e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 6, June 2026 | DOI 10.17148/IMRJR.2026.030603

EEG-Based Epileptic Seizure Detection Using Machine Learning Techniques

Jangala Rajesh[1] Dr Tirumala Paruchuri[2]

PG Scholar, Department of Computer Science and Engineering, Malla Reddy Engineering College(A) Hyderabad, Telangana[1]

Professor, Department of Computer Science and Engineering, Malla Reddy Engineering College(A) Hyderabad, Telangana[2]

Abstract: Epileptic seizures are sudden neurological events that require timely detection and intervention. Electroencephalogram (EEG) signals, due to their non-invasive nature, are widely used for seizure analysis. This paper presents a complete automated seizure detection framework using the Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) Scalp EEG Database. The system integrates preprocessing, feature extraction, data balancing, data augmentation, and classification into a robust pipeline. Key features were derived using power spectral density, wavelet decomposition, entropy, and statistical measures. To overcome class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed. Multiple machine learning classifiers, including Support Vector Machine (SVM), Random Forest (RF), Extra Trees (ET), Logistic Regression (LR), XG Boost (XGB), and Light GBM (LGBM) were evaluated. Performance was assessed using 5-fold stratified cross-validation, demonstrating that the integration of spectral and nonlinear EEG features with advanced machine learning methods significantly improves seizure detection accuracy.

Keywords: Epileptic seizure detection; Electroencephalogram (EEG); Machine learning, Feature extraction; Children's Hospital Boston- Massachusetts Institute of Technology (CHB-MIT).

I. INTRODUCTION

Epilepsy is a chronic neurological condition characterized by recurrent, unprovoked seizures that result from sudden, abnormal, and excessive neuronal discharges in the brain. Epilepsy is the most common chronic brain disease and affects people of all ages. More than 50 million people worldwide have epilepsy; nearly 80% of them live in low- and middle-income countries. An estimated 70% of people with epilepsy could be seizure free if properly diagnosed and treated [1]. However, about three quarters of people with epilepsy in low-income countries do not get the treatment they need, and this rises to 90% in some countries. In many such countries, many health professionals do not have the training to recognize, diagnose and treat epilepsy

In this study, we propose a complete EEG-based seizure detection framework that addresses these challenges. The system includes preprocessing, multi-domain feature extraction, and class balancing using SMOTE. Multiple machine learning classifiers, including SVM, RF, ET, LR, XGB, and LGBM are evaluated to determine their effectiveness in seizure detection. Using the CHB-MIT EEG database and 5-fold cross-validation, our method achieves high accuracy and efficiency, demonstrating strong potential for reliable real-time clinical use.

II. LITERATURE REVIEW

EEG based seizure detection has received significant research attention due to its ability to capture brain activity directly, yet the complex and non-stationary nature of EEG signals makes reliable classification difficult. Deep learning approaches have been widely explored to address this problem. Dişli et al. [5] proposed a Continuous Wavelet Transform based depthwise CNN model that extracts time–frequency representations from EEG signals to identify seizure patterns. Although the method achieved strong detection performance, it required large training data and high computational cost, limiting practical deployment.To better understand the overall research landscape, Saadoon et al. [6] reviewed machine learning and deep learning techniques used for seizure prediction. Their study highlighted persistent issues such as data imbalance, inter-patient variability and limited interpretability of models, indicating that accuracy alone is insufficient for clinical usage. Ensemble learning has therefore been explored as an alternative. Al-Adhaileh et al. [7] introduced a multi-model classification approach that combines multiple classifiers to improve robustness and reduce individual model bias, demonstrating more stable performance on EEG datasets.

III.PROPOSED METHODOLOGY

Epileptic seizure detection from EEG signals is a complex process that requires careful handling of high-dimensional, noisy brainwave data. Raw EEG recordings often contain artifacts caused by eye movements, muscle activity, or environmental interference, which can obscure the subtle patterns indicative of seizures. To develop a reliable,

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 6, June 2026 | DOI 10.17148/IMRJR.2026.030603

reproducible, and generalizable detection model, a structured workflow combining signal processing, data augmentation, and machine learning was implemented. This section outlines the materials used, the dataset characteristics, and the systematic procedures employed to preprocess, transform, and analyze the EEG signals for accurate seizure detection.

