# Epileptic Seizure Detection from EEG Signal Using Machine Learning

Md. Alamgir Hossain (1806186), Adib Md. Tawsif (1806185), Arnob Ghosh (1806184), Md. Rokonuzzaman (1806180)

***Abstract* – Epilepsy is a major neurological disease which is related to sudden seizure and loss of consciousness. But there may be seizures due to various other diseases unrelated to epilepsy. Epileptic Seizure can be distinguished easily by the EEG data of affected person. So, in this project, we create a CAD model for detection of Epileptic Seizure using EEG data. We analyze sets of electroencephalographic (EEG) time series: surface EEG recordings from healthy volunteers with eyes closed and eyes open, and intracranial EEG recordings from epilepsy patients during the seizure free interval from within and from outside the seizure generating area as well as intracranial EEG recordings of epileptic seizures. After training and testing of the ML based model, it is found that the model works well (above 90% efficiency).**

## I. INTRODUCTION

The most prevalent neurological illness, epilepsy, affects 50 million people worldwide, 85% of whom live in poor nations. Globally, 2.4 million new cases are reported each year. At least half of all instances of epilepsy start in infancy or adolescence [1]. Elderly individuals (those above the age of 65) may also have sudden onset [2]. When compared to a healthy person, epileptics have a two- to three-fold increased risk of dying before their time [1]. As a result, research into epilepsy has always been of highest significance in the biomedical arena.

The faulty electrophysiological system of the brain, which results in an abrupt and excessive electrical discharge in a collection of brain cells (i.e. neurons) located in the cerebral cortex, is the cause of epileptic seizures. The cerebral cortex is involved, which results in aberrant motor activities that cause jerky (tonic-clonic) spasms of muscles and joints. The underlying physiology is the changed and inappropriate changes in sensory and motor activity brought on by the hypersynchronous firing of neurons. Epileptic seizures are brought on by an abrupt and enormous surge of energy by the brain cells as opposed to a regulated discharge of electrical energy. Large variations in attributes can be seen in the seizures.

A seizure can range in severity from a slight muscle twitch to violent, sustained convulsions. Seizures that occur frequently or unexpectedly can be harmful and put a person's life in jeopardy [3]. The magnitude of the abnormal electrical discharge, the location in the brain where it occurs, and the spread of that discharge all affect the seizure's features. Human understanding of how the brain works is still insufficient to fully comprehend the characteristics of an epileptic brain [4]. Loss of mindfulness, nearly imperceptible anomalies in movement pattern, extremely minor muscle twitching, disruptions in the visual, auditory, gustatory, and emotional sensations, among many others that are frequently difficult to identify manually, are possible transitory signs of epilepsy.

## II. EPILEPTIC SEIZURE DETECTION AND EEG

### *A. Methods for detection of epilepsy*

Given that 10% of people worldwide have one seizure in their lifetime, seizures may not always be caused by epilepsy. Chemical imbalances like low blood sugar, low oxygen, abnormal sodium, calcium, and potassium in blood can also cause seizures. These pseudo seizures, psychogenic, or cryptogenic seizures—also known as non-epileptic seizures—are episodic occurrences and unrelated to epilepsy. They are aberrant electrical impulses in the brain, nevertheless. Therefore, it may be difficult to identify non-epileptic pseudo seizures among the likelihood of seizures caused by primary and/or subsequent epilepsies.

Clinically, we can speculate that epilepsy may be the reason if there are two or more unexplained seizures. If epilepsy is the cause of the seizure, early detection of the condition is crucial for starting treatment with antiepileptic medications, which will enhance the quality of life and safety of epileptic persons. The main cause of physical risks due to accidents, such as drowning (if it happens while swimming or taking a bath), burns, and head injuries, is the unpredictability of the seizures.

Clinical epilepsy diagnosis needs a thorough medical history, neurological exams, blood testing, and occasionally cerebrospinal fluid studies to rule out potential secondary causes, such as the previously stated metabolic abnormalities. To look for any structural abnormalities (such as tumors, aberrant blood arteries, and ischemia) in the brain that may be producing seizures but aren't necessarily linked to epilepsy, imaging techniques like CT (Computerized Tomography) or MRI (Magnetic Resonance Imaging) scans may be utilized. The study of EEG data, however, is the most often used and reliable diagnostic technique for the discovery of epilepsy.

