Enhancing Epilepsy Care in Resource-Constrained Settings through Streamlined EEG Data Analysis

Romel Thommeykutty [1], Anwesh Roy[1] and Asha K[2] [0000-0003-0677-2784]

[1] MSc Bigdata Analytics Student, Department of Advanced Computing, School of Information Technology, St. Joseph's University, Bangalore, Karnataka, India.
romelpalam@gmail.com,anweshroy76429@gmail.com

[2] Assistant Professor, Department of Advanced Computing, School of Information Technology, St. Joseph's University, Bangalore, Karnataka, India.
ashak@sju.edu.in

Abstract. Epilepsy, a debilitating neurological disorder characterized by recurring seizures, affects millions of people around the world. This condition poses significant challenges for patients and healthcare systems. Shockingly, there are approximately 65 million cases of epilepsy globally, with varying rates of occurrence in high-income and low-middle-income nations, ranging from 45 to 82 cases per 100,000 people annually. Alarmingly, more than 90% of epilepsy cases in these areas are undiagnosed and untreated, contributing to what is known as the epileptic diagnostic gap. An electroencephalogram, or EEG, is a medical test that captures electrical activity in the brain. EEG enables medical professionals to track brain activity and identify conditions like epilepsy. In countries with low incomes, epilepsy often goes unnoticed because there is a shortage of healthcare professionals and the expenses involved in acquiring EEG recording equipment are quite high. In this paper, an analytical method has proposed for the detection of epilepsy using dataset from Nigeria. The signal data trained the model with various machine learning algorithms, applied dimensionality reduction and implemented hyperparameter tuning. The paper can bridge the gap in epilepsy diagnosis and potentially revolutionize healthcare provision by offering trustworthy and accessible diagnostic procedures.

Keywords: Machine learning, signal processing, epilepsy detection, Principal component Analysis, bigdata analytics.

1 Introduction

Epilepsy is a widespread condition that affects people all around the world. Its prevalence is not limited to Nigeria. Being one of the most populated nations in the world, India also faces substantial difficulties in the diagnosis and treatment of epilepsy. When it comes to cutting-edge medical facilities and diagnostics, India faces a serious challenge of resource limitation similar to those experienced by many other nations. Particularly in rural and underdeveloped locations, access to sophisticated diagnostic equipment for epilepsy, such as high-end EEG machines may be restricted. Furthermore, the challenges associated with EEG recordings are compounded by the inherent complexities, such as the unpredictable nature and vulnerability to discrepancies among observers. To effectively identify individuals at high risk of epilepsy, maximize the use of limited resources, and support community-based healthcare services, it is essential to establish dependable procedures that do not require expert involvement. The objective of the paper is to establish a robust framework that can be effectively utilized in challenging healthcare circumstances. This accomplished through conducting practical research on different data segmentation methods, feature extraction approaches, and implementation of machine learning algorithms. The performances of the classifiers have been analyzed, and an optimum classifier has been found. Principal compound analysis and hyperparameter tuning was used to enhance the performance of SVM and random forest classifiers. EEG relative power band was used as a feature extraction technique.

1.1 Pervasiveness of Seizure Epilepsy in Realtime

In India, the cost of epilepsy diagnosis and treatment is a significant consideration. Access to epilepsy care can be significantly improved by using low-cost or affordable diagnostic techniques and instruments, especially for groups who are economically deprived. India has a thriving telemedicine sector that could benefit from low-cost and portable epilepsy diagnosis technologies. Remote monitoring and diagnosis can help people in remote or underserved areas gain access to healthcare. Local researchers and inventors in India can be inspired by studies using datasets such as the Nigerian dataset to produce low-cost and novel solutions for epilepsy diagnosis and management. This could include the creation of low-cost EEG sensors as well as machine learning methods for early detection. Not much research has been done on the EEG dataset from Nigeria. Hence, the paper addresses not just Nigeria but areas with socio-economic factors analogous to Nigeria. The paper has implemented fine tuning techniques and dimensionality reduction techniques, along with techniques to treat imbalances in the data. This paper addresses the challenge of epilepsy detection in resource-constrained settings, where there is a shortage of healthcare professionals and limited access to expensive EEG recording equipment. It introduces a machine learning model using SVM for epilepsy detection. The model offers a potential solution to improve epilepsy care in low-income countries by leveraging affordable wireless EMOTIV headsets for data collection. The empirical study uses various machine learning algorithms, PCA for dimensionality reduction and hyperparameter tuning demonstrates the potential of advanced analysis techniques in epilepsy classification. The study also highlights the benefits of big data in epilepsy classification, as it allows for the collection of extensive patient records, including EEG data, medical histories, and medication responses. The paper and proposed study are significant because they provide a viable solution for improving epilepsy care in resource-constrained settings via faster EEG data analysis, leveraging machine learning models and big data techniques.

