International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 30, No. 7, July, 2025, pp. 1915-1922
Journal homepage: http://iieta.org/journals/isi

Multiclass Classification of Epileptic Seizure Using Machine Learning

Kavita Sultanpure*, Jayashree Bagade, Deepali Joshi, Preeti Bailke, Chaitali Shewale, Poonam Pawar, Deepali Jadhav

Information Technology Department, Vishwakarma Institute of Technology, Pune 411037, India

Corresponding Author Email: kavita.sultanpure1@vit.edu

https://doi.org/10.18280/isi.300723

Received: 10 June 2025
Revised: 14 July 2025
Accepted: 20 July 2025
Available online: 31 July 2025

Keywords:
epilepsy seizure, electroencephalogram (EEG) signals, power spectral density, wavelet transform, machine learning, hyperparameter tuning, Random Forest, seizure types

ABSTRACT

Epileptic seizures are neurological disorders instigated by sudden, uncontrolled electrical disturbances and activities in the brain, leading to changes in behavior, movements, or feelings, and, in some cases, loss of consciousness. Electroencephalogram (EEG) signals are utilized in the medical field to diagnose epileptic seizures. For effective management and treatment of patients, accurate and timely detection of these seizures is crucial. The paper presents a robust machine learning system to do multiclass classification of epileptic seizures using EEG data. The study uses datasets from Bonn University, which include five different sets representing various brain states of healthy and epileptic individuals. After preprocessing and normalizing the data, the features are extracted using techniques like power spectral density (PSD) and wavelet transforms. Various classification algorithms like Decision Tree, Random Forest, Naïve Bayes, and Support Vector Machine were evaluated through extensive hyperparameter tuning and cross-validation. The Random Forest model emerged as the best classifier, achieving a significant accuracy of 89% in classifying the data into five classes, showing its effectiveness in distinguishing between different classes of seizures. This approach shows significant promise for enhancing the accuracy of epilepsy classification and optimizing treatment strategies.

1. INTRODUCTION

Epilepsy is a recognized global health issue as per the Statistics of World Health Organization (WHO), affecting millions of people worldwide [1]. Unexpected spikes in uncontrollable brain electrical activity are known as epilepsy seizures. These electrical disturbances cause temporary changes in a person's behavior, movements, feelings, or consciousness. These seizures can significantly impact the individual's quality of life with epilepsy, making effective diagnosis and management crucial [2]. Traditional methods for diagnosing epilepsy rely heavily on visual inspection of electroencephalogram (EEG) recordings by trained neurologists, a process that is time-consuming and subjective. Given the critical need for reliable diagnostic tools, there has been growing interest in leveraging machine learning (ML) techniques to classify and predict epileptic seizures from EEG data [3].

The complexity of EEG signals, with their high dimensionality and variability, poses a significant challenge for accurate and timely diagnosis. EEG signals are intricate, non-linear, and often covered by noise, which makes manual analysis a difficult task. This complexity necessitates the use of sophisticated computational methods to discern patterns and anomalies that may indicate the onset of a seizure. Machine learning presents a promising solution to this challenge as it has the ability to process and analyze large datasets [4].

Machine learning algorithms can analyze large amounts of EEG data more efficiently than human experts, potentially offering more consistent and objective assessments. By training on labelled datasets, these algorithms can learn to distinguish between different types of brain activity, including various forms of seizures. This automated analysis leads to more accurate and reliable diagnoses by reducing the likelihood of human error and speeding the diagnostic process [5].

The aim of this paper is to resolve the problem of multiclassification of epilepsy seizures by means of machine learning. Unlike binary classification, which only distinguishes between seizure and non-seizure states, multiclass classification can identify different types of seizures. This is crucial for tailoring treatment plans to the specific needs of patients, as different types of seizures may require different therapeutic approaches. By utilizing advanced ML algorithms and feature extraction techniques, we seek to develop a model that can accurately classify different types of epileptic seizures.

The project involves the use of a comprehensive dataset from the University of Bonn, which includes EEG recordings from various seizure types. To remove noise and standardize the signals, the data is pre-processed. Features are then extracted using methods such as power spectral density (PSD) and wavelet transforms, which help in capturing the essential characteristics of the EEG signals. Machine learning models such as Decision Tree, Random Forest, Navie Bayes, and

Support Vector Machine (SVM) classifiers use these features as inputs.

The goal is to improve the diagnostic process, offering a tool that aids neurologists in making more informed decisions and ultimately enhancing patient care. By providing a robust and accurate classification system, this research aims to contribute to the broader field of medical diagnostics, paving the way for more advanced and automated healthcare solutions. Through rigorous evaluation and validation, we strive to demonstrate the efficacy of our approach and its potential impact on the management of epilepsy.

