Department of Electrical
& Computer Engineering
Faculty of Engineering & Architectural Science

Course Title:	Biomedical Signals
Course Number:	BME772
Semester/Year (e.g. F2017)	F2019

Instructor	Dr. Sridhar Krishnan

Project Title:	Seizure Detection in EEG signal

Submission Date:	November 25th, 2019
Due Date:	November 25th, 2019

Student LAST Name	Student FIRST Name	Student Number	Section	Signature*
Dayyani	Faranak	500692664	03	F.D
Majd	Parnian	500765210	04	PMajd
Taki	Sarina	500773183	04	S.T

*By signing above you attest that you have contributed to this written lab report and confirm that all work you have contributed to this lab report is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a "0" on the work, an "F" in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Student Code of Academic Conduct, which can be found online at: http://www.ryerson.ca/senate/current/pol60.pdf

Seizure Detection in EEG signals

Faranak Dayyani (500692664), Sarina Taki (500773183), Parnian Majd (500765210)

I. INTRODUCTION

The purpose of this project is to detect seizures by analyzing electroencephalogram (EEG) signals. This was done through multiple steps of preprocessing which consist of filtering and denoising the signal, feature extraction and lastly, a machine learning algorithm/component. The electroencephalogram (EEG) signals are recordings of the electrical activity of the brain [1]. This was done by using 23 electrodes/channels in different lobes of the brain. These electrodes are placed on the scalp of the subject and the data is recorded. The EEG signals have a frequency of between 0.5 to 30 Hz. They consist of different bands/sections that each consist of a specific frequency and are correlated to a specific state of the subject. These bands are described in the table below along with their state of the subject and their corresponding frequencies [2]:

Table 1: EEG Bands

Bands	Frequency Range (Hz)	State
Delta (δ)	0.5 - 4	Deep sleep stages
Theta(θ)	4 - 8	Beginning stages of sleep
Alpha (α)	8 - 13	Relaxed mental state, when the subject is at rest with eyes closed.
Beta (β)	13 - 30	Normal waking consciousness

In order to detect seizures, EEG signals have to be analyzed. Epilepsy is one of the most common neurological disorders that affect people of different ages. It is a chronic disease in which a person is diagnosed if they have one or more unprovoked seizures. In other words, the seizure was caused by reasons other than reversible medical conditions [3]. Seizures can cause changes in the pattern of EEG signals. By monitoring a set of EEG data from patients who suffer from epilepsy/have seizures, we are able to detect patterns and classify the EEG signals for categories of normal versus seizure data.

In order to classify the EEG data, the data needs to be preprocessed. EEG signals are highly susceptible to different sources of noise [4] which makes the analysis of this type of signal is very complicated and difficult. EEG signals have very small signal to noise (SNR) ratio. The higher the SNR value, the better specification the signal has, meaning there is more useful information in the data rather than the unwanted data which we call "noise" [5]. Some sources of noise for EEG signals are environmental such as AC power lines, lighting, and electronic equipment. Another noise source is movement artifacts that are caused by muscle contractions (EMG) as well as the cardiac signal (ECG) [4].

In preprocessing, multiple methods were used for denoising the signal such as synchronized averaging, a bandpass filter, band separation, rectification, smoothing of the signal and normalization of the signal. Synchronized averaging is a method of smoothing the signal by taking the average of each point of the signal in the time domain from the 23 channels that the EEG data is being recorded. Next, a bandpass filter was applied to the signal in order to separate the frequency desired by the EEG signals. The desired frequency where all 4 bands of the delta, theta, alpha and beta of the EEG signals fall into, is in the range of 0.5 to 30 Hz. A bandpass filter is a combination of low pass and a high pass filter. The four bands were then separated in order to determine which band the seizure is occurring in. This was obtained by introducing a new parameter called "entropy" into the system. Entropy is a statistical parameter that demonstrates "the amount of uncertainty and randomness in a pattern" [6]. The higher the value of entropy, the noisier the signal is. The entropy value was used to determine which band has the least amount of noise. This was later on used in the machine learning algorithm. The next step was to smooth out the signal further. This was done using the "smoothdata" built-in function in MATLAB which has the sole purpose of smoothing noisy data [7]. This function was used with Savitzky-Golay filtering method which was also a built-in function in MATLAB. Based on a study in Iran University of Science and Technology, applying the Savitzky-Golay filter was proved to be a very effective method in denoising and smoothing the signal [8]. In the end, the normalization was done to reduce the amplitude of the signal for easier visualization of data.

For the feature extraction component of this project, the features that were chosen for seizure detection were entropy, root mean square (RMS) and standard deviation (STD). The RMS and the STD values were measured for the bands with the lowest value in entropy. RMS and STD are statistical measures that help identify certain patterns in a signal.

In the last step, a machine learning algorithm was created using the features proposed above. This paper is mainly focused on the design of an EEG processing system which can distinguish and detect signals that contain seizures versus normal EEG signals. In the methods section, the details of preprocessing, feature extraction, machine learning algorithm and the accuracy of the results will be discussed. The diagram below illustrates the various steps taken in completing this task:

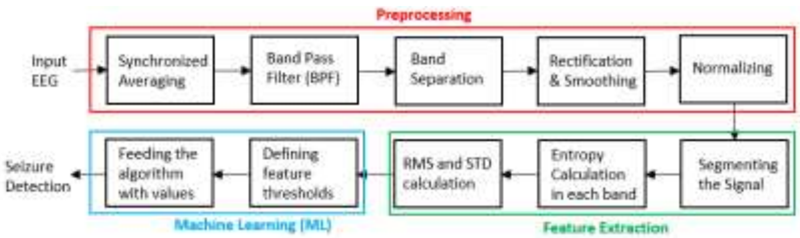

Figure 1: Block Diagram of the system

II. Methods

Preprocessing:

