A NOVEL APPROACH FOR IDENTIFICATION AND ANALYSIS OF EPILEPTIC SEIZURES USING MACHINE LEARNING CLASSIFICATION APPROACHES

Riyazulla Rahman J[1], Dr.Nagaraja S R[2]

Article History: Received: 22.05.2023 **Revised:** 05.06.2023 **Accepted:** 26.07.2023

Abstract

Epilepsy is a chronic neurological disorder distinguished by many types of seizures, some of which are characterised by involuntary repeated convulsions that have a significant impact on the patients' everyday lives. Several strategies have been presented in the literature to detect these types of seizures and identify the patient; however, these tactics fall short in ergonomic difficulties and adequate integration with the exercise equipment. This research work , we developed a machine learning classification algorithm to determine if EEG data indicates a seizure is occurring. To put it another way, the top performing model outperforms random guessing by a factor of 4.3, or "lift." As for the positive classes in the test set, it predicts them with a 100% accuracy rate. It is reasonable to assume this level of performance if this model is deployed in production to predict whether a patient is having a seizure.

Keywords: Epilepsy , Disorder , Seizures, Patients, Detection, Decisions.

---

[1]Research Scholar**,** Assistant Professor , School of Computer Science and Engineering, Presidency University , Bangalore – 560 064 –INDIA
[2]Associate Professor, School of Computer Science and Engineering, Presidency University, Bangalore – 560 064 - INDIA

DOI: 10.31838/ecb/2023.12.6.254

1. Introduction

A seizure, often known as an epileptic seizure, is a period of symptoms caused by abnormally excessive or coordinated neuronal activity in the brain. The effects range from entire body involuntary trembling and loss of consciousness (tonic-clonic seizure) to partial body trembling and varied degrees of loss of awareness (focal seizure) to modest transitory loss of awareness (absence seizure). Although episodes are usually less than two minutes long, returning to normal can take some time. This can result in a loss of bladder control . Convulsions can arise for a variety of reasons.As described in transitory events that can trigger seizures include low blood sugar, alcohol withdrawal, prescription drug addiction, low blood sodium, fever, brain infection, or a concussion. Seizures that occur suddenly and for no obvious reason often result in a lifetime of epileptic episodes because the cause is unknown. Stress and a lack of sleep have been connected to the severity of spontaneous seizures. In epilepsy, the patient has had at least one unprovoked seizure and is at high risk of having more seizures in the future. Conditions such as fainting, nonepileptic psychogenic seizure, and tremor can mimic epileptic seizures but are not epilepsy.

Any seizure that lasts more than a few minutes necessitates rapid medical intervention.Status epilepticus is defined in as a seizure lasting more than five minutes.A first seizure does not usually necessitate long-term anti-seizure medication , unless an underlying problem is detected via electroencephalogram (EEG) or brain imaging].If a patient has only experienced one seizure, it is typically safe for them to complete the entire assessment on an outpatient basis. Many persons who have what appears to be their first seizure have actually had multiple lesser seizures before this one.

Ten percent or more of the population may have had at least one epileptic episode. According to the findings of this study, around 3.5 per 10,000 people have provoked seizures per year, while approximately 4.2 per 10,000 people suffer unprovoked seizures per year.A second seizure happens in almost half of those who have experienced one. At any given time, approximately one percent of the population suffers from epilepsy It is estimated that over 3.4 million people in the United States and over 65 million people worldwide have active epilepsy.

2. Literature Review

Epilepsy is characterised by a person's inability to control their seizures, which manifests as an unpredictable electrical disturbance of the brain. Epileptic seizures can be diagnosed with the help of an electroencephalogram (EEG) recording this abnormality in brain activity. As a result, many ML techniques have been created for the categorization of EEG data. Common ML classifiers include neural networks, decision trees, K-nearest neighbours, and support vector machines [1,2,3,4,5]. Potapov [6] looked into the fact that preprocessing of signals can eliminate crucial information needed for classification in the process of collecting relevant features and suppressing noise. Since the starting classification data impacts the classification accuracy, it is essential that the initial data be taken into consideration for the signal classification procedure.