2. LITERATURE SURVEY AND METHODOLOGY

Dual-tree complex wavelet transform (DTCWT) is a novel automatic method for classifying epileptic seizures from EEG signals. It extracts features from the signals and divides them into seven categories of seizures [3]. The method demonstrated better accuracy than other approaches on the TUH EEG Seizure Corpus (TUSZ) ver.1.5.2 dataset. It highlights the limitations of previous methods like FFT and STFT, which lack time-domain resolution, and emphasizes DTCWT's effectiveness in providing smoother EEG signal representations. The study also stresses the importance of patient-wise evaluation for robustness and suggests future enhancements incorporating clinical features and demographic data. This work represents a significant advancement in automatic seizure classification, aiding neurologists in epilepsy diagnosis and treatment.

The study by Oliva and Rosa [2] presents a unique method for the identification of epilepsy utilising binary as well as multiclass classifiers, with feature extraction from bispectrogram, spectrogram, and power spectrum. The BP-MLP and SMO_Pol algorithms achieved 100% and 98% accuracy for binary and multiclass classification, respectively, demonstrating the great accuracy of the classifiers. Statistical tests and evaluation based on confusion matrices showed competitive results compared to related works. The multitaper method for feature extraction, novel combinations of machine learning and feature extraction techniques, and thorough assessment methodologies are among the primary contributions. The study demonstrates competitive performance in both binary and multiclass classification compared to related works.

The authors [3] delve into the classification of seizure types using EEG signals, employing a dataset from Temple University Hospital Seizure Corpus. They explore various feature extraction methods, including MFCC, Hjorth Descriptor, and ICA, to enhance classification accuracy. Four classes of normal EEG signals are included in the study: generalized non-specific seizures (GNSZ), focal non-specific seizures (FNSZ), and tonic-clonic seizures (TCSZ). By employing Support Vector Machine (SVM) classification, they evaluate different combinations of extracted features and identify the most effective scenario, achieving high levels of sensitivity, specificity, and accuracy in seizure detection.

The authors conduct a comprehensive literature review on epilepsy, focusing on the challenges patients face and the reliance on EEG signals for diagnosis [4]. They emphasize the difficulty and time-consuming nature of manually detecting seizures in EEG data and propose the use of automatic detection frameworks aided by machine learning classifiers. The paper explores various preprocessing techniques, feature extraction methods, and classification procedures used in seizure detection. It also covers the growing importance of accurately identifying seizures, the possible application of machine learning methods, and the trend of incorporating IoT frameworks for remote patient monitoring. The study underscores the need for further research to enhance seizure detection techniques and addresses the challenges and future directions in EEG-based seizure detection.

The authors suggest an integrated method for the three-class categorization of EEG data into normal patients, intermittent epilepsy, and continuous epilepsy that combines a gradient boosting machine (GBM), Symlet wavelet processing, and a grid search optimizer (GSO) [6]. The technique breaks down EEG signals into five sub-bands using fourth-order Symlet wavelets, from which statistical features are taken out for categorization. The proposed SW-GBM-GSO system demonstrated superior classification accuracy and effectiveness compared to support vector machines (SVM) and random forest classifiers. The goal of the project is to maximize detection rates and create a smart mobile application for real-time epilepsy monitoring and diagnosis. It highlights the use of multiple performance indices to assess classification efficacy.

In order to classify EEG data into interictal and ictal states, the article covers many machine learning-based techniques. It highlights the challenges posed by the non-stationary and non-linear nature of EEG signals. Prior research has utilized various machine learning algorithms to interpret and classify these dynamic biomedical signals, aiming for accurate seizure detection [6]. Despite advancements, obtaining complete and discriminative features remains complex. The approach suggested by this study uses both fuzzy-based and traditional machine learning techniques to extract the most unique characteristics from epileptic EEG data. The findings validate the effectiveness of the suggested approach in the context of epileptic seizure detection by showing that Fuzzy Rough Nearest Neighbor (FRNN) and K-Nearest Neighbor (KNN) attain sensitivity, good classification accuracy, and specificity on benchmark datasets.

In order to address the crucial problem of epileptic seizure detection, the authors [7] examine the efficacy of many machine learning techniques for feature reduction, both with and without Principal Components Analysis (PCA). It is found that Random Forest achieves the highest accuracy of 97% with PCA, and both KNN and RF reach 99% accuracy without PCA, highlighting their robustness. The results demonstrate that PCA reduces computational times but also slightly decreases accuracy. This comparative analysis confirms that Random Forest and KNN are highly effective for epileptic seizure prediction.

The paper explores feature selection for differentiating normal and epileptic EEG data using an optimization approach based on a chaotic version of the firefly algorithm [8]. The study introduces the Improved Chaotic Firefly Algorithm (ICFFA) with a Joint Logistic-Tent Map (JLTM) to enhance the balance between exploration and exploitation in multi-objective optimization. By integrating ICFFA with Multi-Class Support Vector Machine (MSVM), the framework achieves high classification accuracy. According to experimental results, ICFFA with JLTM outperforms other cutting-edge techniques in classification performance, with an accuracy of 99.63% for normal data and 98.10% for epileptic data.