An EEG signal usually ranges between 0.5 - 30 Hz in frequency and it consists of different sections/bands that each contain brain activity of specific frequencies. The preprocessing stage is mainly focused on removing the noise from the signal since EEGs are known to one of the noisiest signals. The first step for denoising, was applying a filter called "Synchronized Averaging". The synchronized averaging method is mainly used for smoothing of the signal. This was done by taking the average of each point of the signal in the time domain from the 23 channels that the EEG data was recorded from. The equation used in getting the synchronized average of signals from multiple channels can be shown as below [11]:

$$\bar{y}(n) = \frac{1}{M}\sum_{k=1}^{M} y_k(n) \qquad \text{[Equation 1]}$$

In equation 1 above, M represents the number of signals from different channels. $y_k(n)$ represents one realization of the signals.

The second step was applying a bandpass filter to isolate the 0.5 to 30 Hz of valuable EEG data in the signal. A 6th order Butterworth filter was applied to the signal. A bandpass filter (BPD) is a combination of low pass and high pass filters. In this case, the cut-off frequency for the low pass filter (LPF) and the high pass filter (HPF) was set to be 30 and 0.5 Hz, respectively.

Next, the four bands described in table 1 were separated. Each band corresponds to a specific state in the subject. Seizures can occur in all states, therefore all bands were included for analyzing seizure data. Referring to table 1, the bands delta, theta, alpha and beta have frequency ranges of 0.5-4 Hz, 4-8 Hz, 8-13 Hz, and 13-30 Hz, respectively. Similar to the last step, each of these bands were separated by applying a 4th order Butterworth bandpass filter (BPF). Each band was then rectified in order to calculate the RMS value which can only be taken for positive values mathematically. The rectification process is performed by taking the absolute value of the entire EEG signal.

After going through multiple steps of denoising the signal as explained above, the signals were still difficult to analyze due to the large amount of noise. A smoothing function that is built-in MATLAB was used in order to smooth out the EEG signals for each band. The "smoothdata" function has the sole purpose of smoothing noisy data [7]. Based on a study in Iran University of Science and Technology, applying a filter called "Savitzky-Golay" filter proved to be very effective in denoising signals [8]. After further research, it was determined that the "Savitzky-Golay" filter can be used in conjunction with the "smoothdata" function [7]. Therefore, the two were used together for better reducing the noise in the EEG signals.

The last step in preprocessing was the normalization of the signals. This is often done in order to reduce the amplitude of the signal for better visualization of EEG data which leads to better analysis [9].

Feature Extraction:

The EEG signals that were used in the training set, testing set, and the validation set, all had a 3600 seconds duration (1 hour). After the separation of the four EEG bands in the preprocessing section, the bands were all segmented. The segmentation period was 120 seconds since that was the longest seizure period observed which leads to having 30 segments overall.

The features that were chosen for analysis in order to detect seizure data are entropy, root mean square (RMS) and standard deviation (STD). Entropy is a statistical parameter that demonstrates "the amount of uncertainty and randomness" in a signal or pattern [6]. The higher the value of the entropy, the noisier the signal is. The entropy values for signals with seizure data were analyzed. The entropy value was used to determine which band has the least amount of noise. This was later on used in the machine learning algorithm. The lower the value of the entropy, the less noisy the signal is. Therefore, the minimum entropy was chosen as one of the features that was used to choose a band for seizure detection in a signal.

The two other features that were used for detecting seizures in an EEG signal are RMS and STD. The root mean square and standard deviation are statistical measures that help identify patterns in a signal. RMS is also a “measure of energy content in a given signal” [10].

After the seizure band was indicated with entropy specifications, the STD and RMS values were taken for all 30 segments in the seizure band. The training set contained 24 EEG signals in which some were normal EEG signals and some were seizure data. The number of normal and seizure EEG signals were chosen randomly. Five trials took place for training the data and the amount of normal and seizure data were selected randomly each time. In the training set, all the seizure periods were specified. Therefore, the seizure segment was chosen and the RMS and STD values for that segment were recorded. This was done for all EEG signals with seizure data in the training set. For the normal signals, after identifying the band with minimum entropy, the average of RMS and STD values were taken for that band. Arrays of values for RMS and STD were created for both normal and seizure data in order to be used in machine learning (ML) algorithm.

Machine Learning and Accuracy Calculation:

In order to perform any machine learning algorithm for detection purposes, the dataset should be divided into either two groups or three groups. The population would be broken in such a way so that 80% is used for "training" and 20% is used for "classification". On the other hand, the dataset would be cut into three groups of 60% ("training"), 20% ("validation"), and 20% ("classification") if the second option is preferred. In this paper, the first logic is applied mainly because it would result in a more diverse dataset for training and increased accuracy.

As it was mentioned previously, the extent of the dataset used in this project was 30 EEG signals in total which 15 of them had seizure portion and 15 were driven from healthy individuals. 80% of the EEG signal dataset (24 signals) was used for the training purposes which were selected randomly. The system was trained based on the feature values of these 24 signals and the definition of healthy and unhealthy was defined for it. The second step was "classification" for which we devoted 20% of our dataset that is 6 EEG signals and performed a "decision tree" classifier (refer to Figure 2) Based on this classifier, the provided EEG signal would be determined as a non-seizure or seizure based on the values of the extracted features defined for these two states. Consequently, the machine learning type implemented for the purpose of this project is "supervised binary ML".

