Epilepsy Prediction using Machine Learning

S Dharshika [1], Sahreen Sajad [2], Sushmitha N [3]

[1,2] Department of Information Science and Engineering, R.V. College of Engineering, Bengaluru, Karnataka, India
[3] Assistant Professor, Department of Information Science and Engineering, R.V. College of Engineering, Bengaluru, Karnataka, India

---***---

Abstract - *Epilepsy has severe impacts on patients including disrupting their social relationships and less mobility. Prediction of the disease can help the patient prevent the onset of seizures with the help of appropriate medication. Since the traditional methods of studying EEG are prone to misdiagnosis, Machine Learning can provide a more accurate diagnosis. In this paper, we aim to survey models to better describe methodologies for a high-precision model to predict epilepsy in patients.*

***Key Words:* Seizure, EEG (Electroencephalogram), Machine Learning, KNN (K-Nearest Neighbour, Logistic Regression, Decision tree.**

1. INTRODUCTION

Epilepsy is a disorder characterized by the patient's frequent seizure episodes. These can be brief or even go undetected. In other cases, seizures are found to showcase constant shaking. Epilepsy occurs because of abnormal functioning in brain activity. WHO defines Epilepsy as "having two or more unprovoked seizures." The disease can have severe complications for the patient from breaking bones to causing accidents. In the case of epilepsy, there is no certain cause for a seizure. Seizures that are caused due to some specific causes are not considered Epilepsy. Epileptic patients usually become victims of social stigma and experience social anxiety more than others.

Epileptic seizures are seen when there is some sort of abnormal activity of neurons in the cortex of the brain. Cortex is the neural tissue outside the cerebrum of the brain and is often called grey matter. This abnormal activity of neurons can be witnessed in the electroencephalogram (EEG) of the patient. In most cases, the underlying cause of the disease remains unknown but could also be due to injuries, stroke, or certain infections.

People carrying the disorder experience frequent bruising or fracturing of bones due to injuries caused by the disorder. The psychological condition of the patients also seems to be affected. Epilepsy patients have higher risks of anxiety and depression due to social stigma as well as lack of self-confidence. Premature deaths also occur thrice higher than the people who do not carry the disorder. Causes of death also include prolonged seizures, accidental falls, or drowning. These can be preventable. Medication can help control the intensity of seizures, but may not always be helpful. In about 31% of cases, the medications are not much helpful. In cases where medication does not provide much help, surgery options may be considered. Epilepsy does not always remain lifelong. The symptoms can reduce to an extent that treatment may no longer be necessary.

About 50 million people worldwide are Epilepsy patients. 5 million people are diagnosed with Epilepsy every year. Countries with better income are less likely to have Epilepsy than those with lesser income. This could be because of the higher tendency in low-income countries to contract birth injuries, malaria, and other diseases.

2. EPILEPSY PREDICTION

2.1 EEG Signals

Electroencephalography (EEG) records brain activity to help understand the brain signals corresponding to the state of a human's brain. When the EEG test is carried out, small sensors are attached to the patient's head which helps to pick up the signals produced by the brain. The signals picked up are recorded by a machine and a diagnosis is made by the doctor.

The EEG diagnosis is a little difficult to carry out and may lead to misdiagnosis in some cases. A trained specialist carries out the test.

In some cases, a normal EEG is not enough to understand which treatment would be the best for the patient. In such cases, there are attempts made to record EEG when the seizure is occurring in the patient. This is referred to as the ictal recording. When EEG signals are taken between seizures, they're referred to as the inter-ictal recording. An ictal recording includes an audio and a video which are synchronized. During such time, medications are usually not given to the patients in case the test is carried out in a hospital. If it's carried out outside the hospital, then the medications are not withdrawn.

