International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

EEG-Based Epileptic Seizure Detection and Classification Using Machine Learning

Pratik Rajendra Saindane[1], Dr. Ayesha Siddiqui[2]

[1]MCA (Computer Application), JSPM University,Wagholi, Pune, India
Email: *pratiksaindane71[at]gmail.com*

[2]Associate Professor, Department of Computer Science, JSPM University, Pune, India

Abstract: ***In this research, there is a new technique for the identification of epileptic seizures using the strength of ML and DL algorithms in EEG signals. Epilepsy seizures are specific neurological disorders that can be characterized through unique features present in Electroencephalography (EEG). Therefore, these are highly credible with researchers. Recently, ML and DL algorithms have proven to be very effective tools in feature extraction and classification in EEG signals. There are many research works, where EEG signals have been converted into images/feature vectors in the time-frequency domain and subsequently classified. On the contrary, this study will focus on classifying EEG signal representation in time series through the use of machine learning classifiers in terms of parameter tuning along with the deep learning technique of one-dimensional convolutional neural network (1D CNN). The primary objective of this study is to find the best classifier, while at the same time emphasizing the importance of some significant parameters like sensitivity, precision, and accuracy, which play an important role in medical studies. UCI Epileptic Seizure Recognition is the dataset used for this study that contains time series data points in EEG signals. On processing, the dataset is fed to various classifiers including XGBoost, Tabnet, RF, and One- Dimensional Convolutional Neural Network (1D CNN), achieving an accuracy rate of 98%, 96%, 98% and 99%, respectively.***

Keywords: XGBoost, TabNet, Deep learning (DL), Machine Learning (ML), Random Forest (RF), Epileptic seizures, 1D CNN, data points, Time series

1. Introduction

Epilepsy ranks among the most widespread neurological disorders and is marked by a condition wherein seizures occur unpredictably and repetitively, affecting the lives of patients regardless of their age. Symptoms of seizures differ substantially, particularly in the elderly population, wherein seizures may be mild and difficult to detect accurately. For instance, children have various risk factors like genetic factors, injuries during birth, and nervous system disorders that predispose them to suffer from the disorder. Seizures, when not dealt with appropriately, cause numerous health challenges such as damage to the brain, which poses risks to patients' lives . The electroencephalogram (EEG) technique is fundamental to the identification and assessment of epileptic seizures as it captures the electrical signals of the brain on a real-time basis.

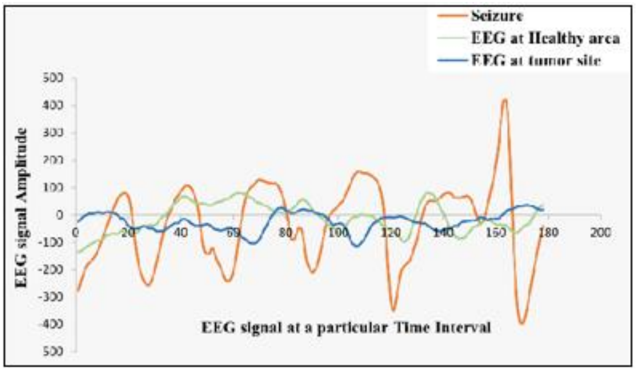
The EEG signals can provide vital information on the abnormal brain patterns which could help determine whether there has been a seizure. However, the process of analyzing these brain patterns is very cumbersome, requires a lot of training, and is prone to making mistakes. Therefore, there is a need for an automated technique to identify and categorize epileptic seizures.

There have been some advancements in the field of signal processing and machine learning, allowing for greater automation in seizure identification. The use of the technique involving the extraction of features from wavelets and nonlinearity has worked well in identifying the peculiarities in the EEG signals. Traditional machine learning techniques such as the K-Nearest Neighbor (KNN) and Random Forest have also been effective in identifying seizure patterns.

Due to the rapid evolution of deep learning technologies, new types of algorithms including CNN, LSTM models, and combinations of these algorithms have been proposed to apply in epilepsy detection. First of all, the benefit of applying deep learning algorithms in detecting seizures consists in the possibility of identifying complicated patterns present in the EEGs automatically without extracting them manually. Second, deep learning algorithms proved to be effective in the differentiation of epileptic seizures from other diseases. Moreover, publicly available databases such as those in UCI Machine Learning Repository and a study on nonlinear dynamics in the brain have resulted in the creation of highly accurate and reliable detection systems. Finally, more advanced algorithms such as the XGBoost model have increased accuracy due to the processing of large amounts of data efficiently.