Furthermore, accuracy calculation has been done through the performance of 5-fold cross-validation on the whole population. This means that all the 30 signals were divided into 5 groups (each having 6 individual signals), 4 of which was used as the training dataset and one group was used as the testing dataset and the accuracy of it was calculated. This procedure was applied 5 times and each time with a new group set as the testing population. Accuracy calculation is based on the values of "True Positive", "True Negative"," False Positive"," False Negative" (Equation 2). "True Positive" is defined as the cases that the machine results in a correct judgment by determining the given EEG as a seizure EEG whereas, "False Positive" is the cases that the EEG signal is actually from a healthy individual but is determined as seizure. Also, "True Negative" are those healthy EEG signals that are correctly determined as healthy by the machine and "False Negative" are those that are concluded as healthy but are in fact, unhealthy based on the ground truth provided to us. These variables as well as "sensitivity" and "specificity" are determined based on a confusion matrix (Table 2 to Table 6). Sensitivity is interpreted as the extent to which a diagnostic test is specific for a particular condition (Equation 4) and Specificity is explained as the quality of state of a medical device being sensitive to "true negative" states (Equation 3). The obtained average accuracy percentage of the 5-fold cross-validation was 66.67%.

$$Accuracy: \frac{(TN+TP)}{(TN+TP+FP+FN)} \qquad \text{[Equation 2]}$$

$$Sensitivity: \frac{(TP)}{(TN+TP+FP+FN)} \qquad \text{[Equation 3]}$$

$$Specificity: \frac{(TN)}{(TN+TP+FP+FN)} \qquad \text{[Equation 4]}$$

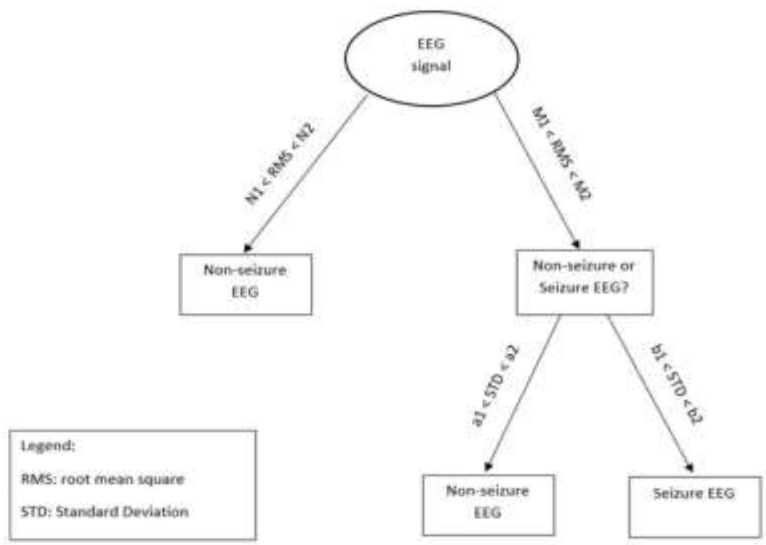

Figure 2: Diagram of the decision tree classifier

III. RESULTS

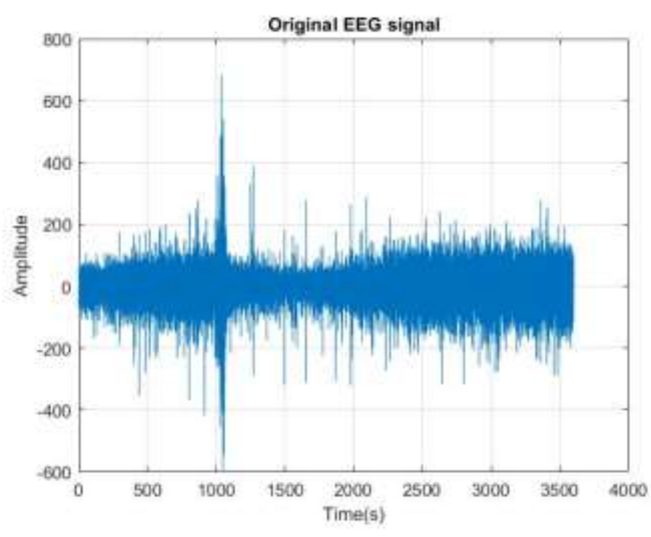

Figure 3: This graph shows the original EEG signal of one electrode of the subject.

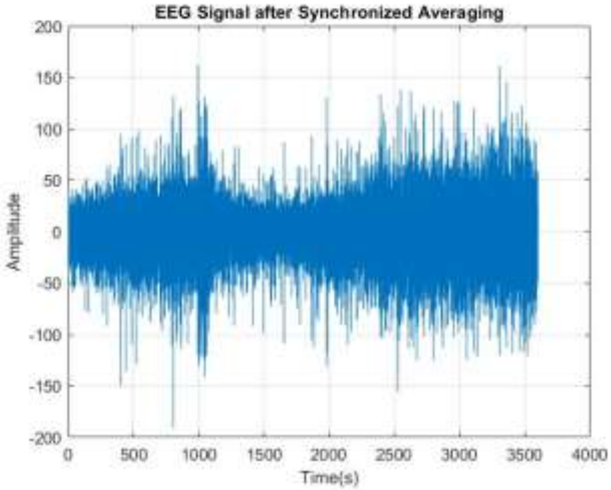

Figure 4: This graph shows the Synchronized average EEG signal of 23 electrodes.

