Predictive Analytics in Healthcare Epileptic Seizure Recognition*

Ashok Bhowmick†
The G. Raymond Chang School
Ryerson University
Toronto, Ontario
ashok.bhowmick@ryerson.ca

Tamer Abdou‡
Arish University
Ryerson University
Toronto, Ontario
tamer.abdou@ryerson.ca

Ayse Bener§
Data Science Laboratory
Ryerson University
Toronto, Canada
ayse.bener@ryerson.ca

ABSTRACT

Introduction Clinical applications of electroencephalography (EEG) span a very broad range of diagnostic conditions. Epileptic seizure is the fourth most common neurological disorder in that. **Related Work** There has been considerable progress in clinical understanding of epilepsy, however many aspects of seizure prevention are still a mystery. Predictive modeling of EEG can provide significant value addition to substantiate the diagnosis of epilepsy. **Methodology** Machine learning algorithms are applied to predict the probability of epileptic seizure using an open source multi-class dataset. **Results and Discussion** Comparing the F-score from different classifiers, it is found that XGBoost gives the best performance in binary classification and Random Forest provides the best performance in multinomial classification. **Conclusion** Our results show that it is possible to predict epileptic seizure with significant accuracy from non-epileptic parameters using a suitable machine learning algorithm. We also observe that binary classification methods have higher prediction accuracy.

CCS CONCEPTS

• **Computing methodologies** → **Classification and regression trees**; *Feature selection*; *Cross-validation*; Bagging; • **Applied computing** → **Health care information systems**;

KEYWORDS

Predictive analytics; Healthcare; Epileptic Seizure; Classification; Cross-Validation

ACM Reference Format:
Ashok Bhowmick, Tamer Abdou, and Ayse Bener. 2018. Predictive Analytics in Healthcare Epileptic Seizure Recognition. In *Proceedings of ACM Conference (Toronto'18).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.475/123_4

Toronto'18, CASCON 2018, Toronto, Ontario, Canada 2018. ACM ISBN 123-4567-24-567/08/06…$15.00 https://doi.org/10.475/123_4

1 INTRODUCTION

Understanding and recognizing epileptic seizure is a difficult task in mental healthcare domain [17][21]. In recent years, predictive analytic has shown that data driven approaches can provide better and smarter prediction of healthcare results, by integrating clinical data sets with machine learning techniques, as a unified system [12].

In this study, we employ machine learning techniques to gain insights about detecting epileptic seizure [1]. Our focus is on analyzing the significant factors that are affecting the seizure detection, as inscribed in each set of Electroencephalography (EEG) data that has various conditions in the actual measurement. Predictive analytics would be a useful approach since the occurrence of epileptic seizure is highly unpredictable and it is very difficult to probe the brain in real time in an outpatient setting. A predictive model that uses the measurements of symptomatic evidences, therefore, may be a useful tool to effectively diagnose such cases. We employed a supervised learning algorithm to predict the probability of seizure cases. Although there are studies that use conditional methods to probe the brain in conventional electroencephalogram, our proposed approach shows how different types of probing contribute towards the diagnosis of epilepsy by creating different classes in an EEG.

The outcome of this study may help practitioners in measuring the most relevant measurements to predict epileptic seizure and reduce the cost of diagnostic investigation. Our research question in this study is: How to predict epileptic seizure from non-epileptic parameters?

The remainder of the paper is organized as follows: Section 2 summarizes the background of the problem and related work. Then, we introduce and describe the details of overall methodology in Section 3. In Section 4, the results are presented along with a discussion on our findings. Threats to validity are addressed in Section 5 followed by a conclusion in Section 6.

2 RELATED WORK

Electroencephalography (EEG) is a technique that is widely used to measure the functional activity of human brain. It is a noninvasive electrophysiological monitoring method that records the electrical activity of the brain by placing electrodes along the scalp. It measures the brain activity as received at the surface of the brain in terms of voltage fluctuations resulting from ionic currents within the neurons of the brain. It is a graphic display of the difference in

[1]The work in this paper is conducted in the Capstone Project Course of the Certificate in Data Analytics, Big Data, and Predictive Analytics at Ryerson University.

voltages from two sides of brain recorded over time, representing a proportional measure of activity. Hence, EEG data is displayed as a continuous time series waveform of very small voltage signals, normally of the magnitude of a microvolt. Though the waveform are recorded at the surface of the brain (cortex), it is influenced by the activities that are deep underneath the cortex [6][17][18][26].

The recording of EEG data is done by placing small metal plates as electrodes at specially designated places on the scalp. Voltage signals are measured with respect to corresponding reference electrodes by attaching electrodes individually or as preconfigured cap [10][23]. The measured voltages at a given time can either be positive or negative. In an analog machine, the signal is recorded as continuous pattern and in group, they are referred as the brain waves. Modern machines digitize the signal using an ADC which can be stored, displayed and manipulated using a computer. The rate at which the signal is digitized is called the sampling rate, which typically ranges between 50 to 300 Hz [6][18][22].

Electrical signals are created when electrical charges move within the central nervous system. Neural function is normally maintained by the ionic gradients established by neuronal membranes. Sufficient duration and length of small amounts (in microvolts) of electrical currents of cerebral activity gives rise to an electric potential. A diffusion membrane potential normally exists through an efflux of potassium ions maintaining an electrochemical equilibrium. This cerebral potential, based upon the intrinsic electrophysiological properties of the nervous system, is recorded at the time of measurement [4][26]. The signal is commonly sinusoidal where peak to peak magnitude varies between 0.5 to 100 microvolts (in healthy brains). The frequency of the signal may typically vary between 0 to 150 Hz. The signals are normally collected instructing the subjects to keep their eyes closed and be in relaxed situation. The waves are categorized into four basic groups viz. 1) Delta: 0.5 to 4 Hz, 2) Theta: 4 to 8 Hz, 3) Alfa: 8 to 13 Hz and 4) Beta: all greater than 13 Hz [6][23].

