EPILEPTIC SEIZURE RECOGNITION USING MACHINE LEARNING ALGORITHMS

C.JAMUNADEVI

Assistant Professor(SrG)

Department of Computer Technology

PG

Kongu Engineering College

Perundurai, Erode,

Tamilnadu

[Jamu83@gmail.com](mailto:Jamu83@gmail.com)

ARUL P

Research Supervisor, Associate Professor,

Department of Computer Science,

Government Arts College (Affiliated to Bharathidasan University, Trichy-24) Tiruchirappalli-620 022, India

DINESHKUMAR M D

PG Scholar

Department of Computer Technology

PG  
Kongu Engineering CollegeErode,India  
[dineshkumarmd02@gmail.com](mailto:dineshkumarmd02@gmail.com)

Abstract— The project entitled “Epileptic Seizure Recognition Using Machine Learning Algorithms” is used to identify the epileptic seizure at the early stage. Epileptic seizure is a significant health concern affecting millions worldwide, necessitate precise and timely identification for effective medical intervention and management. The present system for epileptic seizure recognition often relies on traditional methods, including expert analysis of electroencephalogram (EEG) data and manual review. While these methods have proven effective, they are labor-intensive, prone to subjectivity and may not be suitable for real-time monitoring. This gap lead to the proposed machine learning-based approach which offer potential seizure recognition, allowing for more timely and accurate intervention. This study propose a novel approach for epileptic seizure recognition by combining dimensionality reduction techniques, namely Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), with the K-Means clustering algorithm, then integrate this unsupervised clustering approach with four popular machine learning algorithms: Support Vector Machine classifier, Logistic Regression, Naive Bayes classifier and Random Forest classifier to enhance the performance of seizure recognition. Each cluster is labeled as either seizure or non-seizure based on the majority class within it. The method involves the development and implementation of these algorithms within the context of epileptic seizure. The primary objective is to explore the feasibility and performance of these classifiers, aiming to enhance the efficiency of seizure recognition system. The work emphasizes the potential of machine learning in automating the identification of epileptic seizures, ultimately contributing to improved patient care and medical diagnosis. The improved accuracy of 95% is achieved by Random Forest model.

Keywords— Epileptic seizure recognition, Electroencephalography signals, Machine Learning techniques, Human activity identification.

# INTRODUCTION

The project aims to predict whether a patient has occurred seizure or non-seizure by using EEG signals, Epileptic seizures are a challenging and prevalent neurological condition that affects millions of individuals worldwide. Characterized by sudden, recurrent, and uncontrolled electrical disturbances in the brain, seizures can lead to a range of debilitating consequences, including injury and a reduced quality of life. Early and accurate detection of epileptic seizures is crucial for providing timely medical intervention and improving patient outcomes. Machine learning techniques used to the difficult issue of epileptic seizure recognition. To investigate how different machine learning techniques, such as supervised and unsupervised methods, can be employed to process and classify electroencephalogram (EEG) data and other relevant physiological signals. By leveraging machine learning's ability to discern complex patterns and relationships within data, aim to develop a robust and reliable system for the automatic detection of epileptic seizures.

The main goal is to showcase how machine learning can significantly enhance the precision and effectiveness of identifying epileptic seizures, thereby advancing patient care and enhancing the quality of life for those with epilepsy.

# LITERATURE STUDY

The study [1] proposes highlights the importance of AI and ML in advancing healthcare, particularly in early disease detection, focusing on epilepsy. Despite challenges due to limited data, recent machine learning developments hold promise for revolutionizing seizure prediction, aiming to mitigate its adverse effects.

The study [2] concentrates on automating the classification of EEG signals into Normal, Interictal, and Ictal categories for detecting epilepsy. It employs diverse features extracted through techniques such as Wavelet, Hilbert, and approximate entropy, with prioritized features aiding in distinguishing between cases. The automated classifier shows enhanced performance, indicating a departure from traditional approaches in biomedical applications.

The study [3] tackles the task of detecting epileptic seizures from EEG signals by employing a blend of fuzzy-based and conventional machine learning techniques. It attains enhanced accuracy, sensitivity, and specificity on well-established datasets such as Bonn and CHB-MIT, with Fuzzy Rough Nearest Neighbor (FRNN) and K-Nearest Neighbor (KNN) producing the most favorable outcomes.