### *B. EEG analysis for epilepsy detection*

Through the electrodes positioned on the scalp, the EEG directly captures the electrical activities of the cerebral cortex. The dendrites of the neurons close to the cortical surface are really measured for their electrical potentials. Analysis of the recorded EEG signals can be used to learn more about the characteristics of epileptic seizures. EEG patterns that are different from those in a typical EEG signal recorded from a normal non-epileptic person can be found in signals taken soon before and during seizures.

The transition from the normal stage to the pre-seizure and epileptic seizure stages is accompanied by a change in the electrical behavior of the brain. The connection of neurons in the epileptogenic zone declines throughout the pre-seizure phase.

After that, the circuit is disconnected from the epileptic neurons. The EEG signal changes as a result of these modifications. The isolated epileptic neurons become dormant and gradually lose the inhibitory guidance from the surrounding neurons. As a result of the sudden increase in neuronal discharge, seizures happen. The EEG signal's variability is further increased as a result. The emergence of such aberrant electrical activity is related to an increase in entropy. Sleigh et al. [6] have shown that the entropy variations in the EEG measure entropy changes taking place within the cerebral cortex of the brain. Less autonomous systems and functions are functioning in the brain during epilepsy. Therefore, by looking at the EEG recordings, epileptic seizures may be discriminated from non-epileptic ones clearly because non-epileptic seizures have normal EEG readings.

To do this, an effective EEG signal analysis technique that can yield the most details on the health of the brain must be used. It is always advantageous to identify these activities from the EEG data using various computer approaches by extracting pertinent aspects from the signals for objective analysis and repeatable outcomes.

This project's goal is to create a Computer Aided Diagnostic (CAD) for the detection of epilepsy. Few studies present methods to classify all three stages, namely normal, pre-seizure, and seizure. Some studies concentrate on detecting epilepsy by classifying only the normal and seizure stages (two-class problem) (three-class problem). The next stage is to identify and maybe anticipate the onset if it has been determined that the episode is caused by epilepsy. For this, a Machine Learning-based model has been created. Due to its innovative use of the Data Splitting approach, this model has the advantage of being able to make a prediction based on EEG data of any length.

### *C. Methods of EEG analysis*

EEG analysis during the 1970s involved analyzing the EEG waveform using descriptive and heuristic techniques. Over time, several techniques have been employed to examine a number of minute variations in the EEG signal. According to the authors, the majority of the techniques can be divided into four primary categories: time domain, frequency domain, time-frequency domain, and nonlinear techniques. [5] We have used frequency domain methods like Higher Order Filtering and time-frequency domain methods like Continuous Wavelet Transformation (CWT) in our project.

## III. METHODOLOGY

### *A. Data Collection*

For the project, the dataset provided by Andrzejack et al. [6] was used. The dataset has fiive folders denoted A–E each containing 100 single-channel EEG segments of 23.6-sec duration. These segments were selected and cut out from continuous multichannel EEG recordings after visual inspection for artifacts, e.g. due to muscle activity or eye movements. In addition, the segments had to fulfill a stationarity criterion. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme described in **Fig1**

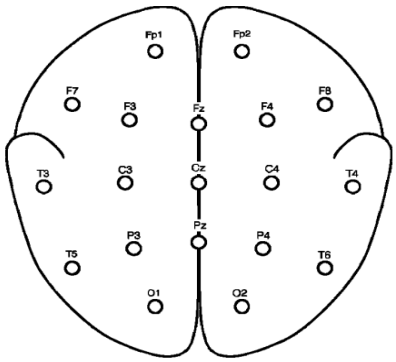


FIG. 1. Scheme of the locations of surface electrodes according to the international 10-20 system. Names of the electrode positions are derived from their anatomical locations. Segments of sets A and B were taken from all depicted electrodes.

Volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B) respectively. Sets C, D, and E originated from EEG archive of presurgical diagnosis. EEGs from five patients were selected, who had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone of **Fig. 2.**

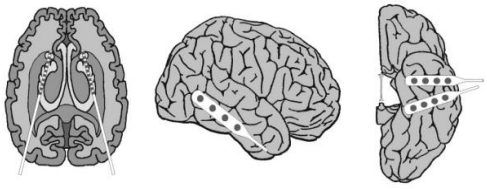

FIG. 2. Scheme of intracranial electrodes implanted for presurgical evaluation of epilepsy patients. Depth electrodes were implanted symmetrically into the hippocampal formations (top). Segments of sets C and D were taken from all contacts of the respective depth electrode. Strip electrodes were implanted onto the lateral and basal regions (middle and bottom) of the neocortex. Segments of set E were taken from contacts of all depicted electrodes.

Segments in set D were recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity. Segments were selected from all recording sites exhibiting seizure activity. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference [omitting electrodes containing pathological activity (C, D, and E) or strong eye movement artifacts (A and B)]. After 12 bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz.

### *B. Data Processing*

The steps required for data processing workflow is given below:


FIG. 3. Steps Related to signal Processing

The signal processing consists of 2 steps.[7]

After collecting the raw EEG data, we performed low pass filtering to remove the high frequency artefacts like EOG, EMG etc. Finite impulse response (FIR) filters are suitable for biomedical signal processing because these filters give output which is proportional to input. We have used FIR Low pass filters of various orders and cut off of **0.1 Hz.** After a thorough comparison, it is seen that for **30 order** filter, the leakage is significantly less.

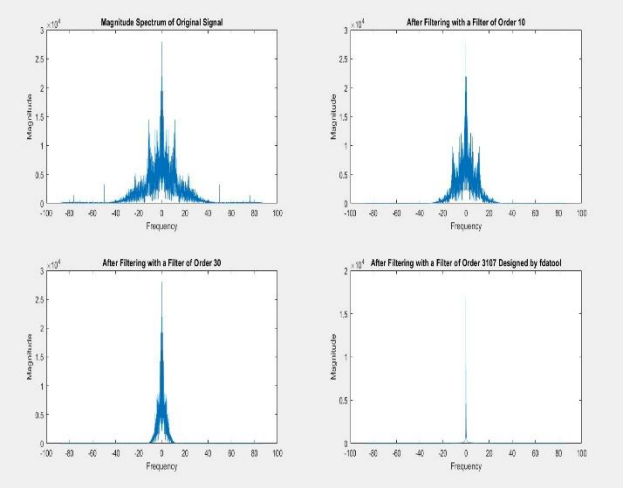


FIG. 4. Output signal after processed with different orders or filter.

Our next step is to use continuous wavelet transform. As the filtered EEG signal, x(t) is high ephemeral, wavelet reconstruction method is applied for noise reduction. The

continuous wavelet transform (CWT) coefficients are found at first using Morse Wavelet Scheme. Then the signal is again reconstructed using the coefficients found earlier.

After the overall signal processing, it is seen that the raw signal and the processed signal.

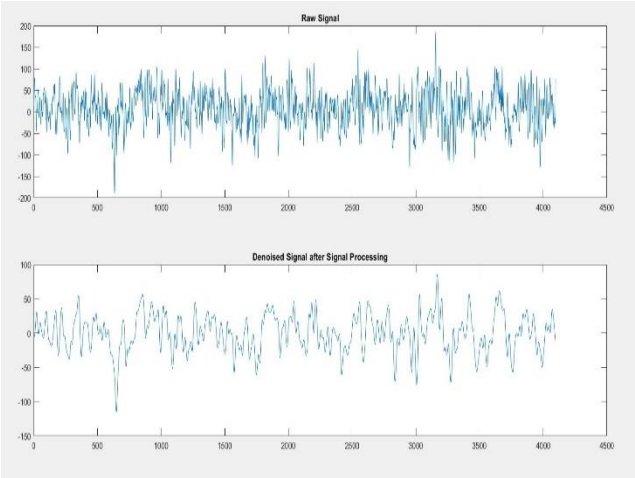


FIG. 5. Output signal after processing with the appropriate filter.

It has been shown the plots that the processed signal is smoother and has less spikes than the original signal.

