Classification of EEG Signals for Predicting Epileptic Seizures

Alba Asllanaj, Mostafa Shuva

***Abstract*—The purpose of this paper is to use pattern recognition methods to predict and classify whether a patient is having an epileptic seizure. The Epileptic Seizure Recognition dataset [5] is used to test and verify the binary classification algorithm. EEG recordings in the files are pre-processed and categorized in five different classes, however, for simplification, we will focus only on two classes: (1) individual is having a seizure, and (0) individual is not having a seizure. Classification methods considered are Support Vector Machines (SVM), Naive Bayesian, and the logistic regression algorithm. The performance of these algorithms are compared by using confusion matrix analysis. In addition, various performance metrics such as error rate, precision, and runtime, were evaluated and compared in each of the experiments conducted. The results indicate that Naive Bayes' classifier generally has the lowest test error score and runtime, SVM classifiers give best performance once the hyperparameters are tuned, and lastly. The logistic regression classifiers perform the worst in all three experiments when used with L_2 regularization, but a significant performance increase can be achieved by using L_1 regularization. The research conducted on the various algorithms determines the best approach in solving the classification of EEG signals problem.**

Keywords*—epileptic seizure detection, Support Vector Machines, Naive Bayesian, logistic regression, binary classification, pattern recognition, feature extraction, elastic net

I. Introduction

Epilepsy affects nearly three million people world wide and can develop on any person or any age. New cases of epilepsy are common among children and those that are over the age of 60. While the causes of epilepsy remains unknown, it can be divided into two groups: brain injuries, and chemical imbalances in the brain. Trauma to the brain from injuries and accidents tend to increase the likelihood of a seizure. Due to the disruptive neural activity that occurs during a seizure, there is the possibility of damage being done to neural tissue and death.

Though this is a problem that affects many people, seizures can be detected with 100% accuracy as shown in [1], [8] and [9]. With the use of bio signals, such as electroencephalograms (EEGs), which are widely used in monitoring brain functionality, advanced opportunities are created to analyze such signals and determine whether an individual is experiencing epilepsy. This is helpful in providing fast care to those affected individuals.

Before continuing, we must define the different types of EGG signals available for our problem. When the EEG signal contains seizure activity, we call the signal ictal. When there is no seizure activity, we will call the EEG signal "seizure-free" or interictal [10].

In this paper, the detection challenges with the EEG signals will be discussed and analyzed. Based on the given dataset, several pattern recognition methods were used to be able to determine if an epileptic seizure is occurring.

Section II of this paper discusses related works and the research we conducted to familiarize ourselves with the methodologies used to generate a solution to this classification problem. The research conducted elaborates on feature extraction from the EEG signals, classification algorithms, as well as preprocessing methods.

In section III we discuss the dataset used, the different classes that come from each data point, followed by an exploratory analysis. Graphical and analytic comparisons of sample data points (seizure vs seizure free) were obtained in section IV, in order to visually determine a pattern or correlation by analyzing features of the frequency spectrum. The EEG waveform, Fourier transform, and the Wavelet transform of the EEG signals are compared.

Section V discusses the proposed algorithms intended to be used for the classification of the EEG signals. The following techniques will be applied using MATLAB: SVM, Naive Bayesian, and logistic regression. In using SVM, all the major kernels will be explored with a focus on the RBF (Radial Basis Function) since the linearity of the data is yet to be determined. In logistic regression the L_1 regularization will be used to select features from the EEG signals – in the results section we will refer this as an elastic net classifier since MATLAB's implementation of this also uses L_2 regularization. The Naive Bayes' classifier will be used as a benchmark, since it is the simplest classifier to implement for a binary classification problem, and our assumption is that it will have the cheapest computational cost.

The experimental evaluation of our methods is discussed in section VI. The three different algorithms – Naive Bayes, Logistic Regression (in addition to Elastic Net), and SVM – are the chosen classifiers for the solution implementation of the EEG signal classification problem. The generalization of each algorithm is discussed in detail, and provides an understanding of the overall advantages and disadvantages of the methods. In addition, we discuss the pre-processed methods and how feature extraction was obtained.

Section VII goes through the results and analysis of our classification methods. The section is separated into the three experiments and dives into the results obtained by the following performance metrics: training & test error, sensitivity, specificity, precision, fall out, F1 score, and runtime. We will see the performance of each algorithm varies based on how the initial data is pre-processed.

Lastly, in section VIII, we go over some of the future works that can be done to expand on the experiments conducted and discussed in this paper.

II. Related Works

Based on the papers we reviewed, features were extracted from the EEG signal through various different methodologies. One research paper in particular discusses the empirical mode decomposition (EMD) to find intrinsic mode functions (IMF) which were used by [9] and [10] to extract features that allow for the distinction between ictal and interictal signals. Another approache to this problem include the use of wavelet transforms as the basis for feature extraction as discussed in [1] and [3]. Furthermore, [8] uses the FFT along with L1 regularization to select the best features from the spectrum of the EEG signal.

