Machine Learning and EEG Signals
Examination of Machine Learning Models for the Classification of Epileptic Seizures

Justin Viveiros
500828996
Toronto Metropolitan University
justin.m.viveiros@ryerson.ca

***Abstract*—The capture and interpretation of electroencephalogram (EEG) signals is an important clinical tool that can help in the diagnosis and treatment of various neurological conditions. Epilepsy syndrome is a neurological condition where afflicted individuals suffer from unprovoked random seizures, those being random electrical activity in the brain. The diagnosis of EEG signals is a useful tool to the diagnosis of epilepsy and there is much interest in developing machine learning tools to assist in the interpretation and classification of seizures from EEG signals. There are a variety of machine learning frameworks that have been created to classify seizure EEGs, differing in feature extraction from using primarily time-domain frequencies to utilizing signal decomposition techniques, and differing in models like support vector machines to convolutional neural networks. A MATLAB implementation was attempted utilizing segmentation sizes of 3 and 10 seconds as well as utilizing a primarily time and frequency based feature set and a signal decomposition based feature set. The results of other machine learning frameworks were successful in differentiating seizures from normal EEG signals. The MATLAB implementation was also successful generating high classification accuracy, with logistic regression being the superior technique and signal decomposition features performing better with longer signal segments. An observation about the MATLAB results was that there was a comparably small data set utilized for testing and validation compared to the results from literature likely leading to a higher classification accuracy that may not be present were a larger feature set used.**

***Index Terms*—epilepsy, empirical mode decomposition, signal decomposition, feature extraction**

I. Introduction

A. Motivations and Problem Statement

The diagnosis and classification of epilepsy is a challenging process and is reliant on a variety of methods and factors including family history, age at which symptoms began, type of seizure, brain imaging, and EEG interpretation [1]. Of particular note, is the importance of EEG signal interpretation which is complicated by the fact that EEGs are by nature nonlinear and nonstationary and that is particularly true for EEG signals during a seizure [2]. There has been a great amount of interest in signal processing techniques that can identify seizures from EEG signals to help in the diagnosis process, particularly in the development of machine and deep learning frameworks that can identify epileptic seizures with high accuracy [3]. This paper will examine some factors that go into the development of a machine learning framework for the classification of seizure EEGs including how signal length, features extracted from EEG signals, and different classifiers affect the accuracy of said frameworks.

B. Background on EEG Signals and Epilepsy

The central nervous system acts as a controller for the functions of the rest of the human body [4]. The brain is made up of millions of neurons which process and transmit electrical signals, in the form of cortical action potentials, to one another and throughout the body [4]. These cortical action potentials of neurons change over time and will depend on various factors like stimuli being experienced or function being carried out by an individual [5]. As a consequence of the cortical action potentials, an electric field is generated which can be measured to gain understanding into brain activity [5].

An electroencephalogram (EEG) is a signal which can be captured that graphically represents the electrical activity of the brain and is captured through the use of electrodes which can be placed either on the scalp non-invasively or invasively [5], [6]. The most common methodology to place electrodes is the international 10-20 EEG system [5]. There is a great amount of utility to being able to gather information on general brain function and as such EEGs have seen wide use clinically in a variety of contexts [6], [7]. One major area of clinical EEG use is in the diagnosis of neurological conditions like Alzheimer's disease, seizures, and epilepsy syndromes [6]–[8]. EEGs also see use in some other areas like mapping of brain function, sleep analysis, and even as inputs for entertainment [6].

Of particular importance to this paper is the use of EEGs for epilepsy diagnosis. An epilepsy syndrome is a neurological disorder characterized by the presence of unprovoked seizures as the main symptoms of the condition [1]. Seizures are the result of abnormal transient rhythmic neuronal impulses in the brain, and usually manifest randomly and in a variety of forms [1]. Diagnosis of epilepsy typically involves a variety of factors including family history, type of seizures, onset age, neurological imaging, and interpretation of EEG signals while a seizure is taking place [1]. Diagnosis is complicated by the paroxysmal nature of the seizures as the EEG signals being examined are nonstationary and nonlinear, making analysis with some signal processing techniques difficult [1]. Typically, EEGs are recorded while a video of an individual is being taken so that the EEG activity occurring at the time of the seizure can be examined [1]. With that in mind, there has been great interest in automatically determining when an epileptic seizure has taken place from the signal itself so that this analysis could be performed outside of clinical environments and potentially with commercially available EEG scanners [1]. To this end, there are a variety of machine learning and deep learning algorithms that aim to classify seizure EEGs automatically to assist clinicians.

