EPILEPTIC SEIZURE DETECTION AND PREDICTION USING DEEP LEARNING

FELIX GEORGE
B160423EC

TONNY JOHN
B150168EC

DEEPAK M
B160615EC

ALEX JOSEPH
B160163EC

BIBIN BABY
B160178EC

ALEX JOHN
B160371EC

Guide
Dr.Sathidevi P S

I. INTRODUCTION

Epilepsy is a neurological disorder in which the affected person experiences abnormal brain activity for a brief period, which causes uncontrollable seizures and loss of awareness. The occurrence of these seizures is neither periodic, nor do they show any noticeable signs before they occur. This makes it very dangerous for them to go about their daily activities, as an unforeseen seizure makes them lose control over what they are doing. For example, people standing near the edge of a tall building may lose control and fall, or a driver having epileptic seizures may lose control of the vehicle and crash, endangering their own life and the lives of people around them. Since the complete cure for epilepsy is not expected in the near future, though tremendous research is being done in that area, a presently viable solution using existing technologies is to build a device that can accurately predict an oncoming seizure (with as low false positives as possible), without the patient being under the continuous monitoring of a doctor. We propose to use Machine Learning algorithms to achieve this objective. We plan to start with the random forest algorithm, which does not make any assumptions on the given data nor assume any particular distribution. This is appropriate since we are analyzing brain signals, which are randomized based on our mental activities. Also, errors in data are compensated for, since the algorithm itself adds errors before processing to analyze it. Depending on the accuracy of this method, we will try on more improvements to the same method or, if found to be low on the accuracy, move to a different method like Long-Short Term Memory algorithm, RNN, CNN, etc.

II. PRINCIPLES

A. EEG

Electrical activity emanating from the brain is displayed in the form of brainwaves. There are four categories of these brainwaves, ranging from the most activity to the least activity. When the brain is aroused and actively engaged in mental activities, it generates beta waves. These beta waves are of relatively low amplitude and are the fastest of the four different brainwaves. The frequency of beta waves ranges from 15 to 40 cycles a second. Beta waves are characteristics of an actively engaged mind.

The next brainwave category in order of frequency is alpha. Where beta represented arousal, alpha represents non-arousal. Alpha brainwaves are slower, and higher in amplitude. Their frequency ranges from 9 to 14 cycles per second. A person who has completed a task and sits down to rest is often in an alpha state.

The next state, theta brainwaves, is of even higher amplitude and slower frequency. This frequency range is generally between 5 and 8 cycles a second. A person who has taken time off from a task and begins to daydream is often in a theta brainwave state. It is a state where tasks become so automatic that you can mentally disengage from them. The ideation that can take place during the theta state is often free flow and occurs without censorship or guilt. It is typically a very positive mental state. The final brainwave state is delta. Here the brainwaves are of the greatest amplitude and slowest frequency. They typically center around a range of 1.5 to 4 cycles per second. They never go down to zero because that would mean that you were brain dead. But, deep, dreamless sleep would take you down to the lowest frequency. Typically, 2 to 3 cycles a second. When we go to bed and read for a few minutes before attempting sleep, we are likely to be in low beta. When we put the book down, turn off the lights and close our eyes, our brainwaves will descend from beta to alpha, to theta, and finally, when we fall asleep, to the delta. Gamma waves are a more recent discovery in the field of neuroscience. Thus the understanding of how they function is continually evolving. To date, its known that Gamma waves are involved in processing more complex tasks in addition to healthy cognitive function. Gamma waves are found to be essential for learning, memory, and processing and they are used as a binding tool for our senses to process new information. In people with mental disabilities, much lower levels of Gamma activity is recorded Frequency range: 40 Hz to 100 Hz

III. Tools used in Analysis

- Artificial Neural Networks: They are computing systems that are inspired by biological neural networks. An ANN has a collection of connected units called nodes that model the neurons. A node that receives a signal process it and can signal nodes connected to it. The signal at a connection is a real number, and the output of each node is computed by a non-linear function of the sum of its inputs. The connections are called edges. Nodes and edges typically have a weight that adjusts as learning progresses.
- Random forests: Random forests are developed to decrease the inaccuracy of the decision tree algorithm. It is a collection of all possible trees from a given data-set, so is the name forest. It consists of building an ensemble of decision trees grown from a randomized variant of the tree induction algorithm. They usually have low bias and high variance, making them very like to benefit from averaging process.