The authors of the work [9] discuss a range of signal processing approaches, including wavelet features, spectral characteristics, statistical parameters, and intrinsic mode functions (IMF), in order to predict epileptic seizures from EEG and ECG signals. Making use of information from the EPILEPSIAE database and Physiobank, the study analyzed features from different EEG electrodes, finding that the left hemisphere (T3, F3, F7) showed higher predictive potential compared to the right hemisphere (T4, F4). According to the analysis, the most useful indicators for predicting the occurrence of seizures were low-frequency IMF PSD regions and IMF spectral peaks, which could be predicted 20–30 minutes in advance.

The paper by Mahjoub et al. [10] proposes a new method for automatic seizure detection using EEG signals, addressing issues of subjectivity and time consumption in visual analysis. Support vector machines (SVMs) are used in classification. Evaluations on a publicly available database reveal high accuracy, sensitivity, and specificity, with MEMD providing the best results, although TQWT and raw data approaches also offer competitive performance with lower computational costs. This method demonstrates promise as an efficient alternative for automatic seizure detection.

The work by Imah and Widodo [11] examines the application of machine learning techniques with emphasis on classification and feature extraction for automatic epileptic seizure diagnosis from EEG data. Using Wavelet Transform (WT) and Principal Component Analysis for feature extraction, the study examines the Random Forest, SVM, Backpropagation, and Generalized Relevance Learning Vector Quantization techniques. The EEG dataset from the University of Bonn, containing signals from healthy and epileptic subjects, was used to classify five classes of brain activity. The results indicate that PCA is the best feature extraction method, with a testing time of less than 0.1 seconds and an accuracy, precision, and recall of 0.9866. GRLVQ performs better than the other classifiers. This demonstrates the effectiveness of GRLVQ combined with PCA in recognizing EEG patterns associated with epilepsy.

Using the Epileptic Seizure dataset from the UCI machine learning repository, the study [12] investigates the application of supervised machine learning and deep learning models for epileptic seizure classification. The results show that ANN achieves a higher accuracy of 98.26% compared to XGBoost, indicating its potential for enhanced epileptic seizure detection performance. The study concludes that the ANN model significantly outperforms traditional classifiers, making it a promising tool for early and accurate seizure detection.

The paper by Wang et al. [13] investigates multichannel EEG signals for epilepsy preictal state prediction, addressing the critical need to predict seizures before they occur. Wavelet packet decomposition (WPD) statistical features, including subband energy ratio, Shannon entropy, norm entropy, and logarithmic energy entropy, are used in the study to extract data. Feature learning and classification are done using a Random Forest (RF) classifier. The WPF+RF algorithm achieves an average classification accuracy of 84.8%, as shown by experimental results on the CHB-MIT EEG database, outperforming traditional statistical features and classifiers like SVM, LDA, and K-NN. This research highlights the potential of WPD-based features and RF in enhancing preictal state prediction accuracy.

The work by Srinivasan et al. [14] stands out by employing Approximate Entropy (ApEn) as a feature for classification using Artificial Neural Networks (ANNs). ApEn is a statistical measure that quantifies the regularity or predictability of time-series data, making it particularly suitable for analyzing the inherent complexity of EEG signals during seizures.

The author Chen et al. [15] of this paper illustrated that the use of ML techniques to analyze the EEG signals can increase the reliability of the seizure detection process. Another work applied the WTSVM on EEG signal, which yielded high sensitivity and specificity of the technique. More recent studies focused on applying deep learning algorithms for the processing of EEG signals and improved feature extraction and classification. Certain research incorporated video event analysis with other methods; others used ECG correlations as a method, which further supports the usage of multi-modal approaches. These works signify the relative efficacy of individual approaches; however, the application of multiple data inputs into a single framework, by use of ensemble methods as suggested in this paper, is expected to exhibit better diagnostic precision.

Epilepsy, which affects approximately 50 million individuals globally, can be identified through several methods with one of the widely used methods being based on EEG. Previous methods of ML for EEG signals used conventional features extracted by humans and were less efficient. Recent developments in deep learning (DL) have integrated categorization and feature extraction into the diagnostic procedure, improving diagnosis accuracy. Comparative analysis of the ML and DL Model, with regards to LSTM network has identified that greater validation accuracy of epileptic seizure can be achieved through use of LSTM that is classified under DL [16]. The use of DL approaches represents a significant advancement in the creation of precise algorithms for automated diagnosis and classification of epileptic seizures, despite the inherent difficulties in interpreting EEG information.

The authors [17] consider the problem of automatically identifying epileptic seizures in EEG data, which can be complicated due to the presence of noise or artifacts. They have developed an automated procedure that can identify seizures and classify them using a range of machine learning algorithms. Discretization was performed during the pre-processing stage and models were developed and tested based on the 70%-30% division of the dataset since the latter was imbalanced. Performance metrics aimed at F1 score in contrast to accuracy. The findings revealed that Random Forest yielded the highest F1 score, with a score of 0.943 for the total primary throughput and the accuracy of 0.977 which is the optimum number for the model to predict and classify the epileptic seizures in the EEG signals.