Interpretation of EEG signal for diagnostics is a complex task that might engage through several decision making steps [17]. Similar pattern can be observed for many different reasons. It is also possible that diverse patterns might point to the same symptom. The signals depend on many different factors and they are unique to individuals. Alpha activity is induced by closing the eyes and relaxation while it is abolished by eye opening or alerting by any mechanism like, thinking or calculating. EEG signal consists of many waves with different characteristics. To deal with this complexity, digitized EEG data is important for critical analysis and interpretation. An abnormal EEG pattern may lead to diagnosis, or it indicates a cerebral dysfunction or it can be something else as well [6][18][23].

Clinical applications of EEG span a very broad range of diagnostics conditions, and epilepsy is the most common one [23]. Epilepsy is a chronic disorder and is distinguished by recurrent unprovoked seizures [25]. A person is diagnosed with epilepsy if they have two unprovoked seizures that were not caused by some known and reversible medical condition like alcohol withdrawal or extremely low blood sugar. The seizures in epilepsy may be related to a brain injury or a family history, but often the cause is completely unknown. The word "epilepsy" does not indicate anything about the cause of the person's seizures or their severity. Many people with epilepsy have more than one type of seizure and may have other symptoms of neurological disorders as well [14][21].

An unusual EEG pattern may or may not be indicative of epileptic seizure because even for the diagnosed epileptic patient, brain shows normal behavior if there is no seizure. On the other hand, unusual activity might also be observed all the time in some patients that could be due to several reasons like, tumor growth etc. [14]. Recording of an *ictal* discharge i.e. EEG pattern when seizure occurs, is often time consuming and laborious in an outpatient setting. *Interictal* discharges, EEG pattern when there is no seizure, could therefore be relevant for analytic to determine the conditions and onset of seizure [9].

The primary motif of this study, as mentioned in the research question, is to learn the hidden patterns within a clinically analyzed past data and explore the possibility of predicting epileptic seizure in an unknown data as an application of the knowledge acquired through the analytic. The structure of an used data with pre-labeled class and the relevant data dictionary could help build a model whereupon the model can learn the numerical relationship between seizure response and non-seizure response as engraved within the data structure created from real life measurements. The developed model would then be able to predict the probability of seizure cases from unknown inputs of similar class values measured in some new cases.

Machine learning (ML) could be the most appropriate tool in such supervised learning framework. Fergus et. al [12] have applied ML on a small data set of 342 records with only 2 classes. They have also tabulated a chart on various binary classification works attempted so far using various classifiers. Acharya et al focused using entropy for seizure detection and seven different classifiers [1][2]. This work is an attempt on a multiclass EEG data having five different classes with 178 attributes through 11500 records. Also, apart from all different classifiers attempted before on various data sets, to our knowledge, XGBoost classifier has been applied for the first time in this work.

3 METHODOLOGY

3.1 Data set

The data set for this project is taken from UCI Machine Learning repository [2]. It is a multivariate time series data having 178 features against 1 categorical attribute through 11500 instances. Each number represents a voltage signal corresponding to the brain activity at a point of time. The original data is composed of five sets denoted by A, B, C, D, E each containing 100 single channel EEG segments of 23.6 seconds duration. Sets A and B consisted of segments taken from EEG recordings on five healthy volunteers using a standardized electrode placement scheme. Volunteers were relaxed in an awake state with eyes open which is denoted as data - A and eyes closed, denoted as data - B, respectively. Sets C, D, and E originated from an EEG archive of presurgical diagnosis. The archive data is selected from five patients all of whom achieved complete seizure control after section and their epileptic zone were correctly diagnosed. The data-D was recorded from within the epileptic zone, whereas data-C was collected from the hippocampal formation of the opposite hemisphere of the brain [3].

[2] http://archive.ics.uci.edu/ml/index.php

The data that is being dealt with here has been reorganized from the original data for machine learning studies as available in UCI Machine Learning repository [24]. Each time series record A through E has been digitized to 4097 data points out of which 23 chunks have been created and distributed into 178 time-features of recorded raw-voltage signals. 5 sets of 100 EEG segments thus creates 11500 instances. The grouping on 23 chunks corresponds to the fact that there are 23 electrodes used in an EEG recording. A class attribute or output (y), is generated as follows:

Class5: Eyes open condition from healthy volunteer, i.e. data-A, EEG is collected keeping the eyes open.
Class4: Eyes closed condition from healthy volunteer, i.e. data-B, EEG is collected keeping the eyes closed.
Class3: Tumor is identified and located but the data is collected at the nonepileptic healthy part of the brain (archive data).
Class2: Tumor is identified and data is collected from the epileptic zone of the brain.
Class1: Recording during seizure activity, i.e. the *ictal* data.

As described, the data is time stamp of recorded voltage in microvolt unit as shown in Figure 4 and Figure 5 for particular instances across 178 attributes. Exploratory examination revealed that there is no missing values in the data set.

3.2 Approach

We employed classification techniques to build a predictive model. The most common approach would be to perform a binary classification of epileptic versus nonepileptic cases, where all of 2 through 5 classes are designated as class value 0 and **Class1** as class value 1. Such analysis would generate a model for predicting the probability of seizure from the parameters of nonepileptic cases. However, it would not reveal the dependencies on individual nonepileptic class so as to distinguish the most relevant measurement that must be recommended to diagnose the epileptic cases. Therefore, the seizure prediction should also be modeled as a multinomial classification. We employed various ensemble methods to predict the probability of occurrence of epileptic seizure.