For the separation of normal and abnormal signals, Harikumar et al. [7] used a genetic algorithm to develop fuzzy logic. In order to determine an appropriate risk for seizure detection, they devised several approaches and binary genetic algorithms.

The bootstrap aggregation method and a variable Q factor have also been presented as a method for diagnosing epilepsy [8]. In addition, Mursalin et al. [9] proposed a unique method for seizure identification in epilepsy by combining a feature selection method and a random forest.

Fuzzy logic has been used in a number of different ways for the purpose of seizure detection in epilepsy [10,11]. These include feature selection, feature extraction, and dimensionality reduction techniques. Thus, numerous investigations address the EEG data, with a variety of studies doing various analyses for epileptic seizure identification. Sharmila et al. [12] developed a paradigm for EEG detection of epileptic episodes in both epileptic patients and control participants. EEG signals are analysed in the designed framework via

linear and non-linear classifiers, both of which rely on the discrete wavelet transform.
Thus, numerous investigations address the EEG data, with a variety of studies doing various analyses for epileptic seizure identification. Nevertheless, it is worthwhile and important to seek out a genuine method of fixing all these problems. This literature survey documents the many ML-based approaches to the identification of epileptic seizures.

3. Seizure Types And Classification

When diagnosing a seizure, doctors consider three factors:

1. Where the seizure begins in the brain (e.g., the beginning)
2. Whether or not the person is aware during the seizure
3. Whether or not the seizure causes movement.

Figure 1 depicts the Seizure Types and Classification

Seizures can be divided into three major groups:

1. Focal Onset
2. Generalized Onset
3. Unknown Onset

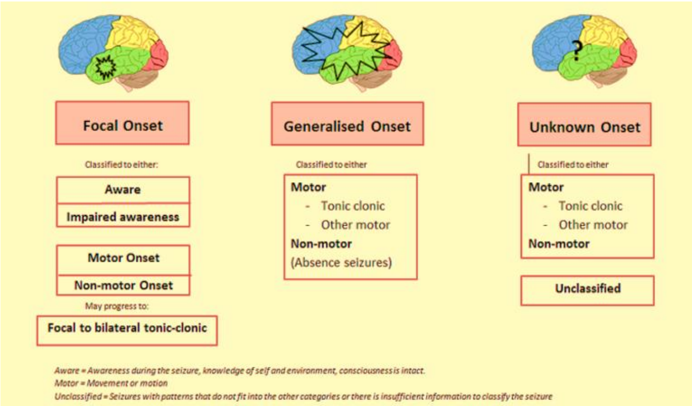

Figure 1. Seizure Types and Classification

3.1 Focal Onset Seizures

A seizure is termed to have a focal onset or focal seizure when it starts in one specific location of the brain. The injury could gradually spread to other areas of the brain. These types of seizures are frequently difficult to detect since they show in ways that can be misinterpreted as anything from intoxication to daydreaming. Around 60% of people with epilepsy get focal onset seizures, often known as focal seizures. There are two types of focal onset seizures based on the state of consciousness of the patient during the seizure. Previously classified as simple partial seizures, the condition is now 2852inimizing as focused awareness, in which the patient is aware of their surroundings but unable to communicate. They are a characteristic of the seizure, but they are sometimes misinterpreted as a warning or an aura (that a more serious seizure is on the way).

A person suffering from focused impaired awareness (also known as a complex partial seizure) may complain of bewilderment, fuzziness, or disorientation. Tonic-clonic seizures cause the victim's muscles to tension and jerk because they originate in one portion of the brain and spread to both sides.

3.2 Generalised Onset Seizures

Generalised onset indicates that the seizure affects both hemispheres (sides) of the brain from the start. As a result, a person may lose consciousness at the commencement of a seizure. Because 2852inimizing2852 onset seizures almost always influence awareness in some way, the labels 'aware' or 'impaired awareness' are not employed. However, they can be further 2852inimizing2852 based on movement: Generalised motor seizure: may entail stiffening (tonic) and jerking (clonic), known as tonic-clonic (formerly known as grand mal) or other motions. Generalised non-motor seizure: These seizures include transient shifts in awareness, gazing, and some may include involuntary or repetitive actions such as lip-smacking.