However, creating seizure detection systems of high accuracy, reliability, and ease of generalization remain challenging tasks to this day. Variations in EEG patterns between different people and artifacts of electrical signals are some of the major obstacles in this process. Therefore, this research project aims at designing a seizure detection system based on machine learning techniques and deep learning approaches.

From the below figure, the X-axis represents the EEG signal at a specific time interval, while the y-axis represents the amplitude of the signal



In this study, various machine learning algorithms are implemented alongside a 1D-CNN classifier to classify the EEG time-series data. This technique enables a thorough comparative analysis of different classifiers for the purpose of understanding the most suitable methods of classifying EEG time-series data for the early detection of epileptic seizures. It is worth noting that the goal of this study goes beyond achieving the highest level of accuracy.

Rather, it aims at demonstrating the researcher's commitment to the needs of the healthcare professionals and their patients. Therefore, this study is focused on the essential metrics that would be important in making a medical diagnosis, such as sensitivity (the capability of detecting seizures) and specificity (the capability of detecting non-seizures) to classify EEG time-series data and predict the occurrence of epileptic seizures. The datasets in this study consist of data points of time series that indicate the value of the EEG signal at a specific time. In addition, the data were analyzed and classified using various techniques, including XGBoost, TabNet, Random Forest, and 1D CNN. The parameters of the classifiers were adjusted based on the dataset to achieve qualitative results.

2. Methodology

In this section, we present the proposed dataset and techniques used for detecting epileptic seizures using machine learning algorithms like XGBoost, TabNet Classifier, Random Forest Classifier with optimized hyperparameters, and 1-Dimensional Convolution Neural Network (CNN).

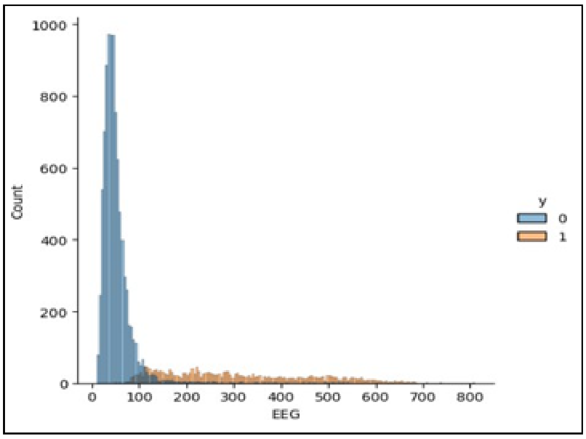
3. Dataset Description

In this work, we have used a publicly accessible UCI Epileptic Seizure Recognition [18] dataset that is obtained from a modified Bonn dataset [19]. The Bonn dataset comprises five separate folders. Each folder contains one hundred files, with each file containing brain activity recorded over 23.6 seconds. The time series data has been segmented into 4097 datapoints, with each point indicating an EEG value recorded at that instance. Therefore, there are a total of 500 subjects with 4097 data points in the Bonn dataset.

The UCI Epileptic Seizure Recognition [18] dataset divides the above-mentioned 4097 data points from the Bonn dataset into 23 segments, each containing 178 data points within 1 second. This process was carried out for all 500 subjects, producing 11500 (23 *500) data records. The purpose of this step was to provide the users with the flexibility to use the dataset for various classification tasks. The dataset consists of five classes, which can be represented as follows:
Class 1: EEG data generated by seizures
Class 2: EEG data obtained from the tumor
Class 3: EEG data obtained from healthy areas of the brain.
Class 4: EEG data acquired from patients who had closed their eyes
Class 5: EEG data obtained from patients with open eyes

All non-seizure class values (2, 3, 4, and 5) have been standardized to 0, while seizure class values were set to 1. The graphical representation of the dataset shown in Figure 2 reveals the difference between the EEG signal data contained in the dataset belonging to two classes - epileptic and non-epileptic. Binary classification is denoted by the label 'y', with y = 0 for non-epileptic cases and y = 1 for epileptic seizures.