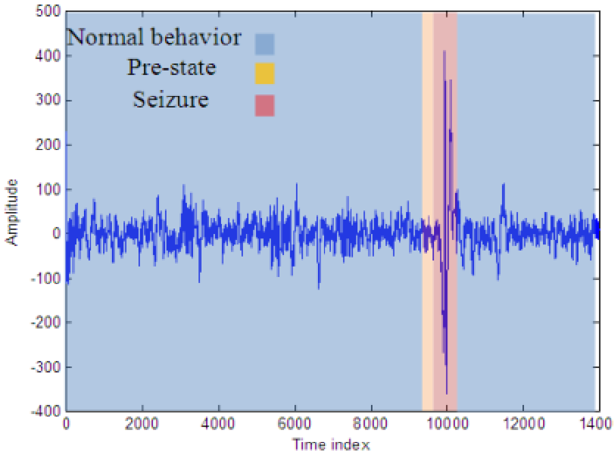

Fig 1: EEG Signal of an Epileptic Seizure [1]

2.2 Related Work

Epilepsy Prediction and Detection have been the focus of research owing to the severe consequences of the disease and that the prediction of the disease could lead to better treatments. Various research papers have been written that propose new methods to predict Epilepsy with higher accuracy and help pre-process the raw EEG data signals. Research popularly involves AI and Machine learning Techniques including Deep Learning. The following includes various related work that has helped us propose a methodology and understand the advancement in the field until now.

Authors in [2] had the objective to predict Epilepsy using Deep Learning techniques. Various datasets have been taken into consideration and a review of various neural networks like CNN is made. The results of the various networks are compared.

Authors in [3] took use of Machine Learning classifiers to predict Epilepsy. This was done using Black Box and Non-Black Box Classifiers. In addition to this, the performance of K-Nearest Neighbours, SVM, and ANN. Etc. is measured and compared.

Authors in [4] had the aim to detect Epilepsy using EEG signals. The paper focuses on comparing the EEG signals of a normal patient to that of an Epileptic patient. Raw EEG waves undergo Fourier Transformation and Short Time Fourier Transformation.

The author in [5] aimed to predict Epilepsy using EEG signals. This paper focused on the problems faced by Deep Learning and Artificial Intelligence in regard to processing the raw EEG signals that may lead to misdiagnosis. Various evaluation metrics are used to display the results and shortcomings of the methods.

Authors in [6] intended to make use of Artificial Intelligence in Epilepsy Detection. This comprised of application of AI including Color Coded Paradigms that could only be studied by experts as well as fully automated analyses that were lesser accurate but could be read by anyone.

Authors in [7] made a comparative study based on EEG and ECG (Electrocardiogram) signals and models were built. Various evaluation parameters were used to compare the results.

Authors in [8] also intended to detect Epilepsy using EEG Signals. This involved the implementation of the Support Vector Machine and Feed Forward Clustering Technique which was done using MATLAB.

Authors in [9] proposed a system that could pre-process EEG signals based on their first and second derivates. The input features are reduced in this model without disturbing the results in any form.

2.3 Methodology

2.3.1 Dataset

The dataset is taken from UCI Machine Learning Repository. 11,500 samples are taken in the dataset. All of these have 158 features each. Five classes are then built to categorize the taken samples referred to as 1,2,3,4,5. Class 5 corresponds to EEG signals when the eyes are open. Class 4 corresponds to EEG signals when the eyes are closed. Class 3 corresponds to if the EEG signal identifies a brain tumor. Class 2 corresponds to the area where the brain tumor is located and finally, Class 1 is the recording of the seizure activity.

A new dataset is obtained from the old one by transforming using discrete wavelet transform. By adjusting the size of the component basis functions, it is simple to estimate a spike in the EEG signal since wavelets have a finite support. Any time-varying signal, for instance, will be divided into smaller uniform functions, referred to as the fundamental functions, using discrete wavelets. A total of 25 features are taken from the dataset that has the highest importance using feature extraction techniques. Random forest features selection technique is used for this purpose. Following are the columns and their importance:

	column	importance
43	43	0.000231
34	34	0.000246
24	24	0.000288
42	42	0.000321
7	7	0.000410
16	16	0.000438
3	3	0.000508
6	6	0.000539
33	33	0.000566
15	15	0.000651
25	25	0.000686
44	44	0.000713
22	22	0.000857
21	21	0.000887
4	4	0.000966
46	46	0.000979
31	31	0.001482
30	30	0.001729
39	39	0.001921
40	40	0.002281
12	12	0.003375
13	13	0.004868
45	45	0.004941
41	41	0.007224
0	0	0.007472

Fig 2: Feature Importance

Following are some of the models that are used along with performance metrics and their comparison:

2.3.2 KNN

This classifier helps to measure the Euclidian distance between the test data and the training data. Here, we do not calculate for the closest image. We look for multiple k-closest images in the training set which is used to predict which class the test image belongs to.