C. Machine Learning Overview

Machine learning broadly defined is the process of training a machine to be able to identify patterns and classify data, with the tools being utilized to teach the machine being called the model [6]. There are a variety of applications for machine learning including: long-term monitoring, event detection, and diagnosis [6]. The entire ML process can be broken up into a few major steps, the first being the training of the ML model. During the training step, features are extracted from a large portion of the dataset being utilized (often 60-80%) and are put into the machine learning model in order for it to learn the patterns that appear in the dataset [6]. The next stage called testing involves utilizing the remaining or a small portion of the dataset, not utilized for training, and passing said data into the model to see if it can accurately classify said data based on the patterns it learned previously [6]. Finally, validation is usually performed by shuffling the dataset in some manner and training and testing the model based on this shuffled data, where the training and testing process is performed multiple times to increase model robustness [6]. There are a variety of methods to perform this including k fold cross-validation and leave-one-out cross validation [6], [9]. Machine learning models can be broken up into three major categories, those being supervised, semi-supervised, and unsupervised models [6]. These categories differ in their use of unlabelled or labeled data [6]. In supervised models, labels are typically available for data and the model aims to learn the various labeled classes and classify test data based on those labels [6]. Examples of this type of model include: Logistic Regression (LR), Linear Discriminant Analysis, and Support Vector Machines (SVM) [6]. Unsupervised models are utilized when labels in data are weak or unavailable and as such the model is not given any information about the classes in the data and instead finds patterns or groups within the data itself without being told specifically what those groups are through labels [6]. Examples of this type of model include clustering techniques [6]. Finally, semi-supervised models are a mix of the two other types and involve the use of a large portion of unlabeled data and a comparably smaller amount of labeled data and is primarily used in medical image analysis where relevant features may be hard to extract [6]. Two ML techniques that will be examined here are the LR and SVM techniques.

D. Logistic Regression

The first type of ML technique that will be examined here is the supervised technique called Logistic Regression (LR) [6], [9]. It is utilized in cases where the classification is of a categorical nature, as an example, data could be pass or fail or healthy or diseased [6]. The technique functions on the basis of a logistic function which takes the shape of a sigmoid curve and is represented mathematically by equation 1, where z is defined by equation 2 and has b's corresponding to logistic regression coefficients and f's representing features extracted from the data [6]. The b values are typically calculated utilizing either the least squares method or maximum likelihood estimation [6]. From the logistic function, data can fall into one of two categories: the first group being lower than 0.5 probability or the second group being higher than 0.5 [6]. This technique is an incredibly simple ML methodology and involves low computational cost, making it ideal for tiny ML applications [6]. A major downside to the technique is that it can only perform two-group classification and is not suitable at distinguishing more than two [6]. In the case of seizure detection, this drawback is not significant as one could classify data as either normal or having a seizure.

$$f(z) = \frac{1}{1+e^{-z}} \tag{1}$$

$$z = b_0 + b_1 f_1 + b_2 f_2 + ... + b_p f_p \tag{2}$$

E. Support Vector Machines

Another popular form of supervised machine learning is the Support vector machine (SVM) [6]. In this technique, data points are separated into two classes through the use of a hyperplane which creates a decision boundary between the two classes, where if a point is on one side of the hyperplane it is considered to be the first class and if it falls on the opposite it would be considered to be the second [6], [10]. The hyperplane in the case of SVM use can be linear or of another shape depending on a kernel function utilized [6]. Often, the data can not be easily split with a linear hyperplane and as such a variety of other kernels like polynomial, gaussian, or radial basis function kernels have been utilized to better separate the classes with a suitable hyperplane shape [6]. The overall goal of the SVM technique is to maximize the margin or the distance from either class to the hyperplane, therefore making classification easier [6], [10]. The SVM technique has some advantages like its memory efficiency as well as its ability to handle nonlinear features and higher dimensional overlapping features better than other techniques [6]. A negative to SVM is a large susceptibility to noise, making it a technique that performs better with an extensive denoising step in preprocessing [6].

II. Materials and Methods

A. Algorithms for ML Epilepsy Classification from Literature

Firstly, one example of a deep learning architecture that has been successful in the classification of seizure and non-seizure EEGs is described in [3]. In addition to being able to classify seizure and non-seizure EEG signals, this framework was capable of diagnosing the seizure signal as either a mild seizure or a severe case [3]. In this paper, a deep learning architecture featuring a convolution neural network (CNN) is utilized in addition to features extracted through the use of empirical mode decomposition [3]. In this paper, EEG signals from two EEG seizure databases were utilized, those being the Bern-Barcelona dataset and CHB-MIT Scalp dataset. The EEG signals were decomposed using the EMD process, however the EMD process was stopped when six IMFs were generated as opposed to allowing the sifting process to complete until no further IMFs could be generated from a residual [3]. The reason for this was that results generated with the first six IMFs generated an optimal classification rate for the deep learning process as opposed to utilizing all IMFs [3]. From these six IMFs, five features were generated to be passed into the CNN. The first feature was the mean value from the IMF subbands given by equation 3, where C is the number of IMF sub bands [3]. The second feature was the energy of the subbands given by equation 4 [3]. The third and fourth features were peak max index feature (PMA) and peak min index feature (PMI), derived by dividing 1 by the maximum or minimum in a particular IMF sub band [3]. The final feature is the linear index feature (LIF) which is computed as the difference between the PMA and PMI divided by the PMI [3]. These external features are computed from the IMFs and then passed into the CNN with the IMF subbands where a second set of features is calculated from the internal layers of the CNN [3].