The study [4] represents an efficient feature extraction method for epilepsy detection from EEG signals, combining DWT, LBPTH, and LBPMAD. It achieves more accuracy in classifying ictal vs. non-ictal EEG signals using SVM and KNN, even with short input signals (512 data points), making it suitable for wearable medical devices.

The paper [5] introduces a novel approach, RDCSAE-IKRVFLN, for epileptic seizure recognition using EEG signals, combining feature extraction with efficient classification. The proposed method is implemented on FPGA for a computer-aided-diagnosis system, demonstrating practicality and reliability in epilepsy detection and recognition.

The paper [6] focuses on enhancing epilepsy detection from EEG signals by applying Discrete Wavelet Transform (DWT) for feature extraction. To improve classification accuracy, it introduces SPPCA and SUBXPCA dimensionality reduction techniques.

The research [7] introduces an automated model for seizure detection that utilizes "sigmoid entropy," a computationally efficient feature extracted from discrete wavelet transforms. Tested on multiple EEG databases, it achieves a seizure detection rate of 85.0%, with low false detection rates and minimal detection delay. Sigmoid entropy proves superior, suggesting its potential as a biomarker for epileptic seizure recognition.

The study [8] introduces the GOA-SVM hybrid model for EEG-based seizure detection, combining Grasshopper Optimization Algorithm and Support Vector Machines with RBF kernel. The proposed approach achieves better accuracy in distinguishing normal and epileptic EEG data, outperforming Particle Swarm Optimization (PSO)-SVM and standard SVM with RBF kernel.

This paper [9] introduces a one-step semi-supervised epilepsy detection system, reducing the labeling cost by leveraging unlabeled data. The neural network strategy enhances decision boundary consistency, resulting in 10.3% and 4.9% higher AUROC curves compared to supervised methods on CHBMIT and Kaggle datasets, respectively.

This study [10] presents an IoT-based model for real-time seizure state recognition from EEG data. A hybrid Genetic Whale Optimization Algorithm (GWOA) and Machine learning for feature selection and classification. This proposed NB-GWOA-DEELM model demonstrates superior performance that classifying seizure states, outperforming other methods while avoiding over- and under-fitting.

This study [11] reviews the use of Machine Learning (ML) in EEG-based seizure detection and classification. It assesses the advantages of DL over traditional ML methods and aims to improve the diagnosis and management of epileptic seizures for better patient outcomes and expert support.

In this study [12], Ultra-wideband radar is used to detect persons through walls, and Kernel Principal Component Analysis (KPCA) is used to extract various features, while Support Vector Machine (SVM) is used for classification. The approach effectively distinguishes four statuses of detection, demonstrating its potential for nonlinear pattern recognition in complex environments.

This paper [13] addresses class imbalance in classification, focusing on enhancing the AdaBoost algorithm by introducing weighted vote parameters that consider not only the global error rate but also the classification accuracy of the positive class. The proposed algorithms, guided by the imbalance index, outperform traditional methods, particularly in terms of the F1 Measure.

The research [14] represents an effective method using EEG data for epilepsy seizure prediction. By employing random matrix theory and patient-specific SVM classifiers, it achieves high accuracy and sensitivity with low false predictions. This approach excels in sensitivity while maintaining interpretability, outperforming existing methods.

This paper [15], Personalized online prospective seizure prediction utilizing six intracranial EEG data is presented, which presents an adaptive pattern learning system. Patient-specific pattern libraries are gradually constructed using a two-level online feature extraction method. The results demonstrate impressive prediction accuracy, with the best scheme achieving 82% accuracy on 10 patients with epilepsy, offering a promising tool for challenging seizure prediction.

# PROPOSED WORK

In our epileptic seizure recognition utilizes a dataset in KAG\_epileptic\_seizure\_recognition dataset is a publicly available dataset of discussions on the online platform Kaggle, specifically focused on discussions related to brain EEG signals recording.