2. Notion of Bigdata in Classifying Epilepsy?

Big data can provide significant benefits for classifying epilepsy by offering access to a vast amount of patient data and enabling advanced analysis techniques. Here are some ways in which big data can benefit epilepsy classification. Big data allows for the collection of extensive patient records, including EEG (electroencephalogram) data, medical histories, medication responses, and more. With this rich dataset, machine learning algorithms can be trained to identify subtle patterns and markers of epilepsy, leading to a more accurate and timely diagnosis. By analyzing historical data, big data analytics can help identify early signs and predictive factors associated with epilepsy. This can lead to earlier interventions and improved outcomes for patients. Big data can facilitate the development of personalized treatment plans by considering individual patient characteristics, such as genetics, lifestyle, and responses to various medications. This can result in more effective treatment strategies and reduced side effects. Pharmaceutical companies can leverage big data to identify potential drug candidates for epilepsy treatment. Analyzing large datasets can reveal novel insights into the molecular mechanisms of epilepsy, potentially leading to the development of new medications.

IoT devices and wearables can collect continuous data from epilepsy patients, such as heart rate, sleep patterns, and movement. Big data analytics can process this information to detect seizures in real-time and alert caregivers or medical professionals. Researchers can access large-scale datasets to conduct epidemiological studies and clinical trials more efficiently. This can accelerate the discovery of new therapies and interventions for epilepsy. Through predictive analytics, big data can help healthcare systems identify high-risk individuals and provide targeted interventions, reducing hospital admissions and healthcare costs associated with epilepsy management.

3. Literature review

Numerous studies have evolved in epileptic detection using EEG signals. A research work depicts the usage of a dataset from Nigeria and applies various statistical and feature extraction methods [1]. Study used TUH EEG seizure corpus to accurately classify multiclass seizure classifications [2]. Another work uses 54-DWT mother wavelets for the analysis of EEG signals along with four machine learning classifiers [3]. The statistical features that are extracted from the EEG signal are divided into seven families and passed through various machine learning classifiers in order to classify the EEG signals as normal or abnormal in order to detect epilepsy. P-1D-CNN is basically a pyramidal, one-dimensional CNN that uses consumes less memory and has a significantly shorter detection time and can be easily implemented in real life clinical areas [4]. The dataset used here is the University of Bonn dataset, which is extensively used for detecting epilepsy. The datasets primarily consist of five sets (A to E) with 100 one-channel instances in each set. In another paper, a machine learning-based model using ANN was proposed [5]. The dataset used here is obtained from the physio-net database of pediatric patients at Children's Hospital Boston. The dataset primarily consists of EEG recordings of female patients with a minimum of 1 and a maximum of 4 seizures recorded using bipolar electrode placement. The sampling process of the EEG recording was done at 256 Hz with a 16-bit resolution. Deep learning models were used to detect epilepsy and a hardware implementation of STFT (Short-Time Fourier Transform) [6]. The deep learning techniques have been applied to the Bonn EEG dataset in order to detect epilepsy. Time-frequency analysis of EEG segments was conducted using STFT with the extraction of frequency bands, where CNN was applied with Bi-LSTM. Genetic algorithms and particle swarm optimization (PSO) were used to improve the parameters of SVM [7]. The paper made use of five sets (A–E) of single-channel EEG segments that were made available to the public. From surface EEG recordings of healthy individuals who were awake and had the eyes open or closed, respectively sets A and B were taken. Pre-surgical diagnosis EEG archives contained sets C and D of seizure-free activity while set E contained seizure activity. Detection of meaningful patterns in EEG signals were used with machine learning algorithms for timely seizure detection [8]. The dataset used is the University of Bonn dataset, which has five folders (A–E), and each folder consists of 100 single-channel EEG segments. Delta, Theta, Alpha, Beta, and Gamma were the five categories used to group the amplitude of brain waves measured at the surface of the skull. The feature engineering techniques that were applied to EEG signals include feature extraction in time and frequency, as well as time-frequency domains using the Butterworth filter, FFT, and wavelet transform. Another piece of research offers a compatibility framework that incorporates dominant channel selection and a unique method for reading XLtek EEG data and integrating local EEG signals from an epilepsy center into the CHB-MIT dataset [9]. A deep-learning model of 1D-CNN, Bi-LSTM and attention is used to test and assess the integrated datasets which contain selected channels. To understand the non-linearity of the data and down-sample the convolutional layer's output a deep-learning model with a number of layers including activation and max-pooling 1D layers, is utilized. Another paper utilizing ResNet and VGGNET, deep learning-based CNN 1D convolutional neural networks, was created [10]. Preprocessing the EEG signals to extract the required properties for categorization was the suggested method.

