duration of 20 minutes. Similarly, any time instance before and after four hours of the occurrence of seizure activity is referred to as "interictal" state. Thus, the 5 class-based anticipation of seizure includes interictal, three preictal (Pre S1, Pre S2, Pre S3), and ictal classes.

Several algorithms have been proposed in the literature for seizure detection using signal processing techniques in time-domain, such as key points-based detection [5] and statistical features-based detection [6]. Similarly, another class of detection methods used frequency-domain features. For example, [7] uses features derived from the Fourier transform of EEG signals for the detection of seizures. There is one more class of methods that jointly uses both time and frequency domain features. For example, [8] uses features obtained from Short Time Fourier Transform (STFT). It can be noted that the aforementioned classes of methods conventionally use those hand-crafted features directly for classifications. Furthermore, the aforementioned methods typically vectorize all the features, and thus they could not encode the neighborhood order information about time or frequency in the feature descriptor.

With the recent advancements in deep learning, there is renewed interest in developing more accurate detection algorithms for epileptic seizure detection [3], [9]–[11], as well as for seizure anticipation [2], [4]. This has resulted in coming up with methods that can extract more high-level features than

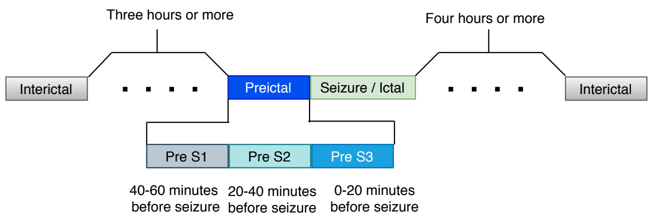

Fig. 1. Illustration of partitioning of an EEG signal into (i) seizure (ictal), (ii) preictal and (iii) interictal states for the purpose anticipation of seizures. Preictal state is further divided into three classes.

the handcrafted features [11], as well as performing a more accurate classification [3].

Despite the increase in accuracies with the application of deep learning approaches, the key challenges to be addressed are the computational complexity in terms of the dimensionality of the feature descriptor, and the deep Convolutional Neural Networks (CNN) with a large number of parameters to be learned. There have been attempts to address these issues, e.g., in [12], initial features are extracted by dividing the spectrogram into multiple blocks followed by computing the local binary patterns of spectrograms of each of those blocks, and then concatenating the histograms of all those local binary patterns. As this has resulted in a very high dimensional feature vector, they have further applied dimensionality reduction techniques to achieve a low-dimensional feature space. This, however, results in losing the critical neighborhood order information both in time and frequency.

The method proposed in this paper addresses the aforementioned issues more elegantly by dividing the feature space into blocks, computing a representative Mean Magnitude Value (MMV) for each block, and then constructing a 3D tensor of MMV. The proposed method results in not only a low-dimensional feature space but also constructing a descriptor that can efficiently encode the neighborhood order in both time and frequency.

The rest of the paper is organized as follows: Section II presents the details of the proposed method. Section III presents the evaluation results. Finally, discussion and conclusions are presented in Section IV .

II. Proposed Method

The algorithm proposed in this manuscript has two key contributions: (i) a compact spectrogram-based feature descriptor that can efficiently represent the EEG data, (ii) a CNN model that learns high-level features from the spectrogram-based feature descriptor, and performs both detection and anticipation of epileptic seizures. These two contributions are presented in detail in the following subsections.

A. Spectrogram-based Feature Descriptor

A spectrogram-based representation simultaneously captures variations in both time and frequency, and they play a vital role in the analysis of non-stationary EEG signals. A spectrogram is obtained by computing the magnitude of Short-Time Fourier Transform (STFT). In order to compute STFT of a signal, Fourier Transform (FT) is computed over a sliding window. The mathematical expression for the STFT of a given EEG signal fragment $x(t)$ is given by:

$$F_{\text{STFT}}(\tau, f) = \int_{-\infty}^{\infty} x(t)\, g(t-\tau)\, e^{-j2\pi f t}\, dt, \tag{1}$$

where $g(t)$ is a window function (e.g., Hanning window), and τ is the time index at which STFT is computed.

