International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0433
e ISSN: 2584-2854
Volume: 03
Issue: 08 August 2025
Page No: 2757 - 2762

Detection of Epileptic Seizures Using ML Algorithms Enhanced by ADASYN and SMOTE Sampling Techniques

Sanagavarapu Sunitha[1], Dr. Umadevi Ramamoorthy[2], S. Sunitha[3], Dr. DVNS Murthy[4], B. Rama[5], K. Srujana[6]
[1]Research Scholar, SSCS, CMR University, Bengaluru, Karnataka, India.
[2]Associate Professor, SSCS, CMR University, Bengaluru, Karnataka, India.
[3]Associate Professor, Computer Science, BBCIT, Hyderabad, Telangana, India.
[4]Director, Statistics, BBCIT, Hyderabad, Telangana, India.
[5]Assistant Professor, Statistics, BBCIT, Hyderabad, Telangana, India.
[6]Assistant Professor, Mathematics, BBCIT, Hyderabad, Telangana, India.
Emails: *sanagavarapu.sunitha@cmr.edu.in[1], umadevi.r@cmr.edu.in[2], amruthabala2725@gmail.com[3], dvns42@gmail.com[4], botturama99@gmail.com[5], srujanasujji424@gmail.com[6]*

Abstract
Machine learning algorithms play a crucial role in healthcare and medical diagnosis applications within the computer-aided research domain. Continuous seizures, which are sudden bursts of electrical activity in the brain, are a hallmark of epilepsy, a neurological condition of the brain. To detect epileptic seizures, this study monitors EEG (Encephalography) signals and turns them into a dataset. Then, using ADASYN (Adaptive Synthetic Sampling) and SMOTE (Synthetic Minority Oversampling Technique) to balance the data, the dataset is subjected to algorithms like Linear Regression, Support Vector Machine, Regressor, Logistic Regression, Decision Tree, KNN (k-nearest neighbor), and Random Forest. This paper investigates the Complete EEG dataset, i.e., Epileptic Seizure Recognition from Kaggle. Based on historical data, these models assist us in identifying epileptic episodes. Every model trained on the dataset produced accurate values. It was recognized that Random Forest with the SMOTE Model gives better accuracy for the given EEG datasets.

Keywords: *Epilepsy, EEG, Linear and Logistic Regression, Decision Tree, Random Forest, Support Vector Machine and Regressor, KNN, ADASYN, SMOTE*

1. Introduction

"Health is Wealth". Now a days, Health is essential for all. Due to unhealthy habits and irregular timing intervals people are suffering with many health issues, in these the common problem is "Epilepsy". However, 3.5% of people are suffering with epileptic abnormalities. EEG (Encephalography) is the ordinary problem-solving method to show Epilepsy [1]. Various causes might cause Epilepsy like Influence, Head Trauma, Brain conditions, Infectious Diseases, and Developmental Disorders [2]. We cannot predict because of this reason the person suffered from Epilepsy. EEG is a method to record an electrogram of the artless electrical movement of the brain (Neurons) [3]. The individual Neuron electric potential picked EEG. EEG recording is taken via locating electrodes on the scalp by operating gel. Each electrode has its particular name itemized by the International 10 – 20 System. EEG also helps to identify Seizure Disorders, Sleep Disorders, Brain tumors, Brain Injuries, Brain Infections, Stroke, Attention Disorders, and Behavioral delays in children. In recent years, Machine Learning algorithms have shown outstanding ability in automating the detection of Epilepsy with Seizures and other brain disorders. Presently advanced deep learning

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0433
e ISSN: 2584-2854
Volume: 03
Issue: 08 August 2025
Page No: 2757 - 2762

techniques, especially Convolution Neural Networks, play a vital role in this area.

2. About the Data Set

2.1. Overview of the Dataset

The "**Epileptic Seizure Recognition**" dataset has taken from the Kaggle, and it is **not balanced.** It contains EEG recordings as numerical values.