3.3 Types of Generalised Onset Seizures

This category includes a wide range of seizures. These are some examples:

1. Absence
2. Tonic-Clonic
3. Myoclonic jerks
4. Tonic
5. Atonic seizures
6. Clonic jerks

3.4 Unknown Onset Seizures

It is impossible to identify whether or not a seizure is focal or minimising if its onset is unknown. The seizure type may shift from minimising to focal onset when more information is acquired or as further testing is performed. A diagnosis of an epileptic seizure may be made, however it may be impossible to determine the specific type of seizure that occurred. Inexperience with seizures or an atypical presentation of symptoms could be to blame.

4. Problem Statement

About 1.2% of the U.S. population (3.4 million individuals) has epilepsy, and over 65 million people throughout the world suffer from this disorder of the central nervous system. In addition, around one in every twenty-six persons may develop epilepsy. There are several different types of seizures, each with its own unique set of symptoms include convulsions, coma, and confusion. Even more subtle visual indicators of seizures include patients' inability to respond or respond blankly for brief periods of time. Injuries such as falling, biting one's tongue, or losing control of urination or defecation might result from a seizure's onset. Therefore, these are some of the reasons why seizure detection is so important for those who are under medical supervision and who are prone to seizures. Epileptic seizure detection using biometric differences had previously proven possible. The EEG is a crucial source of data for this purpose since it measures electric brain activity, which can be used to detect epileptic seizures. However, there is no record keeping in the server or the display in the current system

5. Proposed Methodology

Seizure detection architectures can detect ongoing seizures and deliver 2853inimizing seizure data to clinicians for use in controlling epilepsy. Closed-loop seizure detection systems may also be able to deliver rapid therapy in response to seizures at their earliest clinical beginnings, 2853inimizing the severity of headaches and maybe arresting seizure spread. A seizure detection device should be able to confirm or deny the presence of a seizure that is currently taking place.

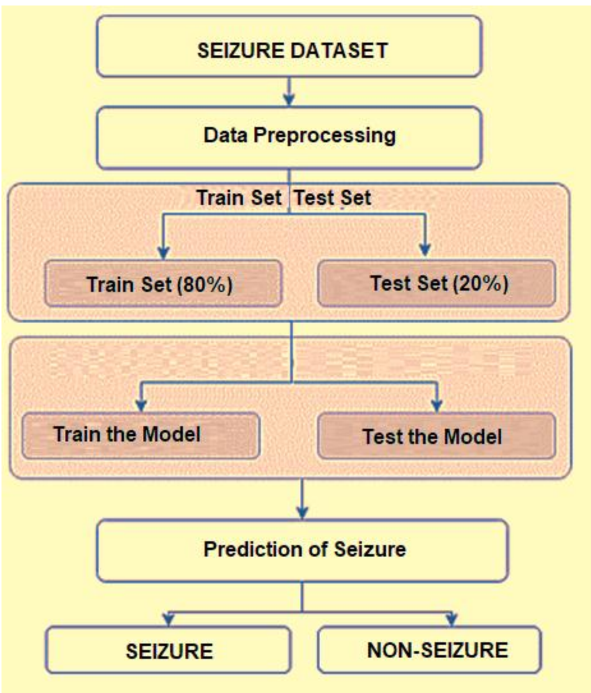

Figure 2. Proposed Methodology for Epileptic Attack Detection

6. Machine Learning Classifiers

6.1 KNN

K-Nearest Neighbours is a fundamental classification technique in Machine Learning. It is classified as supervised learning and is widely used in pattern recognition, data mining, and intrusion detection. It is extensively used in real-world scenarios because it is non-parametric, which means it makes no underlying assumptions about dat.. We are provided prior data (also known as training data) that categorises coordinates into groups based on an attribute.