Nonetheless, it is also necessary to explore that there is actually only 5 classes inherent within the data. That would reaffirm the class description in the data set. Unsupervised learning methods can be applied to the whole data set without the given class attribute so as to reveal how the numbers group by themselves in terms of various distance criteria. Hierarchical clustering study is also of interest in this regard to explore if the data has any embedded structure. Data cleaning and subset structuring are done using R [20]. Exploratory analysis is done using Python. The coding for classification models are done using Python as well. For replication purposes, a documented implementation of our proposed algorithm has been uploaded, made public and can be found on GitHub [3].

3.3 Performance Measures

To validate the employed classifiers in our study, we used 10-fold cross validation, as it is one of the most commonly used techniques that allows for computing predictions in an accurate way and helps in reducing the bias estimation [27]. Afterward, the performance

[3]https://github.com/ashokbhow/Capstone-Project

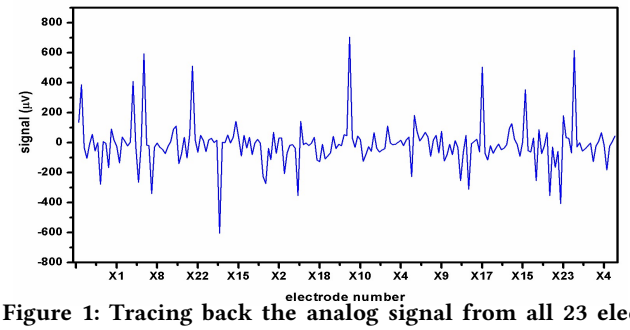

Figure 1: Tracing back the analog signal from all 23 electrodes

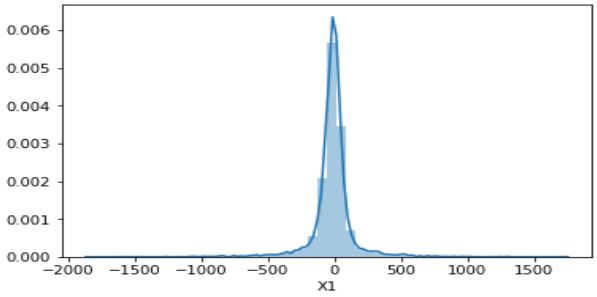

Figure 2: Histogram and distribution over attribute X1.

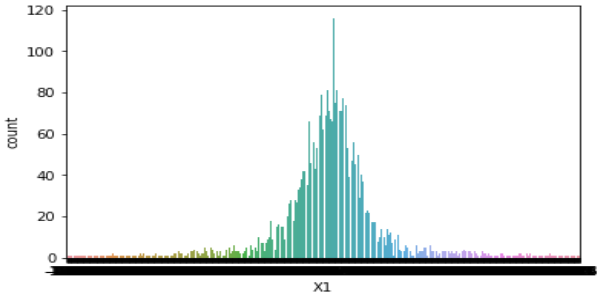

Figure 3: Distribution of voltage values by counts in attribute X1 over 11500 instances

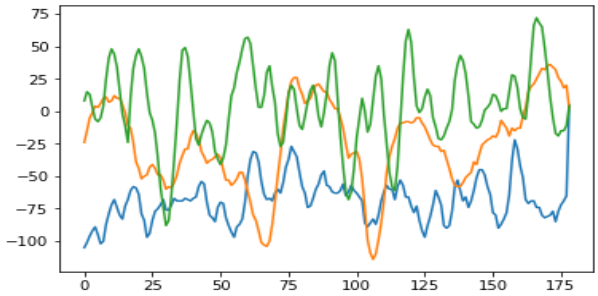

Figure 4: Variation across the features regenerating the timestamp. Class1 in blue, Class2 in orange; Class3 in green.

of the multinomial and binary classifiers has been measured using Recall and Precision, as follows [5]:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (1)$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (2)$$

Where True Positives are the items that have been correctly retrieved by the employed classifiers, False Positives are the items

that have been incorrectly classified, and False Negatives are the ones that have not been retried by the classifier. In order to have a sum-up and represent the precision and recall with one single value, we computed the harmonic mean of recall and precision as follows [5]:

$$F - measure = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \quad (3)$$

Confusion matrices have been generated for all classifiers, in order to compute the Precision, Recall and F-measure values. Table 1 shows an example of the confusion matrix that is generated for XGBoost classifier.

Table 1: Confusion Matrix for XGBoost Classifier

		True diagnosis		
		Positive	Negative	Total
XGBoost	Positive	2976	29	3005
	Negative	43	747	790
	Total	3019	776	3795

In addition, ROC plots have been generated, as shown in Figure 15. These plots show the relationship between the true positive and the false positive rates, and the diagonal shows the expected performance of the classifier. We used AUC measure to compute the area under the ROC, as an area of 1 presents a perfect classifier [11]. Results are shown in Figure 16.

