



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023



Impact Factor: 8.423



Analysis of Best Machine Learning Classifier for Epilepsy EEG Signal

Ruthresh M , Prasanth B, Ashwin Raj M , Meenatchiselvan M, Mamtha G N

Department of Electronics and Communication Engineering, Faculty of Engineering and Technology Jain(Deemed-to-be-University), Jakkasandra, Bengaluru, India

Assistant Professor Department of Electronics and Communication Engineering, Faculty of Engineering and Technology Jain(Deemed-to-be-University), Jakkasandra, Bengaluru, India⁵

ABSTRACT: This study attempts to produce the best classification approach to detect the epilepsy EEG data. Among the different classifiers, including random forest, decision trees, XG boost, and others. Without compromising performance, machine learning classifiers can classify EEG data, detect seizures, and highlight pertinent meaningful patterns. Using a sizable dataset of labelled EEG signals derived from epileptic patients, this classifier's creation entails substantial training and validation. Then, using various classifier types, many classification models are generated. The classifier's performance is assessed in terms of sensitivity, specificity, and overall accuracy, and it is contrasted with current state-of-the-art methods for epilepsy detection. The outcomes show the advantage of the suggested classifier in terms of its capability to precisely identify epileptic.

KEYWORDS: Epilepsy detection, Machine learning classifiers.

I. INTRODUCTION

The neurological condition epilepsy is characterised by aberrant brain activity that results in seizures or episodes of strange behaviour, feelings, and occasionally loss of consciousness. Epilepsy may come from a genetic condition or an acquired brain lesion, such as a trauma or stroke. Epilepsy can strike anyone at any time. Men and women of different ages, races, and cultural backgrounds can develop epilepsy. There are many different types of seizure symptoms. During a seizure, some epileptics just stare blankly for a brief period of time, while others constantly jerk their limbs or legs. Epilepsy is not always indicated by a single seizure. In general, an epilepsy diagnosis requires at least two unprovoked seizures that occur at least 24 hours apart.

In the US, it affects 1.2% of the population (3.4 million people), and it affects over 65 million people worldwide. Furthermore, 1 in 26 persons will experience epilepsy at some point in their lifetime. The most recent survey found that more than 2% of people globally suffer from epilepsy, with 85% of those people living in developing nations. Epilepsy has a negative impact on people's daily lives and productivity. There are thought to be 2.4 million new cases worldwide each year. The majority of the time, experts use EEG signals to diagnose epilepsy. EEG signals are divided into two categories: intracranial and scalp EEG. Using the international standard 10-20 system, scalp electrodes are positioned on the scalp surface to record scalp EEG. The 10-20 system, also known as the International 10-20 system, is a method for describing and applying the position of scalp electrodes during an EEG test, polysomnography sleep study, or unpaid lab research that is widely accepted on a global scale. This technique was established to maintain consistent testing procedures, ensuring that the results of a subject's study (whether clinical or research) could be collated, replicated, and successfully analysed and compared using the scientific method.

II.RELATED WORK

This section will discuss a literature review of works that have been done by various authors in the fields of machine learning for the detection of epileptic seizures. These techniques include K-nearest neighbors, logistic regression, Stochastic Gradient Descend, Naive Bayes Classifier, Random Forests, Decision tree, XG booster, Gradient booster, Extra tree classifiers.

[1] Milind Natuet.al., observed that features like energy, skewness, and entropy are the most commonly used features; however, it is essential to optimize the feature vector to reduce the burden of the classifier and retain accurate results. It is very hard to decide which classifier is most optimum, so to conclude a point about the classifier selection, these are tested and evaluated with multiple datasets. The literature found that earlier research scholars have used different techniques: SVM, KNN, and ANN. The main limitation of these classifiers is that they are incapable of providing a suitable explanation for patterns and logic rules for the hidden models. However, it is observed that a random forest classifier yields high-accuracy results for seizure detection.

[2] Daniela Gifu et.al., proposed the first idea was to predict the potential epilepsy surgery using Support Vector Machine (SVM) and Naïve Base classifiers, already used for depression. These methods provide significant opportunities for discovering information in clinical text and enables the clinical scientist to explore novel approaches to care. There are a few years since the first system for medical records have been released on the market, these types of systems helped the healthcare field to establish a base for different directions in modern medical informatics and played an important role for EMR system. Many approaches for medical text processing are well-known, based on ML and Deep Learning (DL).