The basic approach to this problem is to select the proper features of the EEG signal to get the best classification. In order to select such features, the EEG signal must be decomposed. Based on earlier papers, the most common way of decomposing the EEG signal is by empirical mode decomposition to find the intrinsic mode functions [6]. We can represent the EEG signal in terms of its IMFs, $d_k(t)$, as follows,

$$x(t) = m_k(t) + \sum_{k=1}^{N} d_k(t) \tag{1}$$

where $m_k(t)$ is called the residual and $d_k(t)$ is an amplitude and frequency modulated (AM-FM) function. This decomposition fully characterizes the original signal $x(t)$ in terms of its IMFs and residual function. Since this is the most popular approach, we will attempt to use the EMD to select features in order to classify ictal and interictal EEG signals.

A newer approach utilizes wavelets to decompose the EEG signal ([1],[3]). The empirical wavelet transform (EWT) as proposed in [7] provides another method, similar to EMD, to decompose a signal. Unlike EMD which is purely an algorithmic approach to signal decomposition, EWT is based on mathematical theory and offers an analytic approach to generating the filter banks necessary to decompose the signal.

III. Dataset

The dataset that is being using was obtained from the University of California – Irvine (UCI) repository [5]. This set was distributed by the School of Mathematical Sciences at Rochester Institute of Technology. The files are pre-processed and restructured, from a dataset that contained EEG recordings for epileptic seizure detection. Each data point is an integer, which represents the value of the EEG recording at that given sample.

The data was collected by placing electrodes based on the international 10-20 system on subjects that are healthy and subjects that are diagnosed with epilepsy [2]. Figure 1 shows the placement. The EEG signal represents voltage fluctuations in the brain, so we can think of each sample $x_i[n]$ as a discrete signal representing voltage. That is, for each value of n, the sample $x_i[n]$ has units in terms of voltage. This piece of information helps provide an intuition for why the waveforms analyzed in Section IV appear the way they do, since a seizure is an irregularity in the brain's neural behavior.

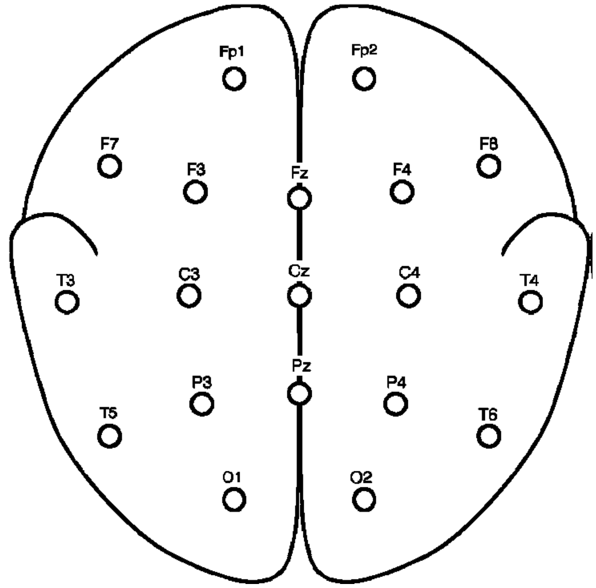

Figure 1: Placement of electrodes for collecting EEG data [2].

There are 4097 data points that were recorded from 500 individuals. From these individuals data was collected in the time span of 23.5 seconds, sampled at 173.61 Hz, and each second contains 178 data points. Figure 2 shows how the dataset is formatted, prior to being modified for this project.

The response variable y is the output of the input vector containing 178 data points for one second. The classes of y are represented by the values $\{1, 2, 3, 4, 5\}$, where:

- 5: Eyes open, while recording data
- 4: Eyes closed, while recording data
- 3: Identified where the region of the tumor was in the brain and recording the EEG activity from the healthy brain area
- 2: Recorded the EEG from the area where the tumor was located
- 1: Recording of seizure activity

Each of the above categories contains 2300 samples, for a total of 11500 total sample points. A sample in this data set can be though of a function $x[n]$ that is defined for $n \in [1, 178]$. Note that we are representing x as a discrete signal since data provided was sampled from an EEG signal. The original EEG signal is $x(t)$ and the relationship between the EEG signal an the discrete signal provided in our dataset is,

$$x[n] = x(nT), \tag{2}$$

where $T = 1/f$ and f is defined as the sampling frequency.