  Random forest algorithm works well even if the data-set is noisy or even if it contains missing values. Unlike the decision tress, the selection of variables and the number of selections of variables are completely randomized in each different tree in the bagging procedure, which prevents overfitting of the complete model whereas a single tree might have high variance and low bias. Random forests give the idea on the contribution of each variable to the model. This helps in getting a deeper understanding and interpretation of the data-set and model. Random forests are difficult to analyze mathematically, but the theoretical methods of applying random forest algorithms works well in practice
- Convolutional Neural Networks: They are a type of Artificial Neural Network that has one or more than one convolutional layer. These layers can either be wholly interconnected or pooled. Before passing the result to the next layer, the convolutional layer uses a convolutional operation on the input. Due to this convolutional operation, the network can be much deeper but with much fewer parameters. Due to this ability, convolutional neural networks show very useful results in image and video recognition, natural language processing, and recommender systems. A typical CNN has Convolutional Layers, Pooling Layers, and Fully Connected Layers.

  The Convolutional Layer extracts features by convolving the inputs with different feature extraction masks. The Pooling Layer is primarily used to help reduce the computational complexity and extract prominent features.

It is also referred to as the downsampling layer. It helps in controlling overfitting. The convolution and pooling layers would only be able to extract features and reduce the number of parameters from the original images; to generate the final output, we need to apply a fully connected layer to generate an output equal to the number of classes we need. The Fully Connected Layer is the essential ANN with an input layer, few hidden layers, and an output layer. Therefore the output from the previous layers must be flattened before feeding it into the Fully Connected Layer.

- Wavelet transforms: Since sine waves oscillate forever and are not localized in time, abrupt changes in the time domain cannot be represented by the Fourier transforms well. So, to identify and characterize data with abrupt changes, we use wavelet transform. Wavelet transforms also enable us to obtain a simultaneous time-frequency analysis and also visualize it. A wavelet is a finite duration decaying signal with zero mean. There are many types of wavelets, like the Mexican hat, Morse, and Amor, that are used to represent different classes of data. To construct the wavelet transform of a signal, we do two operations on these wavelets Scaling and Shifting. Scaling refers to time scaling scaling by a factor of 2 means compressing the time-limited wavelet to half its original time length. The different scaled wavelets are shifted moved in time along the full-time length of the signal and compared with the original signal by correlation at each stage of the shift to determine the amplitude of each scaled wavelet in the time scale. Here we use Continuous wavelet transform -CWT with the type of wavelet taken as Morse. The output is a 3D array of magnitude vs. scale vs. time. The scale multiplied by the sampling frequency gives the actual frequency values. We get a representation of what the correlation of a particular scaled wavelet is with the signal at a particular point of time. To represent the cwt, we generally plot the time in the x-axis, Frequency (scale x Sampling frequency) in Y-axis and the magnitude in z-axis

IV. IMPLEMENTATION

A. DATASET AND CHALLENGES

One of the sections of the dataset containing 23-channel peg signal monitoring epileptic patients for periods of interval 1 or 4 hours each. This interval is too small for multiple seizures to occur, or no seizures at all. This causes a major problem with the dataset that will reflect in any machine learning model we try to implement since all will learn to predict or are biased to predict any given sample as no seizure. This is because it only see few samples with seizure. EEG data is sampled at 256Hz, which makes the dataset extremely large and difficult to try new algorithms due to long processing time required. We tried to solve all the mentioned problems with the dataset using various methods discussed below.

B. OUR APPROACH

- We avoided all the files with no seizure data for training and testing purposes to avoid complete biasing of models to predict no seizure. We used the different samples of the same patient's data to test the model since the correlation between the data will help to get a better result than using uncorrelated data. We started with the random forest since the dataset is in the amplitude domain where the values have no relations with previous occurrences where a neural network will perform poorly. This method gave good performance with high accuracy but with a poor F1 score.

- To overcome the problem due to imbalanced dataset, we used SMOTE (Synthetic Minority Oversampling Technique) oversampling method. This made all the classes in the dataset of equal samples, which prevents biasing the model to predict classes with more samples. This gave us better results than using the dataset alone for training and is verified by an increased F1 score and still maintaining the accuracy.

- We also made a completely different approach since the classes we interested in are minority in size. We used outlier dictation to predict seizure points. This method didn't work well since it is unable to distinguish or to construct boundaries to predict seizure in amplitude domain.