[3] Mohammad Khubeb Siddiqui et.el., observed a major challenge is to detect seizures correctly from a large volume of data. Machine learning classifiers are well-suited for accurate seizure detection in EEG signal datasets due to their ability to handle the complexity of the signals. However, it is essential to carefully choose appropriate classifiers and features. In this regard, the mentioned paper suggests that a decision forest, specifically a black-box decision forest consisting of an ensemble of decision trees, is the most effective approach.

[4] Mahmood A. Jummah et.al., the average sensitivity of the method was 74% showed a nonlinear approach to classify EEG seizure segments by offering nonlinear decision functions with support vector machine (SVM). As a result, the accuracy average was 91.84%. An automatic seizure detection method was proposed by Yuan in 2018 by using nonlinear features for EEG records to construct the features vector, then feeding them into a single-layer neural network that was trained by extreme learning machine (ELM) algorithm. Finally, the output of the neural network was passed through a post-processing step to make the decision either ictal or interictal. The average sensitivity was approximately 91%. Yuan introduced a novel approach in 2014 for automatic seizure detection using intracranial EEG. In this method, the EEG records were divided into segments of 4 seconds in size and subjected to wavelet transform decomposition at five different levels. To generate a feature vector, diffusion distances were utilized.

[5] R. B. Messaoudet.al., observed an important issue regarding a large number of studies in this field are reproducibility and comparability. The research community needed a common and solid framework that would allow to assess the statistical validity of the results and to compare different studies. The authors established the foundation of a framework by introducing two crucial ideas: the seizure prediction horizon (SPH) and the seizure occurrence period (SOP). The SPH represents the minimum time interval between an alarm and the start of a seizure, also known as intervention time (IT). On the other hand, the SOP refers to the period following the SPH in which the seizure actually occurs. It is important to note that when comparing the results of different studies, the performance measurements can only be considered comparable if they are evaluated using the same SPH and SOP values, and on the same database. Given the large variability of the choices for these parameters in the literature it is often hard to assess and compare studies results. And yet, some very recent studies seem to ignore this framework and only present their classification accuracies as a proof of the superiority of their method. Thus, even if these studies show outstanding results claims to achieve a sensitivity of 99.6%

[6] Fauzia P. Lestari et.al., the EEG data was from chb-mit database, downloaded from <https://physionet.org/pn6/chbmit/>. These EEG data collected at the Children's Hospital Boston. The data consists of EEG recordings from pediatric subjects with intractable seizures. The objective of the study was to evaluate patients who had discontinued their anti-seizure medication, observe their seizures, and determine their suitability for surgical

intervention. A total of 11 patients, consisting of 4 males and 7 females aged between 1.5 and 22 years, were included in the study. EEG recording data were collected, and each .edf file contained one hour of digitized EEG signals. The signals were sampled at a rate of 256 samples per second with a 16-bit resolution. The files contained 23 EEG signals, and the recordings followed the International 10-20 system of EEG electrode positions and nomenclature. The data were divided into training data and validation data for analysis and evaluation. The training data consist of 71 Seizure files and 339 Non seizure files, while validation data consist of 12 Seizure files and 10 non seizure files.

[7] DwiSunaryonoet.al., observed the features were inputted to Bayes Net, Support vector machine (SVM), artificial neural network (ANN), logistic regression (LR), and functional tree (FT) for classification. Used min amplitude, max amplitude, zero-crossing rate, skewness, kurtosis, mean, median, mode, energy, entropy, variance, signal-to-noise, coefficient of variation, standard deviation, mean and power directly extracted from EEG signals to classify epilepsy. The researchers employed the Kruskal-Wallis nonparametric test to select the features for classification purposes. They utilized an artificial neural network trained with particle swarm optimization (PSO-ANN) for the classification task. In addition, lagged Poincare plot parameters extracted directly from EEG signals were used as features for epilepsy detection, as previously reported. These features were then fed into k-NN (k-Nearest Neighbors) and probabilistic neural network (PNN) algorithms for classification, achieving an impressive accuracy of 98.33%. This highlights the increasing popularity of machine learning algorithms in the classification of electroencephalography (EEG) signals for the purpose of epileptic seizure detection.

[8] L. Torlay et al., The utilization of machine learning algorithms for classifying electroencephalography (EEG) signals to detect epileptic seizures is gaining significant popularity. These algorithms possess the capability to autonomously learn from EEG data and make predictions, aiding doctors and researchers in more precise and timely identification of seizures. This advancement contributes to improved accuracy and efficiency in seizure detection. In addition, the use of machine learning algorithms can help improve access to healthcare for communities that may not have access to specialized medical care. This literature review specifically examines the XGBoost algorithm, recognized as a highly potent and efficient tool for classifying epileptic seizures. A selection of pertinent papers utilizing XGBoost for seizure classification is reviewed, encompassing their discoveries and contributions to the field. The findings of this review indicate that XGBoost is a valuable asset in the classification of epileptic seizures; however, additional research is necessary to enhance and optimize its performance further. In particular, there is a need for more research on the use of XGBoost in different populations and settings, and on the development of more sophisticated and customized machine learning models for epileptic seizures classification.