To automatically predict epileptic seizures, research applies deep learning to the spectral analysis of EEG signals [11]. The method entails a number of processes, including filtering, segmenting the EEG signals into short duration segments, and spectraldomain transformation. The one-dimensional Convolutional Neural Network (CNN) was used in a study to automatically detect epilepsy seizures in EEG readings [12]. The CNN model was composed of a number of hidden layers, including an embedding layer, dropout layer, 1-D convolutional layer, max pooling-1D layer, vanilla hidden layer, and single unit output layer. Positive integer values were transformed into fixed-sized vectors by the embedding layer, while the dropout layer worked to minimize over-fitting by randomly removing 20% of the neurons. A study had conducted using 318 patients' EEG signals from publicly accessible EEG datasets from Guinea-Bissau and Nigeria and achieved a high accuracy of 82.818% [13]. To identify epileptic seizures, decision tree (DT), k-nearest neighbor (KNN), naive Bayes (NB), support vector machine (SVM), and artificial neural network (ANN) classifiers were utilized [14]. A tailored convolutional neural network (CNN) for feature extraction and classification tasks was suggested and assessed. Fivefold cross-validation was used to generalize classifier performance. EEG signals were divided into various folds, and the performance of per-class classification was calculated using confusion matrix. When it came to identifying epileptic seizures from EEG signals, the suggested Convolutional Neural Network (CNN) classifier performed better than PCA-based machine learning classifiers with CNN's classification accuracy ranged from 99.50% to 100% for cases with two classes and 98.48% for cases with three classes, and 96.32% for cases with five classes.

4. Methodology

4.1 Data acquisition

The dataset used in the research paper is based on the people of Nigeria, where the data was collected using the EMOTIV headsets which record brain activity in a very detailed way with a sampling rate of 128 times per second and a high accuracy of 16-bit resolution. Several participants were asked to sit for five minutes while wearing a wireless headset. The two most important sensors were common modsense (CMS)S) and another known as driven right leg (DRL). The CMS acts as a reference point, getting accurate data, and the DRL sensor is used to cancel unwanted noise in the background. The electrodes were placed at anterofrontal (AF3, AF4, F3, F4, F7, F8), frontocentral (FC5, F6), occipital (O1, O2), parietal (P7, P8) and temporal sites (T7, T8).