Training: AUC: 0.992, Accuracy: 0.640, Recall: 0.281, Precision: 0.998, Specificity: 0.999, Prevalence: 0.500
Validation: AUC: 0.975, Accuracy: 0.840, Recall: 0.241, Precision: 0.958, Specificity:0.997, Prevalence: 0.208

2.3.3 Logistic Regression

Logistic Regression will help to give a linear relationship between two images. This will in turn help to calculate the true and the false values. If the model is separated based on positive and negative elements, the model works better.

By imposing a linear function on the input features by projecting the sample points onto a line, logistic regression analyses the input characteristics. The logistic regression line is produced by maximizing the log of likelihood, which is the same as maximizing the likelihood, by the linear function, which is achieved by adding the log of the likelihood of each sample point. In a binary classification model, the best fitting function would indicate that the probability of the positive class would be very near to 1 (100 percent), and the probability of the negative class would be very close to 0.

Training: AUC: 0.628, Accuracy: 0.666, Recall: :0.538, Precision: 0.722, Specificity: 0.793, Prevalence: 0.500
Validation: AUC: 0.975, Accuracy: 0.840, Recall: 0.241, Precision: 0.958, Specificity:0.997, Prevalence: 0.208

2.3.4 Support Vector Machine

They select the decision boundary that optimizes the distance from the nearest data points of all the classes, and SVMs vary from other classification techniques. The maximum margin classifier or maximum margin hyperplane is the name of the decision boundary produced by SVMs. The accuracy achieved is 96%.

2.3.5 Naïve Bayes

The number of parameters required for naive Bayes classifiers is linear in the number of variables (features/predictors) in a learning task, making them extremely scalable. Instead of using a costly iterative approximation, as is the case with many other types of classifiers, maximum likelihood training may be performed simply by evaluating a closed-form expression that takes linear time.

Training: AUC: 0.982 Accuracy: 0.930, Recall: 0.890, Precision: 0.967, Specificity: 0.970, Prevalence: 0.500
Validation: AUC: 0.985, Accuracy: 0.960, Recall: 0.916, Precision: 0.893, Specificity: 0.971, Prevalence: 0.208

2.3.6 LSTM

Recurrent neural networks are a type of long short-term memory. The output from the previous phase is sent into the current step of a 5 RNN as input. It addressed the issue of long-term RNN dependency, in which the RNN can predict words from current data but cannot predict words held in long-term memory.

RNN's performance becomes less effective as the gap length rises. By default, LSTM may save the data for a very long time. It is utilized for time-series data processing, forecasting, and classification.

A total of 50 epochs was given to achieve an accuracy of over 95%.

Graphs for loss and accuracy are plotted to record how these kept changing with the number of epochs.