2.2. Class Distribution

The dataset labelled as **five classes**:

- **Class 1** – Epileptic seizure activity
- **Class 2** – EEG from tumour-affected brain region
- **Class 3** – EEG from healthy brain region in tumour-affected patient
- **Class 4** – EEG with eyes closed
- **Class 5** – EEG with eyes open

The dataset has **11,500 samples**.

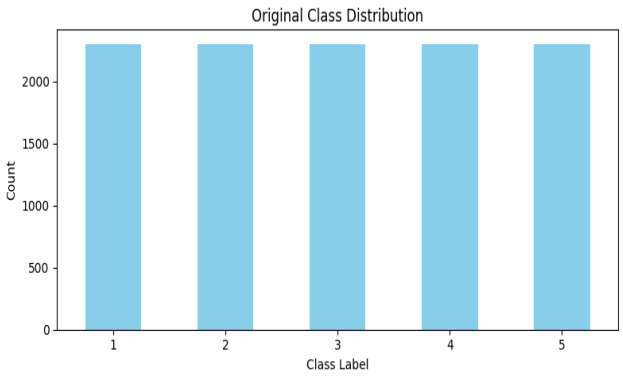

Figure 1 Original Class Distribution

2.3. Objective of the Analysis

The dataset contains EEG recordings of to develop accurate forecasting models to predict epileptic seizures. (Figure 1)

2.4. Variables in the Dataset

Each sample consists of 179 EEG signal values and one class label y (column 180).

3. Literature Review

This paper examines automated Epileptic seizures detection methods overall review [4].This paper studied the different EEG datasets from different databases and applied different machine learning algorithms on that dataset to predict seizures [5].This paper gives How to read brain data and apply mining techniques for epileptic Seizure detection [6].The authors of this paper identified seizure detection on the CHB-MIT database using EMD (empirical mode) with feature extraction [7].In addition to proposing an alternative architecture based on pre-trained data sets and easily identifiable seizures, the authors of this article describe how to translate age signals into visuals [8]. To process and classify SVMs with an accuracy of 94.88, the authors of this research developed a system that included feature extraction and classification of Taylor Fourier rhythm-specific models and a filter bank [9].In this paper author provided various epileptic seizure detection techniques using biomedical signals [10].The author describes best evaluation model by using feature extraction to evaluate epileptic seizure detection using EEG [11].This work focuses on seizures identification automatically with the help of EEG and different pattern recognitions with segmentation and extractions [12].Author explained different machine learning classifiers to detect epileptic seizures.EEG data on CNN with SMOTE it tells us to maintain sampling on EEG to regularize [13].

4. Methodology

4.1. Data Collection

This dataset has 11500 rows *180 columns all together. It is grief with an imbalance problem, such as:

• Samples are not equal in each class.

• This work is carried by binary classification problems: Seizures in Class 1 versus non-seizures in Classes 2–5.

• In the class imbalance, Class 1 (seizures) is underrepresented in comparison to the total number of non-seizures.

4.2. Preprocessing

First step: unnecessary columns are removed. In this dataset, remove "Unnamed" column. Next checks for missing values. Separate feature and target variable (y) for seizures. Apply StandardScalar for normalization as mean=0 and Standard Deviation = 1. The dataset contains 11500*179 records. In this, 178 are features, and the 179th column (y) is the targeting variable.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0433
e ISSN: 2584-2854
Volume: 03
Issue: 08 August 2025
Page No: 2757 - 2762

4.3. Feature Selection

In this, features are selected using a correlation analysis between each feature and the target. Generally, the top 10 features were selected, with the highest correlation to the target. The following Figure 2 & Figure 3., the bar plot shows 10 features associated with y.