In the research [18], the authors focus on the possibility of positive changes in the healthcare field concerning patients and organizational aspects as endowed by AI. They describe the application and benefits of using AI for large medical databases for diagnosing diseases and finding adequate treatments. However, achieving complete AI inclusion in human roles is still a process that will last a number of decades due to several issues. This paper highlights the potential of utilizing AI in healthcare alongside pointing out its limitations, while also suggesting that more studies should be done to determine the effectiveness of the technology on health sectors. AI has generally been seen as the catalyst for upcoming advancements in the healthcare industry.

Epileptic seizure detection plays a crucial role in the effective treatment and management of epilepsy. Accurate and timely classification of seizure types can significantly enhance patient outcomes. This study focuses on developing a robust machine learning framework to classify different brain states using EEG data from the Bonn University dataset. Preprocessing the data, extracting features, and applying different machine learning classifiers are all part of the methodology.

2.1 Data collection

The first step involves the collection of EEG signal data, which is stored in five distinct datasets labeled as Sets A, B, C, D, and E. Each set holds 100 text files, with each file representing a different EEG recording, each with a duration of 23.6 seconds [19]. These datasets are crucial as they provide the foundation upon which the classification models will be built and evaluated [3]. The subsets represent different brain states:

I. Set_A(Z): Surface EEG recordings from healthy volunteers with eyes open.
II. Set_B(O): Surface EEG recordings from healthy volunteers with eyes closed.
III. Set_C(N): Intracranial EEG recordings from epilepsy patients during seizure-free intervals in the hippocampal formation of the opposite hemisphere.
IV. Set_D(F): Intracranial EEG recordings from epilepsy patients during seizure-free intervals in the epileptogenic zone.
V. Set_E (S): Intracranial EEG recordings from epilepsy patients during seizures.

2.2 Data loading and preprocessing

To begin the analysis, it is essential to verify the presence of files within each dataset folder. This verification is done by listing the contents of each folder to ensure no files are missing. Subsequently, a custom function load_data_from_folder is used to read the data from each file and load it into numpy arrays. This function is designed to handle various file extensions, such as the .TXT extension used in Set C. After loading the data, the shapes of the datasets are printed to confirm successful loading and to facilitate debugging.

2.3 Data normalization

Normalization is a critical step to ensure that the data has a consistent scale, which improves the performance of machine learning algorithms. Using StandardScaler from sklearn. preprocessing, the data is normalized so that every feature has a standard deviation of one and a mean of zero. This process helps in mitigating the effects of varying scales and units in the dataset.

$$X_{normalized} = \frac{X - \mu}{\sigma}$$

2.4 Feature extraction

Feature extraction is a crucial step in our methodology, as it transforms raw EEG signals [20] into a set of meaningful features that can be used for classification. For this project, we used power spectral density (PSD) and Wavelet Transform to extract features from the EEG signals. Below, we provide a detailed explanation of each method, along with necessary formulas and derivations.

2.5 Model training and hyperparameter tuning

The pre-processed dataset was divided into training and testing subsets, maintaining an 80-20 split ratio to ensure robust model evaluation. Hyperparameter tuning was conducted using GridSearchCV with cross-validation. This technique exhaustively searches through a specified parameter grid to identify the optimal combination of hyperparameters, ensuring the best model performance based on cross-validation metrics.

Key performance indicators like accuracy, precision, recall, and F1 score were used to assess each classifier's performance on the testing subset. The model that performed the best was the classifier that showed the highest accuracy on the test set.

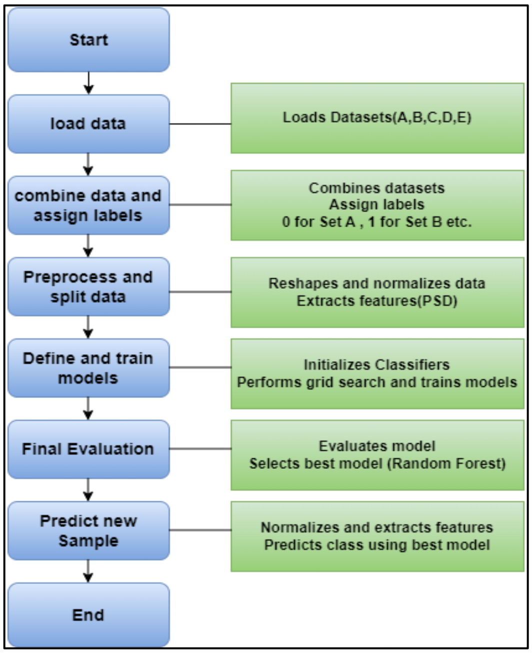


Figure 1. Proposed machine learning pipeline for epileptic seizure classification using EEG data