In the dataset, only the response variables that are classified as '1' indicate that the individual is experiencing an epileptic seizure. For our binary classification problem, we will break the response variable y into the following two classes:

- 0: the individual does not experience a seizure ($y = \{2, 3, 4, 5\}$)

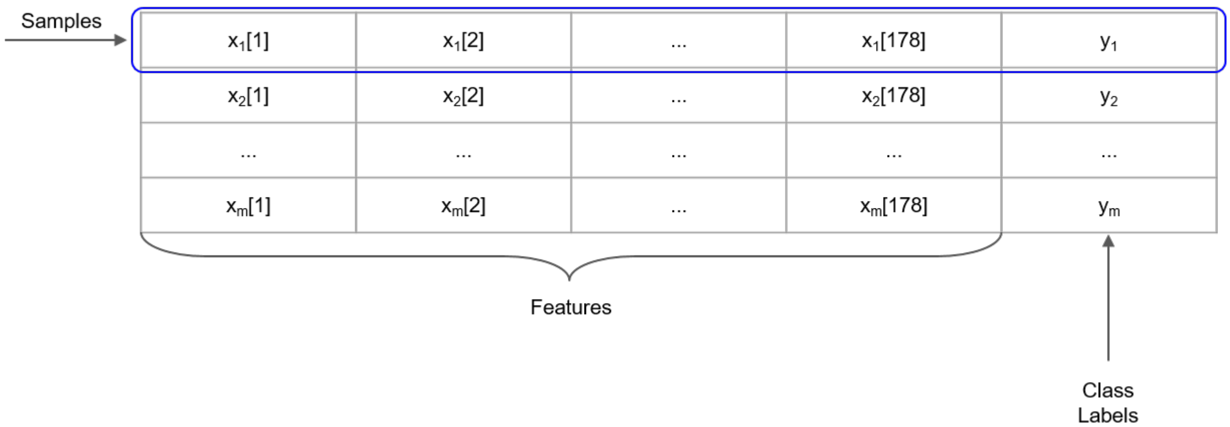

Figure 2: Each row of the dataset contains a sample (x_i, y_i). The first 178 columns of the dataset represent values of the EEG signal, which can be thought of as the components of the input vector x_i. The final column is the class label for each input vector.

- 1: the individual is experiencing a seizure ($y = \{1\}$)

Based on how the data is provided, our project will model a function $y = f(x)$ where the output will simply be the class the EEG signal belongs to. The function f will treat each EEG sample, x_i, as a signal and attempt to identify key features that will help determine the proper class label output y.

IV. Exploratory Analysis

An initial analysis was done on the dataset acquired for this project. In the analysis, the following three items were considered:

1) waveform of the EEG signals
2) Fourier transform of the EEG signals
3) Wavelet transform of the EEG signals

When comparing just the first three samples in our dataset, we observe a major different in the two sets of data. EEG signals taken during a seizure episode display higher peak amplitudes than seizure free EEG signals, as seen in Figure 3. To mathematically quantify this difference, the energy was measured and compared between the two types of EEG signals. From signal processing, we can compute the energy of a discrete time signal – the original data from the EEG recordings were analog, however the data was then sampled at 173.61 Hz – by the following equation,

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2, \tag{3}$$

where the upper and lower bounds in the finite case are restricted to the interval $[0, N-1]$.

Computing the energy of the signal gives us a parameter to characterize EEG samples showing seizure activity and those that do not show seizure activity. We can similarly calculate

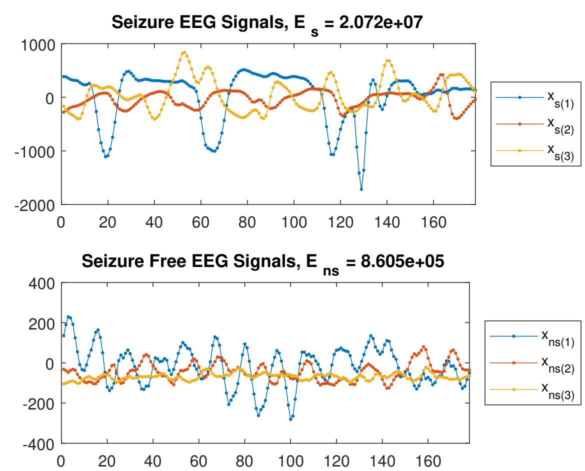

Figure 3: Comparison of the first 3 sample seizure and seizure free EEG signals. Note that the mean energy = of the seizure EEG signal is $E_s = 2.072 \times 10^7$ and the mean energy of the seizure free EEG signal is $E_{ns} = 8.605 \times 10^5$.

the power of the EEG signal as follows,

$$P = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x[n]|^2. \tag{4}$$

However, the power of any finite EEG signal will be equal to zero, per the definition above. So, it makes more sense to characterize our dataset using the energy metric. Doing so, we can conclude seizure activity is characterized by high energy compared to no seizure activity. This, intuitively, matches our understanding that during a seizure episode there is abnormally excessive brain activity with neurons firing constantly and also

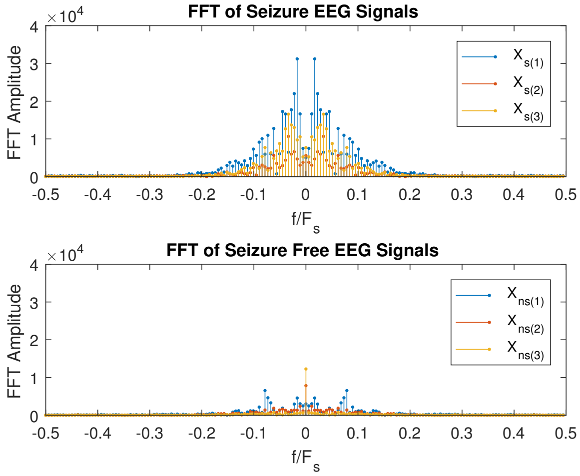

Figure 4: Comparison of the Fourier transform of the first 3 sample seizure and seizure free EEG signals. The same axes were used to provide an easy visual comparison of the difference between FFT amplitudes. Notice that the spectrum seems to be relatively sparse.

matches the plots in Figure 3.