- The dataset, when visualized using an edf viewer software, we are easily able to identify the parts in the dataset when the seizure has occurred since there is a rise in the frequency and the amplitude of the signal data. Since Convolutional Neural Networks show very effective results in image and video recognition, we fed the signal data into a typical Convolutional Neural Network. The trained model(without undersampling the dataset) gave a below-average accuracy, which can still be raised slightly by choosing the hyperparameters-number of layers and filter sizes more appropriately.
- Though the learning algorithms may extract more features from the time vs. amplitude data during training, we wanted to aid the process by making use of characteristics in the frequency domain. Starting with the basic approach of the Fourier transform, we analyzed the power spectral density of the signals in the frequency domain. We separated different bands of frequency to see if the peak in the time domain corresponding to the seizure would be constituted by one of the bands or at least mostly constituted by one. But we found the seizure peak constituted all different frequencies from 0Hz to 32Hz. and also since the basis used in Fourier transform is a sine wave of an infinite period, the Fourier transform of the data taken in parts may not produce the same results as the Fourier transform of the complete data taken together. What we needed was a method that can represent what frequencies occurred at what times. We decided to use the wavelet transform.

C. RANDOM FOREST CLASSIFIER

This must have been done to eliminate the possibility that the algorithm classifies one of the periods based on the occurrence of another period before/after it in the order of index (simply saying, time). That is, we want the random forest to know that a particular set of data represents the interictal period not just because it comes after the pre-ictal period, but mainly because of its own characteristic EEG features. The dataset was separated into a training set and test set. The training set is what we use to train the random forest. It consists of the EEG data as the classification parameter, and the markings are also given to match the output. The random forest trains such that the features of EEG at a time instant are matched with the marking of that time (both markings and EEG of the training data). The training may take minutes or hours, depending on the size of the data. The data was massive, but considering the power of the system allotted, the training took only minutes. Now, once the training is over, we input the test data just the EEG readings from the same patient, but from a different time interval. The output should be markings 0s, 1s or 2s corresponding to each reading of EEG, and it should match the actual markings given along with the test data. We initially trained with raw data i.e., the actual EEG data sampled at 256 Hz without any pre-processing.

The training data contained interictal and pre-ictal periods. But using this raw data, the prediction was not good. It predicted seizures in post-ictal regions and also predicted pre-ictal periods when there were none in the data. The results can be seen below.

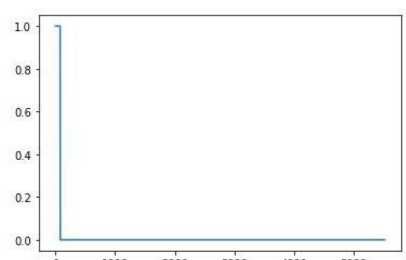

Fig. 1. Actual Markings

D. RANDOM FOREST CLASSIFIER AFTER APPLYING SMOTE

(Synthetic Minority Oversampling Technique)
Here we used an oversampling method to increase the number of samples in the under sampled classes. We used the same training and testing

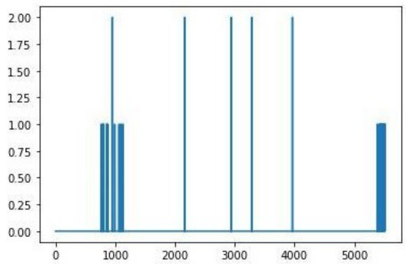

Fig. 2. initial predictions using Random Forest

```
[[5359   18    5]
 [  99   39    0]
 [   0    0    0]]
```

Fig. 3. Confusion Matrix. The rows represent actual values and the columns represent the predicted values. For example, row 0, column 2 - 5 means 5 zeros were predicted as 2's

dataset that we used for the former method since it gave a better comparison.

As we can see the interictal points are marked correctly, the number of false positives were reduced significantly but predicted some pre-ictal points even though the test set we used had no markings of pre-ictal data.

This increased the F1 score of the model significantly to 72 than the previous 40. We got an accuracy of 98 with a precision of 68 and 74 as recall. These were the maximum values we got after a lot of fine-tuning the model and was unable to get more satisfying results in working with a dataset in the amplitude domain.