## Data Pre-Processing

To perform data preprocessing by dropping unnecessary columns, transforming the target variable into a binary format (albeit redundantly), and standardizing the feature values through feature scaling using StandardScaler. These preprocessing steps are essential for preparing the data for machine learning tasks, improving model performance, and ensuring the data with a suitable format for further analysis or modeling.

### Data Cleaning

In this process to removes an unnecessary column labeled 'Unnamed: 0' from the DataFrame. This action helps clean the data and eliminate redundant information.

### Target variable transfer

To transform the target variable, denoted as `y`, into a binary format. Specifically, it utilizes a lambda function to convert all values equal to `1` into `1`, while converting all other values into `0`. It is worth noting that this transformation is applied twice in the code, with the first transformation being commented out. This redundancy should be addressed for code efficiency.

### Feature Scaling

Feature scaling, a prevalent preprocessing method in machine learning, aims to ensure that features possess comparable scales, thereby potentially improving the performance of diverse machine learning algorithms.

## Dimensionality Reduction

### Reducing the number of variables to a lower-dimensional space while maintaining important aspects of the original data is the goal of dimensionality reduction, which is applied to the dataset using methods like Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE).

### Both PCA and t- are employed to decrease the dataset's dimensionality to two dimensions in order to prepare it for viewing or additional analysis. While PCA focuses on preserving the overall variance of the data, t-SNE aims to retain local structures and relationships between data points. Each technique is selected according to the particular requirements of the analysis or visualization work, and each has strengths of its own.

### 

## Methodology

### Support Vector Machine

Finding the optimal decision boundary or line to partition n-dimensional space into classes is the aim of the support vector machine method. This will enable the prompt assignment of following data points to the correct category.

. = c (The point resides on the decision boundary)

. >c (positive samples)

. <c (negative samples)

. – c ≥ 0

Putting –c as b, to get

. +b ≥ 0

Hence,

Y =

SVM are used to perform classification tasks on two different datasets. The SVM is initiated with a 'sigmoid' kernel, which represents one among various mathematical functions employed to map the various input data into a higher-dimensional space.

The SVM classifier is then trained using labeled training data (`pkx\_train` and `tkx\_train`) to learn how to separate different classes in the data. After training, the classifier uses the learned knowledge to make predictions on new, unseen data (`pkx\_test` and `tkx\_test`).

### Logistic Regression

When performing binary classification tasks, the goal is to predict one of two possible outcomes (yes/no or 0/1) using statistical techniques. It is applied to perform classification between two distinct datasets. The `fit` method is employed to train the model using training data, enabling it to understand the relationship between features and target labels.

P(Y=1|X) = 1 / (1 + e^-(β0 + β1X1 + β2X2 + ... + βn\*Xn))

• P(Y=1|X) is the probability of the target variable Y being 1 (success) given the input features X.

• β0, β1, β2, ..., βn are the coefficients associated with each input feature.

• X1, X2, ..., Xn are the values of the input features.

• e is the natural logarithm's base.

Subsequently, to `predict` method is used to make predictions on new data, enabling the classification of data points into one of two classes. A target variable representing the binary outcome to predict (e.g., whether a patient has occur seizure or non-seizure) is accompanied by one or more input features (independent variables) used in this recognize process.

### Random Forest Classifier

Renowned for its resilience and capacity to manage intricate datasets, One effective ensemble method in machine learning is the Random Forest algorithm. To maximize a Random Forest model's performance, it is essential to adjust a number of hyperparameters prior to training. Of these hyperparameters, node size, tree count (sometimes called n\_estimators), and the number of sampled features are the three that stand out as being very important.

The first hyperparameter that regulates this process is the node size. It determines the minimum number of samples needed to split a node when building each decision tree in the Random Forest. The algorithm can't produce too complicated trees that might overfit the training set if we set the node size appropriately. Poor generalization on unknown data results from overfitting, which happens when a model identifies noise in the training set instead of the underlying patterns.

The second parameter that establishes how many decision trees to include in the Random Forest ensemble is the tree count hyperparameter, which is sometimes referred to as n\_estimators in Python implementations. A higher tree count usually results in more diversified predictions and lower bias risk, which enhances the model's accuracy and robustness. A balance must be struck depending on the particular dataset and computer resources available, as adding too many trees can significantly raise computing costs.