Next, we consider the frequency spectrum of the EEG signal given by the discrete Fourier transform as defined by,

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-\frac{2\pi i}{N} kn}. \tag{5}$$

We plot the magnitude of the Fourier transform and discard the phase information, since plotting the phase did not give any compelling insight into the dataset. As seen in Figure 4, the seizure free EEG signal does not exhibit high amplitudes as compared to the signal containing seizure activity. In addition, the frequency analysis shows that when seizure activity is present, a wider range of frequencies are impacted.

Aside from the Fourier transform, another metric we can consider is the Power Spectral Density (PSD) which is a measure of signal power across different frequencies. From Figure 5, we see that the roll off of the PSD with respect to frequency is greater when there is seizure activity.

A spectrogram measurement of the seizure versus seizure free signal also confirms the notion that during a seizure, the amplitude of the signal in the frequency domain is amplified over a broader frequency range, as shown by Figure 6.

Since some research in this field used wavelets to solve the seizure detection problem [1],[3], we also analyzed what the wavelet decomposition of the dataset looks like. In particular, we computed the EWT to get an idea of what the decomposed EEG signal looks like. The EWT is a method of computing an adaptive filter bank based on the signal itself instead of using a pre-defined basis like the Fourier transform or the discrete wavelet transform. In order to do this, we used the EWT toolbox presented by [7].

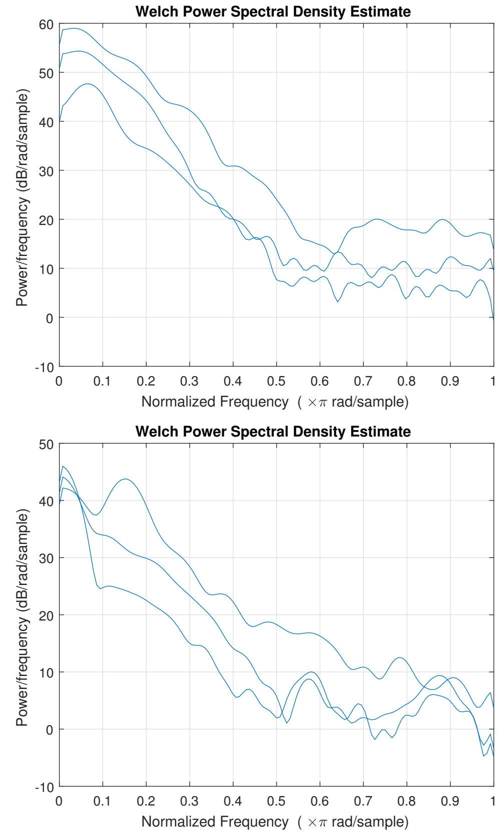

Figure 5: Power spectral density of seizure signals (top) and seizure free signals (bottom).

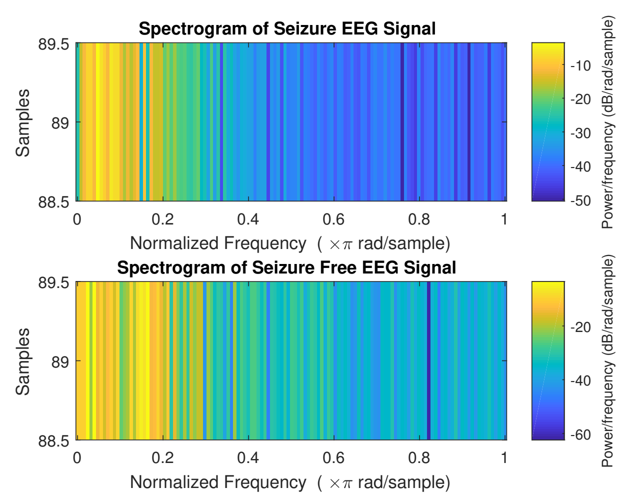

Figure 6: Comparison of the spectrogram of a seizure and seizure free EEG signals.