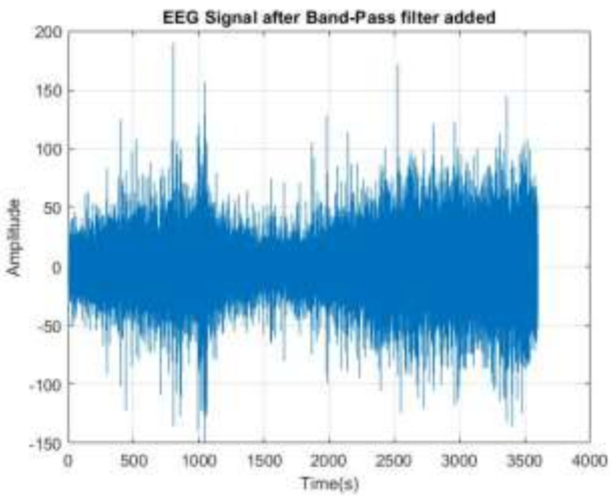

Figure 5: This graph shows the Band-pass filtered EEG of synchronized average EEG.

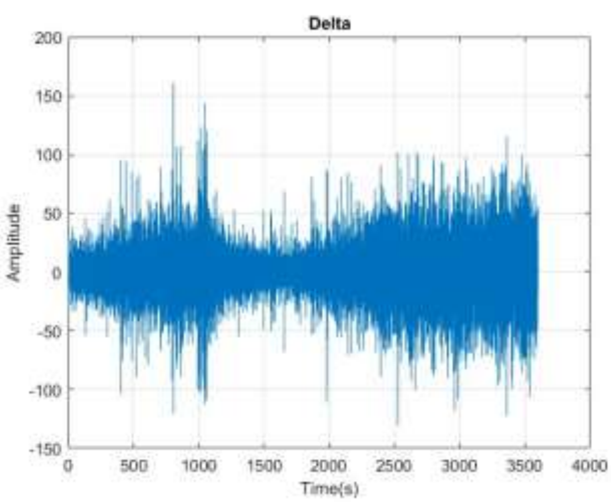

Figure 6: This graph shows the Delta band signal of the filtered EEG signal.

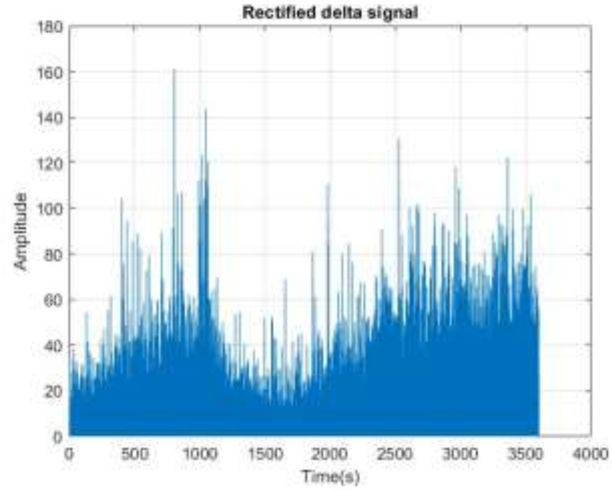

Figure 7: This graph shows the Rectified Delta band signal.

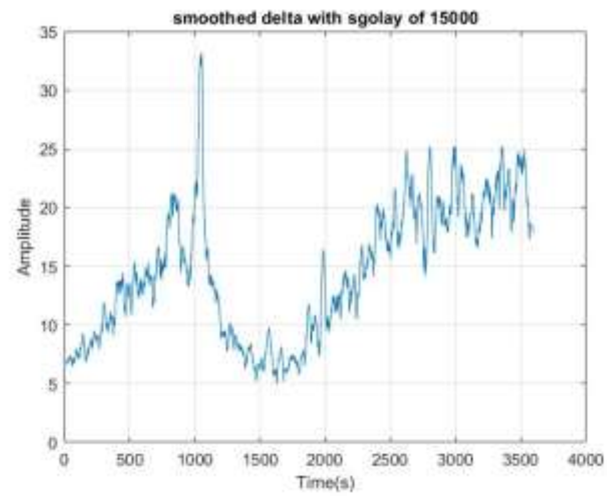

Figure 8: This graph shows the Smoothed Delta band signal which is smoothen with Sgolay of 15000.

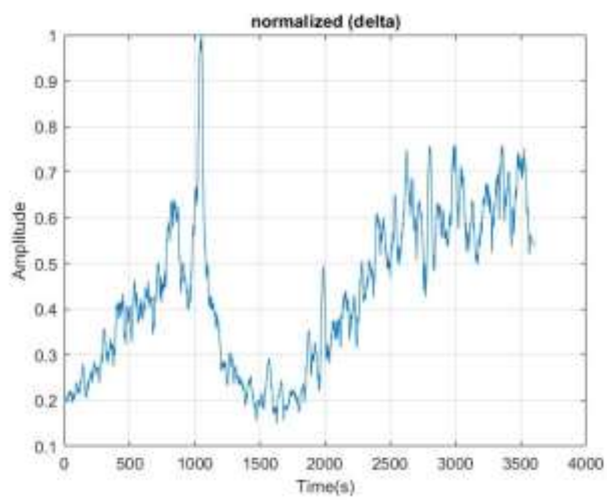

Figure 9: This graph shows the Normalized version of the smoothed Delta band signal.