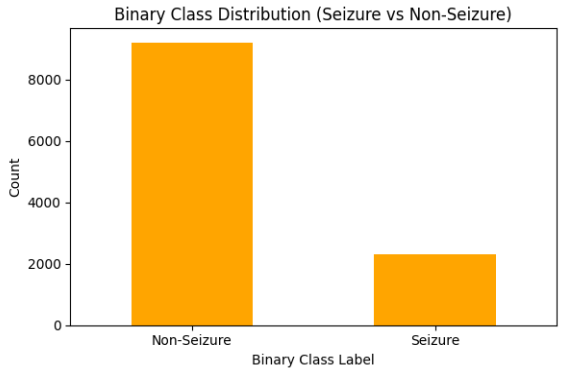

Figure 2 Binary Class Distribution

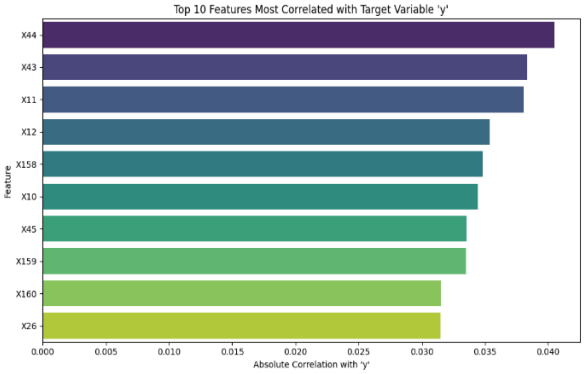

Figure 3 Feature selection for epileptic seizures

4.4. Handling Imbalance

To address this imbalance, the following techniques are commonly used to balance the dataset before training machine learning models.

- **SMOTE** (Synthetic Minority Over-Sampling Technique)
- **ADASYN** (Adaptive Synthetic Sampling)

4.4.1. SMOTE (Synthetic Minority Over-Sampling Technique)

- **Purpose:** To create artificial samples for the minority class to balance the distribution of classes. SMOTE balances the dataset by synthetically generating seizure samples to match the non-seizure count. Figure 4 displays the binary class distribution for the samples.
- **How it works:** SMOTE selects a minority class sample. It locates the classes k closest neighbors. It generates a synthetic sample along the line segment connecting the two samples after choosing one of their neighbors at random. Until the desired balance is reached, the process is repeated. It reduces overfitting compared to simple duplication. Works well with continuous features.

4.4.2. ADASYN (Adaptive Synthetic Sampling)

- **Purpose:** Like SMOTE, ADASYN generates synthetic samples for the minority class, but it focuses more on samples that are harder to learn. After applying, the binary class distribution is shown in Figure 5.
- **How it works:** ADASYN identifies minority samples that are misclassified or lie near the decision boundary. It generates more synthetic samples for these "harder" cases. The number of synthetic samples is adaptively determined based on local density. It focuses on improving classifier performance by targeting difficult samples. Enhances decision boundaries.

After SMOTE and ADASYN, the dataset Class Distribution

Each class is perfectly balanced with 2,300 samples:

- Class 1 (Seizure): 2300
- Class 2–5 (non-seizure): 9200 (combined)

4.5. Classification /Analysis & Validations

Split the dataset into train and test as 80 and 20. Apply different machine learning models are to be applied on the dataset. Here, classification models like Linear, Logistic, Support Vector, Decision Tree, Random Forest and k-nearest Neighbor. These above model metrices are presented in the results and discussions.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0433
e ISSN: 2584-2854
Volume: 03
Issue: 08 August 2025
Page No: 2757 - 2762