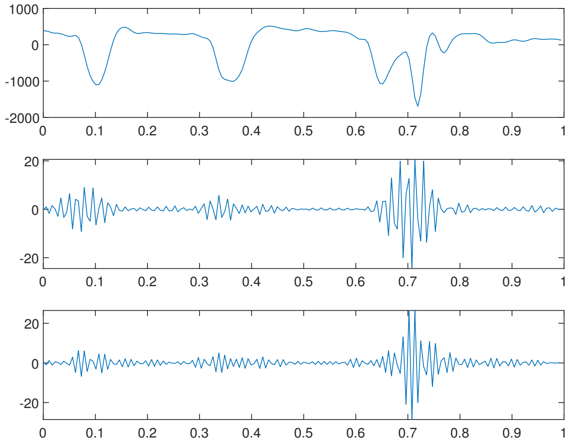

Figure 7: Empirical wavelet decomposition of the ictal EEG signal. This particular sample only has 3 EWT components.

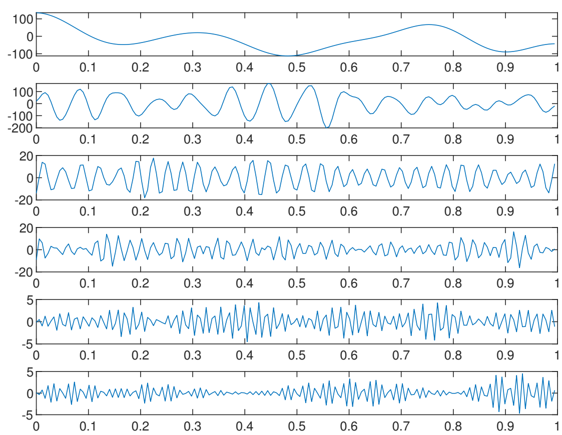

Figure 8: Empirical wavelet decomposition of the seizure free EEG signal. This sample has 7 EWT components, but we are plotting the first 6.

Based on all the exploratory analysis done, we believe it will be helpful to take features from the frequency spectrum, since there are clearly visible patterns that we can use in classifying ictal and interictal EEG signals.

V. Proposal

This project will propose an algorithm to classify the EEG signals using SVM, Naive Bayesian, and logistic regression techniques that will be implemented using MATLAB functions. The features will be determined from the FFT, EMD, and wavelets of the EEG signal. We will solve the binary classification problem using the previously mentioned methods and measure the robustness of the classifier using various metric (i.e. accuracy, precision, runtime, etc.).

A. *Support Vector Machines*

Support Vector Machines (SVM) is a type of algorithm used in pattern recognition for linearly or non-linearly (kernel function) separable patterns with the use of an optimal hyperplane. The data points that lie closer to the hyperplane are called support vectors, hence the name Support Vector Machines. SVMs aim to split the dateset in its feature space using a hyperplane while maximizing the margin. We will aim to construct various SVM classifiers and determine which of the following kernels works best with our data [4]:

- Linear: $K(x_i, x_j) = x_j^T x_j$
- Polynomial: $K(x_i, x_j) = (\gamma x_j^T x_j + r)^d, \gamma > 0$
- RBF: $K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2), \gamma > 0$
- Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_j^T x_j + r)$

In the above equations γ, r, and d, are the kernel parameters that need to be defined.

Using an RBF kernel we need to set two additional parameters: a C parameter, and a γ parameter. The default value for the C parameter is 1.0, but can be changed to better fit the data. Low C values have a softer margin by allowing misclassification, whereas higher C values leave very little room for error which can lead to overfitting – this is essentially a regularization parameter. Similarly, for small γ values, i.e. $\gamma = 2^{-5}$, there is less complexity, which allows for some misclassifications. With larger γ values, i.e. $\gamma = 2$, we risk overfitting.

Figure 9 represents a binary classification problem that illustrates the issue of overfitting. To avoid overfitting the data, an exhaustive grid search will be used in conjunction with K-fold cross-validation to determine the best pairing of C and γ.

To identify the best C and γ values for the data that is being categorized, we can iteratively test different parameter values. This is one of the drawbacks to using SVM since choosing the correct kernel and parameter values can be computationally intensive and highly dependent on iterative trials.

B. *Naive Bayes'*

Naive Bayes is a basic probability algorithm that can be used for classification problems. In this paper, Naive Bayes will be used as a Linear Classifier, with binary classification of $y = 0$ or 1. To construct this algorithm we will need Bayes' Theorem,

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{6}$$

- **prior**: the independent probability of the class $p(y)$
- **likelihood**: conditional probability of x given y, $p(x|y)$
- **posterior**: conditional probability of y given x, $p(y|x)$

Since we are dealing with binary classification, the Bernoulli distribution will be used in the model. The Naive Bayes

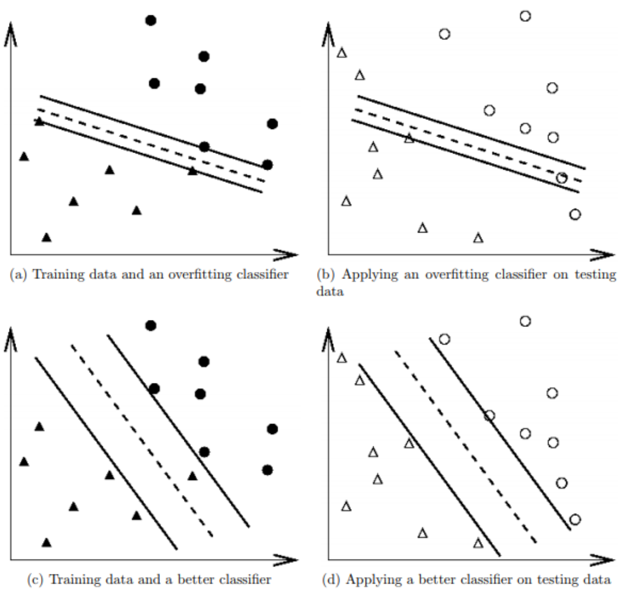

Figure 9: Comparison of SVM classifiers [4].