$$Mean(M) = \frac{\sum_{i,j=1}^{N} |IMF_{i,j}|}{C} \tag{3}$$

$$Energy(E) = \sum_{i,j=1}^{N} |IMF_{i,j}|^2 \tag{4}$$

A second example of an algorithm which aims to perform seizure classification from EEGs is discussed in [10]. The dataset in this case is made up of EEG samples where a single sample has two EEG signals, those being signal x and signal y, where the two signals are recorded from adjacent EEG signals [10]. In this implementation, signal decomposition in the form of EMD is again utilized to generate IMFs from EEG signals, in this case being limited to 10 IMFs instead of the 6 from the previous algorithm [3], [10]. Once the IMFs have been generated, the Hilbert transform is performed on the IMFs to generate the analytic signal of the various IMFs [10]. From the analytic signal representation, the instantaneous frequency can be calculated from each analytic signal and is shown in equation 5, where $\phi(t)$ is the instantaneous phase from the analytic signal [10]. The first feature used in this machine learning algorithm is the average variance of instantaneous frequencies (AVIF) of an IMF, the average being the average between the x and y of the signal pair [10]. The second feature calculated for this algorithm was average sample entropy (ASE), which is a modification of approximate entropy and is defined as the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar for the next given point, where self matches are omitted [10]. These two features were used in a least-square support vector machine (LS-SVM) and is expressed mathematically in equation 6, where $k(x, x_i)$ is the kernel function [10]. Three kernels were tested in the case of the machine learning algorithm, those being a linear kernel, a polynomial kernel, and finally a radial basis function (RBF) kernel [10]. Of particular note for this paper was the fact that the third IMF was found to be statistically the most useful across both features, and the five features of ASE for IMF3, IMF5, and IMF6, along with AVIF for IMF1 and IMF3 were applied to the classifier [10].

$$\omega(t) = \frac{d\phi(t)}{dt} \tag{5}$$

$$f(x) = \text{sign}[\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b] \tag{6}$$

B. MATLAB Implementation

In terms of MATLAB implementation, the major goal of the implementation is to examine how the accuracy of a machine learning model is affected by factors like the size of the signal being utilized, the features being extracted from said signal, and finally the type of classifier being used. The dataset utilized in this MATLAB implementation was the CHB-MIT Scalp EEG database available from PhysioNet, which is utilized in many other seizure detection machine learning procedures [3], [11]. This dataset contains EEGs recorded from 22 subjects, 5 being male and 17 being female using the standard 10-20 EEG electrode positions and most individual signals were recorded over the course of an hour and the signal sampling rate was 256 Hz [11]. In the case of some recordings, seizures were observed over the course of the recording session and were labeled [11]. In terms of the signals chosen: 10 seizure signals were utilized and 10 normal signals were utilized from 10 of the patients from the total pool.

The first step to the MATLAB implementation was the extraction of the seizures from the overall larger EEG recording, this was done by using the label available from the dataset, after which the first 30 seconds of that seizure was clipped from the larger seizure signal. The signals which did not have seizures had their first 30 seconds utilized, so as to have some consistency in signal length and to assist in the following signal segmentation step. The next step was the segmentation of the signal which was done in two ways to examine how they impacted the final result. Firstly, signals were segmented into 3 second parts, which resulted in a 30 second signal being split into 10 even portions, which would allow for the examination of how a short window would affect machine learning. The second set of segmentation was performed at a 10 second window, which split each of the signals into 3 segments of equal size.

After the segmentation step, the next step was the feature extraction techniques. This was split into two main techniques, the first being made up of primarily time domain and frequency domain while the second was made up of features gathered from signal decomposition techniques and the joint time-frequency domain. The purpose behind these features being extracted was to test how beneficial signal decomposition and joint time-frequency analysis was to machine learning for EEGs. Signal decomposition techniques and joint time-frequency analysis techniques like Wavelet decomposition, spectrograms, empirical mode decomposition, and the hilbert-huang transformation, are understood to better deal with nonlinear and nonstationary signals like EEGs and are capable of extracting data that would otherwise be obfuscated in other signal processing techniques, and the expected outcome would be that these methods would provide better features overall than other simpler feature extraction methods.

