Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary | Peer Reviewed | Open Access
Volume: 03, Number: 03, Pages: 84 - 88, Year: 2020

Comparative Analysis of Classifiers for Prediction of Epileptic Seizures

Shehzaib Shafique[1], Saba Sarfraz[1], Usman Qamar Shaikh[1], Aamash Nadeem[2] and Zia Ur Rehman[1]

[1]Biomedical Engineering Department, Riphah International University, Islamabad, 44000, Pakistan
[2]Computing Engineering and Built Environment Department, Ulster University, Belfast, BT15 1ED, United Kingdom

Corresponding author: Shehzaib Shafique (Email: shehzaibshafique@gmail.com).

Abstract- **Epilepsy is a neurological disease in which people suffer from seizure attack and lose the normal function of brain. Almost 50 million people have epilepsy in the world due to which it has become the most common neurological disease. Early prediction of epilepsy helps patients to avoid epilepsy and live normal life. Many studies have been conducted for the early prediction of epilepsy. However, selection of the most appropriate classifier has always been a question that needs to be resolved. In this study, we are using six classifiers of machine learning which are KNN, Naïve Bayes, Linear Classification Model, Discriminant Analysis Model, Support Vector Machine and Decision Tree, to find the best classifier for the prediction of epileptic seizures, in term of accuracy. Dataset from "Kaggle" was used. Preprocessing and cross-validation of the data was carried out for training and testing of classifiers. The results depict that Naive Bayes classifier has a better average accuracy of 95.739% as compared to other classifiers. The future work of this study is to implement suggested model in real time, so that the workload of medical members could be reduced.**

***Index Terms*—Classifiers, Epilepsy, Epileptic seizures, Epileptic seizures detection, Machine learning algorithms, KNN, Naïve Bayes, SVM.**

I. INTRODUCTION

Epilepsy is a neurological disorder which is characterized by abnormal functioning of brain, seizures of unusual behavior, sensation, and loss of awareness. According to WHO, almost 50 million people suffer from epilepsy in world and because of this it has become the most common neurological disease globally. Any person can develop epilepsy. There is no age and gender restriction in developing epilepsy. Even child can suffer from epilepsy. There are many reasons of epilepsy, but the most common causes of epilepsy are loss of oxygen, stroke, head injury, brain tumor and any other type of brain damage.

Seizures have different symptoms. Some people stare blankly, and some people twitch their legs and arms during seizures. Based on single seizure, a person cannot claim of having epileptic seizures. For the diagnosis of epilepsy, almost two unprovoked seizures are required. Epilepsy can be treated with medication or surgery [1].

70% of the people suffering from epilepsy can live seizures free life if epilepsy get diagnosed and treated properly. Early detection of epilepsy helps in giving enough time that it can be avoided, and person can remain normal in his daily life. In this era of technology, it has become possible for us to detect epilepsy before its onset and avoid it by giving medicines to patients. Since the technological era, epilepsy detection is very concerned topic among researchers. Many researches have been carried out to predict it by using machine learning and avoid epilepsy before its onset. Machine learning gives a new way to predict epilepsy. Although this technique is in a testing phase but still it will help patient to live their normal life and it will decrease the workload on medical workers. Many researchers used different machine learning algorithms to predict epilepsy. In the study of Haotian Liu [2], the authors used online available data set of EEGs and used six machine learning algorithms and 3 deep learning algorithms to detect epilepsy. They did normalization, binarization and cross validation of data. After getting accuracy from different algorithms, they made conclusion that random forest and Gradient Boosting Decision Tree are better than other because they gave accuracy greater than other. But there were some limitations in their study, as they said in paper "Firstly, we find epilepsy in the signal of one second but in real study it has more clinical meaning to identify seizure in real time base". Secondly, they did not make any prominent efforts in refinement of neural network. In the study of Bandarabadi [3] the authors used data of EEG of 24 patients. After preprocessing of the signals, they used SVM model for epileptic seizures prediction. Their model gave 75.8% accuracy with false production rate of 0.1 per hour. The signal collected from scalp gave slightly less accuracy than the signal collected from the intercranial region. In their study they also said that intercranial EEG gives more clear but localized epileptic information than the EEG recorded from scalp. But the authors concluded that intercranial EEG is only restricted to the area of surgery or area of placement of electrode, but scalp EEG give generalized spatial view of brain activity. In their study, the authors mentioned one of the very interesting things to reduce artifact. They said that by regularization of output of classifier, motion artifact can be reduced. They also said that to get the good results, in real time, from the predictor they made, the predictor must be personalized to each patient. This means that the predictor must be trained on each patient's data.

