Machine Learning Algorithm to Detect Neurological Disorder

J. V. Alamelu[1*], and *M.* D. Nandeesh[1]

[1]Department of Electronics and Instrumentation Engineering, Ramaiah Institute of Technology, Bengaluru, India

Abstract. Abnormalities of brain is the main concern for healthcare industries, if it is not recognized or treated may lead to sever brain death. Neurological disorder such as epilepsy is slower and requires constant diagnosis and treatment. With the advent of recent technology physiologists are using x-rays, EEG, CT and MRI for early detection of such disorders. The diagnosis is accomplished by implementing machine learning techniques which aid in proper treatment. In this paper signals are acquired using electroencephalogram, a technique, that consists of parameters to classify different neurological disorders such as anxiety, epilepsy spikes etc., For this purpose, Machine Learning algorithms is implemented to categorize, which performs the classification of different brain abnormalities especially on epilepsy. Accuracy seems to be 96.1 under Extra Tree Classifier. This act as the assisting platform for doctors, aimed at effective treatment.

1 Introduction

Electroencephalogram (EEG) has turn out to be one of the best and important tools employed to diagnose various brain neurological abnormalities by physicians. Over past few decades EEG have been used by physician or neurologists mainly due to its low cost, minimum effort and availability of machine to diagnose several disorders in brain such as epilepsy, seizure disorder, sleep disorder, brain tumors etc., Epilepsy is one of the identified as life threatening and common brain disorder disturbing human beings. According to the WHO during 2020, five crore patients are suffering from epilepsy in united states [1]. Major cause of brain diseases are seizures, tumors, stroke, trauma, inflammation etc., The general technique to detect seizure by using non-invasive technique with multichannel recording electrical brain activities using EEG. Neurological disorders generally affect the brain, spinal cord as well as nerves system found throughout the body. Feature based on entropy for automatic detection of epileptic seizure is implemented by discrete wavelet transform for analysis of EEG signal. To improve the diagnosis accuracy a combination of different learning models is utilized to overcome the error of individual technique. This algorithm has shown a better accuracy and better effective tool for real time application. DWT has remained generally used in analysis of EEG signal based on automatic detection of diseases such as seizures. To obtain better accuracy and minimum computational cost, wavelets and decomposition plays an important role. This technique has provided a method for settings of DWT affect and its enactment suitable for detection of seizure [2]. Epilepsy datasets from patients can be used in detection of seizure as well as prediction algorithms, but main drawback is reproducibility as well as generalizability of algorithm makes it difficult in findings proper recognition. These studies have recognized seventeen features that finds datasets that are unique from each other [3]. Epilepsy detection and diagnosing epilepsy is tedious as well as obviously time-consuming. Timely and accurate detection is crucial to initiate proper drugs as well as therapy that will minimise the complication of seizures. Feature extraction based on CNN is used to detect different segments. Results has demonstrated that frequency domain signals have better results in CNN application [4]. Classification and combination of signals of EEG for detecting epileptic seizure is implemented by utilizing CNN techniques. The required features are extracted features using ANN, SVM, naive Bayes and K-nearest neighbours and classified [5]. Improved feature by utilizing CNN-SVM method has produced better classification in detecting epileptic using EEG detection with better accuracy, when compared to SVM and RF method. These consequences specified that deep learning techniques with instinctive feature extraction could improve the classification when compared using traditional technique [6]. Real-time methodology for detecting epilepsy using EEG signal using STFT and google net CNN has produced better accuracy and sensitivity[7],[8]. This technique utilizes a sliding window method for real time application [9]. Expressive deep network model based on ResNet for an efficient classifies seizure types from EEG signals [10]. ResNet technique using deeper network is used for detecting seizure. Results exposed a better performance in both F1 score and accuracy [11],[12]. Deep learning approaches can be used for better results using feature extraction and channel selection. Seizer detection and EEG classification using time series data has been implemented by using models such as fully connected neural networks [13], recurrent neural networks (RNN) [14], and convolutional neural networks (CNN) [15].