A. Proposed Methodology

The proposed seizure detection framework, as illustrated in Figure 1, is a multi-stage pipeline designed to convert preprocessed EEG signals into actionable diagnostic insights. Each stage of the proposed seizure detection pipeline serves a specific purpose in ensuring accurate and robust classification of EEG signals.

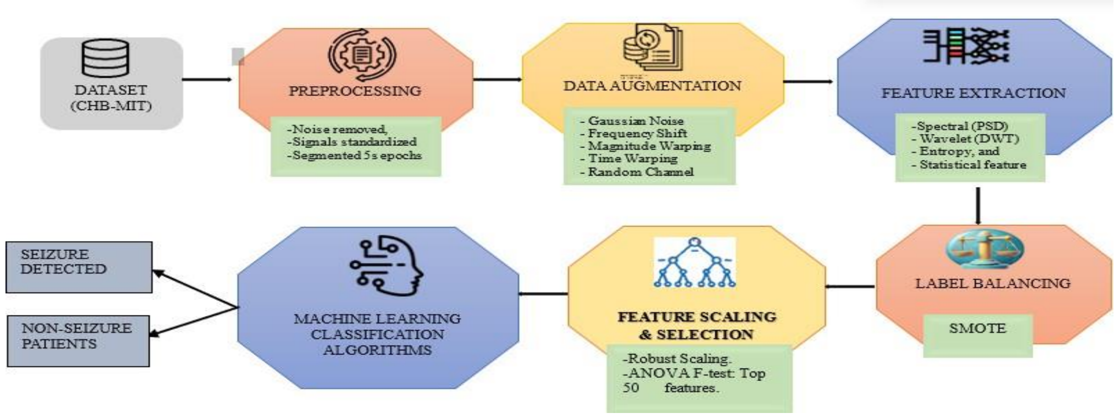


Fig.1. The proposed framework for EEG-based epileptic seizure detection

B. Dataset and Organization

The study uses the CHB-MIT Scalp EEG dataset [11] as described by Table 1, which contains long continuous brain signal recordings from children suffering from epilepsy. The data was recorded in a real hospital environment, so it includes normal activity, sleep patterns, and noise along with seizures. All signals were recorded at a sampling rate of 256 Hz and stored in EDF format. To make the analysis manageable, the recordings were divided into fixed 5-second segments. After segmentation, a total of 29,197 epochs were obtained. Among them, 4,785 segments represent seizure activity and 24,412 represent normal brain activity.

Table 1 Characteristics of CHB-MIT dataset used in the study

Parameter	*Description*
Sampling Frequency	256 Hz
Recording Format	European Data Format (.edf)
Number of Subjects	23
Total EEG files	42 files containing both seizure (ictal) and non-seizure (interictal) data
Segmentation Strategy	EEG signals segmented into 5 second non-overlapping epochs
Total number of Epochs	29,197 epochs
Seizure epochs	4,785 epochs (16.4% in total)
Non-Seizure epochs	24,412 epochs (83.6% of total)

A. Signal Preprocessing

Preprocessing is critical for removing noise, eliminating irrelevant artifacts, and standardizing EEG signals before feature extraction. Several techniques are widely employed in the preprocessing of epileptic EEG signals, including

a) *Re-referencing*: If all EEG channels were available, Common Average Referencing (CAR) was used. This means the average signal of all channels was subtracted from each channel to reduce shared noise and make the brain signals clearer.

b) *Band-pass filtering:* A zero-phase finite impulse response (FIR) filter was applied between 0.5 and 100Hz*:*

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 6, June 2026 | DOI 10.17148/IMRJR.2026.030603

$xf(t) = F^{-1}(H(f). F\{x(t)\})$ (1) where $x(t)$ is the original EEG

Mathematically, if $x(t)$is the EEG signal, the notch-filtered signal $x_n(t)$can be represented as:$x_n(t) = F^{-1}H_{\text{notch}}(f) \cdot Fx(t)$

C. Data Augmentation

To enhance the robustness and generalization of seizure detection models, various data augmentation techniques were applied to the EEG signals. EEG signals are highly variable due to factors such as patient-specific brain activity, electrode placement, environmental noise, and physiological changes. This variability often makes models prone to overfitting and reduces their ability to generalize across different subjects or recording conditions. Data augmentation addresses this challenge by artificially expanding the training dataset and introducing controlled variability, thereby improving the model's ability to recognize seizure patterns under diverse conditions.