### *C. Pre-processing Dataset for Modeling*

Steps related to data pre-processing:

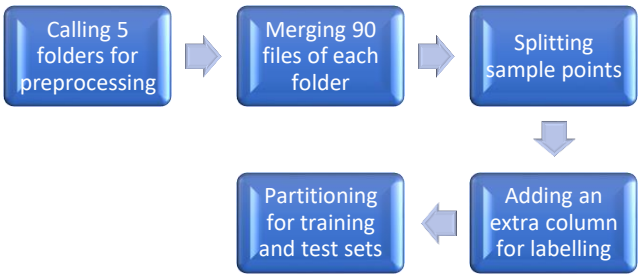


FIG. 6. Steps related to data pre-processing

Five folders have been called for preprocessing datasets. Then by skipping 10 files of each folder rest of 90 files have been merged. As each file contains 23.5s sample points, they need to be split into per second sample points. An extra column has been added to the table for proper labelling and partitioned whole data sets into test set and training set.

For proper modelling a large dataset is needed. Sampling frequency was 173.16HZ and each file has 4098 sample points with 23.5s . So after splitting each second has 178 sample points. Consequently each file has 23 rows and 178 colums. So For 450 files there are 10350 rows and 178 columns. Finally 179$^{th}$ column is introduced for levelling the whole dataset for seizure and non-seizure patients. Then whole datasets were partitioned into training set and testing set.

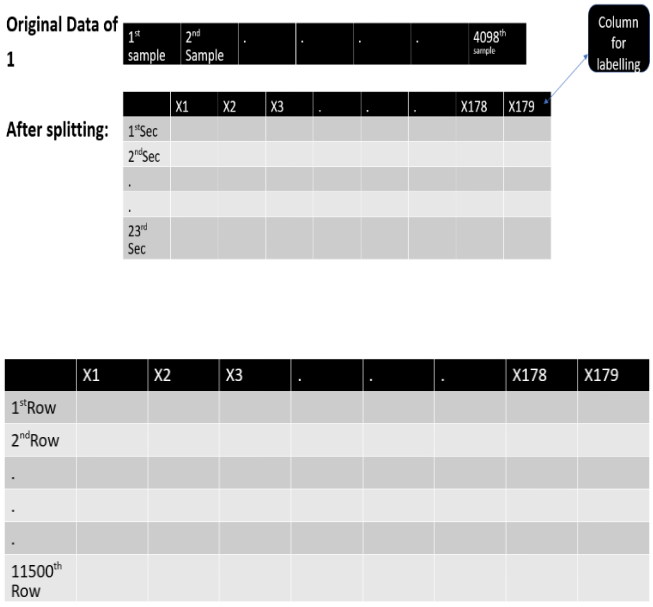


FIG. 7. Data Splitted into per second data

### *D. Algorithm Selection*

#### D.1. Decision tree

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data). In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

Important Terminology related to Decision Trees

1. Root Node: It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
2. Splitting: It is a process of dividing a node into two or more sub-nodes.
3. Decision Node: When a sub-node splits into further sub-nodes, then it is called the decision node.
4. Leaf / Terminal Node: Nodes do not split is called Leaf or Terminal node.
5. Pruning: When we remove sub-nodes of a decision node, this process is called pruning. You can say the opposite process of splitting.
6. Branch / Sub-Tree: A subsection of the entire tree is called branch or sub-tree.
7. Parent and Child Node: A node, which is divided into sub-nodes is called a parent node of sub-nodes whereas sub-nodes are the child of a parent node.Decision trees classify the examples by sorting them down the tree from the root to some leaf/terminal node, with the leaf/terminal node providing the classification of the example.

Each node in the tree acts as a test case for some attribute, and each edge descending from the node corresponds to the possible answers to the test case. This process is recursive in nature and is repeated for every subtree rooted at the new node.