Since the STFT can take complex values, it's magnitude $|F_{\text{STFT}}(\tau, f)|$ spectrum is given by:

$$\sqrt{\left(Re\left\{F_{STFT}(\tau, f)\right\}\right)^2 + \left(Im\left\{F_{STFT}(\tau, f)\right\}\right)^2}, \tag{2}$$

where $Re\{.\}$ and $Im\{.\}$ represent the real and imaginary parts of $F_{\text{STFT}}(\tau, f)$. In the present work, a Hanning window of length 62 samples is used for the window function.

In order to construct a spectrogram-based feature descriptor, inspired by the recent work [12], the input EEG signal is divided into 3 blocks in the time axis, and 19 blocks in the frequency axis. While the time division is uniform, frequency division into 19 blocks is based on their physical interpretation. More specifically, EEG signals are usually interpreted in the range of $0.5-70$ Hz. This range is broadly divided into (i) δ-band of $0.5-4$ Hz, (ii) θ-band of $4-8$ Hz, (iii) α-band of $8-13$ Hz, (iv) β-band of $13-30$ Hz, and (v) γ-band of $30-70$ Hz. Furthermore, similar to [2], each of $\delta, \theta,$ and α bands are subdivided into 3 sub-bands of equal size, and each of the last two bands are subdivided into 5 sub-bands of equal size.

It is common to use statistical features like mean, standard deviation, and skewness to compactly and efficiently represent each block [6]. In this work, we propose to use Mean of the Magnitude Value (MMV) of the spectrogram computed within that block as the statistical feature because of its low computational complexity and efficient overall representation. As mentioned above, the input data has been divided into $18 \times 19 \times 3$ blocks where there are 18 channels, 19 divisions in frequency, and 3 divisions in time. Thus, the block-wise computation of MMV would result in a 3D tensor of size: $18 \times 19 \times 3$. Notice that, on the one hand, the computation of block-wise statistical metric MMV makes the descriptor less sensitive to small local variations, and on the other hand, the construction of a 3D tensor ensures encoding high-level neighborhood prior information in both time and frequency.

Fig. 2 illustrates all the steps described so far for constructing a 3D tensor of feature-descriptor. This feature-descriptor is given as an input to the proposed CNN model for further learning high-level features, and performing classification. The details of the proposed CNN architecture are presented in the following subsection.

B. Proposed CNN Model

The classification accuracies obtained from direct handcrafted features are usually less compared to classification based on high-level features learned from deep learning approaches. Convolutional Neural Networks (CNN) are well