1) Frequency Shift Augmentation

This method involves slightly shifting the frequency components of EEG signals to simulate variations due to electrode placement or individual differences [13]. It helps in making models invariant to such frequency shifts.
Mathematically it is denoted by:

$$x_{\text{aug}}(t) = \mathscr{F}^{-1}(\mathscr{F}\{x(t)\} \cdot e^{j\Delta ft})$$

where F denotes the Fourier transform, F^{-1}denotes the inverse Fourier transform, Δf is the frequency shift applied to the signal.

2) Magnitude Warping

Magnitude warping involves applying a smooth scaling function to the amplitude of EEG signals. This technique simulates variations in signal amplitude due to recording conditions or physiological changes.
Mathematically it is denoted by:

$$x_{\text{aug}}(t) = \alpha(t) \cdot x(t)$$

where $\alpha(t)$ is a scaling factor determined by cubic spline interpolation of randomly generated nodes, $x(t)$ is the original EEG signal at time t.

3) Time Warping

Time warping involves stretching or compressing the time axis of EEG signals to simulate variations in signal duration. This method helps in making models invariant to such temporal variations [14].
Mathematically it is denoted by:

$$x_{\text{aug}}(t) = x(t')$$

where $t'$is the warped time index, determined by a smooth warping function, $x_{\text{aug}}(t)$is the original EEG signal at time t

4) Random Channel Dropping

This technique involves randomly dropping or masking certain EEG channels during training to simulate electrode failures or noisy channels. It helps in making models robust to incomplete or noisy data [15].

IV PERFORMANCE MEASURES

The proposed models were evaluated using 5-fold stratified cross-validation, preserving the seizure-to-non-seizure ratio in each fold. SMOTE was applied only to the training data to address class imbalance, while feature scaling and selection were restricted to the training set to prevent information leakage. Models were trained on the processed training folds and evaluated on the corresponding validation folds, with final results averaged across all five folds.

A. Confusion Matrix Analysis

A confusion matrix is used to evaluate the performance of a classification model by comparing predicted labels with actual labels. It consists of four components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). From these values, important performance metrics are cases are actually seizures. Recall (Sensitivity) shows how well the model detects actual seizure cases. Specificity measures how accurately the model identifies non-seizure cases [27].

B. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve evaluates binary classification performance by plotting the True Positive Rate (Sensitivity) against the False Positive Rate (FPR = 1 − Specificity) across varying decision thresholds. Each point on the curve represents a different classification threshold, illustrating the trade-off between seizure detection sensitivity and false alarm rate. The Area Under the Curve (AUC) provides a single summary measure of classifier performance:

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 6, June 2026 | DOI 10.17148/IMRJR.2026.030603

AUC = 1.0 indicates perfect discrimination, AUC = 0.5 represent classification.
Higher AUC values indicate stronger discriminative capability and robustness. In EEG-based seizure detection, a high AUC confirms that the model reliably distinguishes seizure activity from normal EEG signals across different thresholds [28].

V RESULTS AND DISCUSSION

This section presents the experimental results of the proposed EEG-based epileptic seizure detection framework using the CHB-MIT dataset. The performance of multiple machine learning classifiers is evaluated using 5-fold stratified cross-validation, with emphasis on accuracy, robustness, and the impact of data augmentation and class imbalance handling. The results are analysed to identify the most effective models for reliable seizure detection.

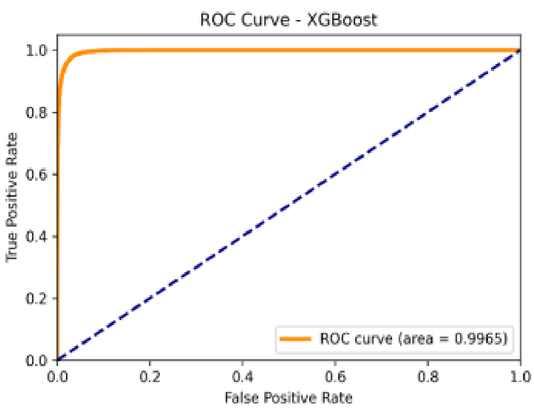
Random Channel Dropping and Magnitude Warping introduce moderate reductions in performance across all classifiers. However, ensemble-based models, particularly boosting methods, remain comparatively stable, demonstrating their robustness to controlled signal variations.