4. Proposed Approach



This research uses an approach involving four classifiers: XGBoost, TabNet, decision tree, and 1D-Convolutional Neural Network (CNN). Prior to the modeling process, some data preprocessing techniques were conducted for the dataset. Specifically, the dataset was split into two sets: 80% training set and 20% validation set.

Figure 3 shows the architecture of the proposed methodology consisting of data processing steps, different classifiers, and performance metrics used in the approach. In the process of feature extraction, some data points were extracted from EEG signals by authors in [18]. In our work, these data points were considered as features and preprocessed to make sure that all features were on the same scale so that some features would not dominate others during the learning process. Furthermore, data points were inputted to classifiers and the performance was calculated based on some metrics as depicted in Figure 3 below.

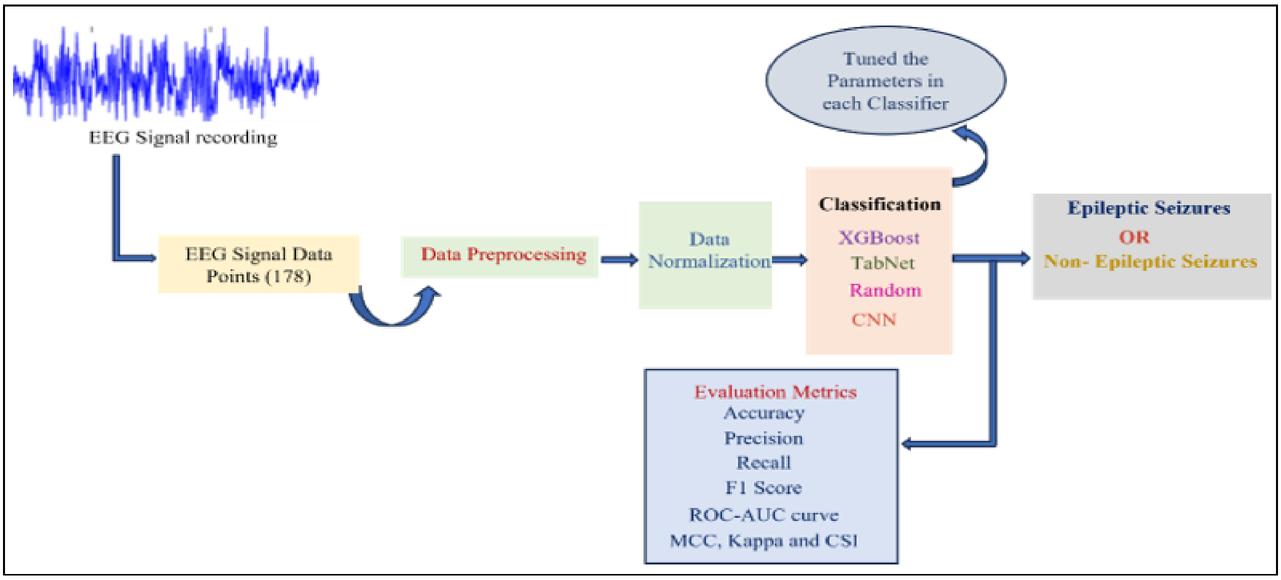


Figure 3: The Block diagram for the proposed method

The configuration of the hardware and software utilized for developing this proposed method is detailed in Table 1. We used Google Colab for both code development and execution purposes.

Table 1: Experimental Configuration

Component	*Specification*
GPU	Google Colab T4 GPU
CPU Operating System RAM Language Development Platform Libraries	AMD Ryzen7 5700u 64-bit OS, Windows 16.0 GB Python 3.8 Google Colab Keras, Pandas, Scikit Learn, Pytorch, TensorFlow.

5. Random Forest Classifier

The Random Forest is yet another machine learning algorithm that is used for ensemble learning, much like the XGBoost. In this method, many decision trees are trained, and then the final output provided by the model is the class, which is the mode of all the outputs from different decision trees. It offers excellent performance and is resistant to overfitting. The use of multiple decision trees will improve the generalization capabilities of the model towards different temporal patterns available in epileptic seizures datasets. This classifier will also help in identifying the important features at various times, enabling us to recognize important variables that lead to seizures. The parameters for the

Random Forest classifier in our case are set as follows: n_estimators are set to 1000 so that it forms a large variety of decision trees, which makes the model more stable and avoids the risks of overfitting. random state is set to 42, and the criterion parameter helps define the function used for calculating the quality of splits in a decision tree. The 'gini' criterion was chosen, which measures the probability of misclassifying any randomly selected object.