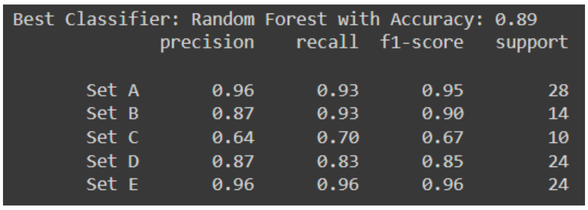
Figure 1 illustrates the machine learning pipeline for classifying epileptic seizures from EEG data. EEG datasets from Bonn University, encompassing various brain states, are first loaded and combined. Preprocessing and normalization techniques ensure data consistency. Feature extraction, using methods like power spectral density (PSD) and wavelet transforms, then extracts relevant characteristics from the prepared data. Following that, a number of machine learning algorithms are trained and assessed, including Decision Tree, Random Forest, Support Vector Machine (SVM), and Naive Bayes. Through hyperparameter tuning and cross-validation, the most effective model is identified. As highlighted in the abstract, the Random Forest model achieved the best performance, reaching an accuracy of 89% in classifying the data into five classes. This flowchart captures the workflow of the proposed system for accurate and reliable classification of epileptic seizures using EEG signals.

3. EXPERIMENTAL SETUP AND RESULTS

A multi-class classification system was implemented to distinguish between five different seizure types (Set A, Set B, Set C, Set D, and Set E) from EEG data. Four distinct machine learning algorithms' performances were evaluated: Random Forest, Support Vector Machine (SVM), Naive Bayes, and Decision Tree. Random Forest achieved the highest accuracy (89%), precision (89%), recall (89%), and F1-score (89%) among the four algorithms. The other algorithms produced the following results:

I. Support Vector Machine: - Accuracy: 58%, Precision: 56%, Recall: 58%, F1 Score: 52%
II. Naive Bayes: - Accuracy: 60%, Precision: 72%, Recall: 60%, F1 Score: 63%
III. Decision Tree: - Accuracy: 81%, Precision: 84%, Recall: 81%, F1 Score: 82%

Figure 2 shows the FEG data classified into five seizures using a Random Forest model.



Best Classifier: Random Forest with Accuracy: 0.89

	precision	recall	f1-score	support
Set A	0.96	0.93	0.95	28
Set B	0.87	0.93	0.90	14
Set C	0.64	0.70	0.67	10
Set D	0.87	0.83	0.85	24
Set E	0.96	0.96	0.96	24

Figure 2. EEG data classified into five seizure types using a Random Forest model demonstrates 89% accuracy

The use of visualization is essential to comprehending the data and the functioning of the model. Sample signals from each dataset are plotted to provide visual insights into their characteristics. Additionally, the wavelet transform of these signals is plotted to analyze their frequency components. If the Random Forest classifier is identified as the best model, the feature importance is plotted to highlight the most significant features used by the model. Figure 3 shows the visualization of features extracted from various EFG signal sets (A-W) using machine learning technique.

3.1 Power spectral density (PSD)

The power spectral density (PSD) estimates a signal's power distribution across frequencies. It provides insights into the dominant frequency components of the EEG signals, which are essential for distinguishing between different brain states.

3.2 Welch's method for PSD estimation

Welch's method is a widely used technique for estimating the PSD of a signal. It entails splitting the signal into overlapping segments, window-functioning each segment, calculating the windowed segment's periodogram, and averaging the periodograms. Welch's method provides better PSD estimates by averaging over overlapping segments, which reduces noise variance compared to a simple periodogram. This method is for analyzing stationary signals where frequency content doesn't change much over time. Welch's method is more robust against noise than raw Fourier methods.

1. Segment the Signal: Divide the EEG signal $x[n]$ into overlapping segments. Let L be the length of each segment, and D be the overlap between consecutive segments.
2. Apply Window Function: Apply a window function $w[n]$ to each segment. Common window functions include Hamming, Hanning, and Blackman windows.

The windowed segment is given by:

$$x_w[n] = x[n] \cdot w[n]$$

3. Compute Periodogram: Compute the Discrete Fourier Transform (DFT) of each windowed segment to obtain the periodogram. The periodogram for a segment k is given by:

$$P_k(f) = \frac{1}{L} \left| \sum_{n=0}^{L-1} x_w[n] e^{-i2\pi f n/L} \right|^2$$

4. Average Periodograms: Average the periodograms of all segments to obtain the PSD estimate:

$$PSD(f) = \frac{1}{K} \sum_{k=0}^{K-1} P_k(f)$$

where, K is the total number of segments.