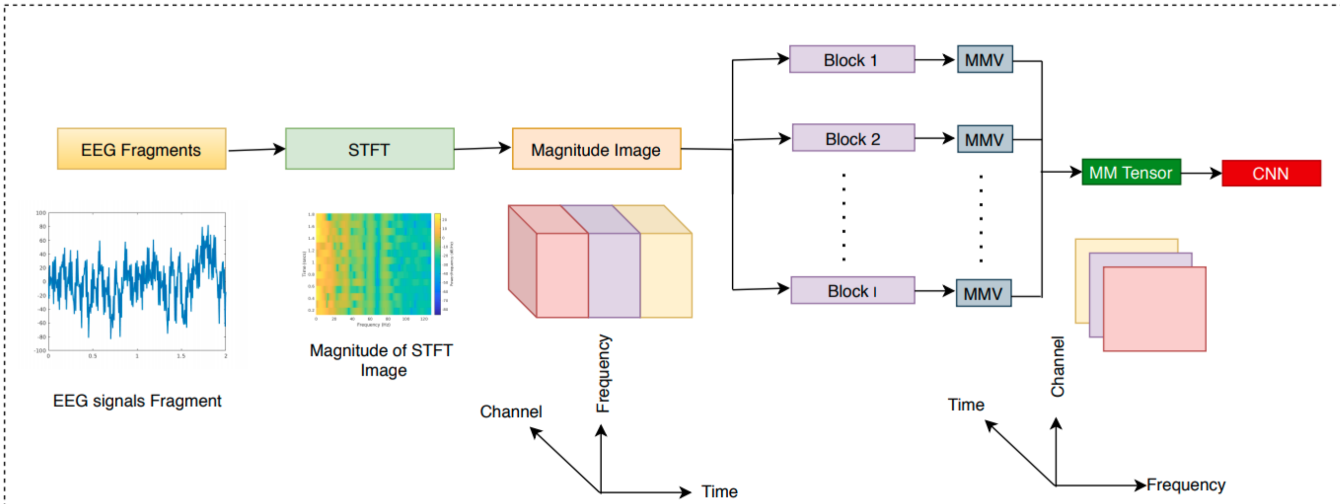

Fig. 2. Illustration of the proposed feature descriptor. Spectrogram of the input EEG fragment is computed and it's magnitude is divided into blocks as described in Section II. A 3D tensor of Mean Magnitude Value (MMV) of these blocks is computed and is given as an input to the proposed CNN model.

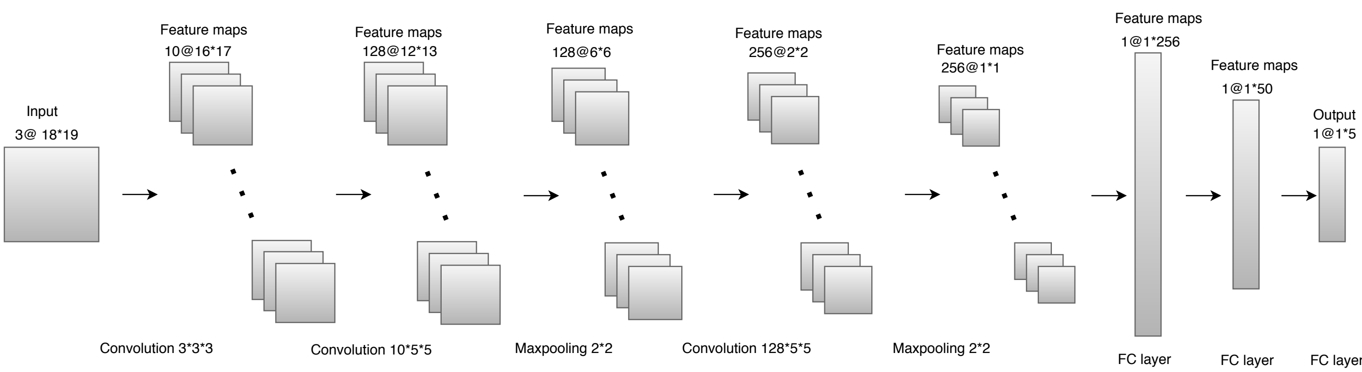

Fig. 3. Illustration of the proposed CNN architecture that learns high-level features and performs detection as well as anticipation of epileptic seizures. It takes as input a 3D tensor of mean magnitude values spectrograms of each block of the EEG signal.

known for the extraction of high-level features of a given data [13]. To this end, this paper proposes a new CNN model that learns high level features with the feature-descriptor presented in the preceding subsection as an input and also performs a classification.

Fig. 3 illustrates the proposed CNN for learning high-level features and performing seizure classification. It consists of a total of 3 convolutional layers, 2 max-pooling layers, 2 fully-connected layers, and an output soft-max layer. At the end of each convolutional layer, ReLU activation function is used. The mathematical expression for ReLu is defined as: $f(n) = \max(0, n)$. Similarly, soft-max activation function is applied at the output layer for converting the values at output neurons into probabilities. The output probability from the soft-max function for i^{th} neuron and j^{th} sample with its output value z_{ij} is given by:

$$\text{Soft-max Function: } p_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^{K} e^{z_{kj}}}, \tag{3}$$

where K is the total number of classes (or number of output neurons). Notice that the K value is 5 in the current evaluations.