* Corresponding author: jvalamelu@msrit.edu

The EEG features for seizure are mainly sharp and spikes waves. Detection using combination of DCT and Deep learning CNN algorithm with better accuracy was obtained [16],[17]. Detection and classification was explored by utilizing support vector machine as well as orthogonal matching pursuit to obtain a better result [18]-[20]. With the advent of the proposed work, diagnosis of neurological abnormalities can be handled at an early stage. This can aid physician to identify suitable diagnosis as well as preventive measures. This work is intended to predict whether the patient suffering from epilepsy. The identification will be derived from the EEG signals and different classification methods are proposed to determine the accurate prediction.

The paper is organized with Materials and Methods, Results and Discussions and Conclusion in the remaining sections.

2 Materials and Methods

The data has been obtained from UCI's machine learning repository in order to perform the crucial task of detecting anomalies in the brain. There are 500 patients in the dataset under consideration, and each patient has 4097 electroencephalogram (EEG) values. Following an equal division of the 4097 data points into 23 chunks per patient, each chunk was converted into a single row in the dataset. The 178 readings in each row are converted into columns; in other words, one second of EEG readings is made up of 178 columns. Overall, there are 11,500 rows and 179 columns, with the patient's status. From the dataset, morphological feature such as spikes peak, duration and its rate has been considered and classified for epilepsy occurrences.

The data pre-processing is shown in Figure 1.

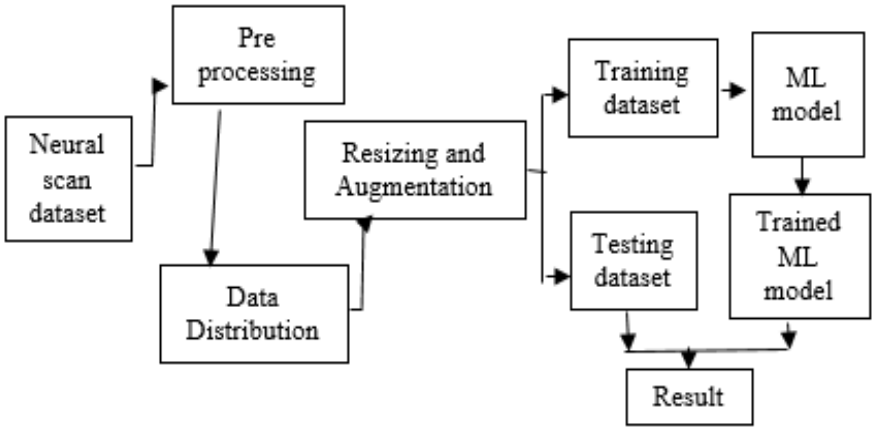

Fig. 1. Block diagram to detect abnormalities in the brain.

After the data pre-processing steps, the dataset for training and testing has been segregated in the ratio of 70:30. Further dataset is modelled using different machine learning techniques with 10 fold cross validation. The current research is to classify different abnormalities using machine learning techniques. In this work K Nearest Neighbour, Logistic Regression, Stochastic Gradient Descent, Naive Bayes, Decision Tree, Random Forest and Extra Tree Classifier have been chosen. Accuracy of the classification from the considered dataset has been compared using different classifiers. The purpose of the study is to obtain better accuracy and AUC, in order determine better model performance, seven different models has been considered to identify it. The work is performed using python language with Scikit-learn.

2.1 K Nearest Neighbour (KNN)

This describes the concept behind MacQueen's K-Means algorithm: When n data points $(x_1, x_{2,\ldots.}x_n\})$ are specified, K cluster centers $(a_1, a_{2,\ldots.}a_n\})$ are chosen so that the square sum of the distances between each data point and its nearest cluster center is as little as possible. This square sum of the distances is known as the objective function Wn, and its mathematical expression is given in Equation (1).

$$W_n = \sum_{i=1}^{n} min_{i<j<k} \left|x_i - a_j\right|^2 \quad (1)$$

2.2 Logistic Regression (LR)

A extended linear model, which is a generalization of the ideas and capabilities of regular linear models, is what logistic regression is. By projecting the sample points onto a line, logistic regression applies a linear function to our input features. Each sample point's log of likelihood is added up to perform the linear function, which then maximizes the log of likelihood—which is equivalent to maximizing the likelihood—to create the best-fitting logistic regression line. In a binary classification model, the best fitting function would predict that the probability of the positive class would be very near to 1 (100%) and the probability of the negative class would be very close to 0.