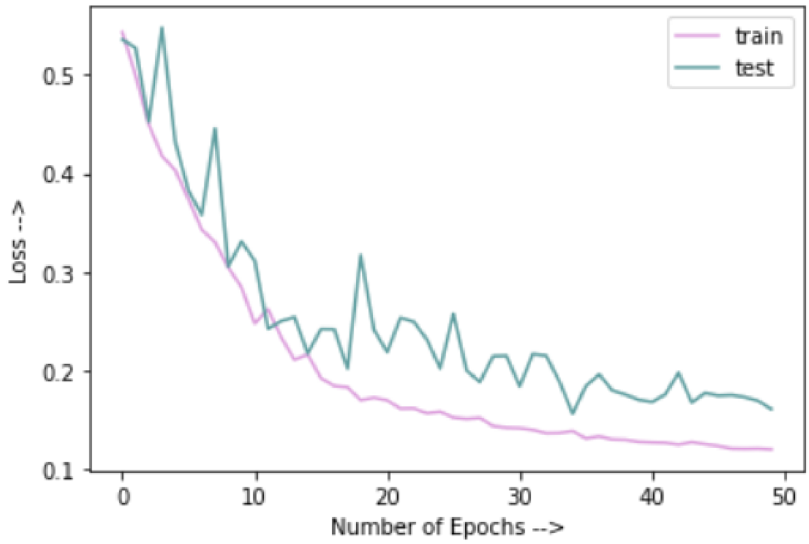

Fig 3: Training v/s Validation Loss

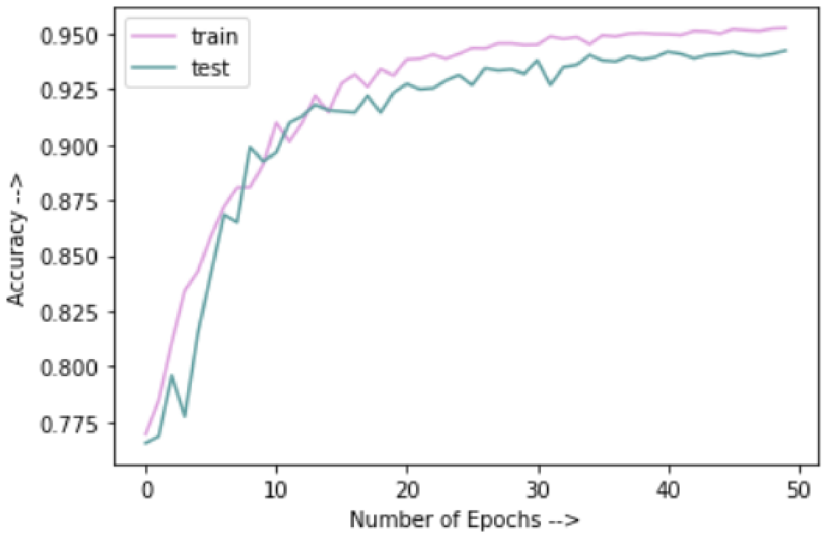

Fig 4: Training v/s Validation Accuracy

2.3.7 Decision Tree

A decision tree often poses a question, then categorizes the sample based on the response. Category or numerical classifications are both acceptable. The tree finishes after the algorithm have split all samples into classes or by satisfying specific criteria of the classifier characteristics.

The classifying method operates by repeatedly dividing data into sub-regions of the same class. A root node, often known as the root, is the top of the tree. There are arrows leading both to and away from internal nodes, sometimes known as simply nodes, which branch out farther. Last but not least, arrows point to Leaf Nodes or just Leaves but not away from them.

Training: AUC: 0.985, Accuracy: 0.982, Recall: 0.964, Precision: 0.999, Specificity: 0.998, Prevalence: 0.500
Validation: AUC: 0.865, Accuracy: 0.909, Recall: 0.853, Precision: 0.744, Specificity: 0.922, Prevalence:0.208

2.3.8 Random Tree

This will be used if a random rule is used to classify the dataset. To simulate the decision tree, error pruning is used so that the data gets randomized. The batch size is taken to be 100.

2.3.9 Random Forest

Random Forest in itself is a collection of tree classifiers that give an average output of the multiple tree classifiers. Randomness could either be the number of rows in the original dataset or the number of columns or branches of each tree in the dataset. To simulate Random Forest, the seed is taken as 1. These represent the number of threads that will be used to construct the forest.

Batch size is taken as 100 and the number of iterations is also taken to be 100. These are the number of trees in the forest. By bootstrapping the set of samples or utilizing an arbitrary number of characteristics at each split, a random forest is created by grouping decision trees that are not connected with one another. As Random Forest blends flexibility with decision tree simplicity, it improves accuracy, which is one of the decision trees' limitations as a classifier.