The input 3D tensor of MMV is first fed to a convolutional layer that has 10 kernels of $3 \times 3 \times 3$ size each. The output of this layer is convolved with 128 kernels having dimensions of $10 \times 5 \times 5$ each. This convolution is followed by max-pooling of kernel size 2×2. The resulting feature maps are further convolved with 256 kernels of $128 \times 5 \times 5$, and is followed by max-pooling of kernel size 2×2. The resulting feature map is flattened. It is followed by 2 Fully Connected (FC) layers and an output layer.

Cross-entropy is used as the loss function, and is given by

$$\text{Cross-entropy Loss: } J = -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{K} y_{ij} \log p_{ij}, \tag{4}$$

where N is the number of samples.

TABLE I

A COMPARISON OF THE PROPOSED METHOD WITH 9 OTHER EXISTING METHODS FOR EPILEPTIC SEIZURE DETECTION.

No.	Authors	Feature Extraction Methods	Channels	Accuracy	Sensitivity	Specificity
1	Rafiudin et. al., (2011) [14]	Energy and coefficient of wavelet coefficient and statistical feature	23	80.16	NA	NA
2	Khan et. al., (2012) [15]	Relative energy and normalizedCoefficients of wavelet coefficients	NA	91.8	83.6	100
3	Kiranyaz et. al., (2014) [16]	Time, frequency, time-frequency domain and non-linear features	18	NA	89.0	94.7
4	Samiee et. al., (2015) [17]	Multivariate textural features extracted by GLCM	23	94.6	70.1	97.7
5	Zabihi et. al., (2016) [18]	Seven features extraction from intersection sequence	23	NA	89.1	94.8
6	Xinghua Yao, (2018) [19]	Using independent RNN to extract feature	17	87	87.3	86.7
7	Hengjin Ke, (2018) [9]	Using a Lightweight VGGNet to extract feature	19	98.1	98.9	97.4
8	Yuan et. at., (2018) [10]	Multi-view feature extraction using Deep learning	23	94.37	NA	NA
9	Tian et. al., (2019) [11]	Multi-domain feature extraction using CNN	23	98.3	96.7	99.1
10	**Proposed Method**	**Using spectrogram-based deep-learning**	**18**	**98.56**	**95.49**	**99.33**

TABLE II

A COMPARISON OF OVERALL ACCURACY OF THE PROPOSED METHOD WITH 2 OTHER STATE-OF-THE-ART METHODS FOR 5-CLASS CLASSIFICATION-BASED SEIZURE ANTICIPATION

No.	Authors	Accuracy
1	Cao et. al., (2019) [2]	86.25
2	Cao et. al., (2019) [4]	87.95
3	**Proposed Method**	**88.65**

III. EVALUATION RESULTS

A. Dataset and Experimental Setup

The evaluations are performed on a publicly available CHB-MIT dataset provided by Children's Hospital Boston and are available on the PhysioNet website [20]. The dataset contains EEG recordings of 23 patients with a total duration of around 884 hours. The recordings are done at 256 Hz (i.e., samples per second) with a resolution of 16 bits per channel. In the current evaluations, 18 channels are used (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, and CZ-PZ).

EEG signals are fragmented to 2 seconds duration. Since the duration of seizure activity in the EEG recordings is relatively very less compared to the non-seizure activity, a total duration of around 15 hours is considered for the evaluations out of which seizure activity is for around 3 hours. In order to further increase samples corresponding to the seizure duration, similar to the earlier works [10], a data augmentation approach is used by adding a 1-second overlap during the segmentation of EEG signals. With these modifications, the final evaluation dataset has $10682, 10678, 10736, 10700,$ and 10700 MMV samples for Ictal, Interictal, Pre S1, Pre S2, and Pre S3 states respectively. The data samples for training and testing are divided into the ratio of $4:1$.

