Machine Learning Based Epileptic Seizure Detection for Responsive Neurostimulator System Optimization

Yangchicheng Shen

1*Westtown school, West Chester, PA 19382, USA
*Corresponding author's e-mail: charlie.shen@westtown.edu

Abstract. This paper proposes a novel method of identifying the time of epileptic seizure happening on patients by employing feature extraction and machine learning-based classification on Electroencephalogram (EEG) signal collected from a closed-loop interface implanted in the brain of patients. The closed-loop device was served as a neurostimulator which introduced stimuli to epilepsy patients when detecting the occurrence of seizure. A set of multiple time- and frequency-domain features are extracted from intracranial electroencephalography recordings of 7 subjects with epilepsy. Trained and tested on the extracted features, an ensemble of machine learning models with parameter tuning achieves an area under the curve (AUC) score of 0.99.

1. Introduction

1.1. Epilepsy
Epilepsy is a central nervous system (neurological) disorder in which brain activity becomes abnormal, causing seizures or periods of unusual behavior, sensations, and sometimes loss of awareness [1]. There are around 65 million people around the world who have epilepsy. It has been considered the fourth most common neurological disorder, with one of the most occurring disease on children. [2]

Traditionally, there have been two main ways to deal with Epilepsy – medicine and surgery. Medicine has been considered the safer method compared to surgery, with patients mainly intake one specific anti-seizure medicine while others take the combination of a few.[1]. However, there are side-effects accompanying by medication therapy such as fatigue, dizziness, weight gain, and even loss of memory, speech, and thinking. What’s worse, about one-third of patients with epilepsy are seizure-medication resistant. This means that they have to seek other methods to treat epilepsy or relieve seizure symptom. Another option is surgery, in which patients undergo a more dangerous process to cure from Epilepsy. Another problem with surgery was that the success rate of Epilepsy surgery is under 35 percent. Many patients still don’t want to undergo the surgery due to the fear and the risk of removing part of their brain and potential side effects.

1.2. Neuro Stimulator
Many people with epilepsy turn to another option, using a neuromodulation device which sends electric current to the nervous system. There are two major approaches to do neuromodulation for epilepsy, vagus nerve stimulation (VNS) and responsive neurostimulation (RNS)(Figure. 2)[3][4]. VNS is a device indirectly attached to patients and send either continuous or scheduled electric pulse to patients. The pulses range from 250 µ A to 2.25mA, with frequency falls between 20 - 30 Hz [5]. RNS was a more responsive device. It has less than 1 mA current, with around 200 Hz. Implanting the device into the brain of patients, it responds by sending electrics towards patients when it detects the seizure events.
happening. Before implantation of RNS, seizure needs to be localized on certain part of the brain. [6] As shown in Figure. 1, seizure represented by the bright spot can be really small and local.

Using electric therapy did provide seizure patients, especially with those not being able to cure from medication therapy and surgery a chance to control seizure events. However, there are pain and risks patients needed to take while using electric therapy. RNS requires patients to suffer from over 250 minutes of electrical pulse each day, suffering from hoarseness, voice alteration, sore throat, and shortness of breath, while patients VNS needs to undergo surgery each 3 to 5 years, with each surgery brings pain to patients and even with the risk of potential intracranial haemorrhage.

1.3. Electroencephalography
RNS requires an analysis over Electroencephalogram(EEG). Electroencephalogram is signals indicating the activity of the brain. It was originally developed and recorded by a German psychologist and psychiatrist Hans Berger. EEG is acquired most usually by using brain machine interface. There are two major types of EEG devices. One is wearable EEG device which places electrodes on human scalps. The other is implanted EEG device which has electrodes implanted inside skull which is also known as intracranial EEG(iEEG). EEG signal is then sampled through electrodes and then amplified and digitized through electric circuits. After sampling and digitizing, multi-channel digital EEG signals will be sent to computer where most post-processing and analysis are done(Figure. 3). However, in neurostimulation device such as RNS, post-processing is executed on the implanted electric circuit. Considering the size of the implanted device and its low heat generation, low power consumption and realtime requirements, EEG signals processing on RNS will be more challenging and restricted. EEG signal is the measurement of the potential difference between detecting electrode and reference electrode. Therefore, EEG signal spatial resolution is determined by electrode number and density.