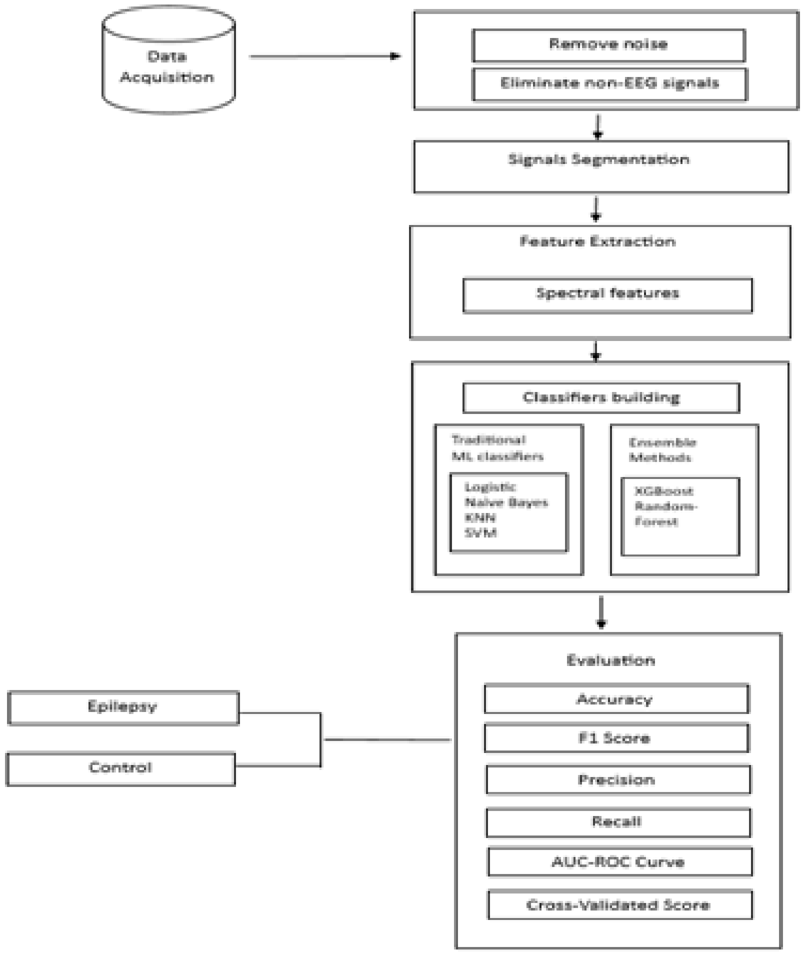

Fig. 1: Proposed framework of Epilepsy Detection

The data was collected for five minutes when the participants were sitting with their eyes closed. Anything unusual was also recorded and stored in a specific format. But the data recorded was not that accurate. Simple methods to smooth out the errors were performed which included the removal of unwanted data points. The quality of the recorded signals was tested, and the bad signals was removed with immediate effect. The head movements are measured through a gyroscope, and if the head movements were improper, then the recording was deleted. Using the process of wavelet analysis, three different wavelets were used to analyze brain activity. Each wavelet was used to divide the brain signals into seven distinct frequency ranges, and the ranges were named delta, theta, alpha, beta, and gamma. It helped analyze brain activity at different time stamps. Figure 1 detailing the proposed framework used in the epilepsy detection.

4.2 Pre-processing

EEG data is loaded from multiple files into Python. Specifically, EEG recordings from subjects with epilepsy ("Epilepsy") and control subjects ("Control") are extracted from separate files. These EEG recordings are saved as data frames. The non-EEG channels from the EEG data are eliminated. The stratified sampling to ensure that the class distribution is maintained. SMOTE is applied to the training data to address class imbalances. SMOTE generates synthetic samples for the minority class (Epilepsy) to balance the class distribution. This step helps improve the model's performance, especially when dealing with imbalanced datasets.

4.3 Signal Segmentation

The data frames containing the EEG data are then converted into MNE-Python Epochs objects. MNE-Python is a library designed for efficient analysis of EEG/MEG data, offering various tools for working with EEG data. The data is transposed and used to generate an MNE Raw Array object. Subsequently, the data is filtered within a specific frequency range (0.1 Hz to 45 Hz). A bandpass filter is applied to the data to retain frequencies between 0.1 Hz and 45 Hz. Fixed-length epochs of 5 seconds duration with a 1-second overlap are created. Any problematic epochs are removed from the dataset. We then combine the epochs obtained from different EEG recordings for both the epilepsy and control groups. This step involves combining data from various subjects and sessions into a single dataset, which is a common approach to EEG analysis. The labels are assigned to the data based on whether the EEG recordings belong to epilepsy or control groups.