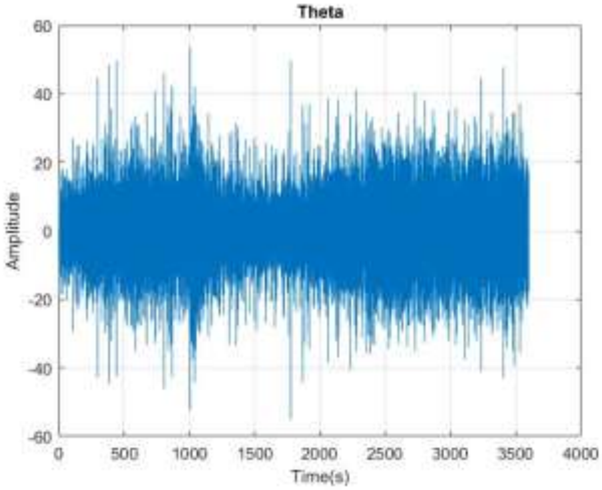

Figure 10: This graph shows the Theta band signal of the filtered EEG signal.

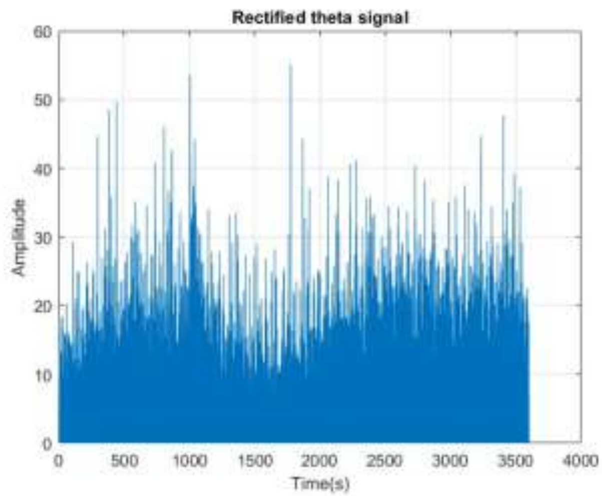

Figure 11: This graph shows the Rectified Theta band signal.

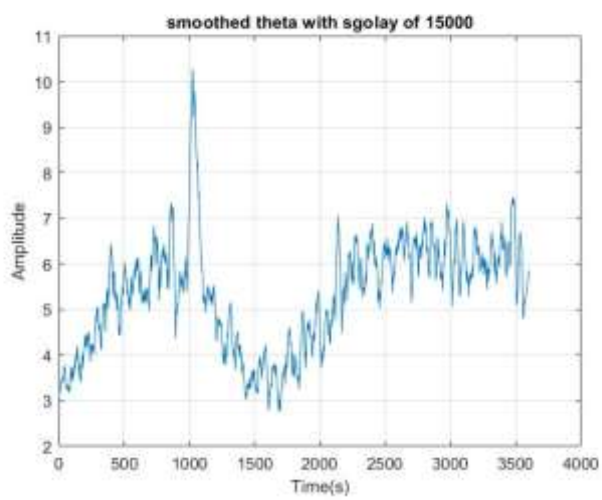

Figure 12: This graph shows the Smoothed Theta band signal which is smoothen with Sgolay of 15000.

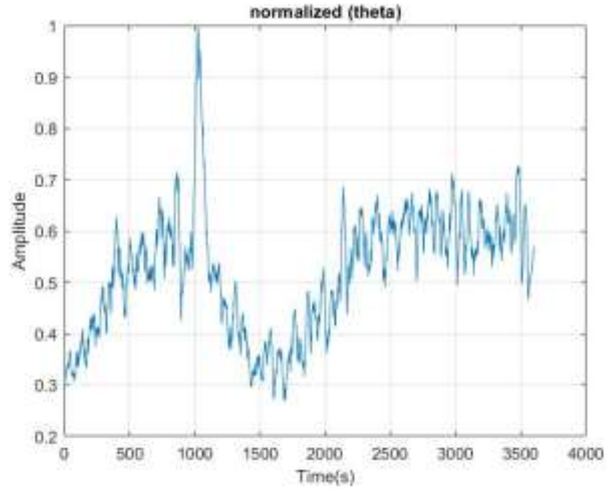

Figure 13: This graph shows the Normalized version of the smoothed Theta band signal.

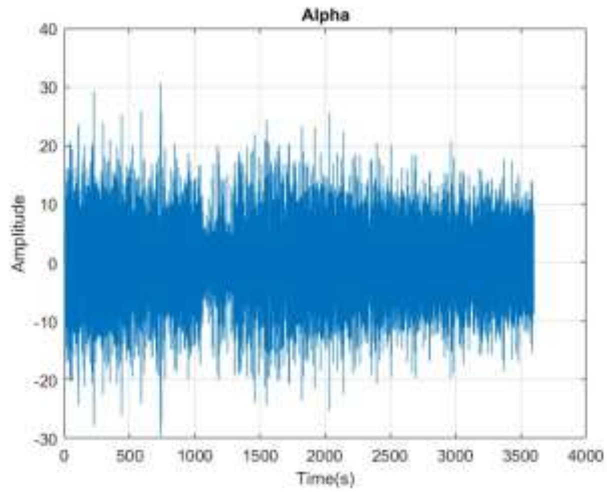

Figure 14: This graph shows the Alpha band signal of the filtered EEG signal.

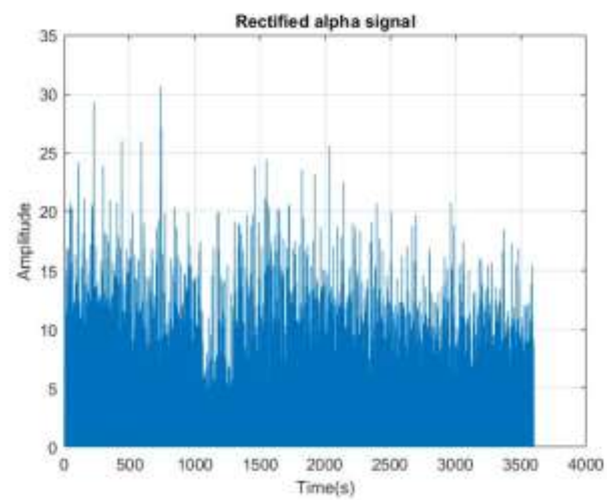

Figure 15: This graph shows the Rectified Alpha band signal.