EEG signal mostly falls into the range of 0 to 45 hertz. It can be separated into five main categories by its frequency: Delta(0-3 Hz), Theta(3-8 Hz), Alpha(8-12 Hz), Beta(12-27 Hz), and Gamma band(27-45 Hz), with increasing frequencies. A typical adult human EEG signal’s amplitude is about 10 µV to 100 µV, when measured from the scalp and is about 10 to 20 mV when measured from subdural electrodes. Noises are also detected along with signals, so signal filtering is needed to be done prior to brain activity analysis. By extracting certain features and classify EEG signal features, we are able to monitor brain activities of humans and even detect certain diseases such as memory loss, Alzheimer and Autism, and epilepsy. [7] Thus, EEG device is not only a good tool widely used for seizure cortex region detection before surgery, but also provides a possibility to accurately detect and predict seizure onsite time and region which can be used in neurostimulation device.
1.4. Related Work

As artificial intelligence gradually become popular in the past few years, many researchers have been trying new machine learning approaches to perform seizure detection and prediction. There have been related works trying to solve the issues in order to alleviate pain for patients. Scientists in 2006 from Texas have tried to use neural net to detect neonatal myoclonic and focal clonic seizures through motion tracking methods, reaching an over 90% sensitivity.[9]. Other mechanisms included logistic regression. In 2005, three scientists from Turkey has managed to use the combination of logistic regression and multi-layer perception neural network to reach a sensitivity of 91% [10]. Also, Artificial Neural Net, or ANN, has been one option for analyzing signals. In 1996, four scientists from John Hopkins University used an ANN with the input of 31 nodes and the output value averages the 8 nodes of output, reaching the sensitivity of 76%. [11] Of all mechanisms, Support Vector Machine, or SVM, has been found the most popular. Five scientists from the University College Cork in Ireland used Support Vector Machine to extract certain features and classify in order to increase the precision of epochs of EEG to 89 percent with one false seizure detection per hour. [12] Furthermore, four scientists from the University of Pennsylvania has used SVM to classify short-time, energy based data to decide whether seizures happened in patients, reaching 97.1% sensitivity. [13][14] In our paper, we used three machine learning methods to detect the length of each seizure – One-class SVM and Tree Ensemble. SVM and Tree Ensemble are two tools used to categorize signals into two categories. SVM is a classifier by maximizing the margin of two categories, sometimes using kernel to perform classification of higher dimension; tree ensemble classifies by breaking a massive decision tree into more smaller trees to reach the purpose.

1.5. Work in this paper

In the past ten years, with the development of machine learning technique and increasing of arithmetic capability, many machine learning based seizure prediction models are proposed. However, most clinical EEG recordings are confidential and therefore there are very few public datasets. Some powerful artificial intelligence models like CNN and RNN didn’t perform very well due to the small training set size. In the two recent seizure prediction competitions [15][16], top ranked algorithms are mostly feature based ensemble model like SVM, random forest and gradient boosted decision tree. Focus of seizure prediction research turn to searching for effective features and prediction model optimization.

In this paper, we proposed a SVM, extreme boosted (XGB) decision tree and KNN based ensemble model with parameter tuning and computation cost reduction. Meanwhile, we also applied phase amplitude coupling as an important feature which was demonstrated as a powerful EEG signal feature for seizure detection and brain activity analysis [17][18][19].

We trained and tested our prediction model based on multiple epilepsy patients who were implanted with different iEEG devices. These iEEG device have different channel numbers and sampling rate. By evaluating our model with these high variant dataset, we could test model performance with different dimension of feature inputs. By reducing high computational cost features and tuning model
hyperparameter, we could finally achieve similar prediction accuracy which provides a possibility of reducing the complexity of prediction model and channel numbers of EEG signal sensing and neuro-stimulation system.