4.4 Feature Extraction

Feature extraction involves extracting EEG power band features from EEG epochs. These features capture the relative power distribution across different frequency bands at each epoch. The resulting feature matrix can be used as input for machine learning models to classify and analyze EEG data based on these extracted features. We define specific frequency bands of interest, such as delta, theta, alpha, sigma, beta, and gamma, along with the corresponding frequency ranges (in Hz). The computing of Power Spectral Density (PSD) of the EEG data in each frequency band is fixed using the Welch method. PSD represents the distribution of signal power across different frequencies. The computed PSD values are normalized by dividing them by the sum of PSD values across all frequencies for each epoch. This normalization step ensures that the features are relative power measures. For each frequency band, the mean PSD values are computed across all channels. This mean PSD value represents the relative power in that frequency band for each epoch. The concatenation of the computed features are framed as a feature vector for each epoch.

4.5 Classification

Classification is done using traditional machine learning classifiers and ensemble methods. In order to enhance the model's performance, fine-tuning its hyperparameters by performing Grid Search with Cross-Validation was done, and its performance was evaluated on the testing data. The classifiers classified EEG data into two classes: Epilepsy and Control, based on the extracted EEG power band features. The performance metrics provide insights into how well the model can discriminate between these two classes.

5. Results and implementation

The metadata for all the patients that included subject.id, recorded period, startTime, session.id, first_condition, remarks, group, and csv. file. In this research, we have divided the participants into two distinct groups, namely ‘epilepsy’ and ‘control’ subjects, based on the ‘Group’ activity present in the metadata file. EEG data has stored for each group separately. The task of feature extraction where the important features are kept, and removed the rest making it more clear and easier to work.

EEG data is collected in raw form, so we use the MNE library in Python to streamline it in a more meaningful way. Filtering is applied to focus on the required frequency ranges. Then the EEG data is divided into short, fixed length segments known as epochs, each lasting 5 seconds with a 1-second overlap. Later, we remove the low-quality epochs to ensure data integrity. Finally, the ones that are well organized and cleaned has received. The list of channel names from the epoch objects is 'AF3', 'AF4', 'F3', 'F4', 'F7', 'F8', 'FC5', 'FC6', 'O1', 'O2', 'P7', 'P8', 'T7', 'T8'. We also use ‘convertDF2MNE’ to reformat the EEG data of both groups (epilepsy and control). The format is a must because MNE requires a special format for detailed analysis of EEG data. The content has move on to label ‘0’ for epilepsy and ‘1’ for control. Six frequency bands are defined, i.e., alpha, beta, theta, delta, sigma, and gamma. To calculate the Power Spectral Density (PSD) of EEG data within a frequency range moved from 0.5 Hz to 45 Hz. PSD is used to measure the power present at different levels of frequency in EEG signal data. A machine learning model has performed PCA to reduce the dimension while retaining 98% of the explained variance. 25 components are retained after PCA is performed.

Table 1: F1- Scores of Controls and Epilepsy

Classifier	Control	Epilepsy
SVM	0.72	0.84

Control, in this context, refers to a healthy person who does not have epilepsy, whereas epilepsy is a chronic neurological disorder that is characterized by recurrent seizures. So here we have taken the f1 score as a measure to distinguish between epilepsy and control classes using the SVM classifier. F1-Score for "Epilepsy" is 0.84, indicating strong performance in correctly identifying epilepsy cases, while the F1-Score for

"Control" is 0.72, suggesting good accuracy in classifying non-epilepsy cases. Both scores balance precision and recall in their respective classes.

6. Performance Analysis

The overall process uses 80% of the data for training and 20% for testing. At first, Random Forest classifier is used in the model, and for hyperparameter tuning, Grid Search with cross validation are performed. The best hyperparameter that best fits the model. The model achieved a cross validated accuracy of 0.827 (approx.). The applied SVM, and the hyperparameters found for this algorithm are 'C'(100), 'gamma'('scale') and 'kernel'('rbf') and the model gave a cross validated accuracy of 82.08% approximately. The SMOTE technique to balance the data is used since the dataset was imbalanced. Table 2 provides the details on the comparison between various ML classifiers performed.