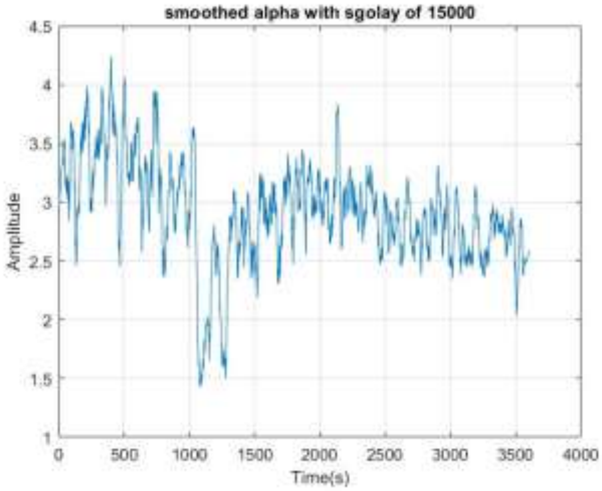

Figure 16: This graph shows the Smoothed Alpha band signal which is smoothen with Sgolay of 15000.

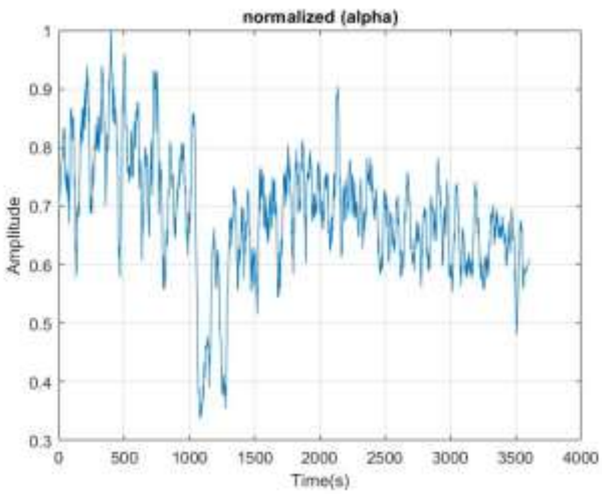

Figure 17: This graph shows the Normalized version of the smoothed Alpha band signal.

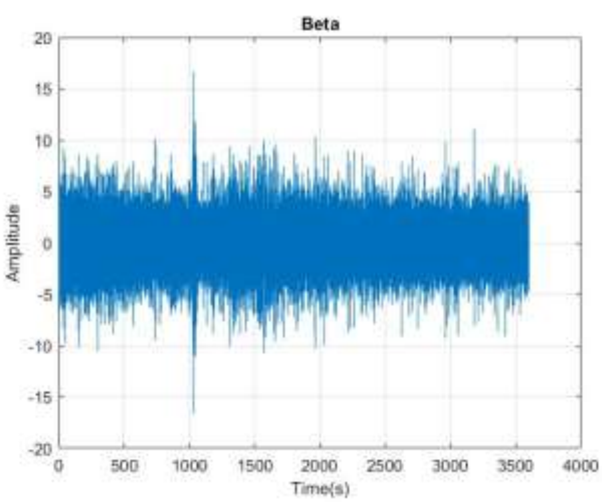

Figure 18: This graph shows the Beta band signal of the filtered EEG signal.

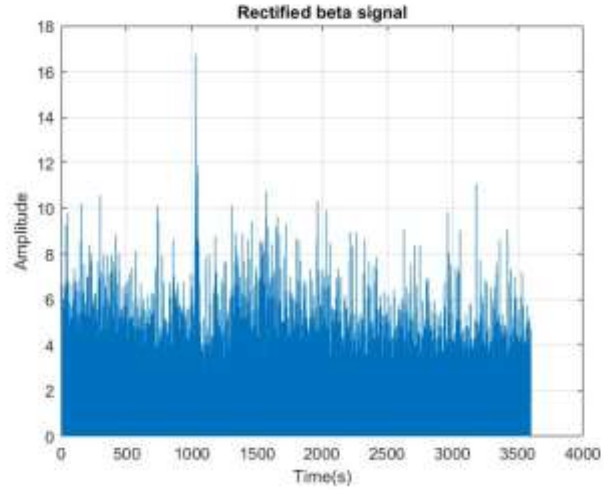

Figure 19: This graph shows the Rectified Beta band signal.

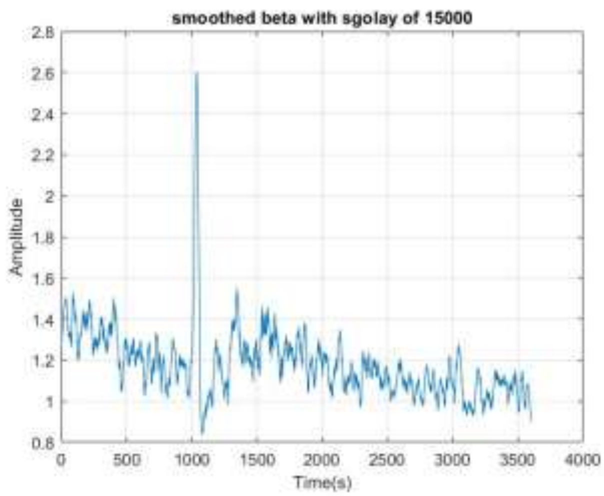

Figure 20: This graph shows the Smoothed Beta band signal which is smoothen with Sgolay of 15000