2.3 Stochastic Gradient Descent (SGD)

A technique termed gradient descent (GD) optimizes a wide range of loss functions for a variety of model classes, including clustering, logistic, and linear regression. GD plots every potential sum of squared residuals in a linear regression model first, then takes its derivative at a selected point to determine the slope. GD then fines the lowest slope of the curve, and in linear regression it's when the derivative is 0. In other models, GD finds the minimum slope by taking initial guesses at big steps and smaller steps when it is closer to reaching

the minimum value; this makes gradient descent very useful when it is not possible to solve for where the derivative equals 0.

2.4 Naive Bayes (NB)

Based on the Bayes Theorem, the aforementioned classification technique makes the assumption that a feature's inclusion in a class has nothing to do with the presence of any other features.

2.5 Decision Tree (DT)

DT is one of the most used non-parametric supervised learning techniques. For both the classification and regression problems, DT learning techniques are applied. Decision trees often operate on the basis of query response patterns. Data is repeatedly separated into sub regions of the same class by the classifying algorithm, and the tree ends when all samples have been classified into pure classes or when certain classifier attribute criteria have been met.

2.6 Random Forest (RF)

In order to create a random forest, decision trees that are uncorrelated with one another are bagged using either a random number of features at each split or bootstrapping the sample set. One of the shortcomings of decision trees as a classifier is their lack of flexibility; RF addresses this issue by combining the simplicity of decision trees with flexibility, which leads to a significant increase in accuracy.

2.7 Extra Tree Classifier (ETC)

It is an ensemble machine algorithm that combines the prediction from many decision trees to give an output.

3 Results and Discussions

The detection and classification of such abnormalities are essential for preterm infant and neonatal not fetal. In this work, detection of epilepsy form the considered dataset has been detected and classified using machine learning procedures. The EEG dataset has been trained and tested to determine the classification accuracy. In this paper accuracy has been considered along with AUC has these two parameters indicates the models performance. The high value of AUC indicates that model performs better. Accuracy inturn indicates the models error statistically. Using extra tree classifier and stochastic gradient algorithm, the accuracy and AUC obtained are tabulated in the table below. The comparison of AUC, Accuracy for different ML models considered in this work is shown in Figure 2, Figure 3 and in Table 1 below. From table it is inferred that ETC provides better accuracy and AUC when compared to other classifiers.

The confusion matrix is shown for the extra tree classifier model, which indicates the performance of the classification in predicting epilepsy. From the results obtained, we were able to plot create the confusion matrix for the Extra Tree Classifier Model and shown in Figure 4. Classification accuracy can be identified form ROC curve that specifies the degree or separable measure. It is shown in Figure 5. running the Extra Tree Classifier model. It is also able to plot the ROC curve for the Extra Tree Classifier, given below. Figure 5 shows the ROC curve plotted for the Extra Tree Classifier model.

Table 1. Comparison of accuracy and AUC for different classifiers.

Algorithms	AUC	Accuracy
K Nearest Neighbour	0.968	83.3
Logistic Regression	0.56	69.1
Stochastic Gradient Descent	0.524	62.3
Naive Bayes	0.983	96.8
Decision Tree	0.855	89.9
Random Forest	0.989	95.6
Extra Tree Classifier	0.992	96.1

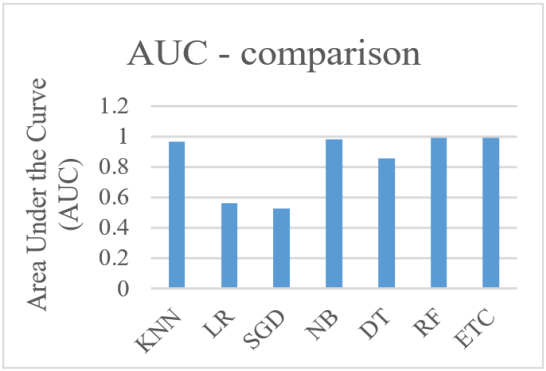

Fig. 2. AUC comparison for different models.