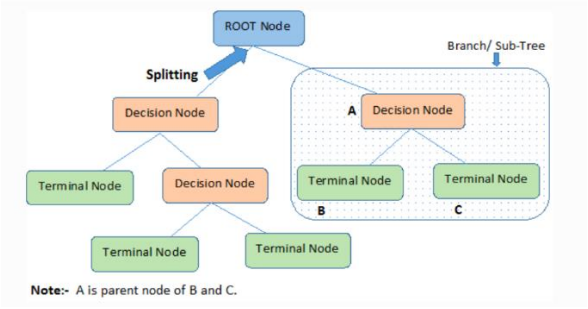


FIG. 8. Block Diagram of Decision Tree algorithm

If the dataset consists of N attributes, then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a complicated step. By just randomly selecting any node to be the root can't solve the issue. If we follow a random approach, it may give us bad results with low accuracy.

For solving this attribute selection problem, researchers worked and devised some solutions. They suggested using some *criteria* like: Entropy, Information, gain, Gini index, Gain-Ratio, Reduction in Variance, Chi-Square.

### D.2. Random Forrest

Random Forest is an example of ensemble learning, in which we combine multiple machine learning algorithms to obtain better predictive performance.

Two key concepts that give it the name random:

1. A random sampling of training data set when building trees.
2. Random subsets of features considered when splitting nodes.

A technique known as bagging is used to create an ensemble of trees where multiple training sets are generated with replacement. In the bagging technique, a data set is divided into **N** samples using randomized sampling. Then, using a single learning algorithm a model is built on all samples. Later, the resultant predictions are combined using voting or averaging in parallel.

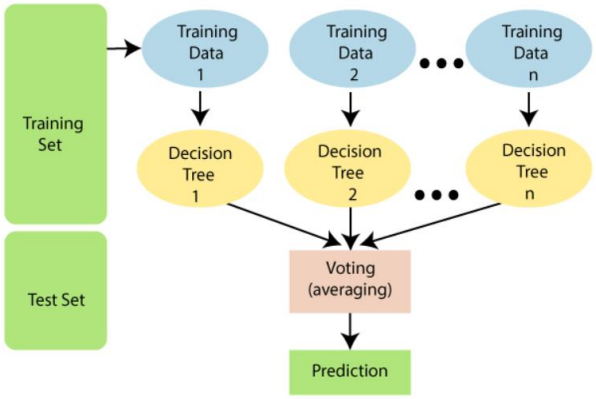


FIG. 9. Block diagram of Random Forrest Algorithm

## *E. Training Machine Learning Model*

### E.1. Decision tree Model

The machine learning model of Decision tree algorithm has been trained with the per second training dataset in matlab. In the training stage, the performance of the model is shown below –

```
Decision tree Accuracy: 92.40 %
Confussion Matrix For per second Data using Decision Tree
stats =
  9×4 table
```

| name | classes | | macroAVG | microAVG |
|---|---|---|---|---|
| "true_positive" | 1597 | 337 | 967 | 967 |
| "false_positive" | 78 | 81 | 79.5 | 79.5 |
| "false_negative" | 81 | 78 | 79.5 | 79.5 |
| "true_negative" | 337 | 1597 | 967 | 967 |
| "precision" | 0.95343 | 0.80622 | 0.87983 | 0.92403 |
| "sensitivity" | 0.95173 | 0.81205 | 0.88189 | 0.92403 |
| "specificity" | 0.81205 | 0.95173 | 0.88189 | 0.92403 |
| "accuracy" | 0.92403 | 0.92403 | 0.92403 | 0.92403 |
| "F-measure" | 0.95258 | 0.80912 | 0.88085 | 0.92403 |

FIG. 10. Performance scores of Decision tree model in training

Here, the accuracy of the model with per second data is 92.4%. The precision is 95.34% and the sensitivity of this model 95.17%. Lastly the F1 score of the model was 95.258%.

### E.2. Random Forrest Model

The machine learning model of Random Forrest algorithm has been trained with the per second training dataset in MATLAB. In the training stage, the performance of the model is shown below –

```
Random Forest Accuracy: 96.32 %
Confussion Matrix For per second Data using Random Forrest
stats =
  9×4 table
```

| name | classes | | macroAVG | microAVG |
|---|---|---|---|---|
| "true_positive" | 1643 | 373 | 1008 | 1008 |
| "false_positive" | 42 | 35 | 38.5 | 38.5 |
| "false_negative" | 35 | 42 | 38.5 | 38.5 |
| "true_negative" | 373 | 1643 | 1008 | 1008 |
| "precision" | 0.97507 | 0.91422 | 0.94464 | 0.96321 |
| "sensitivity" | 0.97914 | 0.8988 | 0.93897 | 0.96321 |
| "specificity" | 0.8988 | 0.97914 | 0.93897 | 0.96321 |
| "accuracy" | 0.96321 | 0.96321 | 0.96321 | 0.96321 |
| "F-measure" | 0.9771 | 0.90644 | 0.94177 | 0.96321 |