Last but not least, the Random Forest's decision tree construction process incorporates a random element through the hyperparameter for the count of sampled features. The algorithm randomly selects a subset of features to examine at each split, rather than taking into account all of the data.

When the relevance of each attribute is determined for each tree, the sum is divided by the number of trees to arrive at:

i=

### Navive Bayes Classifier

The Naive Bayes algorithm is a probabilistic classification method that makes predictions using the Bayes theorem. It is a simple yet effective method for classification tasks, often used in text classification and spam filtering. It's used to train on two sets of data. First, it's trained on the `pkx\_train` dataset with corresponding labels `pky\_train`, and predictions are made on the `pkx\_test`.

P(A|B) = (P(B|A) \* P(A)) / P(B)

Then the categorizer is retrained on a different dataset, `tkx\_train` and `tky\_train`, and used to predict classes for the `tkx\_test` data. The classifier relies on the probabilistic estimation of feature distributions within each class and assigns labels to new data points based on these probabilities.

### Evaluation metrics

The effectiveness of a machine learning model is assessed using evaluation metrics. The evaluation metric selected is determined by the particular task at hand as well as the properties of the data. These are a few typical evaluation measures that are employed:

When evaluating a machine learning model, accuracy is a frequently used metric that counts how many of the model's predictions are accurate. As a ratio of the total number of examples in the dataset to the number of accurately predicted instances, it is computed. A 90% accuracy rate is achieved, for instance, if a model accurately predicts 90 out of 100 cases.

The percentage of Positive samples that were accurately classified as Positive relative to all Positive samples is known as the recall.

Precision (either rightly or wrongly) is defined as the ratio of correctly categorized positive samples (True Positive) to the total number of positively classified samples.

Instead of evaluating a model's overall performance as is done by accuracy, the F1 score focuses on how well it performs inside each class.

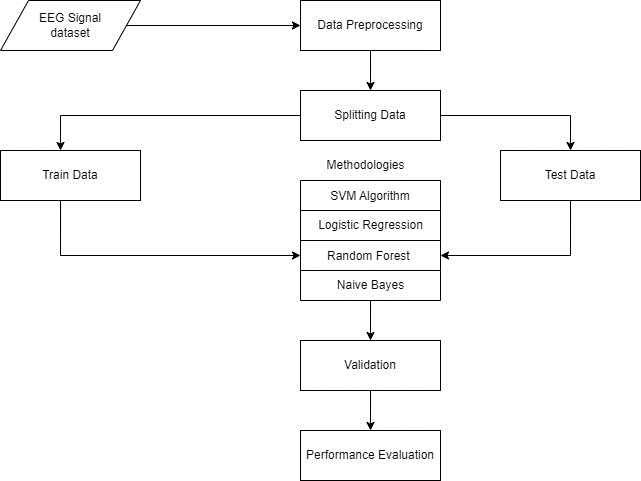
In Binary classification, the following formula can be used to calculate accuracy in terms of positives and negatives:

Accuracy = 𝑇𝑃+ 𝑇𝑁 /𝑇𝑃+𝐹𝑃+𝑇𝑁+𝐹𝑁

Recall = 𝑇𝑃/𝑇𝑃+𝐹𝑁

Precision = 𝑇𝑃/𝑇𝑃+𝐹𝑃

F1 score = 2 \* 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 ∗ 𝑅𝑒𝑐𝑎𝑙𝑙/ 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛+ 𝑅𝑒𝑐𝑎𝑙𝑙



1. System Flow Diagram

Figure 1 displays the system flow diagram process, EEG signal data undergo preprocessing to filter noise, segment into epochs, then split into train and test setsand apply various machine learning algorithms and trained on extracted features from the train set.Model performance is evaluated using metrics like accuracy, precision, and F1-score, refining the process iteratively for better classification.

# RESULT AND DISCUSSION

Epileptic seizure recognition through EEG data signals posts is an important research area, as it can help identify individuals who may have seizure and provide them with support or intervention. Support vector machine (SVM), logistic regression (LR), random forest (RF), and naïve bayes (NB) are commonly used in machine learning algorithms for epileptic seizure recognition.