In the study of Karlık [4] the authors used data of 200 normal people and 200 epileptic patients to train their models. The authors used two extraction methods: wavelet transform vector and auto regressive vector. The authors did cross validation of data and applied different algorithms: K-Means, KNN, Naïve Bayes, ANN and SVM. After finding out their accuracy the authors concluded that the KNN performed well on their data set as it gave greater than 99% accuracy in both extraction methods. In the study of chakraborti [5] the authors used ANN model to predict seizures. In their study they used the scalp recorded EEG signals. The targets for ANN were selected according to dataset. The authors used two ANN architecture in order to perform their methodology: perceptron neural network and back propagation network. The performance of both the neural network was measured in term of time taken; percentage accuracy and mean square error. Both the neural network showed 100% accuracy. The authors also suggested the future work by saying that the model must be applied in real time on patients, instead of data set.

In the study of Jaiswal [6] the authors used two techniques of PCA, sub-pattern PCA (SpPCA) and crossed-pattern PCA (SubXPCA), along with SVM to detect epilepsy. SubXPCA and SpPCA were used for features extraction and SVM model is used for prediction. The authors obtained 100% accuracy for epileptic seizures detection. In the study of Xiang [7] the author first found the fuzzy entropy and sample entropy of all the EEG electrode and then features was extracted to form the eigenvectors. Training samples along with labels were made from eigenvectors. These training sample were then fed into SVM to train model for the prediction of epilepsy. After applying the testing data to SVM model, the result showed that accuracy of fuzzy entropy was greater than sample entropy. The authors also suggested the future work by saying that this model must be used in real time to decrease the workload of medical staff. In the study of Usman [8] the authors aimed to detect the preictal state as it occurs before the onset of epileptic seizures. For this purpose, the authors used scalp recorded EEG dataset to train their SVM model. They used SVM because they compared the accuracy of SVM with other two classifiers and found out that SVM gave best accuracy than others.

Moreover, they succeeded in their aim because their model was able to detect seizure on average 23.6 minutes before its onset. Similarly, another study was carried out to detect epileptic seizures. The authors used online available dataset and used KNN and SVM classifiers after cross validation to detect epileptic seizures. The result shows that SVM gave 100% accuracy in comparison to KNN [10].

In our study, we aim to detect epileptic seizure using machine learning. For this:

- We are using six classifiers and will find that which classifier will give the best accuracy in the prediction of epileptic seizures.
- Patients could also be treated, before the onset of epileptic seizure, by medications.
- Also, our study will help to decrease the workload of medical personal in detection of epileptic seizures before the time.

II. METHODOLOGY

A. DATASET

The source of dataset is "Kaggle" website [9]. From the reference, the original dataset was consisting of five folders, each folder contains 100 files, and each file is the EEG recording of single object for 23.6 seconds. Each signal had 4097 datapoints. 4097 datapoints was reshaped into 23 chunks. Each chunk contains 178 datapoint of brain activity for 1 second. Consequently, the dataset contains 11500 subjects with each subject contain a label of brain activity regarding 178 data point as features. The data has five different labels: 1, 2, 3, 4, 5. each label represent different activity of brain, which are:

1: Seizure. Seizure activity was recorded.

2: tumor. The patient had tumor, and EEG was recorded from tumor/epileptic area.

3: healthy area. The patient had tumor and EEG was recorded from non-epileptic area.

4: eyes closed. The EEG was recorded when the eyes were closed of the patient/subject.

5: eyes open. The EEG was recorded when the eyes were open of the patient/subject.

From all the labels, only 1 represent the seizure activity and rest of the labels represent the non-seizure activity.

B. PREPROCESSING

The data was preprocessed in order to train classifiers. Firstly, we looked for the missing data, but no missing data was found. As it can be seen from Fig. 1(a), 1(b), 1(c), 1(d), and 1(e). The smallest value of data was so smaller than the rest of data. So, to cope-up with this, we normalized the data and set the values in the range of -1 and 1. Moreover, the data had different labels and only label 1 represented the seizure activity. So, we did binarization of data. We classified the data into seizure activity (label 1) and non-seizure activity (label 0).