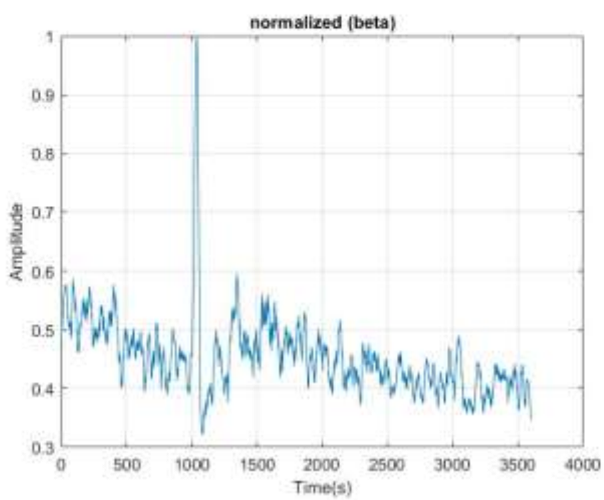

Figure 21: This graph shows the Normalized version of the smoothed Beta band signal.

Faranak Dayyani 500692664
Sarina Taki 500773183
Parnian Majd 500765210

Table 2: This table shows the average value of RMS, Standard Deviation and Entropy for each subject.

Subject	**1**	**2**	**3**	**4**	**5**	**6**
Entropy	2.7996	5.1134	4.2050	4.3166	3.9801	4.2125
RMS	0.3251	0.7552	0.2119	0.595	0.5428	0.6032
Standard Deviation	0.1735	0.1609	0.0219	0.2732	0.2945	0.2308
Band	Theta	Alpha	Beta	Beta	Beta	Beta
Ground truth	Y	Y	Y	Y	Y	Y
Subject	**7**	**8**	**9**	**10**	**11**	**12**
Entropy	4.1723	3.7539	4.1169	4.3912	4.4896	4.5724
RMS	0.5921	0.5625	0.567	0.6001	0.6399	0.7211
Standard Deviation	0.2887	0.3125	0.3071	0.2481	0.2405	0.1496
Band	Beta	Beta	Beta	Theta	Theta	Delta
Ground truth	Y	Y	Y	Y	Y	Y
Subject	**13**	**14**	**15**	**16**	**17**	**18**
Entropy	4.4593	4.0836	4.6109	5.3904	5.4943	5.2474
RMS	0.6151	0.6471	0.6399	0.8015	0.5164	0.5612
Standard Deviation	0.2180	0.2572	0.2888	0.0571	0.0687	0.0604
Band	Beta	Delta	Beta	Alpha	Alpha	Beta
Ground truth	Y	Y	Y	N	N	N
Subject	**19**	**20**	**21**	**22**	**23**	**24**
Entropy	5.3459	5.4691	5.1138	5.1969	5.4042	5.4404
RMS	0.6887	0.7475	0.6437	0.4896	0.7222	0.7304
Standard Deviation	0.06	0.0636	0.0494	0.0569	0.0596	0.064
Band	Beta	Alpha	Alpha	Beta	Alpha	Beta
Ground truth	N	N	N	N	N	N
Subject	**25**	**26**	**27**	**28**	**29**	**30**
Entropy	5.0471	5.1604	4.8185	5.1967	4.8625	5.0035
RMS	0.557	0.5251	0.6714	0.6969	0.5679	0.7179
Standard Deviation	0.0481	0.0702	0.0484	0.0493	0.0526	0.0443
Band	Alpha	Beta	Beta	Alpha	Beta	Alpha
Ground truth	N	N	N	N	N	N

*Note: Y means subject has seizure and N means subject does not have a seizure.

5-Fold cross-validation technique :

Trial 1:

Table 3: This table shows the Confusion Matrix for the first trial.

	Seizure	**Normal**
Seizure	TP = 1	FN = 0
Normal	FP = 0	TN = 4

Trial 2:

Table 4: This table shows the Confusion Matrix for the second trial.

	Seizure	**Normal**
Seizure	TP = 0	FN = 1
Normal	FP = 0	TN = 4

Trial 3:

Table 5: This table shows the Confusion Matrix for the third trial.

	Seizure	**Normal**
Seizure	TP = 0	FN = 3
Normal	FP = 0	TN = 3

Trial 4:

Table 6: This table shows the Confusion Matrix for the fourth trial.

	Seizure	**Normal**
Seizure	TP = 0	FN = 2
Normal	FP = 0	TN = 3

Trial 5:

Table 7: This table shows the Confusion Matrix for the fifth trial.

	Seizure	**Normal**
Seizure	TP = 4	FN = 1
Normal	FP = 1	TN = 1

Table 8: This table shows the accuracy, sensitivity and specificity for all trials.

Trial	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Trial 1	83.34	100	100
Trial 2	66.67	0	100
Trial 3	50	0	100
Trial 4	50	0	100
Trial 5	83.34	80	100
Mean	66.67	36	100

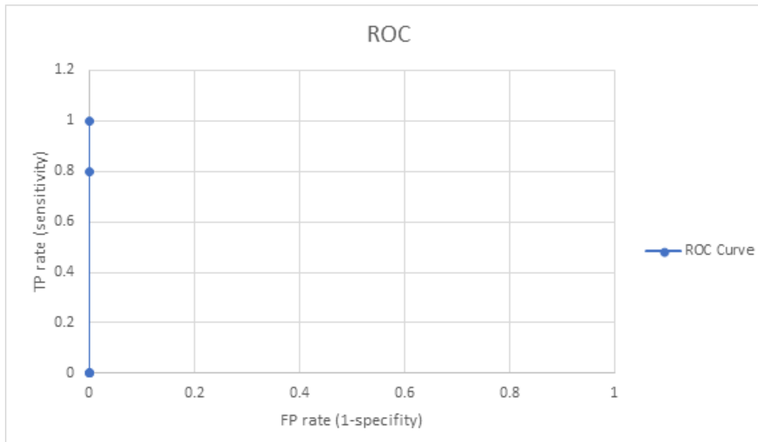

Figure 22: This graph shows the Receiver Operating Characteristic (ROC) curve.