2. Methods

2.1. Dataset
We got our intracranial EEG data from UPenn and Mayo Clinic’s seizure detection challenge. There is also some other popular dataset like Melbourne-University AES-MathWorks-NIH Seizure Prediction Challenge dataset which is now available on Epilepsy Ecosystem [20]. The reason why we use UPenn and Mayo Clinic’s dataset is they have more patient variation and channel number variation. This dataset is from 7 human patients with epilepsy undergoing intracranial EEG monitoring to identify cortex region which can be further resected to prevent future seizures. Intracranial EEG recordings are from the implanted electrodes along anterior-posterior axis of hippocampus, and from subdural electrode grids in different locations [15]. Table I details the recorded iEEG data.

| Subject | Number of Channels | Sample Rate | Record Time   |
|---------|--------------------|-------------|---------------|
| 1       | 96                 | 5000 Hz     | 13m 38s       |
| 2       | 56                 | 500 Hz      | 18m 11s       |
| 3       | 16                 | 500 Hz      | 19m 56s       |
| 4       | 88                 | 500 Hz      | 27m 04s       |
| 5       | 104                | 500 Hz      | 13m 30s       |
| 6       | 88                 | 500 Hz      | 40m 13s       |
| 7       | 96                 | 500 Hz      | 45m 07s       |

Sample rate of these recordings is either 500Hz or 5000Hz and the number of EEG recording channels varies from 16 to 104 among 7 patients. A total number of 3-hour recording is studied in this work. Figure 4 shows a sample of iEEG recording which has two different colour indicating nonictal (blue) and ictal(red) status. Recording data are segmented into 1 second segment with label of ictal or nonictal. A training and testing split with a ratio of 0.8 to 0.2 is performed before further processing and model training.

2.2. Machine Learning Models

2.2.1. SVM: Support Vector Machine, or SVM, is a classifier used to classify the features that are listed in Feature Extraction part of Experiment by maximizing the margin, the distance between each data and the line separating the data into categories (Fig.5). It is based on two assumptions: first, converting
lower-dimension to higher dimension classification problems may simplify the problem; second, the
information closest to the decision surface are those the most important [12].

Assuming that the problem is linearly separable, so that the hyperplane separating the two categories
is defined as:

\[ f(x) = w \cdot x + b \]  

in which w is the normal vector determining the direction of the margin and b is the bias that shifts the
margin in the hyper-plane. Defining that the two categories fell into the hyper-plane of \( f(x) = 1 \) and
\( f(x) = -1 \), the distance between the two can be written as:

\[ d = \frac{2}{||w||} \]  

and d is also known as the margin. Further, the most determining vectors, those closest to the decision
layer, are called the support vector (Fig.5).

In order to categorize the data into two separate categories, labels must be set for each category.
Assuming that we have a set of data, labelling with \( y_i = 1 \) with one category and \( y_i = -1 \) for the other,
we have the following equations:

\[ w \cdot x + b \geq 1 \quad \text{for } y_i = 1 \]  
\[ w \cdot x + b \leq -1 \quad \text{for } y_i = -1 \]  

Thus, the problem can be stated as the following:

\[ \text{maximize } d = \frac{2}{||w||} \]  

for \( y_i w \cdot x + b \geq 1 \) \hspace{1cm} (5)

Or

\[ \text{minimize } d = \frac{||w||}{2} \]  

for \( y_i w \cdot x + b \geq 1 \) \hspace{1cm} (6)

In the prefect scenario, data can be separated linearly into 2 categories without any misclassification,
but more often linear classifier cannot separate the data perfectly because of the traits of data or the
presence of noise. SVM manages this by introducing a non-negative slack variable to represent the
misclassifications. Thus, the previous functions become:

\[ w \cdot x + b \geq 1 - \text{slack variable, for } y_i = 1 \]  
\[ w \cdot x + b \leq -1 + \text{slack variable, for } y_i = -1 \]  

Since we are trying to maximize the margin while minimizing the number of misclassifications, we
ended up with the function of:

\[ \text{minimize } d = \frac{||w||}{2} + C \sum \text{slack variable} \]  

for \( y_i w \cdot x + b \geq 1 \) \hspace{1cm} (11)

In which \( C \geq 0 \) is the trade-off between maximizing the width and minimizing the misclassifications.
It has an inverse relationship with the slack variable. When C approaches infinity, there would be no
misclassification allowed, becoming the hard margin SVM case.