Table 2: Comparison between different ML classifiers

		0	1	Accuracy
SVM	precision	0.74	0.83	0.8
	recall	0.7	0.86	
	f1 score	0.72	0.84	
Logistic Regression	precision	0.69	0.82	0.77
	recall	0.7	0.81	
	f1 score	0.7	0.81	
Naïve Bayes	precision	0.57	0.73	0.68
	recall	0.54	0.76	
	f1 score	0.56	0.75	
XGBoost	precision	0.72	0.84	0.79
	recall	0.74	0.83	
	f1 score	0.73	0.83	
Random Forest	precision	0.73	0.83	0.79
	recall	0.71	0.84	
	f1 score	0.72	0.83	

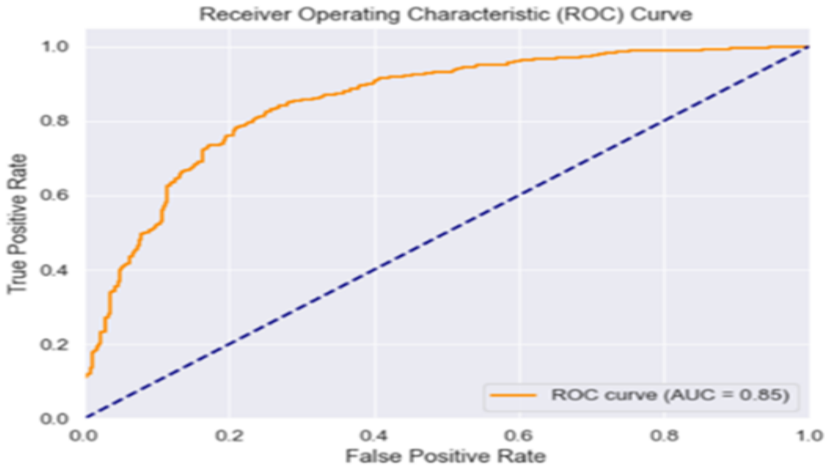

Fig. 2: ROC curve

Figure 2 indicates the ROC curve of SVM and achieved the AUC of 0.85. As compared to RF, SVM is able to more distinctly classify between two classes with a high degree of accuracy.

Table 3: Confusion Matrix for SVM

Confusion Matrix		Predicted	
		Class 0	Class1
Actual	Class0	277	118
	Class1	95	561

Table 3 detailing the confusion matrix, that demonstrates the SVM classifier properly categorizes 277 healthy individuals and 95 patients with illnesses. Additionally, it designated 118 individuals with disease as having no disease and 561 patients without disease.

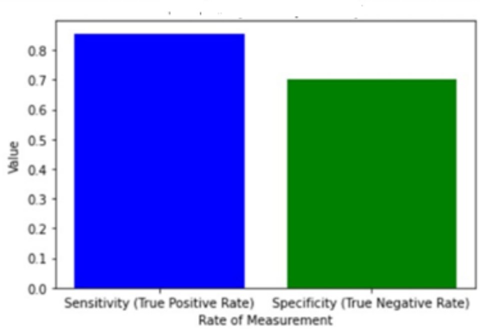

Fig.3: True positive and True Negative Rate of Classification

Sensitivity (TPR) is the proportion of people with epilepsy who are correctly classified as having epilepsy. Specificity (TNR) is the proportion of people without epilepsy who are correctly classified as not having epilepsy. Sensitivity and specificity are important metrics for evaluating the performance of a diagnostic test. A test with high sensitivity is good at identifying people with the condition, while a test with high specificity is good at identifying people without the condition. Figure 3 pointing to the measure of true positive rate achieved 86% and true negative achieved 70%. Based on the evaluation, it's been identified that, SVM classifier have high sensitivity and specificity. This means that the classifier is very good at detecting epilepsy and avoiding false positives. The classifier is more sensitive than specific. This means that the classifier is more likely to correctly identify people with epilepsy than it is to correctly identify people without epilepsy.