According to figure 22 above, Area under the ROC curve (AZ) = 1

IV. Discussion and Conclusion

Epilepsy is amongst one of the most common neurological disorders affecting 39 million people worldwide [12]. Epileptic patients experience seizure which is technically the attacking of neurons and then firing at an abnormal rate and can result in vigorous shaking [13]. Detection of an epileptic patient with machine learning algorithms can be considered to be of paramount importance as it results in a more accurate detection.

Three main steps of preprocessing, feature extraction, and machine learning were executed through our project to result in an algorithmic model for seizure detection. Based on the discussed methods in this report, multiple filters such as moving averaging, bandpass filters, and smoothening were used for the preprocessing step and then the resulted filtered EEG signal was used in the second stage which was feature extraction of STD, RMS, and Entropy. These features were then applied in the machine learning step for detection purposes. The algorithm was used for the identification of seizure in 30 EEG samples 15 of which had a seizure and 15 were driven from healthy individuals. However, 80% of the population was treated for training the machine, and 20% for classification. We were able to achieve an average accuracy of 66.67% with an average specificity of 100% and an average sensitivity of 36%. This perfect achieved number for specificity means that our device can always correctly identify the healthy EEG signals. This is due to the fact that the defined ranges of healthy and unhealthy RMS overlap at one value, resulting in the machine looking into the second extracted feature which is STD. In this case, the values for healthy STD and unhealthy STD are noticeably different and it is impossible for the machine to incorrectly determine the healthy EEGs as unhealthy.

Moreover, one of the limitations which we faced was to obtain the seizure EEG signals from the provided database as most of them either included "NaN" regions where there was no data or were too long or too short as in both cases it would result in an error and termination of the program. The mentioned limitation then resulted in the second challenge we encountered which was the scarcity of the seizured signal. This resulted in us having a seizure threshold based on the RMS, STD and Entropy value of a small dataset and would consequently reduce the robustness of our algorithm. The third constraint for us lied in the pre-processing step as we devoted a huge amount of time to configure which types of filters work best for our project. Consequently, increasing the population of the dataset and training the model with a greater number of EEGs as well as enhancing the algorithm for the early-stage detection of epilepsy would be considered to be a future work of this project as currently our algorithm is only used for the late-stage detection of epilepsy.

REFERENCES

[1] *EEG*. [Online]. Available: https://www.medicine.mcgill.ca/physio/vlab/biomed_signals/eeg_n.htm.

[2]: "Biomedical Sensing and Acquisition", Slides courtesy of Profs. Khademi and Rangayyan, Ryerson University, 2019.

[3] "What is Epilepsy?," *Epilepsy Foundation*, 21-Jan-2014. [Online]. Available: https://www.epilepsy.com/learn/about-epilepsy-basics/what-epilepsy.

[4] G. Repovš, "Dealing with Noise in EEG Recording and Data Analysis", *Informatica Medica Slovenica*, 26-Aug-2010. pp.18-25.

[5] G. Altunian, "What Is Signal-to-Noise Ratio and Why Is It Important?," *Lifewire*, 11-Nov-2019. [Online]. Available: https://www.lifewire.com/signal-to-noise-ratio-3134701.

[6] N. P. Subramaniyam, "Measuring Entropy in the EEG: Sapien Labs: Neuroscience: Human Brain Diversity Project," *Sapien Labs | Neuroscience | Human Brain Diversity Project*, 13-Feb-2018. [Online]. Available: https://sapienlabs.org/measuring-entropy-in-the-eeg/.

[7] "A," *Smooth noisy data - MATLAB*. [Online]. Available: https://www.mathworks.com/help/matlab/ref/smoothdata.html.

[8] H. Azami, K. Mohammadi, and B. Bozorgtabar, "An Improved Signal Segmentation Using Moving Average and Savitzky-Golay Filter ," *An Improved Signal Segmentation Using Moving Average and Savitzky-Golay Filter*, pp. 39–44, Dec. 2011.

[9] Stephanie, "Normalized Data / Normalization," *Statistics How To*, 07-Jan-2018. [Online].Available: https://www.statisticshowto.datasciencecentral.com/normalized/.

[10] Mathuranathan and Dilgush, "Mathuranathan," *GaussianWaves*, 17-Jun-2019. [Online]. Available: https://www.gaussianwaves.com/2015/07/significance-of-rms-root-mean-square-value/.

[11] M. V. Kamath, G. Tougas, S. Hollerbach, R. Premji, D. Fitzpatrick, G. Shine and A. R. M. Upton, "Estimation of habituation and signal-to-noise ratio of cortical evoked potentials to esophageal electrical and mechanical stimulation", Medical and Biological Engineering and Com-puting, 35:343-347, 1997.

[12]"Epilepsy Fact sheet". WHO. February 2016. Archived from the original on 11 March 2016. Retrieved 4 March 2016.

[13] A. H. Mokdad Prof *et al*, "Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *Lancet, the,* vol. 387, *(10036),* pp. 2383-2401, 2016.