In times when both linear and non-linear function cannot perform an outstanding classification, SVM
may import a kernel function that converts the data to a higher dimension to perform a linear
classification. It is constructed with multiple \( \phi \) functions, written in the form of:

\[ k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \]  

Converting from a lower dimension to a higher dimension may simplify the problem as assumed
above.
2.2.2. Gradient Boost Tree: Gradient Boost Tree is based on the model of decision tree, a Machine Learning method that generates a large amount of smaller classifier in order to build a massive model. One decision tree is consisted of nodes, with the parent nodes splitting into more sub-nodes to detail the differentiation. It will finally reach the final node, known as the leaf, for the final decision. (Fig.6)

When training a tree, it will choose from the paths having the least cost based on the cost function we choose. It is using the greedy algorithm, always trying to go along the path with the least cost to minimize the final cost. By recursively doing this, we are finally reaching the leaf with the least cost, as known as the best classifier.

There are two types of decision tree generally: regression tree and classification tree. Regression tree is generally used to generate a numeric data as the answer. It is grown by fitting a regression model to the data, usually using the function:

$$\text{cost } d = \sum (y_1 - y_t)^2$$  \hspace{1cm} (14)

in which $y_t$ is the threshold for calculation and $y_1$ is the data we put in the tree. We continue doing this recursively until each single data is separated.

Classification tree, unlike regression tree, will lead to a statement rather than a numerical data. One way to grow a classification tree is to calculate the information gained in each node. Information gained is calculated by subtracting the original entropy with the conditional entropy. Original entropy is calculated by:

$$H(D) = -\sum \pi_i \log \pi_i$$  \hspace{1cm} (15)

this stands for the randomness of data originally. Conditional entropy is the randomness of the data after the threshold is chosen. Its mathematical expression is represented as:

$$H(D|A) = -\sum \frac{|D_i|}{D} H(D_i)$$  \hspace{1cm} (16)

By subtracting the two, we get:

$$H(D) - H(D|A)$$  \hspace{1cm} (17)

In order to grow a more detailed tree, we would choose the one with the least information gain. This process would happen recursively, until the tree is grown.

However, if every data is separated, this may cause the problem of over-fitting. Thus, we have some rules set for preventing over-fitting. Usually, we would limit the cost reduction to a specific threshold to prevent the cost being too low. We would also limit the minimum size of the leaf node.

It is done by minimizing the loss function. Each tree has the maximum depth it can reach, and each forest has the amount of decision trees. In this study, we set the total number of decision trees 1000, with each having the maximum depth of 4. The final decision tree is made by finding the sum of all smaller decision trees.
2.2.3. **KNN**: K-Nearest neighbor, or KNN, is a rather simpler method. Its fundamental idea is to find the K nearest labelled data to the data waiting to be labelled. In a 2-dimension coordinate, the distance is known as:

\[ d = ||x - xt||^2 \]  \hspace{1cm} (18)

The distance can also be chosen using the kernel function, converting into a higher dimension to calculate the distance. The distance for kernel function is written as:

\[ dk(x, xt) = (k(x, x) - k(xt, xt))^2 \]  \hspace{1cm} (19)

With different numbers of K, there will be different results to the model. KNN usually performs better than SVM when the problem has more than 2 categories. (Fig. 7)

2.2.4. **Model Ensembling**: When using several methods to classify a problem, the result may be more accurate when errors cancel out. Ensembled method is the combination of various base learner in some structure to reach a more optimized result. The way of ensembling mainly consists of averaging the value of all the base learners, bootstrap resampling, bagging, boosting, Adaboost, stacking, and cascading. Averaging is the most often used strategy, and it is done by averaging the classifier’s output. The output will be number will be either 0 or 1, and the final result will be treated as 1 if the averaged outcome will be more than 0.5, vice versa. For an even number of classifiers, there is a possibility to reach the point 0.5, and in this case, people have to categorize 0.5 into 1 category.