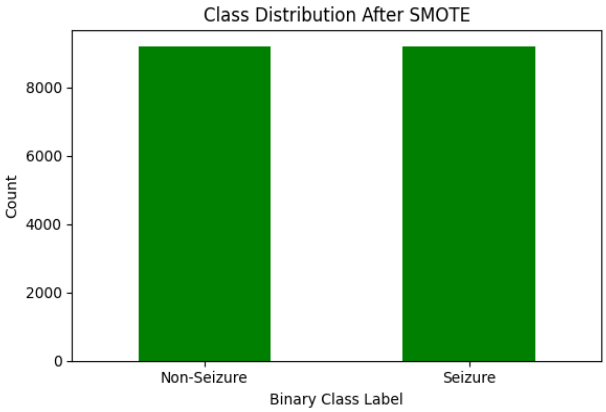

Figure 4 Balanced class distribution

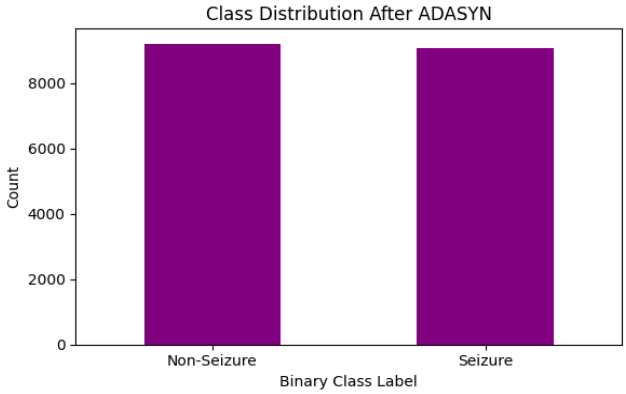

Figure 5 Balanced class distribution with SMOTE with ADASYN

Table 1 Results of the Classifiers

Model	Accuracy %	Precision %	Recall %	F1 Score %	AUC (Area Under Curve) %
Logistic	63.5	66.1	52.4	58.5	61.2
Support Vector Machine	97.8	96.8	92.3	94.5	99.7
Decision Tree	94.4	88.2	83.2	85.6	90.2
Random Forest	97.6	96.7	98.6	97.7	99.8
K - nearest neighbor	92.5	99.7	63.2	77.4	92.4
Linear	81.9	98.0	10.5	19.0	54.1
Support Vector Regressor	97.9	98.4	90.97	94.5	95.3

5. Results and Discussions

5.1.Results

The recital of various classification models used in this paper to detect Epilepsy. The Evaluation is based on key metrics such as accuracy, precision, recall, F1 score, and AUC. (Table 1) These metrics provide a comprehensive understanding of how well each model fits the data and its predictive accuracy. The results are as follows:

5.2.Discussions

The Logistic Regression model's performance is okay, and it failed to achieve good recall value. With a high recall, AUC, and accuracy, the Random Forest does remarkably well in detecting seizures. The support vector machine and regressor performed very well, but less than the random forest model. The K-nearest neighbor with excellent precision and poor recall, which means it misses many seizure cases. The decision tree performs well overall. The linear equation model is not the most effective model for identifying epileptic seizures because of its incredibly low recall and F1 scores.In all the above

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0433
e ISSN: 2584-2854
Volume: 03
Issue: 08 August 2025
Page No: 2757 - 2762

models the best model is random forest model with accuracy 97.6%, precision 96.7%, recall 98.6%, F1 score 97.7%, and AUC 99 9.8% to detect seizures. Figure 6 shows Comparing the Performance Metric Values of Various Models Figure 7 shows Each Model Performance Comparison of Five Metrics