The proposed CNN model is implemented in PyTorch [21] and is trained on Nvidia GeForce RTX 2080Ti GPU. Stochastic Gradient Descent algorithm is used for optimizing the loss function. The training procedure includes batch normalization, momentum ($\rho = 0.8$), and learning rate ($\eta = .005$). Experiments are performed with different batch sizes, and a batch size of 32 is found to give good results.

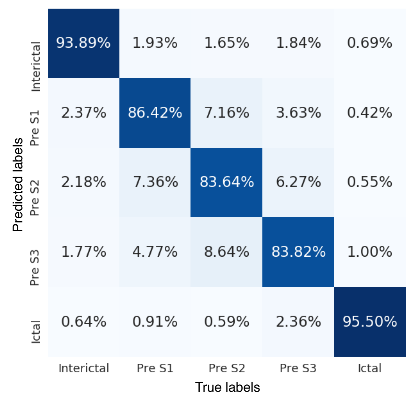

Fig. 4. Confusion matrix for five class classification for seizure anticipation using the proposed method.

B. Performance Analysis

As seizure detection is a binary classification problem, accuracy, sensitivity, and specificity are computed for each method. On the other hand, as seizure anticipation is formulated as a 5-class classification problem, a confusion matrix is computed for it.

The results from the proposed method for seizure detection are compared with 9 state-of-the-art methods. The key aspects of these 8 methods along with their respective values for the aforementioned metrics are summarized in Table I. It can be inferred from these results that the proposed method has outperformed the existing methods in terms of accuracy, and thereby bringing a fair balance between sensitivity and specificity.

For the 5-class based seizure anticipation, the results from the proposed method are compared with two recent methods [2], [4]. Fig. 4 shows the confusion matrix for the proposed method for the seizure anticipation. It can be observed from this confusion matrix that the overall performance of the proposed method is good with an average accuracy of 88.65%. There are however, some misclassifications happening within the preictal region. This is because features in these regions are very similar to each other. Finally, Table. II presents a comparison of the overall accuracies of the proposed method

TABLE III
ABLATION STUDY OF THE PROPOSED DESCRIPTOR. ACCURACY VALUES FOR EACH OF THE 5 CLASSES ARE INDICATED BY C1 TO C5.

Method	Accuracy (in %)					
	C1	C2	C3	C4	C5	Total
2D Descriptor	90.6	83.3	74.5	79.1	95.1	84.5
3D Descriptor	93.9	86.4	83.6	83.8	95.5	88.7

and the other two state-of-the-art-methods. It can be noted that the proposed method is outperforming the existing methods.

An ablation study for feature descriptor is also performed by changing the encoding methods of the spectrogram. Instead of constructing a 3D tensor MMV of size $18 \times 19 \times 3$, a 2D matrix of size 18×57 is constructed by concatenating all the MMV of time-frequency blocks together and classified with CNN having an additional max-pooling layer and a fully connected layer to bring down the dimensions smoothly to output layer; it has resulted in an accuracy of 84.5% for five-class classification instead of 88.65% accuracy with the proposed 3D tensor representation. Table III summarizes class-wise accuracy values obtained from both these descriptors. This experiment has clearly highlighted the importance of retaining the neighborhood information in time and frequency.

IV. DISCUSSION AND CONCLUSIONS

This paper presents a new spectrogram-based deep learning method for the detection and anticipation of epileptic seizures from multi-channel EEG signals. It proposes a new compact 3D tensor descriptor for representing the multi-channel EEG signal. The proposed descriptor encodes the spectrograms of multi-channel EEG signals into a 3D tensor of dimension $18 \times 19 \times 3$ through representative MMV computed for each of these time-frequency blocks of respective channels. This 3D tensor descriptor is constructed in such a way that it retains the valuable neighborhood order information both in time and frequency.