4 RESULTS AND DISCUSSION

4.1 Exploratory Analysis

All features of the data set have been explored. This includes studying the relationship between attributes and instances with respect to their class identity. The data set has been found to be clean and having no missing values. For binary classification study between epileptic versus non-epileptic classes, the class values of **Classes** through **2** to **5** have been assigned as 0 whereas that of **Class1** as 1. Also, binary classification has been studied by choosing only **Class2** or **Class3** against **Class1** in order to compare the effectiveness of **Class2** over **Class3** data in prediction. The class reassignment has been worked out using R whereupon machine learning studies have been performed in Python. Analog EEG data is populated as a time stamp from all the 23 electrodes simultaneously. In this regard, it is interesting to examine the corresponding analog preview out of the digital records. Figure 1 shows such a sample regeneration where the first 200 instances of attribute X1 has been plotted with respect to 23 chunks (see Section 3.1) [24]. The horizontal axis shows the 23 chunks representative of 23 electrodes. The plot designates how the voltages are populated on each electrode. The data that is being analyzed, is prepared from such analog signals recorded by the detector. Salient features of the data are:

(1) There are no missing data and the distribution over any attribute is smooth, however it is non-Gaussian (Figure 2).
(2) The range of values varies with each class (Figure 2 and Figure 3).
(3) Figure 4 and Figure 5 show plots of the instances as grouped by their class identity across all the features (columns). This shows a wide variation in maximum / minimum voltages. The *ictal* data that represents the seizure activity, exhibits larger magnitude.

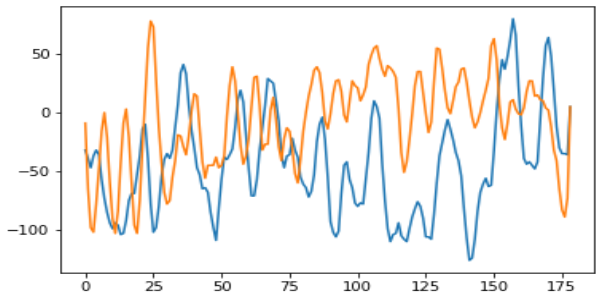

Figure 5: Similar plot for Class4 in blue and Class5 in orange

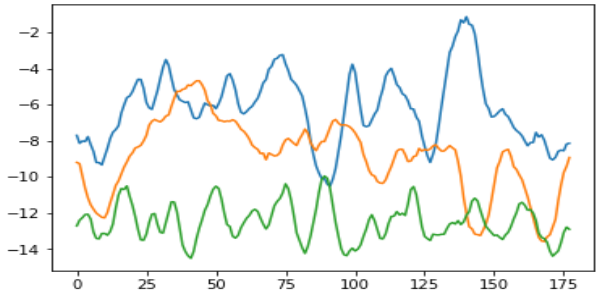

Figure 6: The time-stamp of the mean values of Class1 (green), Class2 (orange) and Class3 (blue). X-axis are the number of attributes and Y-axis shows the voltage in micro-volts.

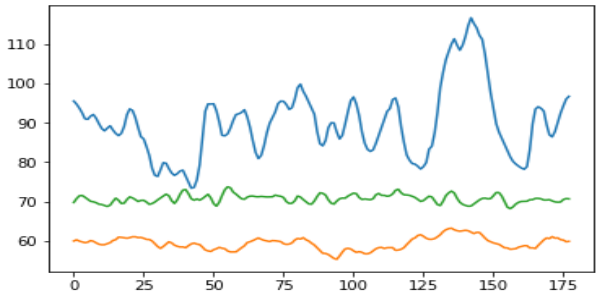

Figure 7: Standard deviation across features for Class1 (blue), Class2 (orange) and Class3 (green).

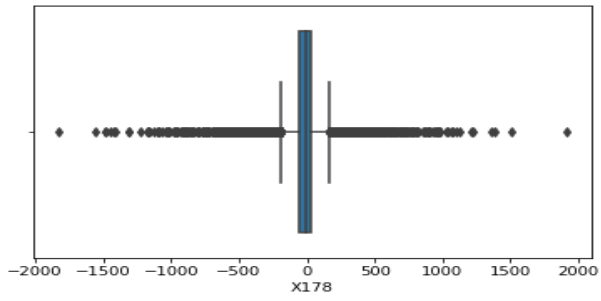

Figure 8: A box-plot of a feature showing apparent outliers.

(4) The variation of mean values by grouping the data in terms of respective class identity, are shown in Figure 6, for 1, 2 and 3 class. The standard deviation, as shown in Figure 7 by same grouping, clearly distinguishes the seizure class (**Class1**) from the rest. It fluctuates substantially over the features maxing at X139.

(5) Apparently, a box plot of any attribute (column) might look like that there are outliers. However, these are maximum/ minimum values of recorded voltages and represent relevant information to determine the degree of seizure. Besides, these data points are well within the distribution curve. Therefore, these data points are not treated as real outliers. Maximum-minimum extent in values under any class, might in fact help to determine the probable degree of damage to the cells due to intense seizure activity. Figure 8 shows a typical boxplot of feature X178 and the variations in maximum values across all the features. It is to be also noted that the maximum value apparently shows a saturation flattening at 2 milivolt, which might be due to the detection limit indicating that the seizure spikes actually shoots over the limit. Such intense electric pulses could be damaging to brain cells which could be a subject of further analytic from the data of clinical investigations.
(6) Similarly, the extent of negative surge is clearly very high in seizure class as compared to that in any non-seizure classes, as shown in Figure 9.
(7) The distribution of mean and standard deviation of **Class1** seizure (*ictal*) data on a linear scale is shown in Figure 10.
(8) There is no correlation between any two features of the data set (refer to R correlation plot: corr-plot.jpg in Github)
(9) There are 2300 instances distributed in each class. The density plot (smooth version of normalized histogram) exhibits almost equal distribution as shown in Figure 11.

4.2 Clustering

As mentioned in Section 4.2, clustering analysis has been carried out by dropping the pre-defined class attribute. Figure 12 shows a scatter plot of each class separately on the same vertical scale. This exhibits a wide extent of variation in values in each class. **Class1**, seizure data has the maximum spread that maximizes in milivolt range. Whereas, all nonepileptic classes are within 100-500 microvolt range. This order of magnitude difference has been prominent to distinguish the seizure data just by such data visualization. Only a couple of values in Class2 show to be comparable to that of **Class1**. However, the data as such, does not reveal any trend. These values are considered as true outliers and are excluded from calculation.