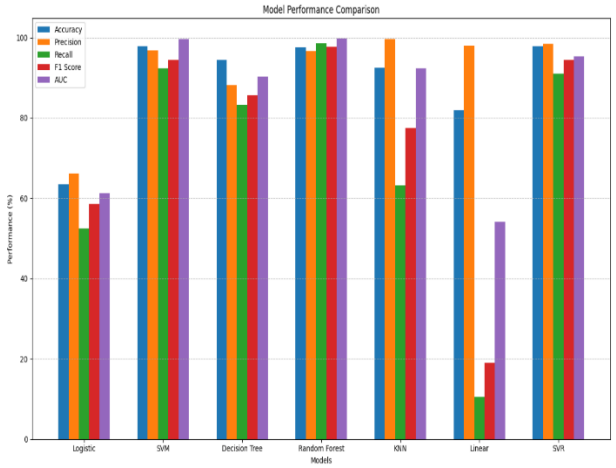

Figure 6 Comparing The Performance Metric Values of Various Models

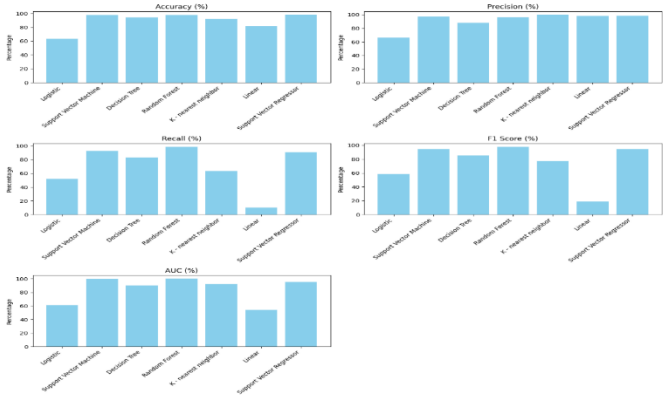

Figure 7 Each Model Performance Comparison of Five Metrics

Conclusion

This study uses the Epileptic Seizures Recognition data from Kaggle to assess several machine learning methods for identifying epileptic seizures. Among these models, the random forest model outperformed the others in every performance metric value. In assessing how well different classification models predict epileptic seizures, the random forest model performs with 97.6% accuracy, 96.7% precision, 98.6% recall, 97.7% F1 score, and 9.8% AUC. In the future, enhanced models such as CNN and Deep Learning models will be able to detect epileptic seizures more accurately.

Acknowledgements

My sincere thanks to my Guide, Dr. Uma Ramamoorthy, for her remarkable assistance and unwavering support throughout this work. My sincere gratitude to all my family members, friends and supporters.

References

[1]. Organization WH (2006) Neurological disorders: public health challenges. World Health Organization, New York

[2]. Reynolds EH (2009) Milestones in epilepsy*. Epilepsia 50(3):338–342. https://doi.org/10.1111/j.1528-1167.2009.02050.x

[3]. WHO (2005) Atlas: Epilepsy care in the world. World Health Organization, Geneva

[4]. Tzallas AT, Tsipouras MG, Tsalikakis DG, Karvounis EC, Astrakas L, Konitsiotis S, Tzaphlidou M (2012) Automated epileptic seizure detection methods: a review study. In: Epilepsy-histological, electroence phalographic and psychological aspects. InTech

[5]. Abbasi B, Goldenholz DM (2019) Machine learning applications in epilepsy. Epilepsia

[6]. Siddiqui MK. Brain data mining for epileptic seizure-detection. Doctoral Dissertation, Charles Sturt University, Australia

[7]. images. In 2020 International Conference on Communication and Signal Processing (ICCSP).

[8]. Jiang, Y., Yao, L. u., & Yang, L. (2022). An epileptic seizure prediction model based on a time-wise attention simulation module and a pretrained ResNet. Methods, 202. https://doi.org/10.1016/j.ymeth.2021.07.006

[9]. Kumar Boddu, R. S., Chakravarthi, D. S., Venkateswararao, N., Chakravarthy, D. S. K., Devarajan, A., & Kunekar, P. R. (2021). The effects of artificial intelligence and

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0433
e ISSN: 2584-2854
Volume: 03
Issue: 08 August 2025
Page No: 2757 - 2762

medical technology on the life of humans. J Pharm Res Int, 33. https:// doi.org/ 10.9734/ jpri/2021/v33i50A33378

[10]. Paul Y (2018) Various epileptic seizure detection techniques using biomedical signals: a review. Brain Inform 5(2):6

[11]. Boonyakitanont P, Lek-uthai A, Chomtho K, Songsiri J (2020) A review of feature extraction and performance evaluation in epileptic seizure detection using eeg. Biomed Signal Process Control 57:101702

[12]. Sharmila A, Geethanjali P (2019) A review on the pattern detection methods for epilepsy seizure detection from EEG signals. Biomed Eng /Biomedizinische Technik. 64(5):507–17

[13]. A review of epileptic seizure detection using machine learning classifiers

[14]. Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Xiaodi Huang &

[15]. Nasir Hussain Brain Informatics volume 7, Article number: 5 (2020)