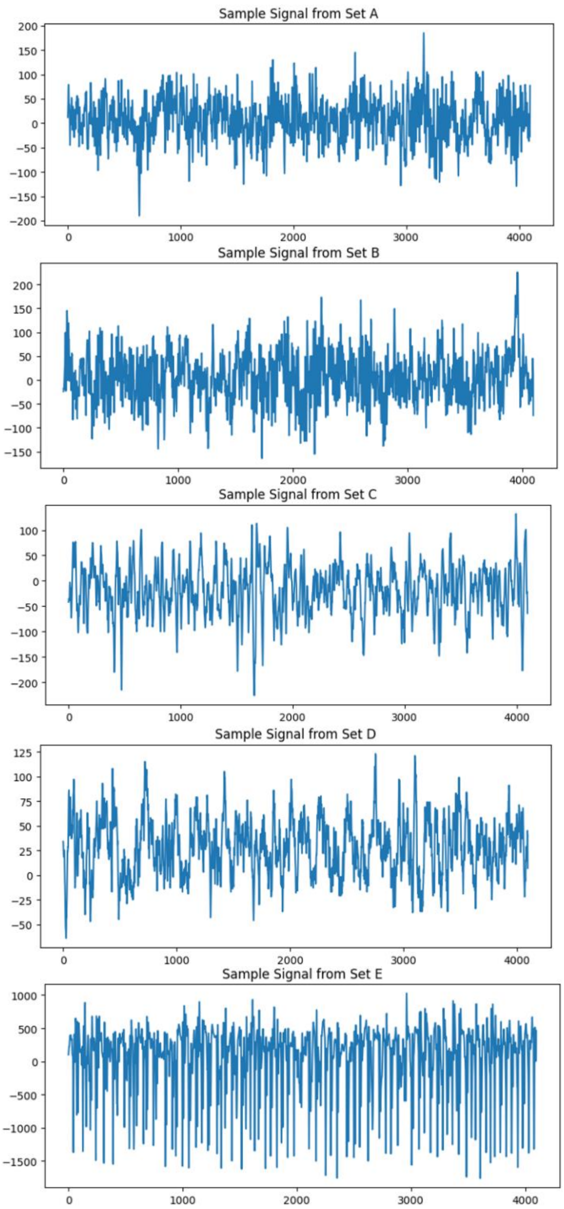


Figure 3. Visualization of features extracted from various EEG signal sets (A-E) using machine learning techniques

The Welch method improves the accuracy of the PSD estimate by reducing the variance compared to a single periodogram.

3.3 Wavelet transform

Another effective tool for feature extraction is the Wavelet Transform, particularly suited for non-stationary signals like EEG. At various scales, it decomposes the signal into components by capturing both time and frequency information.

3.3.1 Continuous Wavelet Transform (CWT)

The Continuous Wavelet Transform (CWT) of a signal $x(t)$ is defined as:

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t)\varphi * \left(\frac{t-b}{a}\right) dt$$

where, $\psi(t)$ is the mother wavelet, a is the scale parameter, b is the translation parameter, and $\psi*$ denotes the complex conjugate of ψ.

The translation parameter b adjusts the wavelet in time, and the scaling value a regulates the frequency resolution.

3.3.2 Discrete Wavelet Transform (DWT)

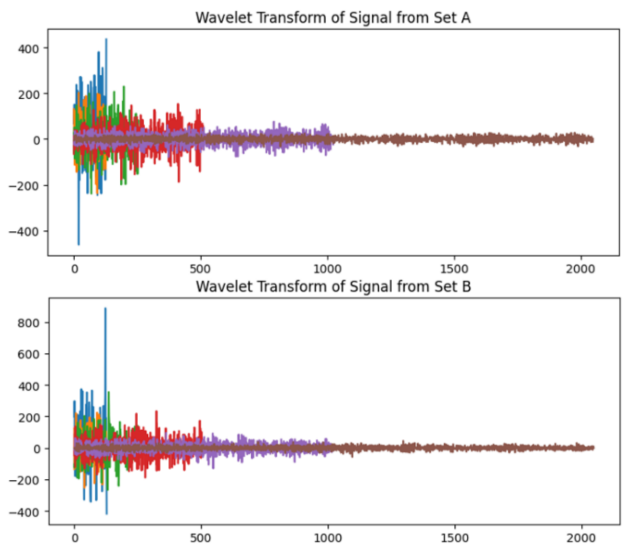
The Discrete Wavelet Transform (DWT), involves discrete scaling and translation steps. The DWT decomposes the signal into a series of wavelet coefficients at different scales.

1. Decomposition: The signal $x(t)$ is passed through a pair of filters: a low-pass filter $g[n]$ and a high-pass filter $h[n]$. This produces approximation coefficients cA and detail coefficients cD:

$$c_A = \sum_n x[n]g[n-2k]$$

$$c_D = \sum_n x[n]h[n-2k]$$

2. Downsampling: The coefficients are downsampled by a factor of 2, reducing the number of coefficients by half.
3. Iteration: The approximation coefficients cA are further decomposed, repeating the process for multiple levels.