EEG signals were utilized for the experiment, which assessed the accuracy of all the approaches. This section will go into great detail on the experiment's findings. The performance of each complex machine learning method SVM, LR, RF, and NB—used to categorize posts in the dataset utilized is evaluated in this work.

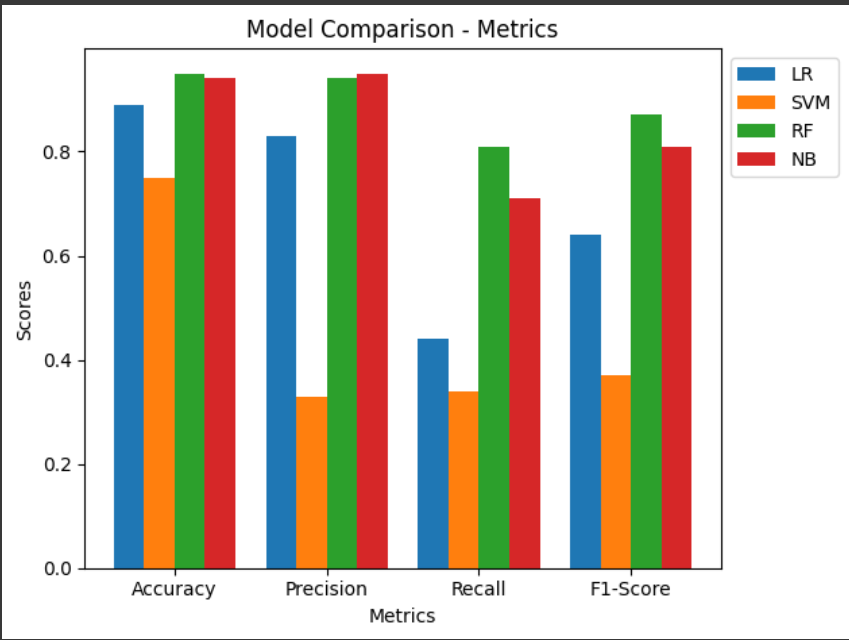
Table 1 displays the experiment's outcomes using this approach. Each of these techniques was put to the test before being put into practice. When performance is tested and parameters like F1-Score, Accuracy, Recall, and Precision are assessed, the true model result is revealed.

1. Performance Review

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| METHOD | PRECISION | RECALL | F1-SCORE | ACCURACY |
| LR | 83% | 44% | 64% | 89% |
| SVM | 33% | 34% | 34% | 75% |
| RF | 94% | 81% | 87% | 95% |
| NB | 95% | 71% | 81% | 94% |

The study found that RF had the highest accuracy (95%), followed by SVM (75%), LR (89%) and NB (94%).

Overall, the findings of the research indicate that RF stands out as the most efficient algorithm for epileptic seizure recognition, followed by SVM, LR, and NB. Nonetheless, it's crucial to acknowledge that the efficacy of these algorithms may be affected by variables such as dataset size and quality, feature selection, and algorithm parameter configurations.



1. Performance Comparsion

In Figure 2, each evaluation metric is compared across different algorithms to illustrate model performance. Figure 2 shows model performance by comparing each evaluation measure across several approaches. Logistic regression (LR) is represented by blue, support vector machine (SVM) by orange, random forest (RF) by green, and naïve bayes (NB) by red.

# CONCLUSION

In summary, Epilepsy seizure recognition offers promising solutions for timely and accurate diagnosis. This study suggests using machine learning algorithms such as support vector machine, random forest, logistic regression and naïve bayes to predict the seizure. The research in this paper shows that random forest predicts a personality characteristic with good accuracy, recall, precision and F1- scores (F1-Score 81.0%, Accuracy 95.0%, Recall 87.0%, and Precision 94.0%).

In this research has yielded promising findings, but more efficient performance is required. This implemented system will be helpful to recognize seizure by using their brain EEG data signal. This study shows that utilizing machine learning to identify the severity of epileptic seizures is both feasible and efficient.