FIG. 10. Performance scores of Decision tree model in training

Here, the accuracy of the model with per second data is 96.32%. The precision is 97.5% and the sensitivity of this model 97.91%. Lastly the F1 score of the model was 97.71%. So the performance of this model was better compared to the decision tree model during the training.

### *F. Create function to predict from complete signal*

As the model has been training with per second data. But it needs to check the performance and get prediction for a whole signal, a function is needed which will talk the whole signal and return its prediction. For this, a function has been created which will take a whole signal of 23.6 seconds consists of 4097 samples. Then it is split the signal into 23 parts and makes it per second signal. Later these 23 signals will pass through the model and return a array consisting of different predictions of size 23x1. From the array the success rate (1s) will be counted and then it will be compared with the threshold level (68%). If the success rate is higher than the threshold, the whole signal will be considered as Seizure. Else, it will be considered and non-Seizure signal.

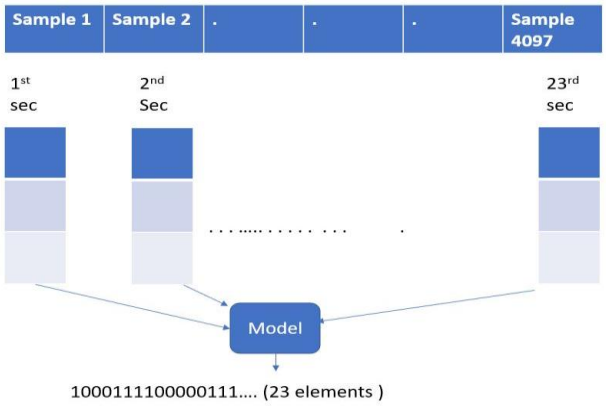


FIG.11. Block diagram of the working of the function

## IV. RESULTS ALANLYSIS

In the splitting section 10 files from each folder for testing were separated. So, the size of the total testing dataset is 50 which contains 40 non-seizure signals and 10 seizure signals. This testing set was totally unseen for the model in training and validation. As two different model were created, the performance of the both model will be analyzed from with this single testing set.

### *A. Testing of Decision Tree Model*

The results of the testing are shown below –

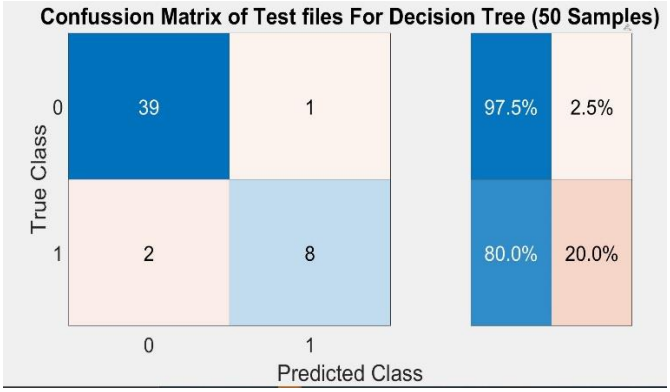


FIG.12. Confusion Matrix of decision tree model on testing set

Confussion Matrix of Test files For Decision Tree (50 Samples)
stats =
9×4 table

| name | classes | | macroAVG | microAVG |
|---|---|---|---|---|
| "true_positive" | 39 | 8 | 23.5 | 23.5 |
| "false_positive" | 2 | 1 | 1.5 | 1.5 |
| "false_negative" | 1 | 2 | 1.5 | 1.5 |
| "true_negative" | 8 | 39 | 23.5 | 23.5 |
| "precision" | 0.95122 | 0.88889 | 0.92005 | 0.94 |
| "sensitivity" | 0.975 | 0.8 | 0.8875 | 0.94 |
| "specificity" | 0.8 | 0.975 | 0.8875 | 0.94 |
| "accuracy" | 0.94 | 0.94 | 0.94 | 0.94 |
| "F-measure" | 0.96296 | 0.84211 | 0.90253 | 0.94 |