6. Convolutional Neural Network

Also, Convolutional Neural Networks can be applied for analysis of various kinds of data besides images. In case of a one-dimensional set of data, 1D convolution can be used in the network. In particular, the 1D-CNN architecture with kernel_size=2 was chosen to analyze short-term features of EEG data. Moreover, using the max pooling technique helps decrease the spatial size of the dataset, while an activation function "ReLU" helps learn complex dependencies between input values. The neural network architecture under consideration contains four convolution layers with a number of 32, 32, 64, and 128 filters accordingly.

The increase in the number of filters in the convolutional layers implies that the process involves hierarchical learning since deeper layers learn more complex and abstract features. Therefore, the architecture allows the model to capture hierarchical features at various levels of abstraction. One should note that the dropout value of 0.2 involved in the current approach indicates that it is a form of regularization. For example, the first FC layer involves 64 neurons with ReLU activation and a dropout of 0.5. The final FC layer contains one neuron with sigmoid activation function. Also, the optimizer used in the model is Adam, and the learning rate and objective function are '0.0005' and 'binary_crossentropy', respectively. Illustrates the architecture of the proposed 1D-CNN model

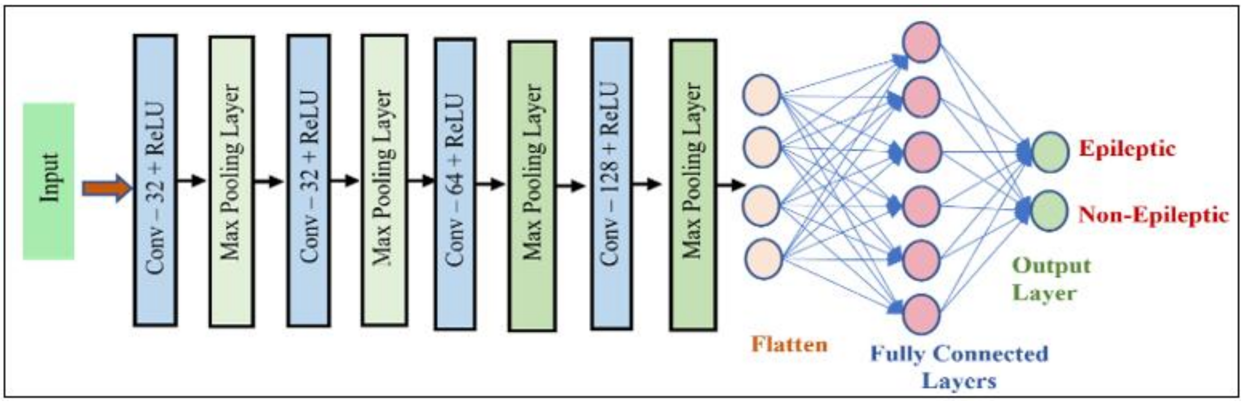


Figure 4: CNN architecture of the proposed method

The following equations below represent mathematically Sigmoid, ReLU, and Binary Cross Entropy Functions in the CNN model.

$$Sigmoid\,(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

where x is the input to the function, e is the exponential whose approximate value is 2.71828 and Sigmoid(x) is the output value whose range of values is from 0 to 1.

$$ReLU(x) = \max(0, x) \quad (2)$$

In this case, when the input value of x is greater than zero, the ReLU activation function will give the same output while when the input value is less than or equal to zero, the ReLU activation function gives output 0. It can therefore be described mathematically as a ramp function accepting only positive values.

$$L(y, \hat{y}) = -(y.log(\hat{y}) + (1 - y).log(1 - \hat{y})) \quad (3)$$

From the above equation, L(y, y^) is the binary cross entropy loss where y is the ground truth having values either 0 or 1 while y^ is the predicted value for class 1.

7. Performance Evaluation and Results

The following evaluation measures are used to calculate accuracy, precision, recall, F1 measure, CSI, MCC, and Kappa to evaluate the performance of the proposed approach for the classification between seizure and non-seizure states. The formulas for these measures are presented below:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FP} \quad (6)$$

$$F1\ Score = 2 * \frac{Precession * Recall}{(Precession + Recall)} \quad (7)$$

$$CSI = \frac{TP}{TP + FN + FP} \quad (8)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)\,(TP+FN)\,(TN+FP)\,(TN+FN)}} \quad (9)$$

TP refers to true positive, which means it correctly identifies positive.