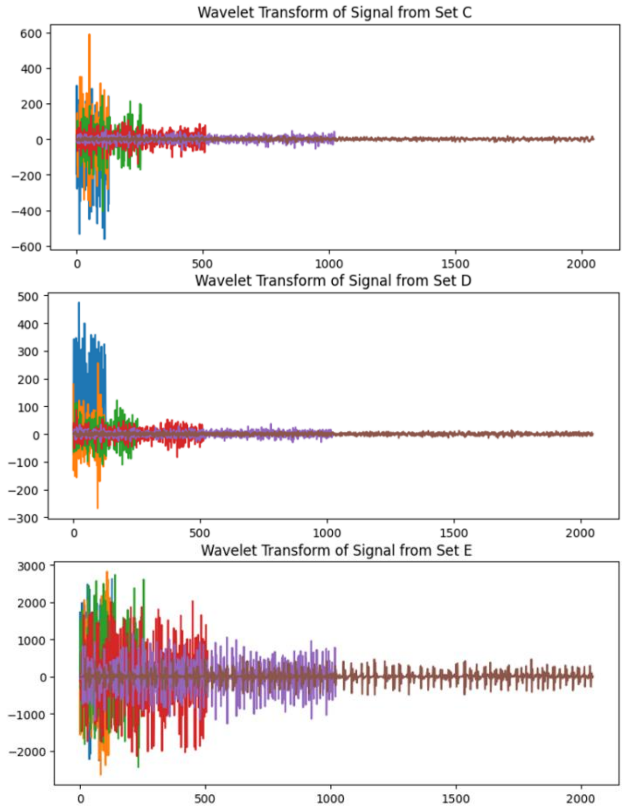


Figure 4. Wavelet transforms of example EEG signals from Sets A-E

The resulting wavelet coefficients shown in Figure 4 capture both the time and frequency characteristics of the EEG signal, providing rich features for classification.

3.4 Machine learning algorithms

Different machine learning techniques were used to classify epilepsy seizures into multiple classes. These include:

Random Forest Classifier (RFC):

A technique for ensemble learning that builds several decision trees and produces the prediction mode. Among the hyperparameters that were changed were the minimum sample size, maximum depth, and number of trees required to divide a node and produce a leaf.

Support Vector Machine (SVM):

A supervised learning model that determines which hyperplane divides data into distinct classes the best. Hyperparameters included various kernel functions, regularization parameters, and gamma values.

Naive Bayes (NB):

A Bayesian probabilistic classifier with high independence assumptions. It was straightforward with minimal hyperparameter tuning required.

Decision Tree Classifier (DTC):

A non-parametric supervised learning technique applied to classification and regression problems. It splits the dataset into smaller subsets based on feature values, and key parameters tuned included the maximum depth and minimum samples required to split a node.

The confusion matrix graph shown in Figure 5 visualizes the performance of a Random Forest model classifying EEG data into five seizure types (Sets A-E). Highlighted diagonal values (e.g., 80 in cell (E, E)) represent correctly classified instances, while off-diagonal values (e.g., 3 in cell (E, A)) indicate misclassifications. This distribution reveals the model's overall accuracy (89%) and potential biases, demonstrating its effectiveness in differentiating seizure types using extracted EEG features.

Figure 5. Confusion matrix graph showing a Random Forest model's performance in classifying EEG data into five seizure types (A-E)

4. CONCLUSION AND FUTURE WORK

This study significantly improved upon previous diagnostic techniques by creating a strong machine learning system for the multiclass classification of epileptic episodes using EEG data. Utilizing a comprehensive dataset from Bonn University, the data was preprocessed and normalized, with features extracted through power spectral density (PSD) and wavelet transforms. The Random Forest classifier outperformed other models, including Support Vector Machine (SVM), Decision Tree, and Naive Bayes, achieving the maximum accuracy of 89% when various classification techniques were evaluated. By developing a robust machine learning approach for the multiclass classification of epileptic episodes with EEG data, this work greatly improved upon earlier diagnostic methods.

The results show that the Random Forest algorithm, in particular, may greatly improve the precision and effectiveness of epileptic seizure identification, providing a trustworthy substitute for manual analysis. This approach speeds up the diagnostic process and reduces the likelihood of human error, providing more consistent and objective assessments. The successful implementation of this system could lead to improved treatment strategies and better management of epilepsy, ultimately enhancing the quality of life for patients. By showcasing the effectiveness of machine learning in handling complicated EEG data, this study advances the field of medical diagnostics and raises the possibility of further developments in automated healthcare systems.

Future work will develop an interactive web-based interface that enables users to upload EEG signals for real-time classification. This interface could preprocess the data, perform feature extraction using the Welch method, and classify the signals using trained models. Implementing this would involve the use of backend frameworks such as Flask or Django and frontend technologies like React or Angular.

Further research incorporates advanced feature extraction techniques such as wavelet transform, short-time Fourier transform (STFT), or empirical mode decomposition (EMD) to capture more precise signal characteristics. Adding nonlinear features such as entropy, fractal dimensions, or Lyapunov exponents could help capture the complex nature of EEG signals, potentially improving classification accuracy.

Implementing model stacking by combining multiple models could leverage the strengths of different algorithms. For example, using Random Forest, SVM, and Naive Bayes as base learners and a meta-learner to enhance overall prediction accuracy could be explored. Additionally, applying boosting algorithms like Gradient Boosting, AdaBoost, or XGBoost to increase model performance by focusing on hard-to-classify instances may be beneficial.