[9] E. Kabir et.al., observed this work is related to several multiclass EEG signal classification techniques in the literature. A statistical framework has been proposed by Suily and Li for multiclass EEG signal classifications. They developed an optimum allocation scheme based on the variability of observation within a group (based on specific time) of the EEG data and selected a representative sample. The representatives were fed to the least square SVM (LS-SVM) classifier instead of taking representative features that may be a limit for further Analysis of best Machine Learning Classifier for Epilepsy EEG Signal Department of Electronics & Communication Engineering, Faculty of Engineering & Technology, JAIN (Deemed-to-be University) 8 consideration of a detection technique. Later, Siuly et al. proposed a sampling-based approach for the classification of multi-category EEG signals.

[10] Prabhpreet Kaur Bhatia and Anurag Sharma et.al., Epilepsy is a perilous neurological disease covering about 4-5% of the total population of the world. Its main characteristics are seizures which occur due to certain disturbance in brain function. During epileptic seizures, the patient is unaware of their physical as well as mental condition and hence physical injury may occur. Proper health care must be provided to the patients and this can be achieved only if the seizures are detected correctly in time. In this dissertation work, a system is designed using wavelet decomposition method and different training algorithms to train the neural network for the classification of the EEG signals. The system was tested and compared with Support Vector Machine (SVM) classifier. The system accuracy comes out to be 99.97%.



S .no	Paper title	Year	Author	Techniques	Result	Outcome
1	Review on Epileptic Seizure Prediction: Machine Learning and Deep Learning Approaches	2022	MilindNatu, MrinalBachute, Shilpa Gite, Ketan Kotecha, and Ankit Vidyarthi	SVM, KNN, and ANN	Precision -96.1% Accuracy-95.8%	KNN achieves better accuracy
2	The Use of Decision Trees for Analysis of the Epilepsy.	2021	Daniela Gifu	Decision Trees Pro Tackle Tool	Accuracy-93.5% Efficiency-88%	decision tree achieves better accuracy with pro tackle.
3	A review of epileptic seizure detection using machine learning classifiers.	2020	Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Xiaodi Huang and Nasir Hussain.	KNN, SVM, Random Forest	Accuracy 100%	100% accuracy in the classifiers
4	Random Forest classifier for EEG-based seizure prediction	2020	R.B. Messaoud	Random Forest	Accuracy-91.5% Sensitivity-90.4%	Process improves significantly the performances leading to an average sensitivity and accuracy with RF
5	Epileptic Seizures Detection Using DCT-II and KNN Classifier in Long-Term EEG Signals	2020	Mahmood A. Jumrah, Ammar Ibrahim Shihab, Akeel Abdul Kareem Farhan,	KNN classifier	Accuracy-95.5%	
6	Epileptic Seizure Detection in EEGs by Using Random Tree Forest, Naïve Bayes and KNN Classification	2019	Fauzia P. Lestari1 Mohammad Haekal1 RizkiEdmi Edison2 Fikry Ravi Fauzy2 Siti Nurul Khotimah1 Freddy Haryanto	Random Tree Forest, Naïve Bayes and KNN Classification	KNN Accuracy-95.5% RF accuracy-86.7% NB accuracy-57.6%	knn perform the best classification based on the accuracy
7	Gradient boosting machines fusion for automatic epilepsy	2018	Dwi Sunaryono	GBM fusion	Accuracy-95.5%	GBMs fusion can improve the performance of a



	detection from EEG signals based on wavelet features					single GBM in classifying EEG signals
8	Machine learning–XGBoost analysis of language networks to classify patients with epilepsy	2017	L. Torlay - M. Perrone- Bertolotti - M. Baciù	xgbooster	Accuracy: 99%	It shows significant potential for classifying patients with epilepsy based on the cerebral region, hemisphere and processing of their language representation
9	Epileptic seizure detection from EEG signals using logistic model trees	2016	E. Kabir	Optimal Allocation Technique	Accuracy-95.5%	OAT and logistics model achieved better classification accuracy
10	Epilepsy Seizure Detection Using Wavelet Support Vector Machine Classifier.	2016	Prabhpreet Kaur Bhatia and Anurag Sharma	SVM	Accuracy 99.97%.	The system accuracy comes out to be 99.97%.