$$d(x, x') = \sqrt{\left(x_1 - x_1'\right)^2 + \ldots + \left(x_n - x_n'\right)^2}$$

$$P(y = j | X = x) = \frac{1}{K} \sum_{i \in \mathcal{A}} I(y^{(i)} = j)$$

6.2 LR

One well-known approach for supervised learning is logistic regression. The categorical dependent variable can be predicted from a set of independent variables using this method. Logistic regression is used to forecast the value of a dependent variable that is categorical. The outcome can only be a categorical or discrete statement. Instead of reporting the exact result as 0 or 1, it displays the probability values that lie between those two extremes.

= e^(b0 + b1*x) / (1 + e^(b0 + b1*x))

6.3 SGD

Gradient Descent is an iterative optimisation procedure that seeks the optimal value (Minimum/Maximum) of an objective function. It is one of the most used ways for modifying a model's parameters in order to lower a cost function in machine learning projects. The basic purpose of gradient descent is to find the model parameters that yield the highest accuracy on both training and test datasets. In gradient descent, the gradient is a vector indicating in the general direction of the function's sharpest climb at a given point.

$$w := w - \eta \nabla Q_i(w)$$

6.4 NB

Nave Bayes is a simple learning technique that use the Bayes rule along with the strong assumption that the attributes are conditionally independent, given the class. While this independence requirement is frequently violated in practise, naive Bayes often produces competitive classification accuracy. With its computing economy and many other desirable characteristics, naive Bayes is commonly used in practise.

P(A|B) = P(B|A) * P(A) / P(B)

6.5 DT

When it comes to supervised learning algorithms, a decision tree is a powerful tool that may be used for both classification and regression. Each internal node represents an attribute test, each branch represents a test outcome, and each leaf node (terminal node) holds a class name, producing a tree structure reminiscent of a flowchart. The tree is constructed by repeatedly subdividing the training data into subsets according to attribute values until a stopping criteria is reached, such as the maximum depth of the tree or the lowest number of samples needed to divide a node.

$$Entropy = -\sum_{i=1}^{n} p_i * \log(p_i) \quad , \quad Gini\,index = 1 - \sum_{i=1}^{n} p_i^2$$

Gain(T,X) = Entropy(T) — Entropy(T,X)

6.6 RF

In the realm of machine learning, Random Forest is a popular supervised learning technique. It can be used to solve machine learning problems including both classification and regression. The name suggests that Random Forest is a classifier that uses many decision trees trained on different parts of a dataset and then averaging the results to increase the dataset's predicted accuracy.In place of a lone decision tree, the random forest compiles the predictions of several smaller trees to arrive at a final prediction. More trees in the forest mean better accuracy and less likelihood of overfitting.

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T}$$

6.7 GB

Gradient boosting is a method that stands out because of how fast and accurate it is at making predictions, especially with big and complex datasets. This algorithm has given the best results in Kaggle competitions and company machine learning solutions. We already know that mistakes are a big part of any algorithm for machine learning. Bias error and variance error are the two main kinds of mistake. The gradient boost method helps us reduce the amount of bias error in the model.

$$F(t)(x) = F(t-1)(x) + \varepsilon h(t)(x).$$

Database

The dataset can be found here in UCI's machine learning repository. The dataset contains 4097 electroencephalogram (EEG) values per subject over a period of 23.5 seconds, with a total of 500 patients. The 4097 data points were then separated into 23 chunks, one for each patient, and each chunk corresponded to one row in the dataset. Each row has 178 readings that are converted into columns; in other words, one second of EEG readings is represented by 178 columns. There are 11,500 rows and 179 columns in all, with the last column containing the patient's status, which indicates whether or not the patient is having a seizure.