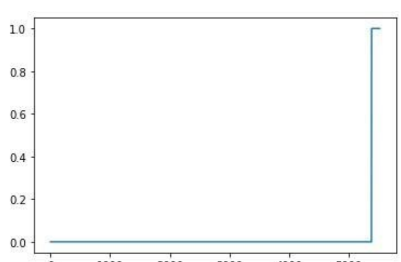

Fig. 4. Actual Markings

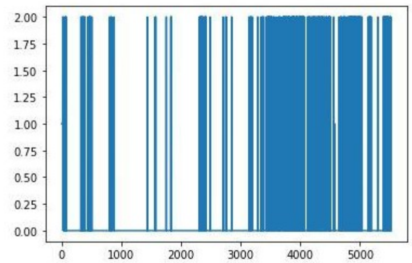

Fig. 5. output after training Random forest classifier with balancing the dataset

```
array([[5037,   55,  350],
       [   7,   62,   10],
       [   0,    0,    0]])
```

Fig. 6. Confusion Matrix. The rows represent actual values, and the columns represent the predicted values. For example, row 0, column 2 - 350 means 350 zeros were predicted as 2's

E. OUTLIER DETECTION

Our dataset was completely unbalanced in a ratio similar to 1:90 or more in most cases. We considered outlier detection because of the same reason. Outlier detection is an unsupervised learning technique to remove the data separated above a few standard deviations. We applied a multivariate outlier detection method, since our data has 23 channels. This is the best method to apply when values distribution cannot be predicted but is computationally expensive.

We applied isolation forest algorithm which predicts a sample as an outlier or not an outlier by using an outlier score, given by fig.4 below. The results of this algorithm were similar to that of simple random forest implementation of the model with an accuracy of 75. This shows no significant result, and the numbers were not showing any promising results.

The failure of the model makes sense since the data is in the amplitude domain and it gives no features to learn from it. This method will produce better results in frequency domain processing since it can easily mark seizure when frequency is above a particular threshold.

F. Convolutional Neural Network

The Convolutional Neural we used had two convolutional layers, each followed by a pooling layer. The fully connected layer had two hidden layers, and the final output layer has three nodes because our dataset had three classes of data - no seizure, before seizure and during seizure. Since the data cannot be simply directly fed into the neural network, we first created a program that created the dataset required for feeding the data into the neural network. The program captured screenshots and saved them into three different folders depending on the key pressed.

First, three folders were created- 0,1,2. A python function named ImageGrab using the python PIL library was used to capture the screenshots. These screenshots were then processed by masking them to get only the region of interest. The signal data - edf file, was put into an edf viewer and was played like a continuous video. The program was made in such a way that screenshots were constantly being taken, being processed and saved in the folder 0. Approximately 5 minutes before the seizure, key 'w' was pressed, and the processed screenshots were now being saved in folder 1. At the exact duration when the seizure started, key's' was pressed, and the processed screenshots were now being saved in folder 2. After the seizure was over, key 'q' was pressed, and the data was again being stored in the folder 0. In this way, we created the training data saved in three folders indicating three different classes.

The images were then read from these three folders and fed into the Convolution Neural Network earlier mentioned, and a trained model was obtained. This model was then used to predict which class -no seizure before a seizure or during seizure; an inputted image belonged to.

G. Frequency Domain analysis

We then tried isolating the frequency components of the data to see if the seizure area were characterized by any particular frequency. The power spectral density of one of the channels is plotted below for this purpose. It can be seen that most of the power is concentrated on the low-frequency components namely, Delta, Theta, and Alpha waves are occupying most of the spectrum. They are present through the data. So we tried separating the higher frequency components using a highpass filter to separate components having frequency¿20Hz. The resulting spectra and the time domain signal are as follows.

5329	303	0
1793	16	0
0	725	0

Fig. 7. Confusion Matrix (without balancing the dataset); The rows represent actual values, and the columns represent the predicted values. For example, row 0, column 1 - 303 means 303 zeros were predicted as 1's