K-means clustering study does not reveal any natural grouping within the data from distance relationship. The data has then been subjected to hierarchical clustering analysis. As observed in Figure 12, large variation in values over three orders of magnitude required the data to be normalized for calculating a possible hierarchical dendrogram which might reveal an embedded class. Figure 13 shows the dendrogram which shows five classes at level 15 and 4 hierarchies of non epileptic classes at the same level. The distance calculation method, WARD is used in the analysis. Single-link, complete-link and average-link methods do not show good result. Nevertheless, as the work is carried out on an open data set, this calculation is done only to justify having five pre-defined classes in the data and the result do not add any input to the classification study discussed in the following.

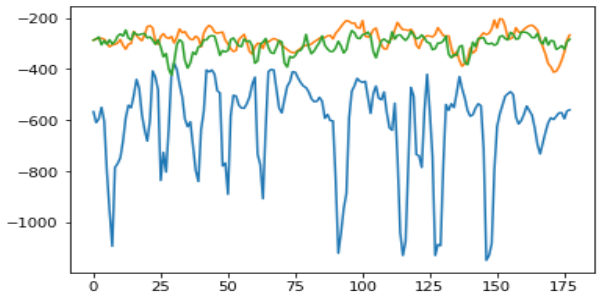

Figure 9: The -ve maximum voltage surge for Class1 (blue), Class2 (orange) and Class3 (green). The reason for + and - voltages is described in Introduction.

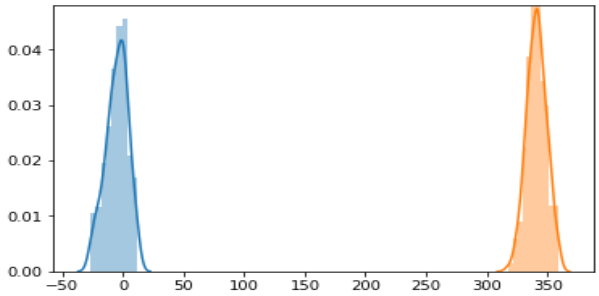

Figure 10: The mean and standard deviation of Class1 data.

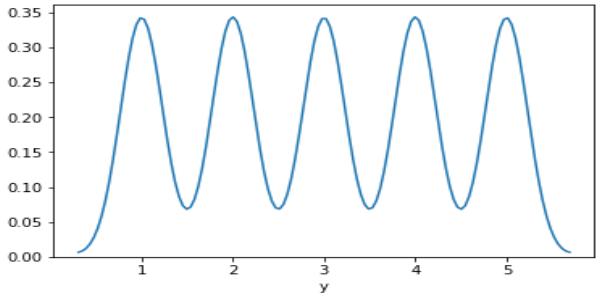

Figure 11: Distribution of instances in each class.

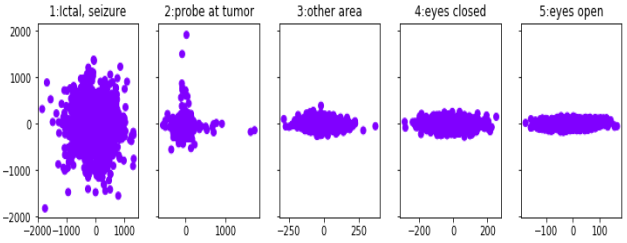

Figure 12: Scatter plots of class values exhibiting their comparative ranges.

4.3 Classification

As it is a multiclass data, not all classifiers are equally effective. The seizure class (**Class1**) is the predicting class and a binary classification can be done with respect to each one of the other non-epileptic classes. This will help to study the seizure predictability of each class individually which might render a comparative scenario of seizure predictability among classes. However, as per the research question, it is required to classify the seizure cases with respect

to all the non-epileptic classes grouped together as one class. To carry out these tasks, the data file has been split into respective separate binary class files using R. Binary classification studies are then carried out using Logistic, XGBoost, Support Vector Machine algorithms and multiclass classification is done using GLMulti [8], Decision Tree, KNN, Random Forest and Multilayer Perceptron (MLP) algorithms [19]. A selection criteria has been applied to select the best features by logistic significance values and principle component analysis (PCA) as well, whereupon the classification algorithms are applied on PCA transforms. Every classification calculations is 10 fold cross validated by creating variable validation sets. Table 2 provides some of the results of various classifications as described in further details below.

4.3.1 Binary Classification. As mentioned earlier, the most important class being **Class1**, the data set is split into two major classes viz. seizure and non-seizure. In this case, classification predicts seizure activity with respect to combined effect of all other classes. For feature selection, two different methodologies have been used. One method is doing a logistic classification of **Class1** vs. all the other classes. where all of **Class2**, **Class3**, **Class4** and **Class5** are set to 0 value and then removing the unimportant attributes by considering the significance score. Then, the logistic classification is carried out considering only the highly significant attributes. About 10 odd features are thus selected, however the accuracy score is not found to be very high.

The effective contribution of all features is then calculated by principal component analysis (PCA). From the variance ratio plot, first 50 components are retained for classification studies as the variance becomes insignificant beyond that. 10-fold cross validation is applied in splitting the data into training and test sets. The accuracy obtained are shown in Figure 14. Some F-measures are given in Table 2.