This 3D descriptor is given as an input to the proposed CNN model to learn further high-level features of the data that are eventually used for detection and anticipation of epileptic seizures. The proposed method has been evaluated for both detection and anticipation, on a publicly available EEG dataset in PhysioNet [20] of 23 patients with a duration of around 15 hours. The results for the detection of seizures are compared with 9 other existing methods. In contrast to the existing descriptors that consider histogram of the local binary pattern [12], or the complete spectrogram features with complex architectures [10], the proposed approach with time-frequency ordered 3D descriptor and with a simple CNN model has resulted in a quite discriminative feature extraction and substantial improvements in both detection and anticipation accuracies. The proposed method has outperformed several state-of-the-art approaches with an overall accuracy of 98.56% in the detection and 88.65% in 5-class classification for the anticipation of epileptic seizures.

REFERENCES

[1] W. Stacey, M. Le Van Quyen, F. Mormann, and A. Schulze-Bonhage, "What is the present-day EEG evidence for a preictal state?," *Epilepsy research*, vol. 97, no. 3, pp. 243–251, 2011.

[2] W. Hu, J. Cao, X. Lai, and J. Liu, "Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2019.

[3] 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, pp. 270–278, 2018.

[4] J. Cao, J. Zhu, W. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked cnns," *IEEE Transactions on Cognitive and Developmental Systems*, 2019.

[5] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals," *IEEE journal of biomedical and health informatics*, vol. 21, no. 4, pp. 888–896, 2016.

[6] D. Gajic, Z. Djurovic, S. Di Gennaro, and F. Gustafsson, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," *Biomedical Engineering: Applications, Basis and Communications*, vol. 26, no. 02, p. 1450021, 2014.

[7] J. Lee, J. Park, S. Yang, H. Kim, Y. S. Choi, H. J. Kim, H. W. Lee, and B.-U. Lee, "Early seizure detection by applying frequency-based algorithm derived from the principal component analysis," *Frontiers in neuroinformatics*, vol. 11, p. 52, 2017.

[8] 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)*, pp. 975–982, 2010.

[9] H. Ke, D. Chen, X. Li, Y. Tang, T. Shah, and R. Ranjan, "Towards brain big data classification: Epileptic EEG identification with a lightweight vggnet on global mic," *IEEE Access*, vol. 6, pp. 14722–14733, 2018.

[10] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 83–94, 2018.

[11] X. Tian, Z. Deng, W. Ying, K.-S. Choi, D. Wu, B. Qin, J. Wang, H. Shen, and S. Wang, "Deep multi-view feature learning for EEG-based epileptic seizure detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 1962–1972, 2019.

[12] M. Li, X. Sun, W. Chen, Y. Jiang, and T. Zhang, "Classification epileptic seizures in EEG using time-frequency image and block texture features," *IEEE Access*, vol. 8, pp. 9770–9781, 2019.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[14] N. Rafiuddin, Y. U. Khan, and O. Farooq, "Feature extraction and classification of EEG for automatic seizure detection," in *2011 International Conference on Multimedia, Signal Processing and Communication Technologies*, pp. 184–187, 2011.

[15] Y. U. Khan, N. Rafiuddin, and O. Farooq, "Automated seizure detection in scalp eeg using multiple wavelet scales," in *2012 IEEE international conference on signal processing, computing and control*, pp. 1–5, 2012.

[16] S. Kiranyaz, T. Ince, M. Zabihi, and D. Ince, "Automated patient-specific classification of long-term electroencephalography," *Journal of biomedical informatics*, vol. 49, pp. 16–31, 2014.

[17] K. Samiee, S. Kiranyaz, M. Gabbouj, and T. Saramäki, "Long-term epileptic eeg classification via 2d mapping and textural features," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7175–7185, 2015.

[18] M. Zabihi, S. Kiranyaz, A. B. Rad, A. K. Katsaggelos, M. Gabbouj, and T. Ince, "Analysis of high-dimensional phase space via poincaré section for patient-specific seizure detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 3, pp. 386–398, 2015.

[19] X. Yao, Q. Cheng, and G.-Q. Zhang, "A novel independent rnn approach to classification of seizures against non-seizures," *arXiv preprint arXiv:1903.09326*, 2019.

[20] A. Goldberger and G. et. al., "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[21] A. Paszke and G. et. al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, 2019.