Classifier has independence assumptions between predictors (knowing the value of one attribute does not necessarily tell us anything about the value of another attribute), and is easy to build since there are no iterative parameters. This is extremely useful when used in large datasets.

Initially, the predictive features, x_i, and the target variable (seizure, or no seizure), y, will need to be isolated. Overall we will build a classifier that can predict the likelihood that a patient will have an epileptic seizure, give the evidence in our data set.

C. Logistic Regression

Another classification algorithm that we will try to use is the logistic regression classifier. We will attempt to solve the problem by using an elastic net to exploit L_1 sparsity in the Fourier spectrum of the EEG signal. More specifically, the cost function we will use is

$$J(\omega) = -\ln p(y|\omega) + \lambda\left[(1-\alpha)\|\omega\|_2^2 + \alpha\|\omega\|_1\right] \quad (7)$$

which uses both L_1 and L_2 regularization to achieve the best classification. Computationally, this also lets us ignore features that do not influence the quality of our classifier. Since MATLAB's implementation of logistic regression utilizes both L_1 and L_2 regularization, classifiers generated using L_1 regularization will be referred to as elastic net classifiers, and the logistic regression classifier designation will be left for classifiers generated using only L_2 regularization.

VI. Experimental Evaluation

A. Algorithms

For this project, we developed four different classifiers:

1) Naive Bayes'
2) Logistic Regression
3) Elastic Net
4) Support Vector Machines

We'll briefly provide the reasoning behind each of the chosen classifiers.

1) Naive Bayes': We chose to do a Naive Bayes classifier primarily because it is a simple MAP estimator. Given the other, more complex, classifiers we will investigate, Naive Bayes provides us with a benchmark to compare robustness metrics with as well as a comparison of computation time.

2) Logistic Regression: Logistic Regression was chosen because we can use different types of regularization parameters with this classifier. In our project use utilize both L_1 and L_2 regularization to classify the EEG signals. During our exploratory analysis, we noted that the Fourier spectrum of the EEG signals were relatively sparse. As a result this problem is a perfect candidate for L1 regularization, since it takes advantage of the sparsity and automatically discards features that are not critical for make the classification.

3) Elastic Net: Since the logistic regression classifier can be generated using either L_1 or L_2 regularization, we decided to also create a classifier using an elastic net. We did this by searching over a parameter space for the regularization parameters that minimized the test set error of the classifier.

4) Support Vector Machine: The last classifier we chose to use is a support vector machine. SVMs are state-of-the-art classifiers are can be applied in many classification problems. For the sake of comparison, we chose to use an SVM as our final classifier to compare its performance (robustness as well as runtime) to Naive Bayes and logistic regression. In our application, we will use an SVM with a radial basis function first, and then we will use MATLAB to optimize the hyperparameters of an SVM (type of kernel, regularization, and gamma value).

B. Pre-Processing Methods

From the technical survey, we discussed multiple ways of preprocessing the EEG signal to help extract features. In our study, we will apply the mentioned algorithms after applying three different preprocessing techniques:

1) the Fourier transform
2) the Empirical Mode Decomposition
3) the Wavelet transform

When applying the empirical mode decomposition, we considered the the sum of highest and lowest order mode function as features. Similarly, for the wavelet transform, the sum of the highest and lowest order wavelet was taken as the feature representation of our data.

In all these cases we are transforming our data into another representation before doing the classification.

VII. Results & Analysis

The results of our analysis are found in Tables 1, 2, and 3. As we can see, despite being a simple algorithm, the Naive

Bayes' classifier performs significantly well in all cases of pre-processed data with a test error of less than 7%. Naive Bayes' is also the fastest algorithm in terms of runtime as it took less than one second of computation time to run. In the following experiments, we will see that the SVM classifiers, being the most complex, has the best performance among the classifiers only when its' hyperparameters are tuned. Logistic regression is a poor choice of classification algorithm in our study, unless L_1 regularization is used along with L_2 to generate an elastic net classifier.

A. Experiment 1: Pre-processing based on the Fast Fourier Transform

In our first experiment we pre-process the EEG signal using the FFT to acquire the frequency spectrum. We then applied our classifiers to the frequency representation of the EEG signal – this is due to the fact that during the exploratory analysis we noticed low frequency amplitude spikes when seizure activity was present.