7. Conclusion

Epilepsy is the second most common neurological disorder after stroke, according to a report from the World Health Organization. People with epilepsy account for about 1% of the world's population. Due to the uncertainty of ictal, epilepsy patients need to take long-term medication, which brings great harm to their bodies and minds. Therefore, the analysis and mining of epilepsy features are helpful to achieve early warning of epileptic seizures, which can not only ensure the personal safety of patients but also remind them to choose emergency antiepileptic drugs. Epilepsy detection is crucial, and diagnosing it has been hard for the underprivileged. It was found that various machine learning methods have been used to detect epilepsy, but not much work has been done using consumer-grade devices. With affordable devices and seamless detection methods, this could become a reality for such sections of society.

References

[1] Nida Alyas, David P. Hastings and Abbas Mehrabidavoodabadi, "A Comparative Analysis of Classifier Performance for Epileptic Seizure Detection Using EEG Signals", International Conference on Pattern Recognition Applications and Methods, DOI:10.5220/0011631600003411.
[2] S. Roy, U. Asif, J. Tang and S. Harrer, "Seizure Type Classification Using EEG Signals and Machine Learning: Setting a Benchmark", 2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB),DOI: 10.1109/SPMB50085.2020.9353642
[3] Mardini, Wail et al. "Enhanced Detection of Epileptic Seizure Using EEG Signals in Combination with Machine Learning Classifiers." IEEE Access 8 (2020): 24046-24055.

[4] Ullah, Ihsan et al. “An Automated System for Epilepsy Detection using EEG Brain Signals based on Deep Learning Approach.” Expert Syst. Applications DOI:10.1016/j.eswa.2018.04.021.
[5] Satarupa Chakrabarti, Aditi Choudhary, Avantika Singh, Raunak Kumar, “A machine learning based method to detect epilepsy”, Inter-national Journal of Information Technology 2018, DOI:10.1007/s41870-018-0088-1.
[6] Sai Manohar Beeraka, Abhash Kumar, Mustafa Sameer, Sanchita Ghosh, Bharat Gupta. “Accuracy Enhancement of Epileptic Seizure Detection: A Deep Learning Approach with Hardware Realization of STFT”, Circuits Systems and Signal Processing 2022, DOI:10.1007/s00034-021-01789-4.
[7] Abdulhamit Subasi, Jasmin Kevric, M. Abdullah Canbaz, “Epilep-tic seizure detection using hybrid machine learning methods”, Neural Computing and Applications 2019, DOI:10.1007/s00521-017-3003-y.
[8]Marzieh Savadkoohi , Timothy Oladunni , Lara Thompson, A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal, doi: 10.1016/j.bbe.2020.07.004.
[9] Mariam K. Alharthi, Kawthar M. Moria, Daniyal M. Alghazzawi, Haythum O. Tayeb, Epileptic Disorder Detection of Seizures Using EEG Signals, doi:10.3390/s22176592.
[10] Dhouha Sagga, Amira Echtioui, Rafik Khemakhem, Mohamed Ghorbel, Epileptic Seizure Detection using EEG Signals based on 1D-CNN Approach, doi: 10.1109/STA50679.2020.9329321.
[11] Kuldeep Singh, Jyoteesh Malhotra, Deep learning based smart health monitoring for automated prediction of epileptic seizures using spectral analysis of scalp EEG, doi: 10.1007/S13246-021-01052-9.
[12] Manoj Kaushik, Divyanshu Singh, Malay Kishore-Dutta, Carlos M. Travieso, A deep learning approach for epilepsy seizure detection using EEG signals, doi:10.18845/tm.v35i8.6461.
[13] R Shiva Shankar, CH Raminaidu, VV Sivarama Raju and J Rajanikanth, Detection of Epilepsy based on EEG Signals using PCA with ANN Model, Doi 10.1088/1742-6596/2070/1/012145.
[14] Joy, A.D., Sarkar, S., & Azad, A.K. (2022). Detection of Epileptic Seizures from EEG Signals Using Machine Learning Classifiers. Bangladesh Journal of Medical Physics.DOI:10.3329/bjmp.v15i1.63560.