The first set of features examined were the signal mean, the zero-crossing rate, the variance of the EEG signal segment, the rms of the signal, and finally the mean spectral entropy of the signal. These features are primarily time and frequency domain based and many of them were utilized in other machine learning frameworks for EEG analysis [12], [13]. The second set of features were extracted utilizing the EMD signal decomposition technique as well as the Hilbert-Huang Transform (HHT) for the gathering of joint time-frequency features. The first feature extracted was the mean value of all IMFs from a particular signal segment, this was calculated following the methodology outlined in equation 1 [3]. The second feature was also discussed in [3] and is the overall energy of the IMFs and was calculated as described in equation 2. The next two features those being mean instantaneous frequency (mIF) and variance of instantaneous frequency (vIF) were gathered from the analytic signal of an IMF, in the case of this implementation, through trial and error using only IMF 1 for these two features produced the best classification result and was the IMF utilized. Instantaneous frequency was calculated in MATLAB through the use of the in-built function hht(), which calculates the hilbert transform and spectrum as well as the instantaneous frequency which then had the mean and variance MATLAB functions performed on it, to generate the mIF and vIF respectively. The vIF feature was chosen due to its performance in [10], while the mIF feature was utilized in [14] within the context of ECG processing for atrial fibrillation. The final feature that was utilized in this implementation was the area of the central tendency measure described in [7]. In this paper, a feature was described that utilized EMD to gather a feature which could differentiate between seizure and normal EEGs [7]. The EEG signals were subjected to the EMD process where their IMFs were calculated [7]. These IMFs were then subjected to the hilbert transform in order to generate an analytic signal of the given IMFs and original signal [7]. An analytic signal representation plot could then be generated from the analytic signal by plotting the real values of the signal against the imaginary values, generating a plot for the IMFs that had a unique starting point and followed a circular motion around that point in a distinct direction of rotation [7]. A central tendency measure (CTM) was calculated for this representation, where a radius around the origin was chosen and the number of points on the representation that fall within the radius is the CTM value and its mathematical equation is given in equation 7 [7]. Points are chosen to be within the radius of the CTM based on equation 8 [7]. Once a CTM value of 95% is reached with a particular radius, that radius is utilized to calculate the area of the representation and is the final feature extracted for the second feature set [7].

$$CTM = \frac{\sum_{n=1}^{N} \partial(d_n)}{N} \tag{7}$$

$$\partial(d_n) = \begin{cases} 1 & (|Rz[n]|^2 + |Iz[n]|^2)^{0.5} < r \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Finally, after features were extracted, the process of training and testing the data was performed utilizing different machine learning classifiers. This was made simple in MATLAB by the use of the MATLAB Classification Learner which is a tool that allows for the simple training and testing of a machine learning framework through the use of a graphical user interface [15]. Of the many classifier options available, the two that were selected were a Linear Support Vector Machine and a Logistic Regression classifier. The purpose of utilizing two options here is it would allow for the opportunity to allow for examination of features across multiple classifiers, in case one set of features happened to perform better with a specific classifier and would test to see which classifier worked best for EEG epilepsy analysis. Training and testing was performed utilizing the Classification Learner tool and a 5-fold cross validation was performed automatically with 20 percent of the data set being reserved for testing purposes. Finally, the results after training were displayed utilizing confusion matrices and a final accuracy, specificity, and sensitivity measure was reported.

III. Results and Discussion

A. Effectiveness of Proposed Algorithms

In terms of the machine learning methodologies discussed in the proposed algorithm sections, all of the methodologies had incredibly high classification accuracy. In the case of the first algorithm discussed, the results of the CNN used in addition to the external features calculated from the IMFs of EEG signals, resulted in a classification accuracy of 99.8% with the Bern-Barcelona data set and an accuracy of 99.7% with the CHB-MIT dataset, which is the same as the one utilized in the MATLAB implementation [3]. These results are incredibly good as the use of the unique deep learning CNN framework has shown to be nearly perfect in classification of epileptic seizures and is the highest of all other methods that were examined here. In addition to being able to nearly perfectly classify seizures, the method also had a 99.2% diagnosis rate for severe epilepsy and a 99.6% diagnosis rate for mild epilepsy cases [3]. Another important factor

TABLE I
OTHER RESULTS

Algorithm	Data Set	Accuracy
EMD - Deep Learning [3]	CHB-MIT	99.7%
EMD - LS-SVM [10]	UPF.EDU	85%
DPE - SVM [16]	UPF.EDU	84%
Exponential Entropy - SVM [13]	Bern-Barcelona	89%
Exponential Entropy - SVM [13]	Ralph Andrzejak	99.5%
Log Energy Entropy, Energy - SVM [17]	Bern-Barcelona	88.14%

to this implementation is the fact that it was designed with the purpose of being able to perform classification accurately across datasets, having only been trained with the Bern-Barcelona set but still achieving a near perfect score with the CHB-MIT set [3]. This ability to perform well outside of the original training dataset is incredibly important as it demonstrates that the method may be very successful in aiding a clinician to make a diagnosis decision in regards to whether a patient is exhibiting epileptic seizures. This example clearly demonstrates the great ability that deep learning techniques have over more traditional machine learning techniques like the one implemented here and in the other papers that were examined.