Based on Table 1, we observe the results of our classifiers when we apply them to the frequency spectrum. Immediately we see that the Naive Bayes' classifier is one of the fastest classifiers for this problem, and the most accurate. However, a well tuned SVM classifier is able to achieve a much better performance with an test error of 0.43%.

The logistic regression classifier, based on L_2 regularization had the worst performance overall. When the logistic regression was modified to use L_1 regularization in the elastic net classifiers we immediately saw an increase in performance – going from 52.93% test error to 1.63% and 0.98% test error respectively. This can be attributed to the L_1 regularization "rewarding" sparse solutions and we previously observed in the exploratory analysis the the frequency spectrum of an EEG signal is sparse and contains peak amplitudes during seizure activity.

B. Experiment 2: Pre-processing based on the Empirical Mode Decomposition

Experiment 2 uses the empirical mode decomposition. In this experiment, each EEG signal was decomposed into its' intrinsic mode functions and the sum of the highest and lowest mode functions was taken as inputs to the classifiers. Using the EMD, we see that the logistic regression still maintains a high test error. The key observation for the elastic net classifiers is that the test error has increased from Experiment 1 to 29.02%. This is because in Experiment 1, the frequency spectrum was sparse, and allowed the classifier to exploit this with the L_1 regularization. In contrast the EMD does not provide a sparse representation of the EEG data and therefore the elastic net's performance is degraded.

Like in Experiment 1, Naive Bayes' performs exceptionally well for being a simple algorithm. The SVM classifier also performs well, once its' hyperparameters are tuned to fit the problem.

C. Experiment 3: Pre-processing based on the Continuous Wavelet Transform

Experiment 3 applies the continuous wavelet transform (CWT) to the EEG data before classification. This experiment was inspired by [3] and [7], where they used the empirical wavelet transform (EWT). For the sake of simplicity, we only used the CWT in this paper.

Using the CWT as the pre-processing method, we see performance similar to that of Experiment 1 across all classifiers. Naive Bayes', once again, performs exceptionally, well achieving a 6.52% test error.

The logistic regression classifier with L_2 regularization performs the worst, just as it did in Experiments 1 & 2. We can conclude that none of these pre-processing methods are optimnal when using a logistic regression classifier with L_2 regularization. When we go to L_1 regularization with the elastic net classifiers, there is an immediate performance increase as the CWT behaves similarly to the FFT where it decomposes the EEG signal into its' frequency representation.

The SVM classifier also performs well once it is tuned with optimal hyperparameters. As seen in Experiments 1 & 2, before tuning the SVM the test error is significantly high. This is partly attributed to the SVM algorithm being a more complex algorithm that requires additional parameter tuning in order to be fully taken advantage of.

VIII. Future Work & Conclusion

The overall goal for this project was to develop an algorithm using Naive Bayes', logistic regression, and SVM as classifiers to determine whether a patient will have a seizure based on EEG data. From the research conducted during our exploratory analysis, we tested each classifier using EEG data that was processed via the FFT, EMD, and CWT. The results of our experiments proved that though Naive Bayes' is a rather simple algorithm, in most cases it can provide the best performance in the shortest amount of time. Logistic regression, with L_2 regularization, performed poorly across all three experiments and leads us to believe a more robust pre-processing method must be needed for feature selection. With L_1 regularization, however, the elastic net implementation of logistic regression performed well in cases where there were sparse features, namely when we classified based on the frequency spectrum of the data. The SVM classifiers performed the best, in all three experiments, but was heavily depended on the choice of hyperparameters. When optimal hyperparameters were selected, the SVM classifier was able to drive the test error below 4%.

Future works related to this problem could attempt to perform additional pre-processing to the data. We have seen that classifying based on frequency features generally provide the lowest error for this problem. In our study only relatively simple classification algorithms were considered. Selecting the optimal hyperparameters for the SVM classifiers took roughly one hour, but we observed that in all the experiments SVM performed the best once the proper hyperparamenters were chosen. A more complicated algorithm that should be considered is a neural network using multilayer perceptrons – as this would allow for more tunable parameters and could potentially help drive the classification error to zero.

Algorithm		Performance Metrics Based on Test Dataset						
	Training Error	Test Error	Sensitivity	Specificity	Precision	Fall Out	F1 Score	Runtime (seconds)
Naive Bayes'	0.0231	0.0304	0.9815	0.9589	0.9551	0.0411	0.9681	0.685
L_2 – Logistic Regression, $\alpha = 0.5$	0.4927	0.5293	1	0	0.4707	1	0.6401	1.049
L_1 – Elastic Net, $\alpha = 0.5$	0.0111	0.0163	1	0.9692	0.9695	0.0308	0.9830	2.546
L_1 – Elastic Net, $\alpha = 0.1$	0.0079	0.0098	1	0.9815	0.9796	0.0185	0.9897	8.396
SVM, rbf, C = 1, $\gamma = 1$	0	0.5293	1	0	0.4707	1	0.6401	1.832
SVM, linear, C = 0.001, $\gamma = 109.9486$	0.0014	0.0043	0.9954	0.9959	0.9954	0.0041	0.9954	0.675

Table 1: Results of Experiment 1.