After binarization, we performed cross validation. The data set was split into ten folds, one-fold consisted of 10% of data and it was used as testing data and other folds was used as testing data. After cross validation, we applied different machine learning classifiers, which will be discussed in later section.

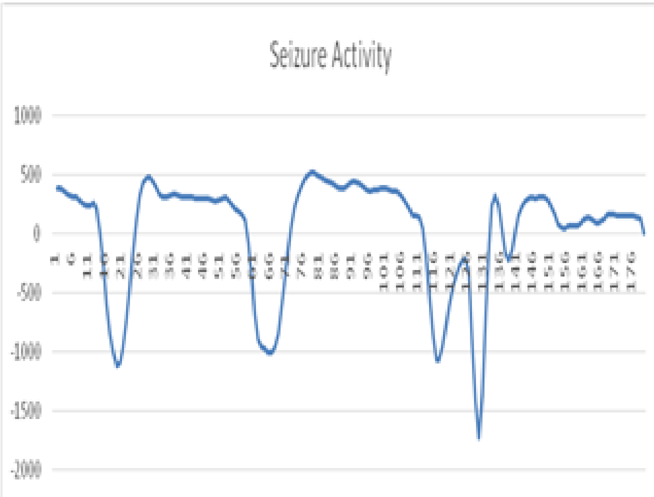


FIGURE 1(a): EEG signal recorded during Seizure activity

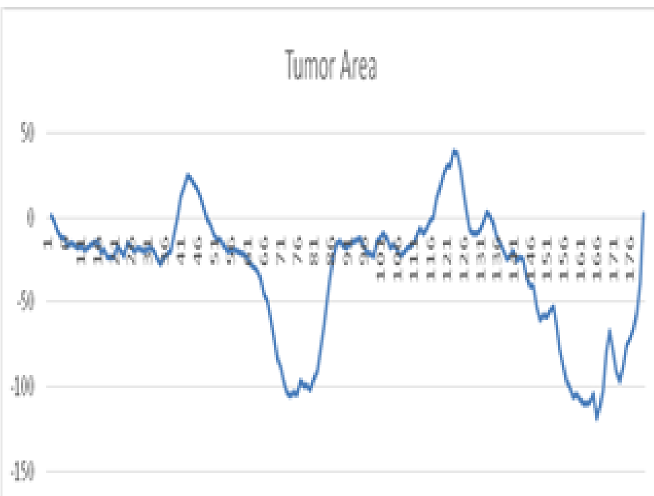


FIGURE 1(b): EEG signal recoded from Tumor area of brain

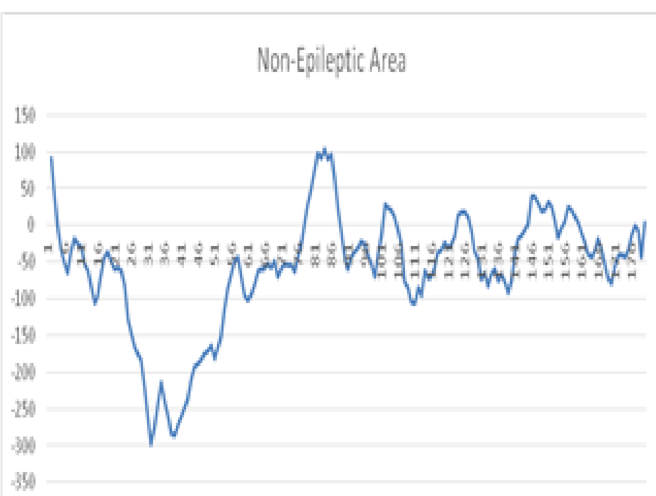


FIGURE 1(c): EEG signal recoded from non-epileptic area of brain

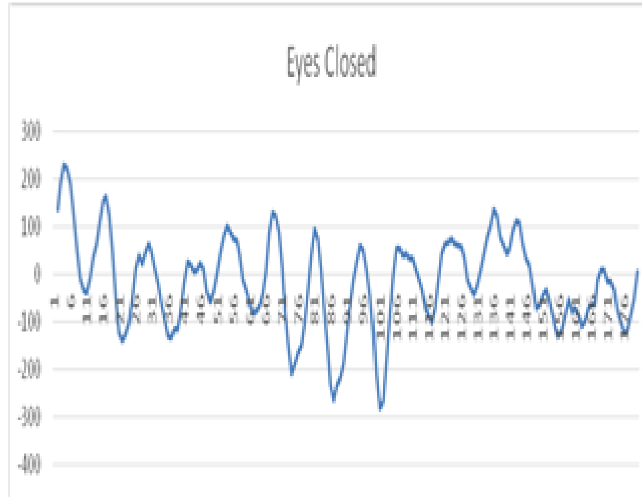


FIGURE 1(d): EEG signal recorded when eyes were closed

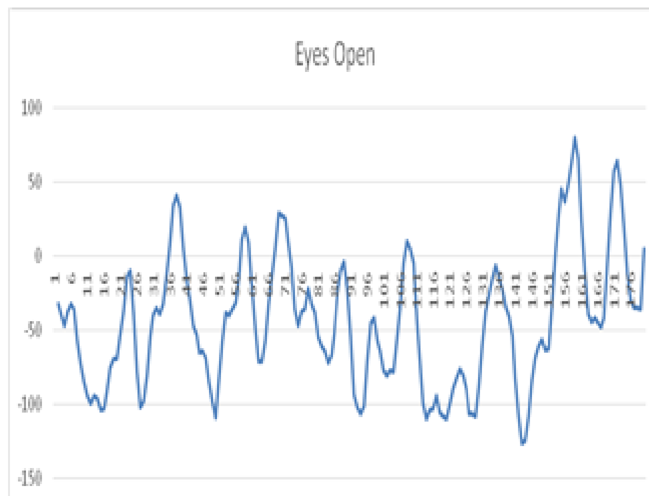


FIGURE 1(e). EEG signal recorded when eyes were open

C. CLASSIFIERS

Classification is machine learning and statistics technique of supervised learning in which the computer program learns from input data and based on this learning it identifies new observations. This collection of data may be clearly bi-class or multi-class. In our study we used 6 classifiers which are KNN, Naive Bayes, Linear classification model, Support vector machine and Decision tree.

1. KNN (K-Nearest Neighbor)

KNN is a commonly used algorithm, mostly used for predictive problems in regression and in classification, but more widely in classification problems. KNN is a general technique that maintains all available cases and classifies new cases based on a similarity measure. It has been already used in data processing and pattern recognition as a non-parametric tool in the early 1970's. To access any strategy, following factors such as the performance review facility, time to evaluate and the power of predictions must be looked at. The algorithm takes and uses a lot of labeled points to learn how to mark certain points. It looks at the labeled points nearest to the new point, which is its closest neighbors, to mark a new point. In our analysis EEG data set is identified by the plurality votes of its neighbors with cross validation.

2. Naive Bayes

A Naïve Bayes classifier is a type of probabilistic machine learning and discriminant analysis which is used for classification tasks. Naïve Bayes is not only a one algorithm, but it is a group of algorithms in which all members of group share a common concept, i.e. each pair of features being identified is independent of each other. The classifier's crux is based upon the Bayes theorem. Naive Bayes algorithms commonly used in the study of opinions, spam analysis, content based etc. These are fast and simple to execute, but they must be independent of the predictors. Usually, the predictors are dependent in real life which hinders the classifier's efficiency.

3. Linear Classification Model

A linear classifier works by making a classification decision based on the value of a linear property combination. The parameter of linear classification model can be determined by generative and discriminative method. A generative method is a statistical technique for the joint distribution of probabilities. The second set of techniques involves discriminative methods, seeking to optimize performance efficiency on a training dataset.

4. Support Vector Machine

Support Vector Machine (SVM) is a type of machine learning and a flexible supervised machine learning algorithm, used for both classification and regression. SVMs were initially developed in the 1960s but then updated in 1990. Like other machine learning methods, SVMs have their own special method of execution. They are recently increasingly common due to their ability to manage several continuous and categorical variables. It is mainly used in labeling problems though.