The results for the machine learning methodology discussed in [10] that utilized EMD to generate ASE and AVIF from IMFs were also very good. In the case of these results, IMF 3 was taken to be the overall most significant IMF in terms of distinguishing between seizures and nonseizure signals, with IMFs 3, 5, and 6 being used to calculate ASE features and IMF 1 and IMF 3 being used to calculate AVIF features that were utilized in a LS-SVM classifier [10]. This paper also examined the difference between the kernels available in a LS-SVM system [10]. In terms of the three kernels examined, it was determined that a RBF kernel was the superior kernel compared to a linear kernel which in turn, outperformed the second order polynomial kernel [10]. In the case of the results of the optimal classifier, this algorithm managed to achieve a classification accuracy of 85% [10]. While lower than the deep learning architecture described in [3], the overall results of the AVIF and ASE method was incredibly good and is good enough to be a potential help to a clinician in the diagnosis of epilepsy. Finally, the result of some other algorithms are displayed in Table 1 and these will form the basis of comparison for the MATLAB implementation results.

B. MATLAB Implementation Results

When examining the results of the MATLAB implementation overall, the MATLAB implementation can be considered to be a success. Table II demonstrates the overall accuracy, specificity, and sensitivity of the various classifiers, segment sizes, and feature sets for the validation of the models done with 80% of the data, while Table III shows the same results but for the testing performed with the remaining 20% of the data. In terms of overall accuracy, the results were fairly good with test accuracy ranging from 83.3% to 100%, sensitivity ranging from 66.7% to 100%, and specificity ranging from 95% to 100%. In addition to these results, confusion matrices for the test results are displayed in Figure 1 through Figure 4, with the confusion matrices corresponding to the perfect tests and 10 second segments being omitted.

While these results were very good, there are some caveats to the overall methodology that are important to note, especially if one were to compare these results to the results of the proposed ML algorithms discussed above. Firstly, the overall size of the data is particularly small with the entire set being made up of 20 EEG signals of 30 seconds each. Many of the machine learning methodologies discussed in literature use upwards of thousands of samples in order to train and test their machine learning algorithms [3]. This small amount of data being used means that the overall accuracy of this implementation would be higher than a larger data set. This is clearly seen when examining the results from the 10 second segments to the 3 second segments overall. The 10 second segments result in 60 total data points to the training and test set of features, while the 3 second segments leads to 200 total data points and what this means for testing is that the 10 second segments are only testing 12 samples, which is probably why it achieves an accuracy of 100% in all but one test. Another factor that likely skews the results is the fact that most studies including those who use the same dataset, utilized the entire group of patients, while this implementation only makes use of the EEGs from 10 patients.

In terms of the expectation before the results for the implementation were calculated, it was generally expected that feature set 2, which combines EMD and HHT in order to generate features which should not be affected by nonstationarity, would perform better overall. This was expected to be particularly noticeable in the 10 second segments, where one would expect non stationary components to be more prevalent compared to the shorter 3 second segments, due to the fact that over time one would expect more deviation in statistical characteristics. In the case of the test data it was found that feature set 2 outperformed feature set 1 in 2 of the 4 cases, performed the same in 1 of the four cases, and performed worse in one case. In the case when feature set 2 performed worse, it was when utilizing a 10 second segment with SVM classifier, it should be noted however that this is the only of the three 10 second segment tests where the test accuracy performed under 100%, while not a large sample size of tests, due to the small amount of total data points being tested, this could be an outlier where running a test again could yield a 100% result. Despite feature set 2 performing better than feature set 1, this again may be the result of the overall low testing data size, rather than the superiority of the second feature set at classifying longer signal segments. When examining the validation accuracy for 3 second segments and 10 second segments, the relationship of feature set 2 performing better with longer signals holds true with feature set 2 outperforming feature set 1 with 10 second samples with accuracies of 81.3% and 85.4% for SVM and LR classifiers using feature set 1, with the same classifiers increasing to 83.3% and 89.6% respectively when using feature set 2.

TABLE II
VALIDATION MATLAB RESULTS

Features	Seg.	Classifier	Acc.	Sens.	Spec.
1	3	SVM	86.9%	85.0%	88.8%
1	10	SVM	81.3%	75.0%	87.5%
2	3	SVM	85.0%	80.0%	90.0%
2	10	SVM	83.3%	75.0%	91.7%
1	3	LR	86.9%	86.3%	87.5%
1	10	LR	85.4%	83.3%	87.5%
2	3	LR	85.6%	80.0%	91.3%
2	10	LR	89.6%	83.3%	95.8%

Another expectation was that feature set 1 would perform better in the 3 second segments compared to the 10 second segments because the features of the first feature set are time and frequency features not designed to deal with nonlinearity or nonstationarity in particular. This was not the case when looking at the test data, probably for the small data in the 10 second test set, but accuracy was higher for the 3 second segments when examining the validation accuracy for both LR and SVM. Another factor to consider in results is that in terms of specificity and sensitivity, the specificity is in general higher than the sensitivity for both classifiers, both feature sets, as well as for both segmentation times.