TN is a true negative, which means correctly identifies negative.

FP is referred to as false positive which means incorrectly identified positive or Type 1 error.

FN refers to a false negative, which means wrongly identified negative or Type 2 error.

$$Cohen's\ Kappa = \frac{P_o - P_e}{1 - P_e} \quad (10)$$

Where, P_o is the relative observed agreement and P_e is the expected agreement.

8. Real-Time Application

When implementing the proposed method for practical applications, it needs to be taken into account that it will take some time to calculate the data. If the calculations take too much time, then it is unrealistic to implement them on a real-time basis. Hence, it is important to choose the right window size to make sure the computational requirements of the model match its real-time requirements. Moreover, the delay parameter needs to be considered during the implementation of the seizure detection from the EEG.

In the case when the large sliding window is utilized for the calculations, the delay parameter will be large as well. In other words, delay is defined as the time between the beginning of a seizure and its identification through the algorithm. Therefore, it can negatively impact the usefulness of the detection since the delay occurs. To avoid delay in real-time seizure detection, the 1.35 s sliding window was implemented in this study.

As it was previously pointed out, finding the eigenvalues was taking quite some time. On the contrary, feeding the Google-Net CNN algorithm with the STFT spectrum proved to be a good way to reduce the computation time required for solving the problem. This approach allowed the process to be applied in real-time scenarios as well. The processing time in this case equals to 0.02 seconds.

9. Challenges and Limitations

In our experiment, we made use of UCI epileptic seizure recognition dataset that comprised extracted data from the Bonn University dataset available in the form of .csv files as compared to the actual signal data. There is a possibility that some nuances and characteristics might be lost while extracting data from the original one. Our model is dependent on the preprocessed data, which implies that it performs well because the preprocessing techniques effectively utilize the information provided by the original EEG signal.

However, there is the risk that our model does not perform well because we did not manage to find all possible combinations of features and characteristics when tuning the parameters and thus obtaining the best possible results. There may be certain areas of the feature space where we could achieve even better results if we had performed our analysis more extensively. However, the current performance of the model does not ensure its effectiveness in real-life situations because we are working with preprocessed data.

Although the above developments are remarkable, some difficulties exist in Epileptic Seizure Detection. Some of the challenges include variability in seizure patterns. There is a significant variation in seizure patterns among patients with epilepsy. It is difficult to construct an algorithmic approach that can detect all types of seizures. In addition, the variation in EEG signals makes it hard to apply one algorithmic technique in seizure detection. A customized approach for each patient increases the effectiveness of seizure detection, yet it is difficult due to variations in seizures. Interdisciplinary approaches involving neuroscientists, physicians, and machine learning experts are necessary to solve these challenges.

10. Future Directions

Multimodal analysis using diverse input data such as EEG signals, ECG signals, accelerometry signals, and others gives more insight into the patient's health state. The use of deep learning techniques can improve seizure detection by incorporating prior knowledge from related tasks. Domain adaptation can be another promising technique in this context, which uses information from relevant domains. In particular, researchers should explore explainable AI techniques to make epileptic seizure detection models explainable. Continued technological advancements and increased interdisciplinary collaborations make the future of epileptic seizure detection promising.

11. Conclusion

In our research paper, the machine learning and deep learning algorithms were used to predict the occurrence of epileptic seizures using EEG signal data. The choice of the parameters for such classifiers as XGBoost, TabNet, Random Forest, and 1D CNN algorithm was carefully made. Our contribution is that we managed to create an algorithm that will not only make predictions about epileptic and non-epileptic seizures accurately but will take into account such measures as precision, recall, and f1 score, which have medical significance, although other researchers may neglect them. It is necessary to note that by using the precision, recall, and f1-score, we underline the importance of both the identification of positive cases and negative cases in the medical sphere.

We thus introduce a novel evaluation criteria system that includes different aspects of the model's performance. Despite similar accuracies obtained by authors who used similar classifiers, our research reveals better precision, recall, and f1 scores. Such comparison underlines the uniqueness and significance of our research results, demonstrating their substantial superiority compared to previously conducted work in the field. The identification of seizure events is extremely significant for timely interventions and individualized patient care plans and our results could be highly useful in Epileptic Seizure Detection.