Training: AUC: 0.997, Accuracy: 0.964, Recall: 0.943, Precision: 0.985, Specificity: 0.985, Prevalence: 0.500 6
Validation: AUC: 0.990, Accuracy: 0.958, Recall: 0.937, Precision: 0.869, Specificity: 0.963, Prevalence: 0.208

2.4 RESULTS

Various performance metrics are taken into consideration to find out which model performs the best. A threshold value is set to 0.5. In order to compare, two of the performance metrics are taken for visualizing i.e, AUC and accuracy. Following is the output obtained:

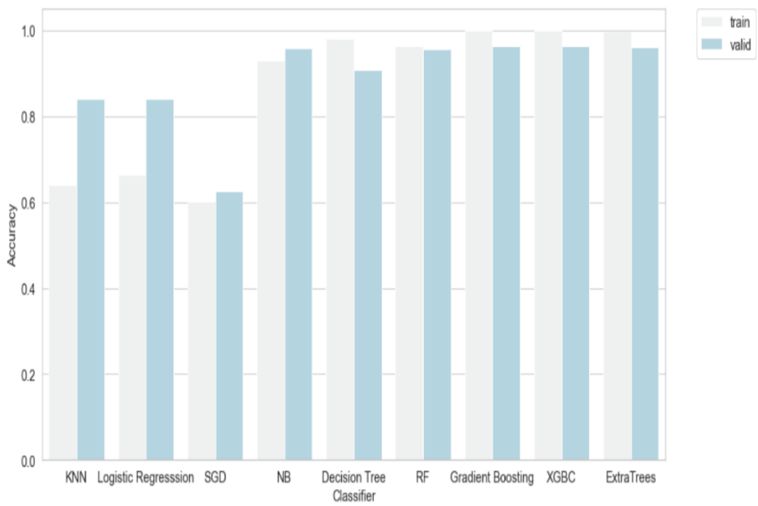

Fig 5: Accuracy of each model

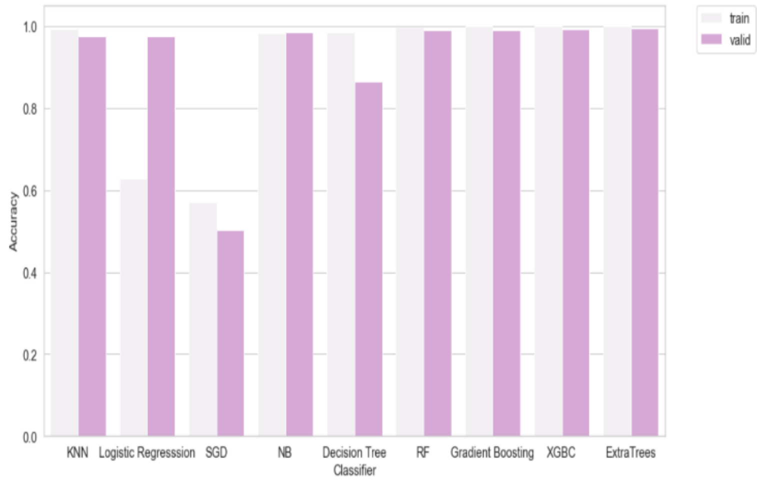

Fig 6: AUC of each model

3. CONCLUSIONS

The prediction of Epilepsy is still undergoing research and no such system has been built yet that is 100% reliable to be used by experts in the medical field. Besides, these systems require some sort of technical expertise as well as medical expertise which may or may not be always present.

The usage of EEG signals is preferred over ECG signals but these require a better form of preprocessing. The raw data form needs better handling. Epilepsy cases are more prevalent in countries that don't have a great economic setup, as such, the need is to deploy such systems in these countries where Epilepsy is growing every year and medical treatment is almost negligible.

This paper proposes models to better predict Epilepsy in patients. As more cases are emerging every year, this paper aims to help researchers better choose a model that can help patients who are prone to become Epileptic so as to provide better treatment and health facilities.

The patient's EEG signals would help the system to predict if they have Epilepsy or not. The model is expected to train using the dataset of various epileptic and non-epileptic patients so as to help predict consecutive seizures in epilepsy.

The goal of this work is to help patients with the disorder as they usually become prone to various physical injuries like accidents or falls. Since the social stigma around such patients who suffer severe episodes of seizure is high, the self-confidence of such patients seems to be lacking and becoming the cause of other problems like depression and anxiety. We aim to help predict the neurological disorder in susceptible patients so that the right course of treatment is given at the right time so that it helps to reduce or maybe, even prevent the disorder.