Finally, when examining the two classifiers against one another, the performance between the two was generally fairly similar. In terms of overall performance however, the LR classifier outperformed the linear SVM in 3 of the 4 cases and performed equally in the 10 second test using feature set 1. Of particular note are the fact that the LR classifier outperformed the SVM in both of the 3 second tests, which is likely the more significant result when compared to the 10 second segment tests due to the larger testing size. The LR classifier also outperformed the SVM in validation accuracy as well in all four of the cases. From these results, it could be concluded that, at least in terms of this dataset, LR performance is superior to SVM performance.

When examining where the implementation could be improved, the most important factor would be expanding the number of signals utilized, both in regards to the size of the data as well as incorporating more patients from the total pool and ensuring that one patient is not over represented in the data, as this likely acted as another factor that skewed the results. When comparing the results of the implementation to the results from other machine learning implementations in literature, the results were fairly comparable, however fair comparisons between the results are complicated by the fact that some results utilized different data sets and on average the number of signals utilized were much larger.

C. N's and V's of Biomedical Signals

An important factor to any biomedical signal processing application is the 4 N's and V's of biomedical signals. Those N's and V's are: non-stationary, non-gaussian, non-linear, non-shortterm, variety, velocity, veracity, and volume [18]. These properties will be examined in the context of the EEG signals utilized in the MATLAB implementation discussed above.

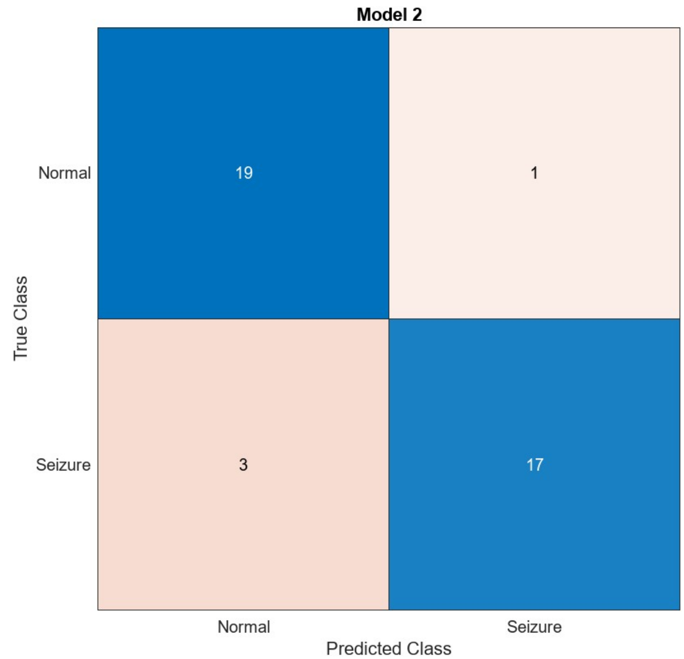

Fig. 1. Test Confusion Matrix of Feature Set 1 with 3 Second Segmentation and LR Classifier

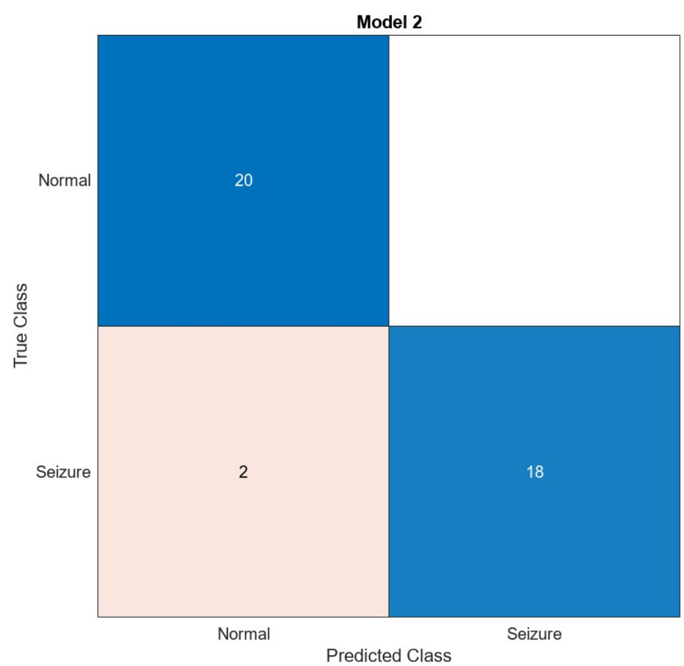

Fig. 2. Test Confusion Matrix of Feature Set 2 with 3 Second Segmentation and LR Classifier