5. Decision Tree

Decision tree is the most widely used effective classification and prediction method. A Decision tree is a tree-like flowchart in which each internal node signifies a check on an element, each branch reflects a test result, and each leaf node (terminal node) carries a class name. Tree dependent approaches allow predictive models with high accuracy, reliability and evaluative ease. With exception of linear models, they are very effective at mapping non-linear connections. They are versatile to address any question (classification or regression) at hand.

6. Discriminant Analysis Model

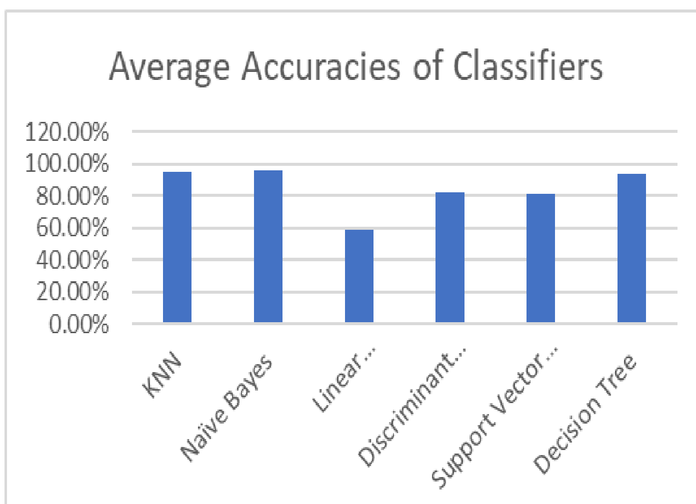
Discriminant analysis model is a type of classifier and a tool for classifying. It is a method used to detect the data from the research when the dependent variable is divided into different categories and the outcome variable is in nature duration. It implies that data is collected by different classes based on specific Gaussian distributions

III. RESULTS

In this section, we will discuss the results of 6 machine learning classifiers, mentioned above, and will compare the corresponding accuracies. We made ten folds of dataset in MATLAB and performed cross validation.
All classifiers yielded ten accuracies and average accuracies were computed. The average accuracies of all classifiers were more than 81% except linear classification model, which gave 58.339% accuracy. Among the six classifiers, KNN and Naïve Bayes inspired us by giving average accuracy of 95.165% and 95.739% respectively.
Moreover, the average accuracy of Naïve Bayes depicts that this model is best for seizure prediction as it has the highest accuracy.

TABLE I: Average accuracy of classifier with Std. deviation

Model	Accuracy	Standard deviation
KNN	95.165%	0.40075543
Naïve Bayes	95.739%	0.523186398
Linear Classification Model	58.339%	3.178721789
Discriminant Analysis Model	82.086%	0.84843902
Support Vector Machine	81.573%	0.942557772
Decision Tree	94.260%	0.4152117



GRAPH 1: Shows the average accuracy of classifiers

IV. DISCUSSION

K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), and Naive Bayes (NB) are the popular classifiers in machine learning and data mining. In this paper we introduced a model of machine learning for the prediction of Epileptic seizures. Following algorithms are used: K-Nearest Neighbor (KNN), Naive Bayes, Linear Classification Model, Discriminant Analysis Model, Support Vector Machine and Decision tree. Naive Bayes has shown better classification accuracy than the other classifiers for the classification of seizure and non-seizure EEG Signals. Firstly, we divided our data into two categories: the seizure activity and non-seizure activity mentioned in research as label 1 and label 0, respectively.

Cross-validation was performed on the data. Cross validation is a statistical method for estimating the abilities of models of machine learning. To make our research easy and interesting the classification accuracy was analyzed using ten-fold cross validation. We split it into ten folds for cross validation. For cross validation we divided our dataset into ten folds where one fold was used as testing data and other folds were used as training data. After that, the above-mentioned machine learning algorithms were applied to find the best classifier in terms of corresponding accuracies. All classifiers gave ten accuracies and then average accuracy was computed. Our experimental findings are listed in Tab. I. The Naive Bayes classifier performed better and gave best average accuracy for the epileptic seizures' prediction, which is 95.739%. The results in the TABLE I show that among the six classifiers, Naive Bayes has a higher average accuracy of 95.739% and K-Nearest Neighbor (KNN) shows the second-best average accuracy with a value of 95.165%, Decision Tree shows the third best average accuracy of 94.260%. Discriminant Analysis Model and Support Vector Machine show the average accuracy of 82.026% and 81.573% respectively. Linear Classification Model shows the lowest average accuracy of 58.339%.