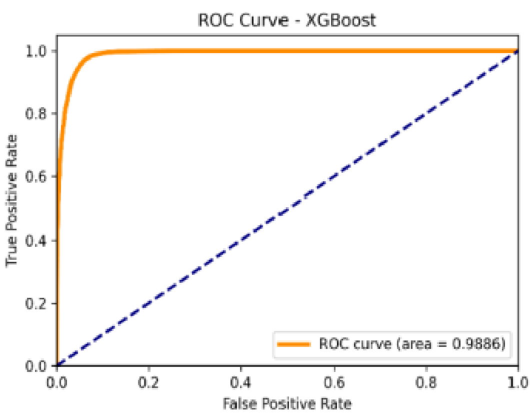
Table 2 Classification Performance (%) of All Classifiers Under Data AugmentationTechniques

Augmentation	Metric	LR	SVM	ET	RF	LGBM	XGB
Frequency Shift	Accuracy	91.44	94.14	93.18	95.47	97.13	**97.19**
	Precision	88.52	91.20	90.46	93.27	95.77	**95.83**
	Recall	95.23	97.70	96.55	98.01	98.61	**98.66**
	Specificity	87.65	90.58	89.82	92.93	95.64	**95.71**
Gaussian Noise	Accuracy	91.50	94.20	93.21	95.25	96.69	**96.74**
	Precision	88.77	91.38	90.64	93.33	95.35	**95.34**
	Recall	95.01	97.61	96.36	97.46	98.17	**98.28**
	Specificity	87.98	90.80	90.05	93.03	95.21	**95.20**
Random Channel Dropping	Accuracy	88.38	91.47	90.19	94.03	96.04	**96.14**
	Precision	84.08	87.26	86.98	91.55	94.54	**94.61**
	Recall	94.69	97.12	94.53	97.02	97.73	**97.86**
	Specificity	82.07	85.82	85.85	91.05	94.36	**94.43**
Magnitude Warping	Accuracy	90.93	93.31	91.72	94.35	95.33	**95.47**
	Precision	88.16	89.97	89.27	92.00	93.03	**93.08**
	Recall	94.56	97.51	94.83	97.14	98.01	**98.25**
	Specificity	87.31	89.12	88.60	91.55	92.66	**92.69**
Time Warping	Accuracy	74.64	81.78	79.22	81.55	83.23	**83.73**
	Precision	70.73	77.63	74.43	76.64	79.30	**79.65**
	Recall	84.07	89.29	89.01	90.76	89.93	**90.60**
	Specificity	65.21	74.27	69.42	72.33	76.52	**76.85**

A. *ROC Curve Analysis of XGBoost*

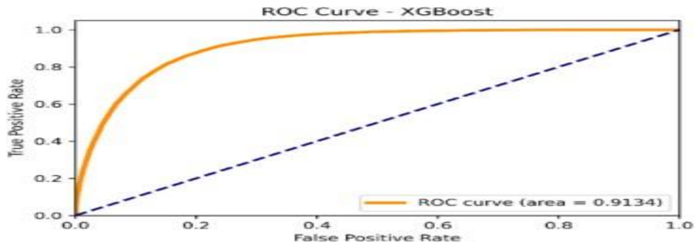
Figure 2(a–e) shows the ROC curves of the XGBoost classifier under different augmentation strategies. The baseline model achieves the highest AUC (≈0.98), while Frequency Shift and Gaussian Noise maintain strong discrimination (AUC ≈0.97). Random Channel Dropping and Magnitude Warping show moderate degradation, and Time Warping yields the





lowest AUC (≈0.84), indicating sensitivity to temporal distortion.

(a) **Frequency shift augmentation** **(b) Gaussian noise augmentation**



(c) Time warping augmentation

VI CONCLUSIONS AND FUTURE WORK

This study presented a comprehensive machine learning–based framework for epileptic seizure detection using EEG signals from the CHB-MIT Scalp EEG Database. The proposed pipeline integrated signal preprocessing, multi-domain feature extraction, class balancing using SMOTE, feature selection, and rigorous evaluation through 5-fold stratified cross-validation. A total of 29,197 EEG epochs were processed, with seizure and non-seizure samples balanced to ensure unbiased training. Among the tested classifiers, XG Boost achieved the highest accuracy of 97.95%, closely followed by Light GBM at 97.85%. The very low cross-validation variance further demonstrated the stability and reliability of the framework.