Figure 14 also shows the results of XGBoost and Support Vector Machine (SVC) classifications. The F-measure, as shown plotted at the inset in Figure 14, for XGBoost classifier is found to be the highest in this data set. For 67/33% train/test cross validation on first 50 PCA transforms, the accuracy obtained is 98.1% with a selectivity of 99.03%. Classification by support vector machine is applied on the scaled data. The F-measure of SVC is however comparable to that given by XGBoost. It is important to note that the prediction accuracy with respect to **Class3** is superior than that with **Class2** in both the classifiers. This suggests that **Class3** i.e. when the probe is placed on the healthy tissue of the brain at the opposite lobe of the tumorous zone, is a better predictor for seizure cases than **Class2** data which is obtained by probing at the tumorous zone while there is no active seizure.

The receiver operating characteristic (ROC) curves of Logistic classification, bench marking seizure at threshold probability >= 0.5, supports this conviction to a certain extent (Figure 15). Nevertheless, the trade off between benefit (TP) and cost (FP) does not exhibit very substantial difference between the influences of **Class2** and **Class3** in logistic classification. The use of higher classifiers, like XGBoost and SVC to examine the effectiveness in predicting seizure activity from non-seizure data are therefore justified. As epileptic seizure is highly unpredictable in occurrence, a good predictive model from available EEG measurements could be a significant value add in substantiating diagnosis.

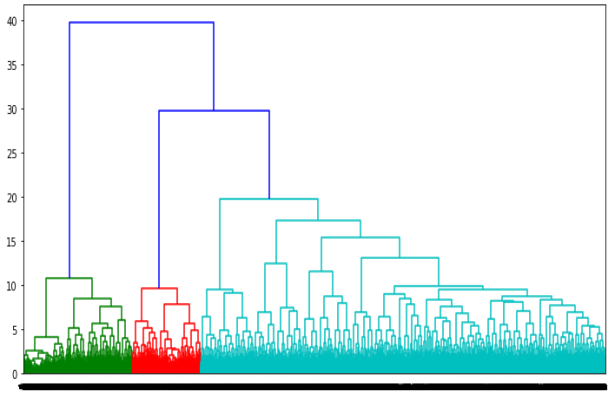

Figure 13: Hierarchical clustering using method = WARD, revealing 5 classes at level 15.

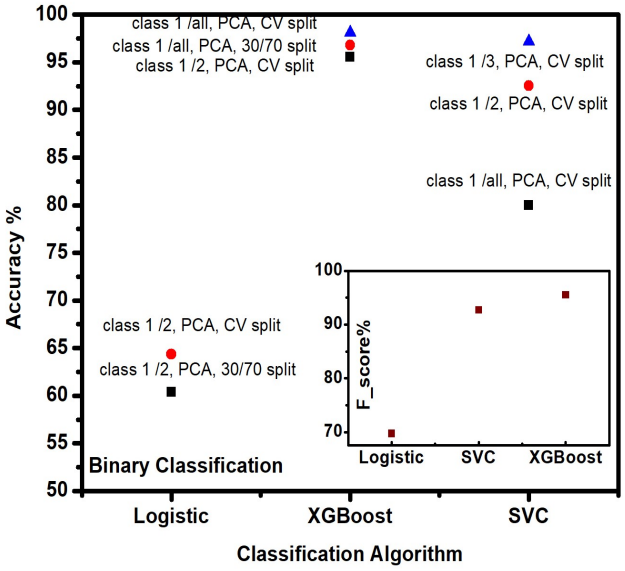

Figure 14: Accuracy ration of binary classification

Table 2: Seizure prediction results of various classifications

Classifier	Accuracy%	Precision	Recall	F-score%
Multinomial Classification				
KNN	57.2 (k=1) ; 41.4 (k=30)			
Decision Tree	54.08	0.85	0.79	81.9
Random Forest	75.69	Ensemble method		
MLP	71.3	0.94	0.939	93.9
Binary Classification				
Logistic	61.4	0.58	0.81	67.6
XGBoost	98.1	0.98	0.99	98.5
SVC	97.17	0.98	0.96	96.7

4.3.2 Multinomial Classification. Multinomial classification has been carried out using k Nearest Neighbor (KNN), Decision Tree (DT), Multi-Layer Perceptron (MLP) and Random Forest algorithms [19]. Figure 16 shows the accuracy obtained by different multinomial classifiers and the corresponding significance scores are given in

Table 2. In KNN, the cross validation accuracy decreases monotonically with the increase of k till K=30 and the highest accuracy is given for k=1 only. This reflects that there is actually no specific distance relationship that might govern the numbers in the data set. Decision Tree classifier does not produce any better accuracy either. The confusion matrix for Decision Tree model is shown in Figure 17. For seizure cases (**Class1**), the recall = 79.2% and precision = 85% for an accuracy of 54%. The ensemble method, Random Forest (RF) exhibits comparably better prediction accuracy for **Class1**, of approximately 75.7% and the average absolute error of 0.58 degrees.

Classification by Multilayer Perceptron (MLP) is done on feature scaled data. The classifier is extra sensitive to feature scaling. A deep net of 100 hidden layers with 200 nodes each (200x100), activation function = RELU and stochastic gradient descent method (solver = ADAM in Python), gives an accuracy of 71.32% with 94% precision and recall value of 93.9% for **Class1** (seizure class). Therefore, the F-measure for MLP becomes 93.9%. This shows that in multinomial classification of seizure class, Random Forest and MLP are comparable predictors.

R-GLMulti has also been tried for multinomial classification [8]. R-source code provides the details on GLMulti calculation. However, the computation takes excessively long time to converge if more than 4 features are selected together. By that, principal component scores (in R) were chosen as features by noting the AIC-scores and a best IC profile is shown is Figure 18. Final best value is provided by selecting PCA components 1, 17, 29 and 38 and the best model is given by the following equation:

$$y = 1 + Comp.17 + Comp.29 + Comp.38 + Comp.29 : Comp.1 + Comp.29 : Comp.17 + Comp.38 : Comp.17 \quad (4)$$

By this, the data is effectively reduced as 4 dimensional principal components for multinomial classification.