FUTURE WORK

Further research is needed to the future, epilepsy seizure detection is poised to make significant strides by harnessing emerging technologies and innovative methodologies. One key area of focus lies in advancing real-time monitoring systems, enabling swift detection and response to seizure events to mitigate risks and improve patient safety. Additionally, there's a growing emphasis on integrating data from diverse sources, such as EEG, wearable devices, imaging, and clinical records, to provide a comprehensive understanding of seizure dynamics and enhance detection accuracy.

Edge computing has the potential to improve seizure detection algorithms by processing data closer to the source, which will lower latency and hasten the timing of interventions. Personalized models tailored to individual patients' seizure characteristics are another promising avenue, aiming to improve algorithm adaptability and performance across diverse patient populations.

Long-term monitoring research endeavors seek to uncover patterns in seizure activity over extended periods, offering insights into triggers, responses to treatment, and the development of personalized care plans. Concurrently, enhancing privacy and security measures is critical to safeguarding sensitive patient data amidst the proliferation of digital health technologies.

Furthermore, involving healthcare professionals, including neurologists and epileptologists, in the development and validation of seizure detection algorithms ensures clinical relevance, usability and seamless integration into existing healthcare frameworks.

Collectively, advancements in these areas hold the potential to revolutionize epilepsy seizure detection, ultimately leading to improved patient outcomes and quality of life.

##### References

1. Praveen A,Andhale,Dr Varsha H. Patil (2022) Machine Learning Approach for Predicting Epileptic Seizures using EEG Signals Expert Syst. Appl., 161, 113676.
2. Anugraha A, Vinotha E, Anusha R, Giridhar S, Narasimhan K (2017) A machine learning application for epileptic seizure detection. ICCIDS 2017: IEEE International Conference on Computational Intelligence in Data Science.
3. Mahjoub C, Jeannes RLB, Lajnef T, Kachouri A (2020) Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods. Biomed Eng/Biomedizinische Technik 65:33–50
4. Tang, L.; Xie, N.; Zhao, M.; Wu, X. (2020) Seizure prediction using multi-view features and improved convolutional gated recurrent network. IEEE Access.
5. Mursalin M, Zhang Y, Chen Y, Chawla NV (2017) Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing 241:204–214.
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 19:219.
7. Al-Hadeethi, H.; Abdulla, S.; Diykh, M.; Deo, R.C.; Green, J.H. (2020) Adaptive boost LS-SVM classification approach for time-series signal classification in epileptic seizure diagnosis applications. Expert Syst. Appl., 161, 113676.
8. Elghamry, Darweesh,M.S.;Ismail,T. (2020) Comparative Analysis of Various Machine Learning Techniques for Epileptic Seizures Detection and Prediction Using EEG Data, IEEE: New York, NY, USA.
9. A. Subasi, J. Kevric, M. Abdullah Canbaz, (2017) Epileptic seizure detection using hybrid machine learning methods, Neural Comput. Appl. 1–9 Springer.
10. Atal DK, Singh M (2019) A hybrid feature extraction and machine learning approaches for epileptic seizure detection. Multidim Syst Sign Process:1–23.
11. Jiang, Y.; Chen, W.; Li, M. (2020) Symplectic geometry decomposition-based features for automatic epileptic seizure detection. Comput. Biol. Med., 116, 103549.
12. Li, P.; Karmakar, C.; Yearwood, J.; Venkatesh, S.; Palaniswami, M.; Liu, C. (2018) Detection of Epileptic Seizure Based on Entropy Analysis of Short-Term EEG. PLoS ONE, 13, e0193691.
13. N. Sriraam, S. Raghu, (2017) Classification of focal and non-focal epileptic seizures using multi-features and SVM classifier, J. Med. Syst. 41 160.
14. Gupta V, Pachori RB (2019) Epileptic seizure identification using entropy of FBSE based EEG rhythms. Biomed Signal Process Control 53:101569.
15. Ma, M.; Cheng, Y.; Wang, Y.; Li, X.; Mao, Q.; Zhang, Z.; Chen, Z.; Zhou, Y. (2021) Early Prediction of Epileptic Seizure Based on the BNLSTM-CASA Model. IEEE Access, 9, 79600–79610.