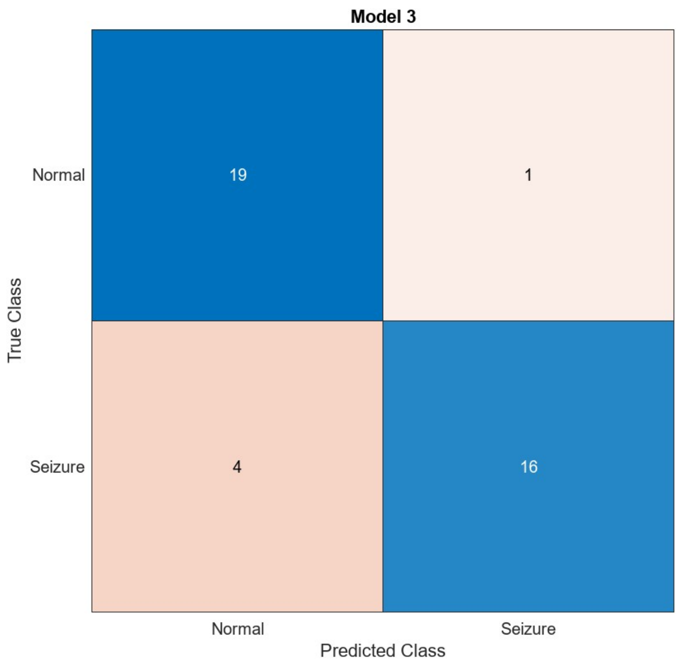

Fig. 3. Test Confusion Matrix of Feature Set 1 with 3 Second Segmentation and Linear SVM Classifier

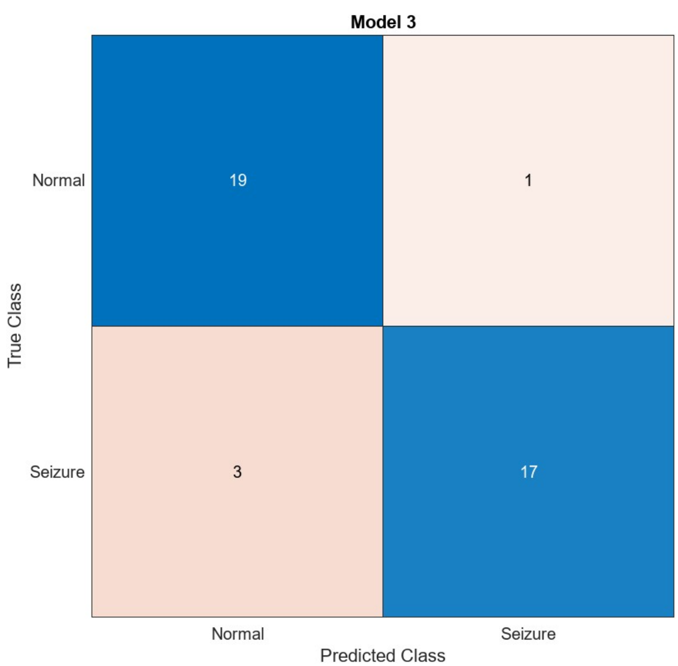

Fig. 4. Test Confusion Matrix of Feature Set 2 with 3 Second Segmentation and Linear SVM Classifier

TABLE III
TESTING MATLAB RESULTS

Features	Seg.	Classifier	Acc.	Sens.	Spec.
1	3	SVM	87.5%	80.0%	95.0%
1	10	SVM	100.0%	100.0%	100.0%
2	3	SVM	90.0%	85.0%	95.0%
2	10	SVM	83.3%	66.7%	100%
1	3	LR	90.0%	85.0%	95.0%
1	10	LR	100.0%	100.0%	100.0%
2	3	LR	95.0%	90.0%	100.0%
2	10	LR	100.0%	100.0%	100.0%

Firstly, two important properties are the non-shortterm and volume properties of signals. In terms of the length and volume of data utilized to perform the MATLAB implementation, the data could be considered to be relatively short term at 30 seconds per signal and 7961 samples a signal. The volume and short term data property comes into play primarily in the segmentation of the signal where the smaller volume and more short term segments of 3 seconds performed better in terms of the classification of seizure signals compared to the longer term 10 second segmented signals in the case of the first feature set, when examining the validation results. There could be a variety of reasons for this but in terms of the first feature set, the shorter segments with less volume of data means that there is likely to be less of a non-stationary factor to the data as compared to the larger data, thereby leading to better results specifically for the first feature set which is less tailored to the analysis of nonstationary and nonlinear data. One could also assume that from the results, the use of the entire 30 second signal undergoing feature extraction may lead to more of an accuracy loss. This relationship was reversed for the results with the second feature set, when examining the LR classifier, where the added non-stationarity from the increased length may not have been as much of a factor.