REFERENCES

[1] Mostafa I. El Sayeid, Entessar Gemeay, Salah Khames, Turkey Alotaiby, Saleh A. Alshebeili, Fathi E. Abd El-Samie, "Statistical Analysis of EEG Signals in 7 Wavelet Domain for Efficient Seizure Prediction." American Journal of Biomedical Engineering.using sentiment", March 2016

[2] Shoeibi, A.; Khodatars, M.; Ghassemi, N.; Jafari, M.; Moridian, P.; Alizadehsani, R.; Panahiazar, M.; Khozeimeh, F.; Zare, A.; Hosseini-Nejad, H.; Khosravi, A.; Atiya, A.F.; Aminshahidi, D.; Hussain, S.; Rouhani, M.; Nahavandi, S.; Acharya, U.R. Epileptic Seizures Detection Using Deep Learning Techniques: A Review. Int. J. Environ. Res. Public Health 2021, 18, 5780.

[3] Siddiqui, M.K., Morales-Menendez, R., Huang, X. et al. A review of epileptic seizure detection using machine learning classifiers. Brain Inf. 7, 5 (2020).

[4] S. Gupta, S. Bagga, V. Maheshkar and M. P. S. Bhatia, "Detection of Epileptic Seizures using EEG Signals," 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), 2020, pp. 1-5, doi: 10.1109/AISP48273.2020.9073157.

[5] K. Rasheed et al., "Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review," in IEEE Reviews in Biomedical Engineering, vol. 14, pp. 139- 155, 2021, doi: 10.1109/RBME.2020.3008792.

[6] Kaur, Taranjit & Diwakar, Anirudra & Kirandeep, & Mirpuri, Pranav & Tripathi, Manjari & Chandra, PSarat & Gandhi, TapanK. (2021). Artificial Intelligence in Epilepsy. Neurology India. 69. 10.4103/0028-3886.317233.

[7] Yang, Yikai & Truong, Nhan & Maher, Christina & Nikpour, Armin & Kavehei, Omid. (2021). A comparative study of AI systems for epileptic seizure recognition based on EEG or ECG. 2021. 2191-2196. 10.1109/EMBC46164.2021.9630994.

[8] Bhatia, Prabhpreet & Sharma, Anurag. (2016). Epilepsy Seizure Detection Using Wavelet Support Vector Machine Classifier. International Journal of BioScience and Bio-Technology. 8. 11-22. 10.14257/ijbsbt.2016.8.2.02.

[9] Brari, Zayneb & Belghith, Safya. (2021). A novel Machine Learning approach for epilepsy diagnosis using EEG signals based on Correlation Dimension. IFACPapersOnLine. 54. 7-11. 10.1016/j.ifacol.2021.11.018.

[10] R.G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, C.E. Elger Indications of nonlinear deterministic and finite dimensional and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state.

[11] Samiee, K., Kovacs, P., and Gabbouj, M. "Epileptic seizure classification of EEG time-series using rational discrete shorttime fourier transform," 2019 IEEE

[12] S Syed Muhammad Usman, Shehzad Khalid, and Muhammad Haseeb Aslam. "Epileptic seizures prediction using deep learning techniques," 2020 IEEE

[13] Nhan Duy Truong, Levin Kuhlmann, Mohammad Reza Bonyadi, Damien Querlioz, Luping Zhou, and Omid Kavehei. "Epileptic seizure forecasting with generative adversarial networks," 2019 IEEE

[14] Levin Kuhlmann, David B Grayden, Fabrice Wendling, and Steven J Schiff. "The role of multiple-scale modelling of epilepsy in seizure forecasting," 2015 American Electroencephalographic Society

[15] Djoufack, Laurent & Tchiotsop, Daniel & Atangana, Romain & Louis-Dorr, Valérie & Wolf, Didier. (2021). Classification of EEG signals for epileptic seizures detection and eye states identification using Jacobi polynomial transforms-based measures of complexity and least-square support vector machine. Informatics in Medicine Unlocked. 23. 10.1016/j.imu.2021.100536.