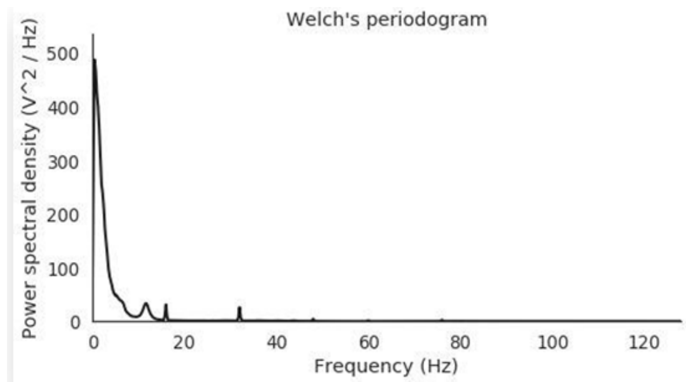

Fig. 8. power spectral density of one of the channels - channel 16

As can be seen from the markings, the seizure (Marked as 2) is characterized by a sudden peak appearing in the time domain signal (marked with green brackets). While predicting this, time domain information might be made use of. However, continued filtering of different passbands made us realize that all frequency components constituted the characteristic peak/spike shown, and in almost equal proportions. Thus, we might conclude that frequency selective analysis might not be the approach to best results though we are moving forward with more processing in the frequency domain. Further, we will try to analyze the wavelet domain characteristics of the signal.

1) Wavelet domain analysis: Though the Fourier-frequency domain analysis of the signals did not produce distinguishable features or made any fre-

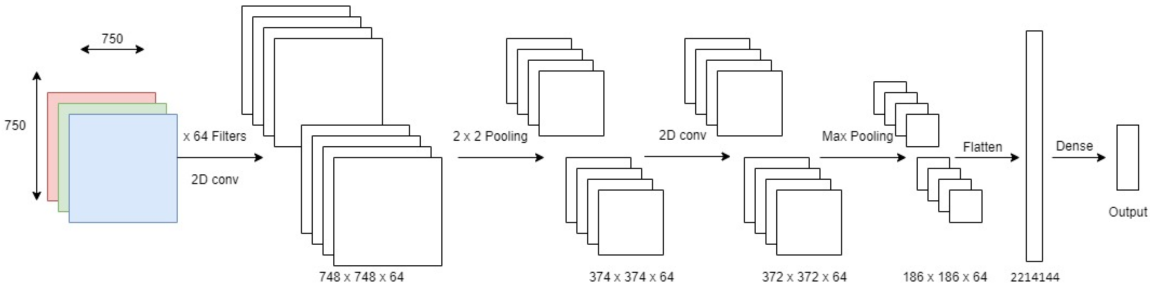

Fig. 9. CNN Architecture

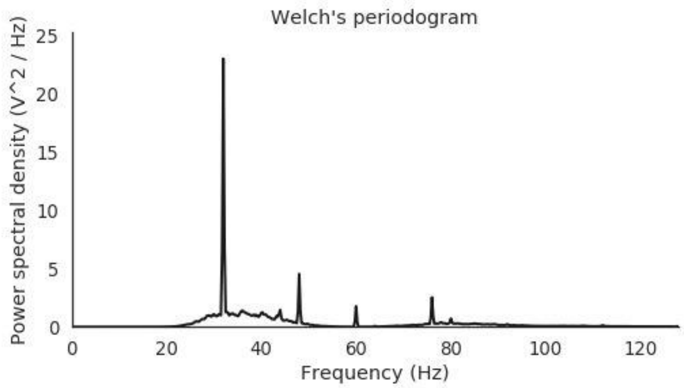

Fig. 10. PSD after highpass filtering

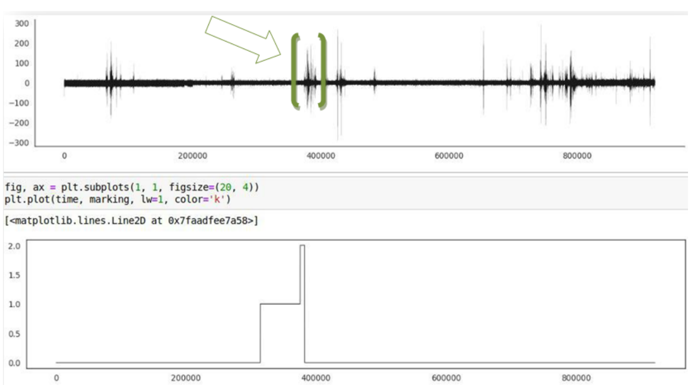

Fig. 11. Time domain representation of highpass signal. The bottom figure shows seizure markings - 0- post-ictal, 1 - pre-ictal, 2 - interictal