2.2.5. **Cross-validation**: Cross-validation is used here in order to maximize the use of data. Each time during the process of training and testing, one patient would be used for the training set, the other two would then be used as the validation set. This would expand the amount of data used for training the model given a limited data on Epilepsy. Performing cross-validating can also alleviate the problem of over-fitting. (Fig. 8) In this case, we performed K-Fold cross validation with K set to 3. This means that cross-validation has been performing for three times.
2.2.6. Our Approach: We first do pre-processing for each iEEG data recording by down-sampling the data and filtering noises. Next, we extracted the features listed below in II, which can be separated by three categories: frequency domain, time domain, and statistical and information theory features. All these features are used in processing the actual data.

Next, we use a combination of the three methods mentioned above: SVM, KNN, and Gradient Boost Tree. We run each model with the ictal and non-ictal data we have, creating three models that can be used in further testing process. During the process of training the three models, we use cross-validations to ensure enough data set is used.

We selected an ensemble of eXtreme Gradient Boosted (XGB) decision trees, 20 Nearest Neighbours (20NN) and Support Vector Machine (SVM) with a Radial Basis Function kernel for our seizure state classification tasks. An ensemble of 1,000 trees with a maximum depth of 4 was used for the XGB model, and the hyper-parameters of the SVM model were tuned and optimized for each patient. Given the small number of seizures for some patients in our dataset, we chose to divide the iEEG data into 3 blocks and use a 3-fold cross validation method to train and test the classifier. The performance of the classifier was measured by area under the curve (AUC), sensitivity, and specificity.

Sensitivity, or the true positive rate, measures how well a classifier is how well the classifier identifying those with seizures. It is defined as:

\[ Sensitivity = \frac{TP(t)}{TP(t) + FN(t)} \]  

in which TP is the number of correctly identified seizures, and FN is the number of incorrectly identified data with seizure labels.

Specificity measures how well the classifier correctly identified the non-seizure data. It is defined as:

\[ Specificity = \frac{TN(t)}{TN(t) + FP(t)} \]

In which TN is the number of correctly identified non-seizure data, and FP is the incorrectly identified data with non-seizure labels.

AUC indicates the ability of the classifier to distinguish between two labels. It is calculated by measuring the area under the receiver operating characteristics, the curve that shows a relationship between Sensitivity and Specificity.

3. Experiment

3.1. Feature Extraction

Features most common and related to analysing EEG is extracted to train the classifiers. Using Fourier transform, we are able to get the frequency domain features from the power spectrum density, known as the PSD. As for the features from the time domain, we are able to directly get the feature from the EEG signal we got as well as its first and second derivative. Also few statistical Feature is picked in order to further analyse the data. Furthermore, to subtract cross channel information, cross-channel feature is also used for model training.
Table 2. Set of Investigated Features

| Feature Type     | Feature Name                      |
|------------------|-----------------------------------|
| Time Domain      | Line length                       |
|                  | Mean energy                        |
|                  | RMS amplitude                      |
|                  | Hjorth Activity                    |
|                  | Hjorth Mobility                    |
|                  | Hjorth Complexity                  |
|                  | Mean Teager energy                 |
|                  | Mean curve length                  |
| Phase            | -amplitude coupling                |
| Frequency Domain | Delta band energy                  |
|                  | Theta band energy                  |
|                  | Alpha band energy                  |
|                  | Beta band energy                   |
|                  | Low-Gamma band energy              |
|                  | Gamma band energy                  |
|                  | High-gamma band energy             |
| Statistic        | Kurtosis                           |
|                  | Skewness                           |
| Cross Channel    | Eigenvalues of correlation coefficient matrix of each channel’s time domain signal |

Most features listed above are commonly used in seizure detection and signal analysis. A typical time domain feature series is Hjorth parameters: Hjorth Activity, Hjorth Mobility and Hjorth Complexity. Hjorth Activity describes the variance of the time function. Hjorth Mobility represents the mean frequency of the power spectrum which is represented by the variance of the first derivative of time function over the Hjorth Activity. Hjorth Complexity shows the change in frequency over time which is represented by the Hjorth Mobility of the first derivative of the time function over the Hjorth Mobility of the time function. Mean curve length is the mean value of the line length of the single. Mean teager energy is the value of energy at certain time subtracts the product of the energy of the last time unit and the energy of the next time unit. It represents the tendency of change of energy. Skewness measures the symmetry of the signal. It is measured by mean - median over standard deviation. Kurtose measures whether the signal is rather peak or rather flat.