In [2]:
```
df.head()
```

Out[2]:

	Unnamed: 0	X1	X2	X3	X4	X5	X6	X7	X8	X9	...	X170	X171	X172	X173	X174	X175	X176	X177	X178	y
0	X21.V1.791	135	190	229	223	192	125	55	-9	-33	...	-17	-15	-31	-77	-103	-127	-116	-83	-51	4
1	X15.V1.924	386	382	356	331	320	315	307	272	244	...	164	150	146	152	157	156	154	143	129	1
2	X8.V1.1	-32	-39	-47	-37	-32	-36	-57	-73	-85	...	57	64	48	19	-12	-30	-35	-35	-36	5
3	X16.V1.60	-105	-101	-96	-92	-89	-95	-102	-100	-87	...	-82	-81	-80	-77	-85	-77	-72	-69	-65	5
4	X20.V1.54	-9	-65	-98	-102	-78	-48	-16	0	-21	...	4	2	-12	-32	-41	-65	-83	-89	-73	5

5 rows × 180 columns

```
In [4]: df.head()
Out[4]:
```

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X170	X171	X172	X173	X174	X175	X176	X177	X178	OUTPUT_LABEL
0	135	190	229	223	192	125	55	-9	-33	-38	...	-17	-15	-31	-77	-103	-127	-116	-83	-51	0
1	386	382	356	331	320	315	307	272	244	232	...	164	150	146	152	157	156	154	143	129	1
2	-32	-39	-47	-37	-32	-36	-57	-73	-85	-94	...	57	64	48	19	-12	-30	-35	-35	-36	0
3	-105	-101	-96	-92	-89	-95	-102	-100	-87	-79	...	-82	-81	-80	-77	-85	-77	-72	-69	-65	0
4	-9	-65	-98	-102	-78	-48	-16	0	-21	-59	...	4	2	-12	-32	-41	-65	-83	-89	-73	0

5 rows × 179 columns

Data Preprocessing

The first step in every data mining project is known as "data preprocessing," and it entails translating raw data into a usable format. There is a high probability of errors and other problems in real-world data since it is frequently unreliable, inconsistent, and/or missing in specific behaviours or trends. Preprocessing data has been shown to be effective in addressing these kinds of problems. The purpose of data preparation is to get raw data ready for analysis.

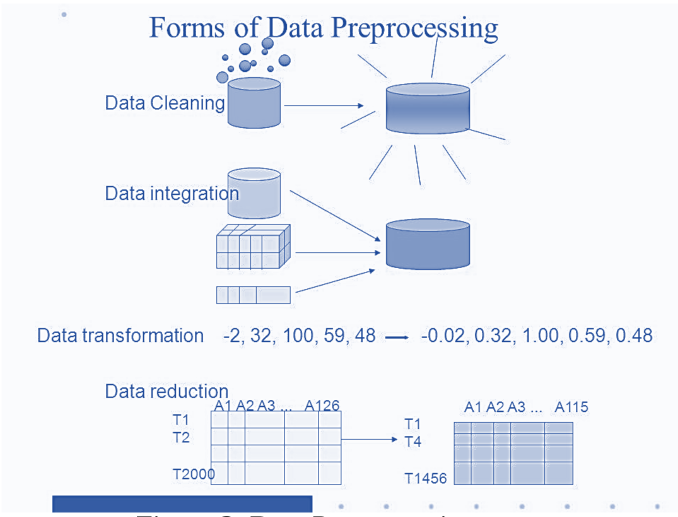

Figure 3. Data Preprocessing

Feature Engineering

Our machine learning algorithm requires a set of numerical variables, thus we must go through the process of "feature engineering" to convert characteristics that are currently in a category or ordinal form. In order for our machine learning algorithm to grasp categorical string variables like one-hot encoding or ordinal variables like label-encoding, we must first process them. Since all of the columns in the epilepsy dataset are EEG readings, there is no need to further enhance the 178 features (columns) that make up the dataset. Depending on the number of samples available, we can typically divide the dataset into training, validation, and test samples in the ratio of 50:25:25, 60:20:20, or 70:15:15. Even a split of 98/1/1 is possible with a very large dataset (hundreds of millions of rows). We need to use the bulk of our dataset for training our machine learning algorithm, which is why we employ a training split. Hyperparameters are fine-tuned and the top performing method is chosen using the validation dataset. Our machine learning model is put to the test on the testing dataset.