Table 1: Summary table of literature survey

From the overview of the literature survey, many researchers have used different types of classifiers such as decision tree, Random Forest , KNN xg booster, etc. They mentioned different types of parameters such As precision, AUC, Accuracy, and f1 Score, but finally, they conclude by taking the accuracy parameters for Detecting Epilepsy EEG signals. Mostly the accuracy gives from 95%-99% some research papers gives 100% Of accuracy

III.METHODOLOGY

Based on our previous literature review, we describe machine learning in this section as the flow of the techniques used for the detection and classification of epileptic seizures using the EEG dataset. We will investigate machine learning methodology.

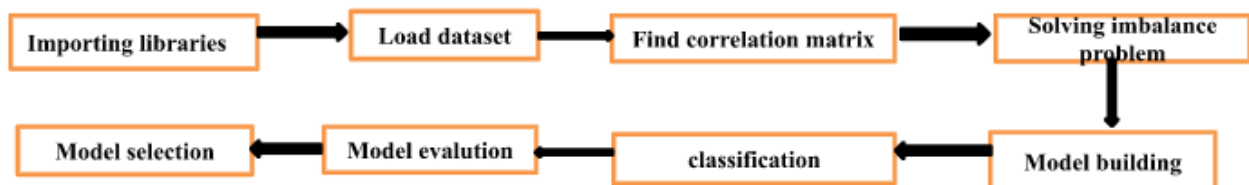


Fig.1: Flow of Machine Learning



The flow diagrams for machine learning methodology shown above are the most effective. This study demonstrates that seizure detection algorithms perform much better depending on the approach used. In ML models, several techniques were utilised, such as K-nearest neighbours, logistic regression, stochastic gradient descend, naive bayes classifier, random forests, decision tree, xg booster, gradient booster, and extra tree classifiers. These models extract distinguishing traits and categorise seizures using conventional statistical algorithms and feature engineering techniques. The effectiveness of these ML models has been useful in understanding how to detect epilepsy, but it is limited by the need for manual feature selection and their inability to detect complex nonlinear correlations in the data.

IV. PERFORMANCE PARAMETERS

Evaluating the performance of a Machine learning model is one of the important steps while building an effective ML model. To evaluate the performance or quality of the model, different metrics are used, and these metrics are known as performance metrics or evaluation metrics. These performance metrics help us understand how well our model has performed for the given data.

1. Accuracy :Ratio of correct predictions to total predictions.Important when: you have symmetric datasets (FN & FP counts are close)Used when: false negatives & false positives have similar costs.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

Where, TP-True Positive , TN-True Negative , FP-False Positive. FN- False Negative

2. Sensitivity/Recall :The ratio of true positives to total (actual) positives in the data.

Important when: identifying the positives is crucial.**Used when:** the occurrence of false negatives is unacceptable/intolerable. You'd rather have some extra false positives (false alarms) over saving some false negatives. For example, when predicting financial default or a deadly disease.

$$\text{Sensitivity or Recall} = \text{TP} / (\text{TP} + \text{FN})$$

3. Precision: The ratio of true positives to total predicted positives.**Important when:** you want to be more confident of your predicted positives. **Used when:** the occurrence of false positives is unacceptable/intolerable. For example, Spam emails. You'd rather have some spam emails in your inbox than miss out on some regular emails that were incorrectly sent to your spam box.**Precision = TP / (TP + FP)**

4. Specificity : Ratio of true negatives to total negatives in the data. **Important when:** you want to cover all true negatives.

Used when: you don't want to raise false alarms. For example, you're running a drug test in which all people who test positive will immediately go to jail.**Specificity = TN / (TN + FP)**

5. F1-Score:Considers both precision and recall. It's the harmonic mean of the precision and recall.

Important when: you have an uneven class distribution.**Used when:** the cost of false positives and false negatives are different. F1 score conveys the balance between the precision and the recall. It is higher if there is a balance between Precision and Recall. F1 Score isn't so high if one of these measures, Precision or Recall, is improved at the expense of the other.**F1 Score = 2 * ((Recall * Precision) / ((Recall + Precision)))**

6. AUC (Area Under Curve): The Area Under the Curve (AUC) is a commonly used metric to evaluate the performance of binary classification model, based on the Receiver Operating Characteristic (ROC) curve.

$$\text{AUC} = \sum_{i=1}^{n-1} (\text{TPR}[i+1] + \text{TPR}[i]) * (\text{FPR}[i+1] - \text{FPR}[i]) / 2$$