Previous research shows that PAC can accurately distinguished the ictal state from the interictal (nonictal) state [19]. A synchronisation index (SI) is used to measure the coupling strength between the amplitude and phase based on the instantaneous phase and amplitude calculated from the Hilbert transformation. Synchronisation index was defined as:

\[
SI_{\gamma_{\omega}} = \left| \frac{1}{N} \sum_{n=1}^{N} \exp(i(\phi_{\omega}(n) - \phi_{\gamma_{\omega}}(n))) \right|
\]

(22)

where N is the number of time points during each analysis section. \(\phi_{\gamma_{\omega}}\) which is the phase of the low-frequency band (\(\omega\)), was identified calculated by finding the remainder between \(\phi_{\gamma_{\omega}}\) and the phase of \(\omega\) [19]. In our feature extraction model, we choose theta band (4-8Hz) and high-gamma band (80-150Hz) to calculate phase-amplitude coupling strength.

3.2. Model

In our data processing, ictal data is labelled as 1, and non-ictal data is labelled as 0. Both ictal data and non-ictal data are then scaled and transformed in preparation for the classification process. For both training and testing, 20% of the data is used for each iteration, with 5 iterations in total to train the classifier.
Figure 9. Float diagram of model training from the amplitude of the high frequency band (γ). The SI was For SVM classifier, the data was input separately to train individual SVM classifier for each channel, and by applying the training classifier to validation set we get classifier for each iteration. The final training classifier for SVM is the result of cross-validating the data for three times.

For K-nearest-neighbors, we set the K value as 20, so the data awaited to be categorized will be labelled based on the label of the 20 nearest neighbors.

For the gradient boosted tree, we have in total 1000 trees to function as classifiers, with each tree having the maximum depth of 4. We as well implement cross-validation on the gradient boosted tree classifier.

In total, three classifiers would be averaged out to get the final outcome of seizure detection.

3.3. Results

Features are firstly subtracted and saved to csv files for each 1s segment data using MATLAB 2019a. Classifiers are trained using python 3.7. The whole process is run on a Intel(R) Core (TM) i78700 CPU and 16GB RAM desktop.

Figure 10 shows the training time for each patient. Patient 3 dataset has only 16 channels which costs the least time. However, patient 5 dataset which has 106 channels took less time for training than patient 6 and patient 7 who have less channel number than patient 5.

Figure 11 shows the AUC, sensitivity and specificity across different patients. Our ensembled model achieve very high AUC on all the patient datasets. Our model performs less sensitivity 0.74 on patient 5 than other patients. However, they are all shown to have high specificity which implies high risk of false detection /alarm. On the other hand, they all showed nearly 1 sensitivity which means most seizure behaviour could be detected.

When comparing the three models we used for ensembling our final classifier, we can find that KNN performs the least specificity and sensitivity. And XGB tree model shows the highest specificity and sensitivity score.

Figure 10. Classifier Training Time for each patient
From Table III we can find that prediction accuracy for each patient is not very high. This is due to the high True Negative (TN) rate. Many nonictal segments are predicted as ictal period. However, our main target is to effectively detect seizure and trigger neurostimulator in time. Since our prediction model has very high AUC and sensitivity, differentiating seizure period should be treated as secondary goal compared with seizure prediction accuracy. Thus, even though our model has this drawback, it is still a good model which could be used for clinical neurostimulator.