7. Results And Discussion

Our research has shown that the GB classifier is the most effective and efficient with an accuracy of 100% of the available options in the vast majority of cases. Table 1 depicts the accuracy of various machine learning classifiers. In the initial run, it performed with an impressive 98.4 percent accuracy in a relatively short length of time. Also showing early promise was RF which achieved an impressive 99.50 % accuracy.

Table 1. Performance of ML Classifiers

Model	**Accuracy**	**Recall**	**Precision**	**Specificity**	**AUC**
KNN	62.30	24.80	99.30	99.80	99.10
LR	67.20	53.60	73.60	80.80	64.00
SGD	57.50	53.70	58.20	61.30	56.60
NB	93.10	89.30	96.60	96.80	98.10
DT	97.40	95.40	99.50	99.50	98.20
RF	95.50	92.20	98.70	98.80	99.50
GB	1.000	1.000	1.000	1.000	1.000

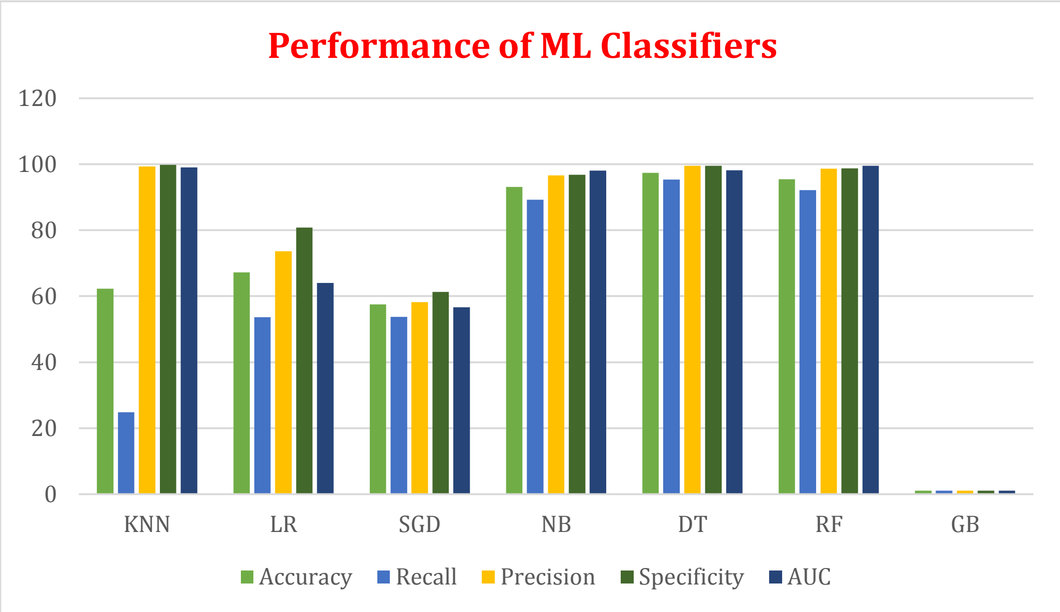