FIG. 13. Performance scores of Decision Tree model on testing set

### *A. Testing of Random Forrest Model*

The results of the testing are shown below –

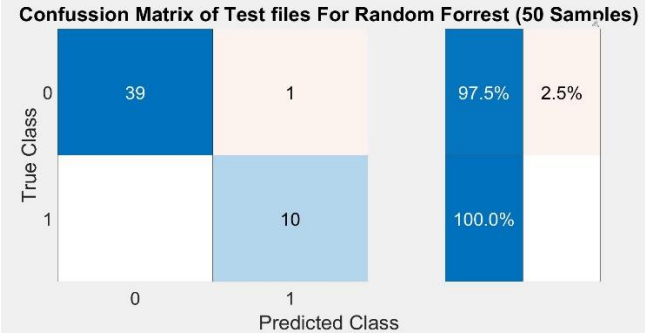


FIG.14. Confusion Matrix of decision tree model on testing set

Confussion Matrix of Test files For Random Forrest (50 Samples)
stats =
9×4 table

| name | classes | | macroAVG | microAVG |
|---|---|---|---|---|
| "true_positive" | 39 | 10 | 24.5 | 24.5 |
| "false_positive" | 0 | 1 | 0.5 | 0.5 |
| "false_negative" | 1 | 0 | 0.5 | 0.5 |
| "true_negative" | 10 | 39 | 24.5 | 24.5 |
| "precision" | 1 | 0.90909 | 0.95455 | 0.98 |
| "sensitivity" | 0.975 | 1 | 0.9875 | 0.98 |
| "specificity" | 1 | 0.975 | 0.9875 | 0.98 |
| "accuracy" | 0.98 | 0.98 | 0.98 | 0.98 |
| "F-measure" | 0.98734 | 0.95238 | 0.96986 | 0.98 |

FIG. 15. Performance scores of Decision Tree model on testing set

## V. CONCLUSION

It can be seen that; Random Forest Classifier gives better accuracy in tests. As the data set contains more non-seizure data than the seizure data, the model is a bit biased towards non-seizure cases. To get better results, the main aim is to collect more seizure data so that the model becomes balanced. Some other Machine Learning techniques can also be applied to test for the accuracy. As epilepsy is a very fatal disease and often randomly breaks out, the model might help detecting seizures with very less time.

## REFERENCES


[1] A. Guekht, M. Brodie, M. Secco, S. Li, N. Volkers, and S. Wiebe, "The road to a World Health Organization global action plan on epilepsy and other neurological disorders," *Epilepsia*, vol. 62, no. 5, 2021, doi: 10.1111/epi.16856.

[2] R. E. Ramsay, A. J. Rowan, and F. M. Pryor, "Special considerations in treating the elderly patient with epilepsy," in *Neurology*, 2004, vol. 62, no. 5 SUPPL. 2. doi: 10.1212/wnl.62.5_suppl_2.s24.

[3] D. Buck, G. A. Baker, A. Jacoby, D. F. Smith, and D. W. Chadwick, "Patients' experiences of injury as a result of epilepsy," *Epilepsia*, vol. 38, no. 4, 1997, doi: 10.1111/j.1528-1157.1997.tb01733.x.

[4] L. D. Iasemidis, D. S. Shiau, J. C. Sackellares, P. M. Pardalos, and A. Prasad, "Dynamical Resetting of the Human Brain at Epileptic Seizures: Application of Nonlinear Dynamics and Global Optimization Techniques," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 3, 2004, doi: 10.1109/TBME.2003.821013.

[5] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowledge-Based Syst.*, vol. 45, 2013, doi: 10.1016/j.knosys.2013.02.014.

[6] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 64, no. 6 I, 2001.

[7] R. Mukherjee, S. S. Dhar, and K. Tara, "Prediction of Disorder of Brain using EEG Signal Processing in MATLAB GUI Platform," 2018. doi: 10.1109/CEEE.2017.8412847.


s