### Table 3. Classification Results

| Patient | Number of Channels | Prediction Accuracy |
|---------|--------------------|---------------------|
| 1       | 96                 | 0.62                |
| 2       | 56                 | 0.7                 |
| 3       | 16                 | 0.62                |
| 4       | 88                 | 0.59                |
| 5       | 104                | 0.72                |
| 6       | 88                 | 0.77                |
| 7       | 96                 | 0.79                |

### 4. Discussion

This paper proposes an ensemble model to combine three classifiers, SVM, XGB, and KNN together to formulate the detection of seizures. Through extracting features from time domain, frequency domain, it is able to comprehensively process an EEG in order to get a more accurate result. Furthermore, we perform cross validation, enabling more data to be processed based on a rather limited amount of data we possess and reduce the risk of overfitting. The result shows equally good performance of our training model on different channel number EEG devices and patients. Classifiers wise and patient’s wise comparison are shown in above figure 11 and figure 12.
1) Figure 12. Performance scores of each individual classifier across the patients.

2) Our model training and test is performed only on the training datasets of this Kaggle competition. This is because there is no available label for competition’s test dataset. This limits the possible novelty and possible clinical contribution. Time variation of EEG signal is usually very high even for the same person and sensing device. More training and validation dataset would be of great help to increase model adaptivity for long term seizure detection.

3) From above result table and diagrams, we can find that adding number of EEG channels doesn’t help with increasing sensitivity or reducing specificity. This shows a potential to reduce EEG channel numbers, we could directly work on neurostimulator device power consumption reduction. However, our feature extraction algorithm runs on each channel and large channel numbers will lead to a large number of features. Feature restoring cost and prediction latency will both increase dramatically. Thus, further investigation should be done to find the optimal channel number and corresponding features.

4) Our model suffers the high specificity and nonictal period misclassification issue. Even though different state classification accuracy is secondary to seizure prediction accuracy, misclassification may cause unnecessary suffer to epilepsy patients. Inter-ictal period is usually much longer than ictal period. Training on non-ictal dataset and perform outlier detection like one-class SVM may help us correct false detection of seizure occurrence.

5) Our feature extraction and classification model have great potential to be applied to responsive neurostimulator to suppress seizure. Even though our test validation set size is limited and model performance may decrease when using Kaggle competition test set for validation, our AUC score is comparable with the top algorithms on the leader board. To further address our model on clinical medical devices, many benchmark tests need to be done for their low power and low latency requirements.
5. Conclusion
In this paper, we investigated a set of bio-markers indicative of epileptic seizure events in iEEG recording data and performed feature extraction and classification using machine learning algorithms. An ensemble of Gradient Boost Tree, SVM and K-nearest neighbors performed very well and achieving an average AUC score around 0.99 among 7 patients. The approach to feature extraction and machine learning-based classification presented in this work is applicable in closed-loop responsive neurostimulation systems for various neurological disorders prediction and treatment.

Acknowledgments
The author would like to offer his gratitude for the magnificent assistance of the research mentor, Jianlin Zhou, on this research. I also would like to thank Kaggle UPenn and Mayo Clinic’s seizure detection challenge Data to perform my experiment.