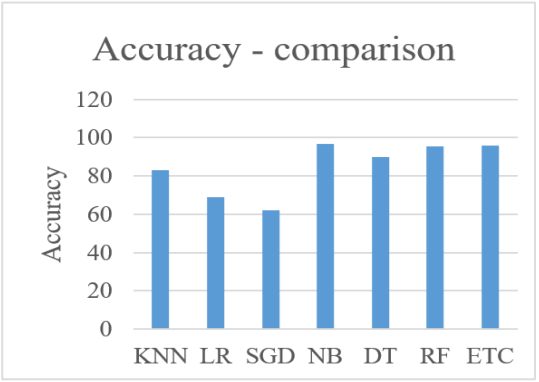

Fig. 3. Percentage accuracy for different models.

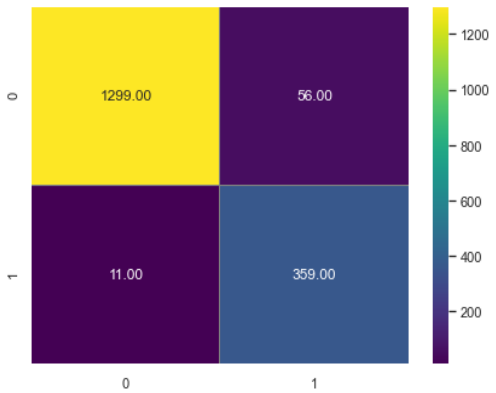

Fig. 4. AUC Comparison for different models.

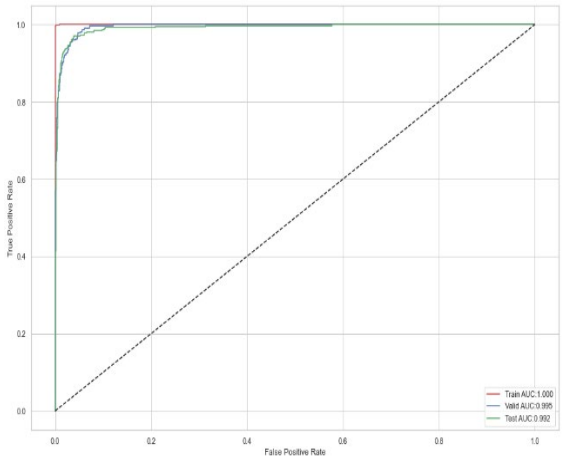

Fig. 5. ROC curve for extra tree classifier.

4 Conclusion

The proposed work has been carried out in detecting epilepsy in a subject at all stages, including the earlier stages. The Extra Tree Classifier gives an AUC closest to 1 and the highest accuracy when compared to the other models. It is successful in reducing the time it takes to provide a similar diagnosis. This also aims in reducing the overall cost of detecting said abnormalities as requires a fewer number of trips to the doctor and reduce the overall hassle for the patient.

This technique will aid the neurologist to detect the disorders and further diagnosis and treatment can be made simpler. In this work K Nearest Neighbour, Logistic Regression, Stochastic Gradient Descent, Naive Bayes, Decision Tree, Random Forest and Extra Tree Classifier have been used and the accuracy of the classifiers has been compared. From the result it has been observed that Extra tree classifier has shown a better performance in both AUC and accuracy which is .992 and 96.1%