Algorithm		Performance Metrics Based on Test Dataset						
	Training Error	Test Error	Sensitivity	Specificity	Precision	Fall Out	F1 Score	Runtime (seconds)
Naive Bayes'	0.0144	0.0207	0.9931	0.9671	0.9641	0.0329	0.9784	0.533
L_2 – Logistic Regression, $\alpha = 0.5$	0.5361	0.5272	0.5242	0.4271	0.4486	0.5729	0.4835	0.623
L_1 – Elastic Net, $\alpha = 0.5$	0.2804	0.2902	0.9746	0.4743	0.6224	0.5257	0.7597	1.043
L_1 – Elastic Net, $\alpha = 0.1$	0.2804	0.2902	0.9746	0.4743	0.6224	0.5257	0.7597	1.043
SVM, rbf, C = 1, $\gamma = 1$	0	0.5293	1	0	0.4707	1	0.6401	1.864
SVM, linear, C = 155.7453, $\gamma = 317.992$	0	0.0380	0.9307	0.9897	0.9877	0.0103	0.9584	1.212

Table 2: Results of Experiment 2.

Algorithm		Performance Metrics Based on Test Dataset						
	Training Error	Test Error	Sensitivity	Specificity	Precision	Fall Out	F1 Score	Runtime (seconds)
Naive Bayes'	0.0750	0.0652	0.9792	0.8953	0.8926	0.1047	0.9339	0.536
L_2 – Logistic Regression, $\alpha = 0.5$	0.4927	0.5293	1	0	0.4707	1	0.6401	0.635
L_1 – Elastic Net, $\alpha = 0.5$	0.0739	0.0685	0.9769	0.8912	0.8887	0.1088	0.9307	7.110
L_1 – Elastic Net, $\alpha = 0.1$	0.0712	0.0625	0.9746	0.8994	0.8960	0.1006	0.9336	6.897
SVM, rbf, C = 1, $\gamma = 1$	0	0.5293	1	0	0.4707	0	0.6401	2.245
SVM, linear, C = 0.001, $\gamma = 6.2014$	0.0666	0.0685	0.9630	0.9035	0.8987	0.0963	0.9298	0.761

Table 3: Results of Experiment 3.

IX. References

[1] M. Ahmad, "Epileptic Seizure Classification using Statistical Features of EEG Signal Epileptic Seizure Classification using Statistical Features of EEG Signal", *International Conference on Electrical, Computer and Communication Engineering*, no. February, pp. 308–312, 2017. DOI: 10.1109/ECACE.2017.7912923.

[2] 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", *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 64, no. 6, p. 8, 2001, ISSN: 1063651X. DOI: 10.1103/PhysRevE.64.061907. arXiv: arXiv:1412.6980v9.

[3] A. Bhattacharyya, V. Gupta, and R. B. Pachori, "Automated identification of epileptic seizure EEG signals using empirical wavelet transform based Hilbert marginal spectrum", *International Conference on Digital Signal Processing, DSP*, vol. 2017-August, 2017. DOI: 10.1109/ICDSP.2017.8096122.

[4] Chih-Wei Hsu, Chih-Chung Chang and C.-J. Lin, "A Practical Guide to Support Vector Classification", *BJU international*, vol. 101, no. 1, pp. 1396–400, 2008, ISSN: 1464-410X. DOI: 10.1177/02632760022050997. arXiv: 0-387-31073-8. [Online]. Available: http://www.csie.ntu.edu.tw/%7B~%7Dcjlin/papers/guide/guide.pdf.

[5] D. Dheeru and E. Karra Taniskidou, *UCI machine learning repository*, 2017. [Online]. Available: http://archive.ics.uci.edu/ml.

[6] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank", *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004, ISSN: 1070-9908. DOI: 10.1109/LSP.2003.821662.

[7] J. Gilles, "Empirical wavelet transform", *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3999–4010, 2013, ISSN: 1053587X. DOI: 10.1109/TSP.2013.2265222.

[8] R. Hussein, Z. J. Wang, and R. Ward, "L1-Regularization Based EEG Feature Learning for Detecting Epileptic Seizure", pp. 1171–1175, 2016.

[9] Z. Roshan Zamir, "Detection of epileptic seizure in EEG signals using linear least squares preprocessing", *Computer Methods and Programs in Biomedicine*, vol. 133, pp. 95–109, 2016, ISSN: 18727565. DOI: 10.1016/j.cmpb.2016.05.002. arXiv: 1604.08500. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2016.05.002.

[10] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions", *Expert Systems with Applications*, vol. 42, no. 3, pp. 1106–1117, 2015, ISSN: 09574174. DOI: 10.1016/j.eswa.2014.08.030. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2014.08.030.