REFERENCES

[1] Dwi Saputro, I.R., Maryati, N.D., Solihati, S.R., Wijayanto, I., Hadiyoso, S., Patmasari, R. (2019). Seizure type classification on EEG signal using support vector machine. In Journal of Physics: Conference Series, 1201(1): 012065. https://doi.org/10.1088/1742-6596/1201/1/012065

[2] Oliva, J.T., Rosa, J.L.G. (2021). Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection. Biomedical Signal Processing and Control, 66: 102469. https://doi.org/10.1016/j.bspc.2021.102469

[3] Albaqami, H., Hassan, G.M., Datta, A. (2022). Wavelet-based multi-class seizure type classification system. Applied Sciences, 12(11): 5702. https://doi.org/10.3390/app12115702

[4] Ein Shoka, A.A., Dessouky, M.M., El-Sayed, A., et al. (2023). EEG seizure detection: Concepts, techniques, challenges, and future trends. Multimedia Tools and Applications, 82(27): 42021-42051. https://doi.org/10.1007/s11042-023-15052-2

[5] Aayesha, Qureshi, M.B., Afzaal, M., Qureshi, M.S., Fayaz, M. (2021). Machine learning-based EEG signals classification model for epileptic seizure detection. Multimedia Tools and Applications, 80(12): 17849-17877. https://doi.org/10.1007/s11042-021-10597-6

[6] Wang, X., Gong, G., Li, N. (2019). Automated recognition of epileptic EEG states using a combination of symlet wavelet processing, gradient boosting machine, and grid search optimizer. Sensors, 19(2): 219. https://doi.org/10.3390/s19020219

[7] Nahzat, S., Yağanoğlu, M. (2021). Classification of epileptic seizure dataset using different machine learning algorithms and PCA feature reduction technique. Journal of Investigations on Engineering and Technology, 4(2): 47-60.

[8] Janga, V., Edara, S.R. (2021). Epilepsy and seizure detection using JLTM based ICFFA and multiclass SVM classifier. Traitement du Signal, 38(3): 883-893. https://doi.org/10.18280/ts.380335

[9] Asharindavida, F., Hossain, M.S., Thacham, A., et al. (2020). A forecasting tool for prediction of epileptic seizures using a machine learning approach. Concurrency and Computation: Practice and Experience, 32(16): e5111. https://doi.org/10.1002/cpe.5111

[10] Mahjoub, C., Le Bouquin Jeannès, R., Lajnef, T., Kachouri, A. (2020). Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods. Biomedical Engineering / Biomedizinische Technik, 65(1): 33-50. https://doi.org/10.1515/bmt-2019-0001

[11] Imah, E.M., Widodo, A. (2017). A comparative study of machine learning algorithms for epileptic seizure classification on EEG signals. In 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, Indonesia, pp. 401-408. https://doi.org/10.1109/ICACSIS.2017.8355065
[12] Rohan, T.I., Yusuf, M.S.U., Islam, M., Roy, S. (2020). Efficient approach to detect epileptic seizure using machine learning models for modern healthcare system. In 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, pp. 1783-1786. https://doi.org/10.1109/TENSYMP50017.2020.9230731
[13] Wang, Y., Cao, J., Lai, X., Hu, D. (2019). Epileptic state classification for seizure prediction with wavelet packet features and random forest. In 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 3983-3987. https://doi.org/10.1109/CCDC.2019.8833249
[14] Srinivasan, V., Eswaran, C., Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. IEEE Transactions on information Technology in Biomedicine, 11(3): 288-295. https://doi.org/10.1109/TITB.2006.884369
[15] Chen, Z., Lu, G., Xie, Z., Shang, W. (2020). A unified framework and method for EEG-based early epileptic seizure detection and epilepsy diagnosis. IEEE Access, 8: 20080-20092. https://doi.org/10.1109/ACCESS.2020.2969055
[16] Chang, H.Y., Yang, S.C., Lan, S.H., Chung, P.C. (2010). Epileptic seizure detection in grouped multi-channel EEG signal using ICA and wavelet transform. In Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris, France, pp. 1388-1391. https://doi.org/10.1109/ISCAS.2010.5537262
[17] Seng, C.H., Demirli, R., Khuon, L., Bolger, D. (2012). Seizure detection in EEG signals using support vector machines. In 2012 38th Annual Northeast Bioengineering Conference (NEBEC), Philadelphia, PA, USA, pp. 231-232. https://doi.org/10.1109/NEBC.2012.6207048
[18] Boddu, R.S.K., Chakravarthi, D.S., Venkateswararao, N., Chakravarthy, D.S.K., Devarajan, A., Kunekar, P.R. (2021). The effects of artificial intelligence and medical technology on the life of human. Journal of Pharmaceutical Research International, 33(50A): 25-32.
[19] Brenner, A., Kutafina, E., Jonas, S. (2018). Automatic recognition of epileptiform EEG abnormalities. Studies in Health Technology and Informatics, 247: 171-175. https://doi.org/10.3233/978-1-61499-852-5-171
[20] Orhan, U., Hekim, M., Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Systems with Applications, 38(10): 13475-13481. https://doi.org/10.1016/j.eswa.2011.04.149