Figure 4. Performance of ML Classifiers

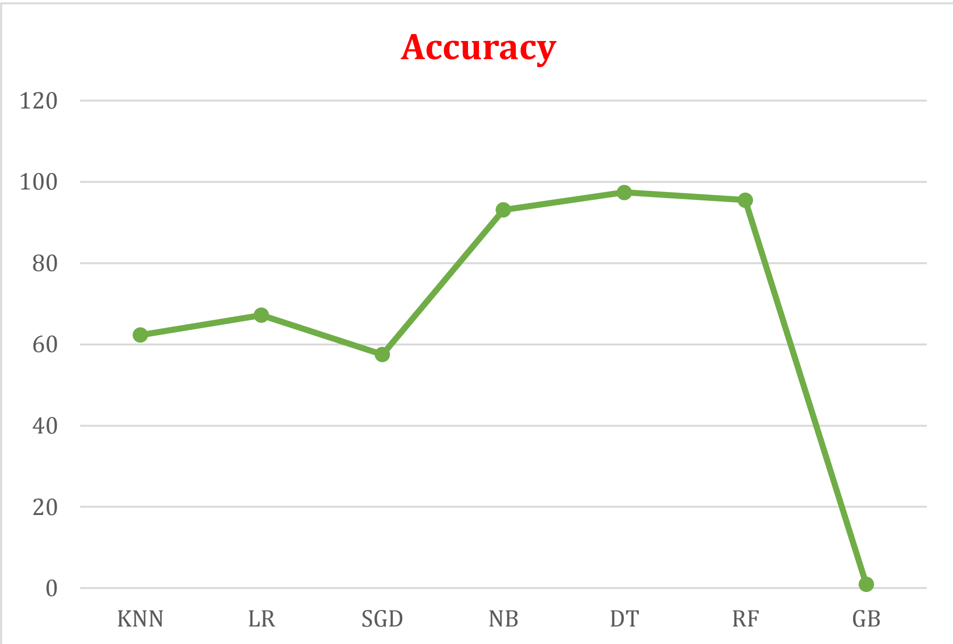

Figure 5. Accuracy

ROC Curve

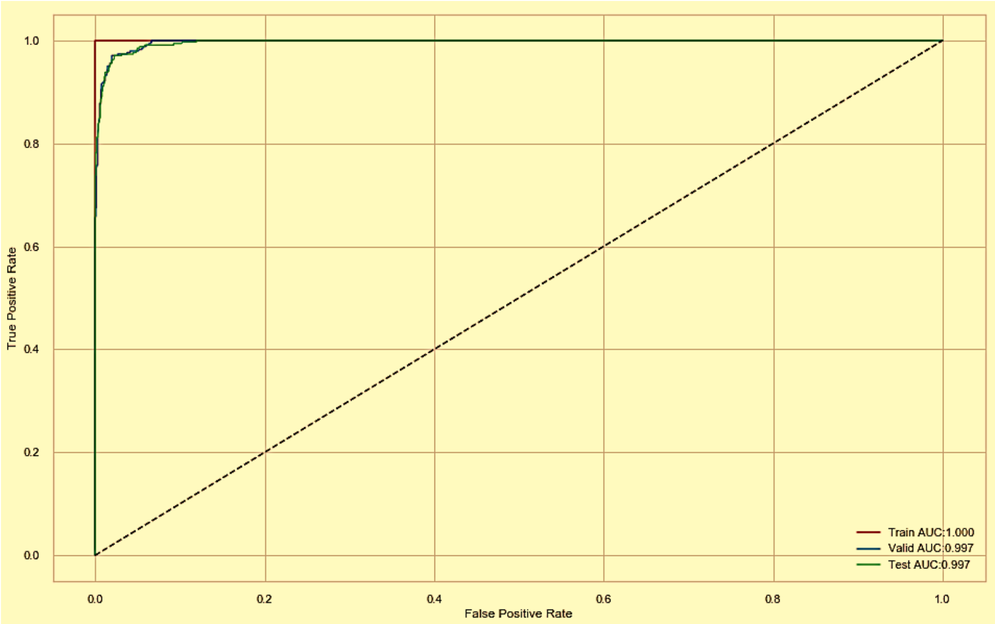

Figure 6. ROC Curve

8. Conclusion

We developed a machine learning classification algorithm to determine if EEG data indicates a seizure is occurring. To put it another way, the top performing model outperforms random guessing by a factor of 4.3, or "lift." As for the positive classes in the test set, it predicts them with a 100% accuracy rate. Figure 4 and 5 clearly depicts the accuracy of various machine learning classifiers. It is reasonable to assume this level of performance if this model is deployed in production to predict whether a patient is having a seizure.

9. References

[1]. 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**, 38, 13475–13481.