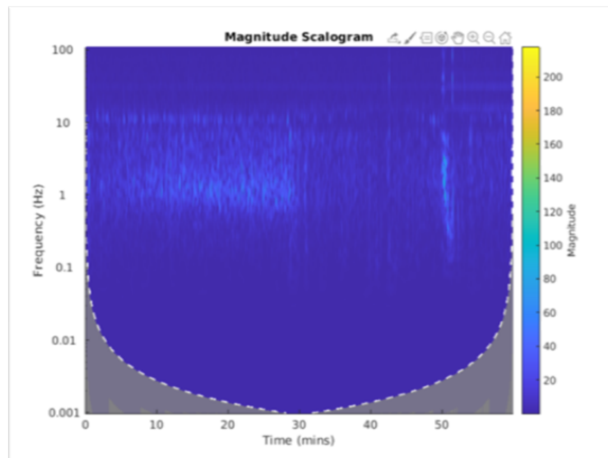

Fig. 12. Channel 1 Here, the seizure region can be identified by two characteristic peaks at 50 minutes and 25 seconds at about 1.41Hz and 2.45Hz.

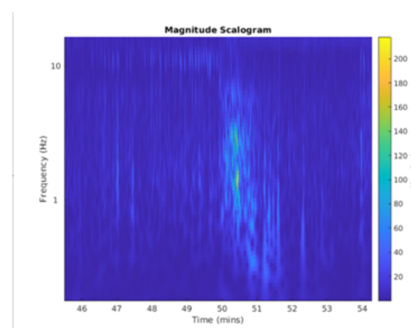

Fig. 13. zoomed-in view of the seizure region that shows the clear peaks (yellow) at the seizure region.

quency corresponding to the seizure period stand out. So we used wavelet transform, specifically continuous wavelet transform using Morse wavelets to see if any distinguishable features can be identified. The results of applying cwt can be seen below. In most of the channels, the region of the seizure is projected up with maximum amplitude, but in some channels, there are areas of similar magnitude in other time locations also. But in all cases, the frequency (scale) of interest lies in the region of nearly 1 2.5Hz. Also the channels show, in that frequency region, another distinguishable characteristic feature before the seizure time (shift) if you consider a narrow horizontal band between 1Hz and 2.5Hz, there is a region of increasing density of higher

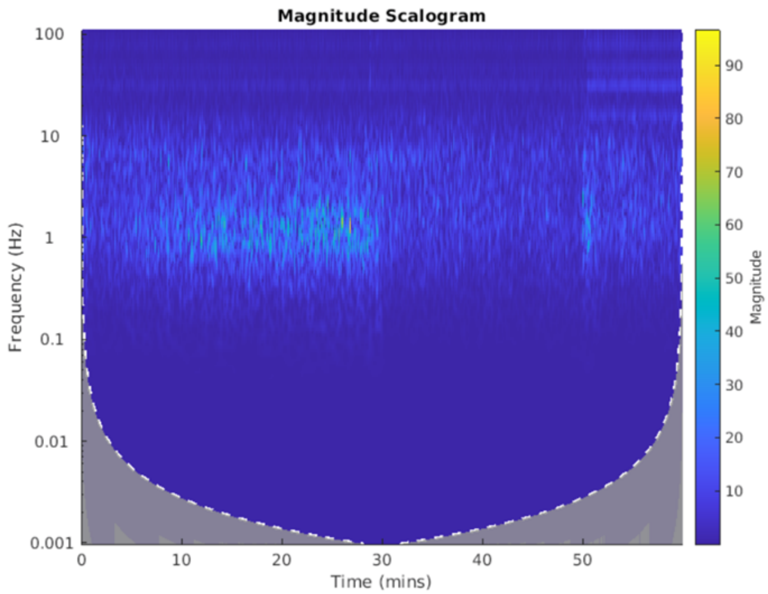

Fig. 14. Channel 4 cwt

magnitudes from 40 minutes before the seizure to 20 minutes before the seizure, followed by a region of decreased (normal) density of higher magnitudes. This could be the defining feature of the pre-ictal period. This feature could be further exploited to predict seizures.

REFERENCES

[1] Syed Muhammad Usman, Muhammad Usman, and Simon Fong, Epileptic Seizures Prediction Using Machine Learning Methods, Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 9074759, 10 pages, 2017.

[2] Turky N. Alotaiby, Saleh A. Alshebeili, Faisal M. Alotaibi, and Saud R. Alrshoud, Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals, Computational Intelligence and Neuroscience, vol. 2017, Article ID 1240323, 11 pages, 2017

[3] Gilles Loupe Understanding Random Forests: From Theory to Practice (2014) Cornell University , arXiv:1407.7502

[4] CHB-MIT Scalp EEG Database:https:physionet.org/content/chbmit/1.0.0/