As a future direction, a combination of several classifiers may be worked out for better multinomial classification [15]. An analysis of entropy may also be relevant [13] in this regard. However, epileptic seizure is highly unpredictable in its occurrence and collection of *ictal* data is a major challenge. Predictive analytics on other types of seizures may also be relevant, for example, deep intoxication, hypoglycemia etc. to compare to that of epileptic seizure.

5 THREATS TO VALIDITY

Internal Validity: We believe the measures in our study are acceptable as we employed two evaluation measures: AUC and F-measure. AUC has been proven to be robust to data imbalance [7], and F-measure that can show varying performance of the classifiers [16]. **External Validity:** The data set that has been used in this study is open accessed and came from one resource [24]. To verify our results and reduce the external threats to validity, we need to apply our proposed methodology on other data sets from other resources and targeting different time frames. **Construct Validity:** To validate the performance of the proposed prediction models, we used K-fold cross validation. This validation technique has been used to estimate the accuracy of predictive models and we employed it to reduce bias in error estimation [27]. As the reorganized data in UCI Machine learning repository [24] has been already pre-labelled, no further labelling scheme is applied to the working data set.

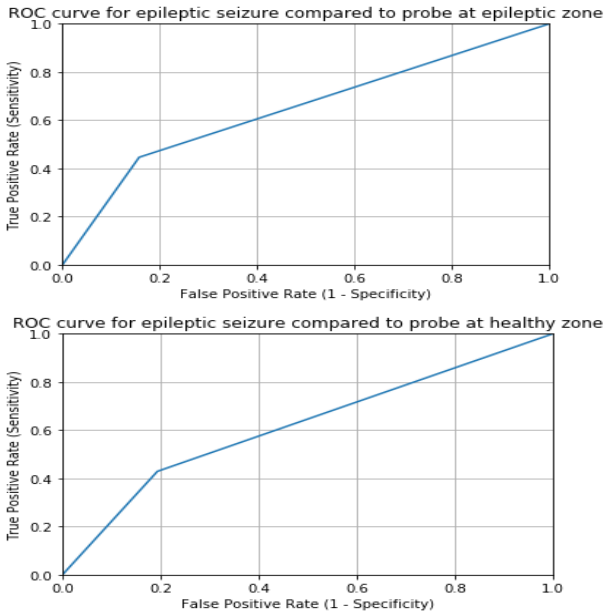

Figure 15: The ROC curves from Logistic classification of Class1 vs. Class2 and Class1 vs. Class3

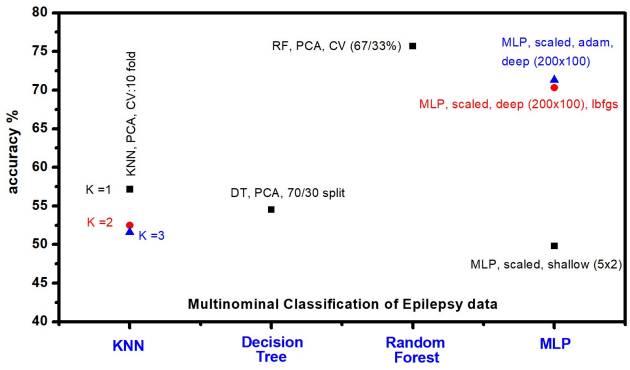

Figure 16: The accuracy of various multinomial classification on the data set.

6 CONCLUSION

The answer to the research question therefore would be that it is possible to train a predictive model for the classification of epileptic seizure from the numerical relationship of epileptic versus nonepileptic cases, in a real life data. From the performance of the model on test cases, it may also be projected that such a model can be universal giving sufficient prediction accuracy and statistical significance (F-measure). In supervised learning framework, the model can be worked out in two ways:

(1) Treating all nonepileptic class values as a single class and then doing a binary classification with respect to the seizure (or epileptic) class. The best prediction with statistical significance in such a model is given by XGBoost classifier.
(2) Doing a multiclass classification where all classes are treated as equally important in seizure prediction, both, RF and

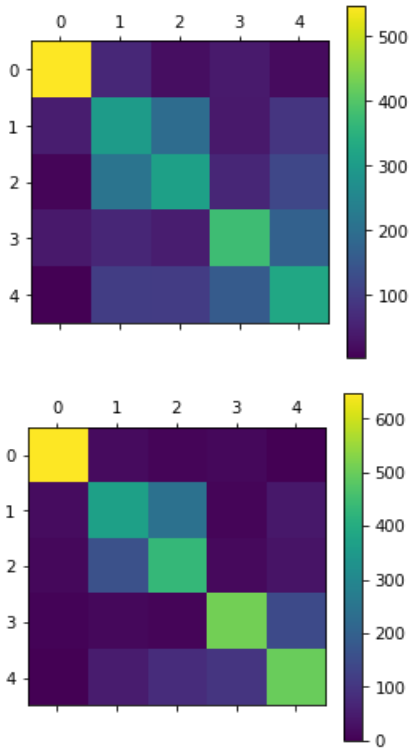

Figure 17: Confusion Matrix: Decision Tree (top) and Multi-Layer Perceptron (bottom).