These results indicate that gradient boosting methods are highly effective for EEG-based seizure detection, providing both robustness and computational efficiency. The proposed framework shows promise for real-time applications and can potentially be adapted for deployment in portable or wearable EEG monitoring systems.

Despite the high performance achieved, several avenues remain open for future research and improvement:

REFERENCES

[1]. World Health Organization, “Epilepsy: Fact sheet,” WHO, Feb. 7, 2025.
[2]. M. K. Siddiqui and M. A. Khan, “A review of epileptic seizure detection using machine learning techniques,” *Brain Informatics*, vol. 7, no. 1, pp. 1–12, 2024.
[3]. Y. Roy, H. Banville, I. Albuquerque, J. Gramfort, T. Falk, and J. C. Pal, “Deep learning-based electroencephalography analysis: a systematic review,” *J. Neural Eng.*, vol. 16, no. 5, p. 051001, 2014.
[4]. J. P. Carvajal-Dossman, F. M. Vargas, and L. A. Pineda, “Retraining and evaluation of machine learning and deep learning models for seizure detection from EEGs,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–12, 2025.
[5]. Dişli, F., Gedikpınar, M., Fırat, H., Şengür, A., Güldemir, H., & Koundal,
D. (2025). *Epilepsy Diagnosis from EEG Signals Using Continuous Wavelet Transform-Based Depthwise Convolutional Neural Network Model.* Diagnostics, 15(1), 84.
[6]. Saadoon, Y. A., Khalil, M., & Battikh, D. (2025). *Machine and Deep Learning-Based Seizure Prediction: A Scoping Review on the Use of Temporal and Spectral Features.* Applied Sciences, 15(11), 6279.
[7]. Al-Adhaileh, M. H., Ahmad, A., Alharbi, F., Alarfaj, R., Dhopeshwarkar, R., & Aldhyani, T. H. H. (2025). *Diagnosis of Epileptic Seizure Neurological Condition Using EEG Signal: A Multi-Model Algorithm.* Frontiers in Medicine, 12, 1577474.
[8]. Jonna, S. T., & Natarajan, K. (2025). *Interpretable EEG-Based Seizure Prediction Using a Hybrid Wavelet CNN-LSTM Model with SHAP.* MM Engineering and Applied Physics.
[9]. Tan, S., Tang, Z., He, Q., Li, Y., Cai, Y., Zhang, J., Fan, D., & Guo, Z.
(2025). *Automatic Detection and Prediction of Epileptic EEG Signals Based on Nonlinear Dynamics and Deep Learning: A Review.* Frontiers in Neuroscience, 19, 1630664.
[10]. D. Thapa and R. Rai, “FREQ-EER: A novel frequency-driven ensemble framework for emotion recognition and classification of EEG signals,” *Applied Sciences*, vol. 15, no. 19, p. 10671, 2025.
[11]. A. Shoeb, "CHB-MIT Scalp EEG Database," PhysioNet, 2009.
[12]. Liao, C., Zhao, S., Wang, X., Zhang, J., Liao, Y., & Wu, X. (2025). EEG Data Augmentation Method Based on the Gaussian Mixture Model. *Mathematics*, 13(5), 729.
[13]. Rommel, C., Paillard, J., Moreau, T., & Gramfort, A. (2022). Data augmentation for learning predictive models on EEG: a systematic comparison.
[14]. Iglesias, G., et al. (2023). Data Augmentation techniques in time series domain. *Neural Computing and Applications*.
[15]. Saeed, A., et al. (2024). Learning from Heterogeneous EEG Signals with Differentiable Channel Reordering.
[16]. N. Ji, et al., "EEG Signals Feature Extraction Based on DWT and EMD,"*Frontiers in Neuroscience*, vol. 13, p. 104, 2024.
[17]. M. Dastgoshadeh, et al., "Detection of Epileptic Seizures through EEG Signals Using Entropy-Based Features and Ensemble Learning," *Frontiers in Human Neuroscience*, vol. 16, p. 1084061, 2023.