References

1. M. Dastgoshadeh, Z. Rabiei, Detection of epileptic seizures through EEG signals using entropy features and ensemble learning. Front. Hum. Neurosci. **16**, 1084061 (2023). https://doi.org/10.3389/fnhum.2022.1084061
2. D. Chen, S. Wan, J. Xiang, F.S. Bao, A high-performance seizure detection algorithm based on Discrete Wavelet Transform and EEG. PLoS One **12**, e0173138 (2017). https://doi.org/10.1371/journal.pone.0173138
3. S. Wong, A. Simmons, J. Rivera-Villicana, S. Barnett, S. Sivathamboo, P. Perucca, Z. Ge, P. Kwan, L. Kuhlmann, R. Vasa, K. Mouzakis, EEG datasets for seizure detection and prediction – A review. Epilepsia Open **8**, 252-267 (2023). https://doi.org/10.1002/epi4.12704
4. M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, J. Xiang, Epileptic seizure detection based on EEG signals and CNN. Front. Neuroinform. **12**, 95 (2018). https://doi.org/10.3389/fninf.2018.00095
5. C. Daftari, J. Shah, M. Shah, Detection of epileptic seizure disorder using EEG signals. Artif. Intell.-Based Brain-Comput. Interface, 163-188 (2022). https://doi.org/10.1016/B978-0-323-91197-9.00006-0
6. R. Du, J. Huang, S. Zhu, EEG-based epileptic seizure detection model using CNN feature optimization, in Proceedings of the 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, Beijing, China, December 21 (2022), 1-6
7. M. Shen, F. Yang, P. Wen, A real-time epilepsy seizure detection approach based on EEG using short-time Fourier transform and Google-Net convolutional neural network. Heliyon **10**, e31827 (2024). https://doi.org/10.1016/j.heliyon.2024.e31827
8. H. Alshaya, M. Hussain, EEG-based classification of epileptic seizure types using deep network model. Mathematics **11**, 2286 (2023). https://doi.org/10.3390/math11102286
9. C. Gómez, P. Arbeláez, M. Navarrete, Automatic seizure detection based on imaged-EEG signals through fully convolutional networks. Sci. Rep. **10**, 21833 (2020). https://doi.org/10.1038/s41598-020-78784-3
10. A. Shoeibi, M. Khodatara, N. Ghossemi, Epileptic seizures detection using deep learning techniques: a review. Int. J. Environ. Res. Public Health **18**, 5780 (2021). https://doi.org/10.3390/ijerph18115780
11. U. Asif, S. Roy, J. Tang, S. Harrer, SeizureNet: Multi-spectral deep feature learning for seizure type classification. Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology, Springer, Cham December 31 (2020), 77-87
12. M.K. Alharthi, K.M. Moria, D.M. Alghazzawi, H.O. Tayeb, Epileptic disorder detection of seizures using EEG signals. Sensors **22**, 6592 (2022). https://doi.org/10.3390/s22176592
13. A. Zarei, B. Mohammadzadeh, Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy-based features of EEG signals. Comput. Biol. Med. **131**, 104250 (2021). https://doi.org/10.1016/j.compbiomed.2021.104250
14. E. Shoka, A. Dessouky, A. El Sayed, EEG seizure detection: concepts, techniques, challenges, and future trends. Multimed. Tools Appl. **82**, 42021-42051 (2023). https://doi.org/10.1007/s11042-023-15052-2
15. S. Saminu, G. Xu, S. Zhang, I.A. el Kader, A recent investigation on detection and classification of epileptic seizure techniques using EEG signal. Brain Sci. **11**, 668 (2021). https://doi.org/10.3390/brainsci11050668

16. P. Boonyakitanont, A. Lekuthai, K. Krisnachal, J. Songsiri, A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. Biomed. Signal Process. Control **57**, 101702 (2020). https://doi.org/10.1016/j.bspc.2019.101702
17. L.V. Tran, H.M. Tran, T.M. Le, Application of machine learning in epileptic seizure detection. Diagnostics **12**, 2879 (2022). https://doi.org/10.3390/diagnostics12112879
18. H. Kode, K. Elleithy, L. Almazedah, Epileptic seizure detection in EEG signals using machine learning and deep learning techniques. IEEE Access **12**, 80657-80668 (2024). https://doi.org/10.1109/ACCESS.2024.3409581
19. D.P. Dash, M. Chakraborty, M. Khorayi, Review of machine and deep learning techniques in epileptic seizure detection using physiological signals and sentiment analysis. ACM Trans. Asian Low-Resour. Lang. Inf. Process. **23**, 1-29 (2024). https://doi.org/10.1145/3552512
20. S. Shanmugam, S. Dharmar, Implementation of a non-linear SVM classification for seizure EEG signal analysis on FPGA. Eng. Appl. Artif. Intell. **131**, 107826 (2024). https://doi.org/10.1016/j.engappai.2023.107826