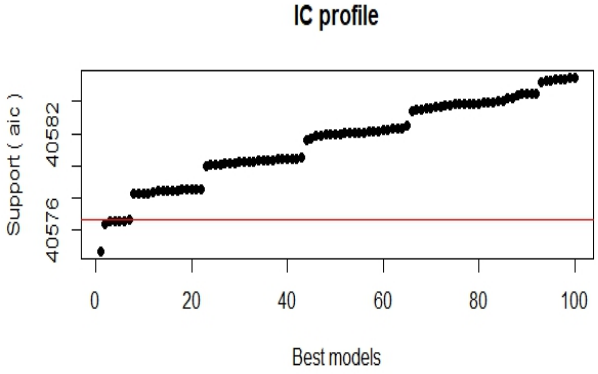

Figure 18: IC Profile for the Best Logistic Regression Model

MLP produce comparable prediction accuracy. The statistical significance of MLP is considerably higher with a deeper hidden layer structure and RELU activation function.

As an additional observation among all nonepileptic classes, it may be mentioned that **Class3**, where the data is collected by probing the healthy part of the brain opposite to the tumorous zone, is a comparative better predictor. By that assertion, **Class2** data seems somewhat redundant and this measurement may not be required. Apart from clinical reasons, probing the tumorous part of the brain might not be essential for seizure prediction, per se.

REFERENCES

[1] Acharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K. H., and Suri, J. S. Automated diagnosis of epileptic EEG using entropies. *Biomedical Signal Processing and Control 7* (2012), 401–408.

[2] Acharya, U. R., Stee, S. V., Swapna, G., Martis, R. J., and Suri, J. S. Automated EEG analysis of epilepsy: A review. *Knowledge Based Systems 45* (2013), 147–165.

[3] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. E. 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 64*, 061907 (nov 2001), 1–8.

[4] Attwood, H. L., and Mackay, W. A. *Essentials of Neurophysiology*. Mosby Inc., 1989.

[5] Baeza-Yates, R., and Ribeiro-Neto, B. Retrieval Information. In *Modern information retrieval*, second ed. Addison-Wesley, New York, USA, 1999, ch. 3, pp. 73–96.

[6] Benbadis, S. R. Abnormal Nonepileptiform EEG. In *Handbook of EEG Interpretation*, second ed. Demos Medical Publishing, New York, USA, 2014, ch. 2, pp. 51–70.

[7] Boughorbel, S., Jarray, F., and El-Anbari, M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE 12*, 6 (2017), 1–17.

[8] Calcagno, V. *glmulti: Model selection and multimodel inference made easy*, 2013. R package version 1.0.7.

[9] Chatrian, G., Bergamini, L., Dondey, M., Klass, D. W., Lennox-Buchthal, M., and Petersén, I. A Glossary of Terms Most Commonly Used by Clinical Electroencephalographers. *Electroencephalography and clinical neurophysiology 37*, 5 (1974), 538–48.

[10] Chernecky, C., and Berger, B. *Laboratory Tests and Diagnostic Procedures*, six ed. Elsevier Health Sciences, USA, 2013.

[11] D'Ambros, M., Lanza, M., and Robbes, R. Evaluating defect prediction approaches: A benchmark and an extensive comparison. *Empirical Software Engineering 17*, 4-5 (2012), 531–577.

[12] Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Khaled, A.-A., and Hamdan, H. A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics 12* (2016), 70–89.

[13] Kannathal, N., Choo, M. L., Rajendra, A., and Sadasivana, P. K. Entropies for detection of epilepsy in EEG. *Computer Methods and Programs in Biomedicine 80*, 3 (june 2005), 187–194.

[14] Kaplan, Peter W. and William O. Tatum, I. Handbook of EEG Interpretation. In *Handbook of EEG Interpretation*, second ed. Demos Medical Publishing, New York, USA, 2008, ch. 4, pp. 97–120.

[15] Lessmann, S., Baesens, B., Mues, C., and Pietsch, S. Benchmarking Classification Models for Software Defect Prediction: A Proposed Framework and Novel Findings. *IEEE Transactions on Software Engineering 34*, 4 (jul 2008), 485–496.

[16] Lipton, Z. C., Elkan, C., and Naryanaswamy, B. Optimal thresholding of classifiers to maximize F1 measure. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8725 LNAI*, PART 2 (2014), 225–239.

[17] Martinerie, J., Adam, C., Le Van Quyen, M., Baulac, M., Clemenceau, S., Renault, B., and Varela, F. J. Epileptic seizures can be anticipated by non-linear analysis. *Nature Medicine 4* (1998), 1173–1176.

[18] Niedermeyer, R., and Lopes da Silva, F. *Electroencephalography: Basic Principles, Clinical Applications, and Related Field*, fifth ed. Lippincott Williams and Wilkins, Philadelphia, USA, 2005.

[19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[20] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[21] Raga, S. V., and Wilmshurst, J. M. Epileptic spasms: Evidence for oral corticosteroids and implications for low and middle income countries. *Seizure, European Journal of Epilepsy 59* (2018), 90–98.

[22] Sanei, S., and Chambers, J. Introduction to EEG. In *EEG Signal Processing*. John Wiley & Sons Ltd,, West Sussex, England, may 2013, pp. 1–34.

[23] Teplan, M. Fundamentals of EGG measurement. *Measurement Science Review 2*, 2 (2002), 1–11.

[24] University of California Irvine. Epileptic Seizure Recognition Data Set. https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition, 2017. Last accessed: 2018-06-28.

[25] William, O. T., and Benbadis, S. R. Epileptiform Abnormalities. In *Handbook of EEG Interpretation*, second ed. Demos Medical Publishing, New York, USA, 2014, ch. 3, pp. 71–96.

[26] William O. Tatum, I. Normal EEG. In *Handbook of EEG Interpretation*, second ed. Demos Medical Publishing, New York, USA, 2008, ch. 1, pp. 1–50.

[27] Wong, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition 48*, 9 (2015), 2839–2846.