7. Prevalence: Prevalence refers to the proportion or percentage of individuals in a population with a particular disease or condition at a specific time.

$$\text{Prevalence} = ((\text{Number of Cases of a Disease or Condition in a Population}) / ((\text{Total Population}) * 100 \%)$$

V. CONCLUSION

This review paper investigated various machine learning (ML) approaches to identify epilepsy seizures. We were able to ascertain the benefits and drawbacks of several ML models for accurately identifying and classifying seizures through a thorough examination of the literature and a comparison analysis of 20 relevant papers. Our main objective was to determine which ML model had the best methodological flow for detecting epileptic seizures. Our findings unambiguously show that machine learning models—particularly XG booster—have excelled at detecting seizures. The

ML models continuously achieved high recall, sensitivity, specificity, sensitivity, and F1 score, proving their effectiveness in correctly classifying seizures and non-seizures. ML performs better and thanks to their strong feature extraction capabilities and their capacity to recognize spatial correlation in EEG signals

REFERENCES

- [1] MilindNatu, MrinalBachute, Shilpa Gite, Ketan Kotecha, and Ankit Vidyarthi “Review on Epileptic Seizure Prediction: Machine Learning and Deep Learning Approaches”, 2022
- [2] Published by Elsevier B.V “The Use of Decision Trees for Analysis of the Epilepsy”, 2021
- [3] Mohammad Khubeb Siddiqui, Ruben Morales Menendez, Xiaodi Huang and Nasir Hussain “A review of epileptic seizure detection using machine learning classifiers”, 2020
- [4] Mahmood A. Jummah, Ammar Ibrahim Shihab, Akeel Abdul Kareem Farhan “Epileptic Seizures Detection Using DCT-II and KNN Classifier in Long-Term EEG Signals”, 2020
- [5] R. B. Messaoud was with Paris Brian Institute, 75013 Paris, France and Engineering school “Random Forest classifier for EEG-based seizure prediction”, 2020
- [6] Fauzia P. Lestari, Mohammad Haekal, RizkiEdmi Edison, Fikry Ravi Fauzy “Epileptic Seizure Detection in EEGs by Using Random Tree Forest, Naïve Bayes and KNN Classification”, 2019
- [7] Elsevier B.V. on behalf of King Saud University “Gradient boosting machines fusion for automatic epilepsy detection from EEG signals based on wavelet features”. Published by, 2018
- [8] Torlay - M. Perrone-Bertolotti -M. Baciú CNRS LPNC UMR 5105, Univ. Grenoble Alpes, 380000 Grenoble, France “Machine learning–XGBoost analysis of language networks to classify patients with epilepsy”, 2017
- [9] E. Kabir School of Agricultural, Computational, and Environmental Sciences “Epileptic seizure detection from EEG signals using logistic model trees”, 2016
- [10] Prabhpreet Kaur Bhatia and Anurag Sharma “Epilepsy Seizure Detection Using Wavelet Support Vector Machine Classifier”, 2016
- [11] Diambra L, Figueiredo J, Malta C. “Epileptic activity recognition in EEG recording”. Phys. A: Stat. Mech. Appl. 1999
- [12] Chen W, Zhuang J, Yu W, et al. “Measuring complexity using FuzzyEn, ApEn, and SampEn”. Med Eng Phys. 2009
- [13] Xie HB, Gao ZM, Liu H. “Classification of ventricular tachycardia and fibrillation using fuzzy similarity-based approximate entropy”. Expert Syst Appl. 2011
- [14] Srinivasan V, Eswaran C, Sriram N “Approximate entropy-based epileptic EEG detection using artificial neural networks”. IEEE Trans Inform Technol Biomed. 2007
- [15] Yuan Q, Zhou W, Li S, et al. “Epileptic EEG classification based on extreme learning machine and nonlinear features”. Epilepsy Res. 2011
- [16] Nicolaou N, Georgiou J. “Detection of epileptic electroencephalogram based on permutation entropy and support vector machine”. Expert Syst Appl. 2012
- [17] Acharya UR, Molinari F, Shree SV, et al. “Automatic diagnosis of epileptic EEG using entropies”. Biomed Signal Process Control. 2012
- [18] Orhan U, Hekim M, Ozer M. “EEG signals classification using the K means clustering and a multilayer perceptron neural network model”. Expert Syst Appl. 2011
- [19] WangMiao D, Xie DC. “Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection”. Expert Syst Appl. 2011
- [20] EEG Database from University of Bonn. Available from: <http://www.epileptologiebonn.de>. June 2013



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details