The real-time detection approach proposed in the research performed remarkably well on the CHB-MIT database, with a success rate of 97.74%, sensitivity of 98.90%, 1.94% false alarms, and a delay of 9.85 seconds using the STFT spectrum. The authors concluded that this technique was a viable choice for use in the real world for seizure detection.

References

[1] U. R. Acharya, "Automated EEG-Based Screening of Epilepsy Using Machine Learning Techniques," *International Journal of Neural Systems,* vol. 22, no. 1, pp. 1250002-1–1250002-15, 2012.
[2] A. Subasi, "EEG Signal Classification Using Wavelet Feature Extraction and a Mixture of Expert Model," *Expert Systems with Applications,* vol. 32, no. 4, p. 1084–1093, 2007.
[3] N. Kannathal, "Entropies for Detection of Epilepsy in EEG," *Computer Methods and Programs in Biomedicine,* vol. 80, no. 3, p. 187–194, 2005.
[4] R. G. Andrzejak, "Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity," *Physical Review E,* vol. 64, no. 6, pp. 061907-1–061907-8, 2001.
[5] S. Ghosh-Dastidar, "A Computerized Classification of EEG Signals Using Support Vector Machine and Neural Network," *IEEE Transactions on Biomedical Engineering,* vol. 55, no. 3, p. 106–112, 2007.
[6] A. T. Tzallas, "Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis," *IEEE Transactions on Information Technology in Biomedicine,* vol. 13, no. 5, p. 703–710, 2009.
[7] R. Hussein, "Epileptic Seizure Detection Using Deep Learning Algorithms," *ACM Computing Surveys,* vol. 52, no. 6, p. 1–34, 2019.
[8] M. Sharma, "EEG-Based Epilepsy Detection Using Machine Learning Techniques," *ACM Transactions on Multimedia Computing,* vol. 16, no. 1, p. 1–24, 2020.
[9] S. Raghu, "Automated Detection of Epilepsy Using CNN Features from EEG Signals," *Computers in Biology and Medicine,* vol. 120, no. 103757, p. 103–112, 2020.
[10] T. O'Shea, "Deep Learning for EEG-Based Seizure Detection and Classification," *Neural Computing and Applications,* vol. 29, no. 5, p. 1–12, 2018.

[11] S. Roy, "Epileptic Seizure Detection Using DWT and Random Forest Classification," *Biomedical Signal Processing and Control,* vol. 49, p. 101–110, 2019.
[12] P. Fergus, "Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques," *BioMed Research International,* vol. 2015, no. 986736, p. 1–17, 2015.
[13] S. Siuly, "Classification of EEG Signals Using Sampling Technique and Least Square Support Vector Machine," *Computer Methods and Programs in Biomedicine,* vol. 104, no. 3, p. 358–367, 2011.
[14] M. Alickovic, "Automatic Classification of Epileptic EEG Signals Using Hybrid Feature Selection and Machine Learning Techniques," *Biomedical Signal Processing and Control,* vol. 45, p. 38–48, 2018.
[15] Y. Yuan, "Epileptic EEG Classification Based on Extreme Learning Machine and Nonlinear Features," *Epilepsy Research,* vol. 96, no. 1-2, p. 29–38, 2011.
[16] H. Daoud, "Efficient Epileptic Seizure Prediction Based on Deep Learning," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 13, no. 5, p. 804–813, 2019.
[17] S. Bhattacharyya, "Automated Detection of Epileptic Seizures Using Multivariate Empirical Mode Decomposition and Machine Learning," *IEEE Sensors Journal,* vol. 18, no. 20, p. 1–9, 2018.
[18] M. Shoeb, "Application of Machine Learning to Epileptic Seizure Detection," *Proceedings of the 27th International Conference on Machine Learning,* p. 975–982, 2010.
[19] J. R. Saab, "EEG-Based Epileptic Seizure Detection Using Machine Learning and Signal Processing Techniques," *Journal of Medical Systems,* vol. 45, no. 7, p. 1–14, 2021.
[20] K. Polat, "Classification of Epileptiform EEG Using a Hybrid System Based on Decision Tree Classifier and Fast Fourier Transform," *Applied Mathematics and Computation,* vol. 187, no. 2, p. 1017–1026, 2007.