More than 50 million people worldwide are diagnosed with Epilepsy. Machine learning is opening doors of new ways for predicting epilepsy. A considerable amount of research has been done to predict seizures using machine learning and to avoid epilepsy before its onset. The medical significance of our method derives from its key characteristics, which include high accurate value of the Naive Bayes classifier and making the epileptic seizure diagnosis ideal with reduced computational burden. For real-time epileptic seizure prediction, the proposed methodology is quick and fast to implement. Epileptic seizure detection based on a machine learning algorithm is more appropriate in real time for a reliable automatic seizure detection system to improve patient care and life expectancy. Examination of an EEG signal is an exhausting process. Visual check for the detection of seizures in EEG signal is difficult and can lead to inaccuracy as well. Therefore, a highly accurate automated seizure detection framework is required. In several studies, precursors of upcoming epileptic seizures were reported using a variety of features abstracted from EEG recordings. The analysis of our methodology is evident from the other methods of machine learning algorithms. The proposed methodology for machine learning will bring a significant increase in the performance of seizure prediction. Through studying the various machine learning algorithms, we may catch patient-specific pathways of seizures.

V. CONCLUSION

Epilepsy is one of the most common and least predictable neurological disorders. The seizures that describe epilepsy are so often unidentified and affect the quality of life of a patient and can cause serious injuries or even death. A lot of work has been done and published to predict the initiation of epileptic seizures using various classifiers and scalp EEG signals. The discussed methodologies including our research will be helpful in faster diagnosis of epileptic seizures and cost reduction in health care. The purpose of our research was to find the best algorithm to predict the epileptic seizure before its onset. For this purpose, we used several classifiers and performed ten folds cross validation. Among the classifiers, Naive Bayes showed the highest average accuracy value of 95.739% with standard deviation 0.523186398 as compared to the average value of KNN which showed the second highest average accuracy of 95.165% with standard deviation 0.40075543 and rest of the classifiers. The future work of our study is to use of our model in real time for prediction of epileptic seizures.

REFERENCES

[1] Guenot M., Surgical treatment of epilepsy: outcome of various surgical procedures in adults and children. Revue neurologique, vol. 160, pp.5S241-0, 2004.

[2] Liu H, Xi L, Zhao Y, Li Z. Using Deep Learning and Machine Learning to Detect Epileptic Seizure with Electroencephalography (EEG) Data. arXiv preprint arXiv:1910.02544. 2019 Oct 6.

[3] Bandarabadi M, Teixeira CA, Rasekhi J, Dourado A. Epileptic seizure prediction using relative spectral power features. Clinical Neurophysiology, vol. 126, no. 2, pp.237-48, Feb 1, 2015.

[4] Karlık B, Hayta ŞB. Comparison machine learning algorithms for recognition of epileptic seizures in EEG. Proceedings IWBBIO 2014. 2014 Apr.

[5] Chakraborti S, Choudhary A, Singh A, Kumar R, Swetapadma A., A machine learning based method to detect epilepsy, International Journal of Information Technology, vol. 10, no. 3, pp. 257-63, Sep 1, 2018.

[6] Jaiswal AK, Banka H., Epileptic seizure detection in EEG signal using machine learning techniques, Australasian physical & engineering sciences in medicine, vol. 41, no. 1, pp. 81-94, Mar 1, 2018.

[7] Xiang J, Li C, Li H, Cao R, Wang B, Han X, Chen J., The detection of epileptic seizure signals based on fuzzy entropy, Journal of neuroscience methods, vol. 243, pp. 18-25, Mar 30, 2015.

[8] Usman SM, Usman M, Fong S., Epileptic seizures prediction using machine learning methods. Computational and mathematical methods in medicine, vol. 2017, Dec 2017.

[9] [Online]. Available: https://www.kaggle.com/harunshimanto/epileptic-seizure-recognition.

[10] Savadkoohi M, Oladunni T, Thompson L., A machine learning approach to epileptic seizure prediction using Electroencephalogram (EEG) Signal, Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 1328-41, 2020 Jul 1.