References
[1] Mayo clinic epilepsy. [Online]. https://www.mayoclinic.org/diseases-conditions/epilepsy/symptoms-causes/syc-20350093;
[2] Scharfman, H. E. (2007), “The neurobiology of epilepsy,” Current neurology and neuroscience reports, vol. 7, no. 4, pp. 348–354.
[3] Epilepsy foundation. [Online]. https://www.epilepsy.com/ learn/treating-seizures-and-epilepsy/devices
[4] Scharfman, H. E. (2007), “The neurobiology of epilepsy,” Current Neurology and Neuroscience Reports, vol. 7, pp. 348–354.
[5] Scherrmann, J., Hoppe, C., Kral, T., Schramm, J., Elger, C. E. (2001), “Vagus nerve stimulation: clinical experience in a large patient series,” Journal of clinical neurophysiology, vol. 18, no. 5, pp. 408–414.
[6] Thomas, G. P., Jobst, B. C. (2015), “Critical review of the responsive neurostimulator system for epilepsy,” Medical devices (Auckland, NZ), vol. 8, p. 405.
[7] Lotte F., Congedo M., Lecuyer A., Lamarche F., Arnaldi B. (2007), “A review of classification algorithms for EEG-based brain–computer interfaces,” Journal of Neural Engineering, vol. 4, no. 2, pp. R1–R13. [Online]: https://doi.org/10.1088%2F1741-2560% 2F4%2F2%2Fr01
[8] Eeg measures of cognition. [Online]: https://hvmn.com/ biohacker-guide/cognition/eeg-measures-of-cognition
[9] Karayiannis, N., Xiong, Y., Frost, J., Wise, M., Hrachovy, R., Mizrahi, E. (2006), “Automated detection of videotaped neonatal seizures based on motion tracking methods,” Clinical Neurophysiology, vol. 23, no. 6, pp. 521–531. [Online]: https: //insights.ovid.com/crossref?an=00004691-200612000-00004
[10] Alkan, A., Koklukaya, E., Subasi A. (2005), “Automatic seizure detection in eeg using logistic regression and artificial neural network,” Journal of Neuroscience Methods, vol. 148, no. 2, pp. 167–176. [Online]: https://www.sciencedirect.com/science/article/pii/ S0165027005001342
[11] amd Ronald, W., Lesser, P., T.Richardson, R., Wilson, K. (1996), “An approach to seizure detection using an artificial neural network (ann),” Electroencephalography and Clinical Neurophysiology, vol. 98, no. 4, pp. 250–272. [Online]: https://www.sciencedirect.com/ science/article/abs/pii/0013469495002774
[12] Temko, A., Thomas, E., Marnane, W., Lightbody G., Boylan, G. (2011), “Eeg-based neonatal seizure detection with support vector machines,” Clinical Neurophysiology, vol. 122, no. 3, pp. 464–473. [Online]: http://www.sciencedirect.com/science/article/pii/ S1388245710006036
[13] Gardner, A. B., Krieger, A. M., Vachtsevanos G., Litt B. (2006), “One-class novelty detection for seizure analysis from intracranial eeg,” Journal of Machine Learning Research, vol. 7, no. Jun, pp. 1025–1044.
[14] Ramgopal, S., Thome-Souza, S., Jackson, M., Kadish, N. E., Fernandez, I. S., Klehm, J., Bosl, W., Reinsberger, C., Schachter, S., Loddenkemper, T. (2014), “Seizure detection, seizure prediction, and closedloop warning systems in epilepsy,” Epilepsy & behavior, vol. 37, pp. 291–307, 2014.

[15] Upenn and mayo clinic’s seizure detection kaggle challenge. [Online]: https://www.kaggle.com/c/seizure-detection/overview

[16] Melbourne university aes/mathworks/nih seizure prediction. [Online]: https://www.kaggle.com/c/melbourne-university-seizure-prediction/overview/timeline

[17] Glim, S., Okazaki, Y.O., Nakagawa, Y., Mizuno, Y., Hanakawa, T., Kitajo, K. (2019), “Phase-Amplitude Coupling of Neural Oscillations Can Be Effectively Probed with Concurrent TMS-EEG,” Neural Plasticity, vol. 2019, p. 6263907.

[18] Tort, A.B.L., Komorowski, R., Eichenbaum, H., Kopell, N. (2010), “Measuring Phase-Amplitude Coupling Between Neuronal Oscillations of Different Frequencies,” Journal of Neurophysiology, vol. 104, pp. 1195–1210.

[19] Edakawa, K., Yanagisawa, T., Kishima, H., Fukuma, R., Oshino, S., Khoo, H.M., Kobayashi, M., Tanaka, M., Yoshimine, T., (2016), “Detection of Epileptic Seizures Using PhaseAmplitude Coupling in Intracranial Electroencephalography,” Scientific Reports, vol. 6, p. 25422, 2016.

[20] Epilepsyecosystem. [Online]: https://www.epilepsyecosystem.org/

[21] Support vector machine - classification (svm). [Online]: http://saedsayad.com/support-vector-machine.htm

[22] Xgboost learns the canadian flag. [Online]: https://luckytoilet.wordpress.com/2018/01/28/xgboost-learns-the-canadian-flag/

[23] 1.6. nearest neighbors. [Online]: https://scikit-learn.org/stable/modules/neighbors.html

[24] Model selection. [Online]: http://ethen8181.github.io/machine-learning/model-selection/model-selection.html