In terms of the nonstationary, nonlinear and variety properties of the EEG signals, the EEG signal could be considered to be nonstationary, nonlinear and have variety as the statistical characteristics of the signal changes over time. This is especially true in the case of seizures, as in the context of the full one hour EEG recordings, the seizures could be considered a portion where the EEG statistical characteristics change somewhat drastically from the normal. In addition, as the EEG reflects brain activity, many factors can affect the overall signal like emotional state or even something as insignificant as eye movement and as such the EEG statistical characteristics will change easily over time. This nonstationary nature of the data is important as it strongly affects the efficacy of the analysis techniques utilized. In the case of the different signal lengths when utilizing an LR classifier, the second feature set increased its performance in validation at higher segment size while the validation accuracy decreased as segment size increased for the first feature set. This is likely due to the fact that the second feature set utilized techniques like EMD and HHT that are adept at analyzing nonstationary and nonlinear signals, like the signals utilized in the MATLAB implementation. As explained above the nonstationary principle is also tied to the length of the signal as there is a larger chance that a shorter signal's statistical characteristics stay consistent and therefore stationary compared to a longer biomedical signal which has more chance of having non-stationary components.

Another particularly interesting signal property to this case would be veracity. In the context of this property, veracity of the machine learning procedures can be considered to be particularly good as when comparing the feature extraction methods, classifiers used, and segmentation, the best result was an overall classification accuracy of 95% with three second segments, Which while lower than the almost perfect classification of seizures discussed in [3], is still a fairly good result.

EEGs can be considered to be non-gaussian in their distribution like most other biomedical signals. Finally, in terms of velocity, the algorithm utilized in this case to classify EEG signals can be considered to be processing a large number of samples over a relatively short time and could be considered high velocity.

IV. Conclusion

When taking both the results obtained from the MATLAB implementation as well as the results gathered from other machine learning models, one can see that machine learning techniques can prove useful in assisting clinicians in isolating seizures from an EEG signal. In addition, there are potential expansions to the machine learning frameworks where classification of seizures can be performed as well as classification of seizure severity or even potentially the classification of various types of seizures from one another.

References

[1] B. F. Shneker and N. B. Fountain, "Epilepsy," *Disease-a-Month*, vol. 49, no. 7, p. 426–478, 2003.

[2] W. Klonowski, "Everything you wanted to ask about eeg but were afraid to get the right answer," *Nonlinear Biomedical Physics*, vol. 3, no. 1, 2009.

[3] R. Srinath and R. Gayathri, "Epilepsy disorder detection and diagnosis using empirical mode decomposition and deep learning architecture," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 11, 2022.

[4] F. Martini and E. F. Bartholomew, *Essentials of Anatomy amp; Physiology*. Pearson, 2021.

[5] S. Krishnan, "Biomedical signals lecture," 2022.

[6] K. Sridhar, *Biomedical Signal Analysis for Connected Healthcare.* Academic Press, 2021. [Online]. Available: http://ezproxy.lib.ryerson.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1951472&site=ehost-live

[7] R. B. Pachori and V. Bajaj, "Analysis of normal and epileptic seizure eeg signals using empirical mode decomposition," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, p. 373–381, 2011.

[8] S.-H. Lee, J. S. Lim, J.-K. Kim, J. Yang, and Y. Lee, "Classification of normal and epileptic seizure eeg signals using wavelet transform, phase-space reconstruction, and euclidean distance," *Computer Methods and Programs in Biomedicine*, vol. 116, no. 1, p. 10–25, 2014.

[9] S. Krishnan, "Machine learning models/algorithms," 2022.

[10] R. Sharma, R. B. Pachori, and S. Gautam, "Empirical mode decomposition based classification of focal and non-focal eeg signals," *2014 International Conference on Medical Biometrics*, 2014.

[11] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, 2009.

[12] T. Staudinger and R. Polikar, "Analysis of complexity based eeg features for the diagnosis of alzheimer's disease," *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011.

[13] F. O.K. and R. R., "Time-domain exponential energy for epileptic eeg signal classification," *Neuroscience Letters*, vol. 694, p. 1–8, 2019.

[14] Z. Huang, Y. Chen, and M. Pan, "Time-frequency characterization of atrial fibrillation from surface ecg based on hilbert-huang transform," *Journal of Medical Engineering amp; Technology*, vol. 31, no. 5, p. 381–389, 2007.

[15] "Classification learner." [Online]. Available: https://www.mathworks.com/help/stats/classificationlearner-app.html

[16] G. Zhu, Y. Li, P. P. Wen, S. Wang, and M. Xi, "Epileptogenic focus detection in intracranial eeg based on delay permutation entropy," *AIP Conference Proceedings*, 2013.

[17] O. K. Fasil, R. Rajesh, and T. M. Thasleema, "Influence of differential features in focal and non-focal eeg signal classification," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 2017.

[18] S. Krishnan, "Digital healthcare and ai lecture," 2022.