[2]. Naderi, M.A.; Mahdavi-Nasab, H. Analysis and classification of EEG signals using spectral analysis and recurrent neural networks. In Proceedings of the 2010 17th Iranian Conference of Biomedical Engineering (ICBME), Isfahan, Iran, 3–4 November 2010; pp. 3–4.

[3]. Subasi, A.; Gursoy, M.I. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Syst. Appl. **2010**, 37, 8659–8666.

[4]. Kumar, Y.; Dewal, M.; Anand, R. Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. Neurocomputing **2014**, 133, 271–279.

[5]. Nicolaou, N.; Georgiou, J. Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines. Expert Syst. Appl. **2012**, 39, 202–209.

[6]. Potapov, P. On the loss of information in PCA of spectrum-images. Ultramicroscopy **2017**, 182, 191–194.

[7]. Harikumar, R.; Narayanan, B.S. Fuzzy techniques for classification of epilepsy risk level from EEG signals. In Proceedings of the TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, Bangalore, India, 15–17 October 2003; Volume 1, pp. 209–213.

[8]. Hassan, A.R.; Siuly, S.; Zhang, Y. Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating. Comput. Methods Programs Biomed. **2016**, 137, 247–259.

[9]. Mursalin, M.; Zhang, Y.; Chen, Y.; Chawla, N.V. Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing **2017**, 241, 204–214.

[10]. Hussain, W.; Qbal, M.S.I.; Xiang, J.; Wang, B.; Niu, Y.; Gao, Y.; Wang, X.; Sun, J.; Zhan, Q.; Cao, R.; et al. Epileptic Seizure Detection With Permutation Fuzzy Entropy Using Robust Machine Learning Techniques. IEEE Access **2019**, 7, 182238–182258.

[11]. Sharmila, A.; Geethanjali, P. DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers. IEEE Access **2016**, 4, 7716–7727.

[12]. \SM. Usman, S. Khalid and MH. Aslam, “Epileptic Seizures Prediction Using Deep Learning Techniques”, IEEE Access, Vol 8, pp 39998-40007, 2020.

[13]. A. Ahmed, W. Ahmad, MJ. Khan, SA. Siddiqui and HM. Cheema, “A wearable sensor based multi-criteria-decision-system for realtime seizure detection”, In 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), July 2017, pp. 2377-2380.

[14]. NE. Cruz, J. Solarte and A. Gonzalez-Vargas, “Automated epileptic seizure detection system based on a wearable prototype and cloud computing to assist people with epilepsy”, In Workshop on Engineering Applications, October 2018, pp. 204-213.

[15]. E. Džaferović, S. Vrtagić, L. Bandić, J. Kevric, A. Subasi and SM. Qaisar, “Cloud-based mobile platform for EEG signal analysis”, In 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), December 2016, pp. 1-4.

[16]. R. Kolb, K. Pineda, M. Curran, T. Skinner, A. Adegbege and BF. BuSha, “Ambulatory Epileptic Monitoring Device”, In 40th Annual Northeast Bioengineering Conference (NEBEC), April 2014, pp. 1-2.

[17]. Y. Liu, S. Sivathamboo, P. Goodin, P. Bonnington, P. Kwan, L. Kuhlmann and Z. Ge , “Epileptic Seizure Detection Using Convolutional Neural Network: A Multi-Biosignal study”, In Proceedings of the Australasian Computer Science Week Multiconference, February 2020, pp. 1-8.

[18]. WL. Mao, HIK. Fathurrahman, Y. Lee and TW. Chang, “EEG dataset classification using CNN method”, In Journal of Physics: Conference Series, Vol. 1456, No. 1, January 2020, pp. 012017.

[19]. NA. Marzuki, W. Husain and AM. Shahiri, “MyEpiPal: Mobile Application for Managing, Monitoring and Predicting Epilepsy Patient”, In International Conference on Advances in Information and Communication Technology, 2016 December, pp. 383-392.

[20]. KVA. Sogamoso and OJS. Parra, “Internet of Things for Epilepsy Detection in Patients”, In International Conference on Cooperative Design, Visualization and Engineering, October 2018, pp. 237-244