Predicting Non-Seizure Regions to Expedite EEG Analysis

Isobeye Daso
idoso@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA

Eliana Mason
elianam@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA

Grace Wu
wugrace@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA

Abstract—An electroencephalogram (EEG) is a test used to evaluate electrical activity in the brain. Neurologists use EEG data to detect abnormal brain activity that may be associated with certain brain disorders such as epilepsy, tumor and stroke. Doing so requires reviewing all hours of a patient’s EEG recording to identify and diagnose seizure activities. Since the length of the recordings can range from several hours to weeks, the demand for specialized expertise often exceeds the supply.

Our project assists neurologists in monitoring epileptic patients for seizures. We focus on identifying regions of the EEG recording that contain no seizure activity and do not need to be reviewed. Doing so will drastically decrease diagnostic variance and neurologists’ time spent on reviewing data. To do this, we identified mathematical features in the EEG data (such as line length, energy, and correlation) and trained multiple classifiers with patient data.

Our primary classifier identifies areas of non-seizure and decreases the time a neurologist needs to spend reading EEG data by 90%. The classified EEG data is displayed on a web application, allowing users to focus only on possible seizure regions, providing a streamlined process for reviewing large volumes of EEG data.

I. INTRODUCTION

Electroencephalogram (EEG) recordings are collected by attaching several electrodes onto the scalp of a patient. The data is specifically used to detect the onset of seizures, which are a common symptom of epilepsy that chronically affect approximately 50 million people worldwide[1].

Yet there are challenges associated with the clinical use of EEG data. The enormous volume of EEG data recordings can range from several hours to weeks. The current clinical practice of seizure detection requires neurologists to manually read through every measured data point while monitoring the behavior of the patients via video recordings. We interviewed three neurologists at the Intensive Care Unit (ICU) in the Hospital of the University of Pennsylvania to better understand their experiences. The neurologists expressed the difficulty of manually detecting seizures because of the variant nature of biological data and the extensive training required. As a result of demand for such human expertise exceeding supply, the neurologists often spend a majority of their day solely devoted to reading EEG data of multiple patients in the ICU.

Our project aims to assist neurologists in monitoring epileptic patients for seizures by decreasing the time neurologists spend reading EEG data. Our project is built upon work of many researchers’ and organizations’ work in seizure prediction. While current state of the art research shows promising results, there remain challenges in seizure prediction. For example, movement artifacts caused by muscle contraction and ocular signals caused by eyeball movements often look similar to seizure activity on an EEG scan. Thus, neurologists use video recordings of the patient to supplement EEG readings in order to detect these differences.

To avoid difficulties of programmatically evaluating video content, we focus on predicting areas of non-seizure instead of predicting occurrences of seizures. Our classifier identifies areas of the EEG data that do not contain seizure content so that neurologists can skip these regions to view more questionable areas that require in-depth evaluation. Therefore, our focus is to ensure that the predicted non-seizure activities are in fact non-seizure, minimizing any false positives of predicting seizure regions as non-seizure regions.

II. RELATED WORKS

Seizure prediction efforts that focused on inspecting clinical EEG data for an onset of seizures began in the 1970s[2][3]. In 1990, Gotman published a technique for automated seizure prediction that scored 76% in prediction accuracy. Many successful seizure prediction algorithms have been reported since the late 1990s[4][5][6][7]. In 2001, Esteller published a seizure prediction algorithm based on line length that achieved an average prediction delay of 4.1 seconds and 0.051 false positives per hour[8]. A total of 111 seizures were analyzed from 10 patients. Afterwards, NeuroPac Inc. published a similar seizure prediction algorithm that achieved 97% sensitivity (percentage of total seizures classified as seizure by the algorithm) at an average prediction delay of 4.1 seconds and 0.051 false positives per hour[8]. In addition, recent studies have shown signal energy as an effective feature for seizure evaluation, including a study of 9 invasive intracranial EEG records that reported sensitivity of greater than 87%[10][11].

III. APPROACH

A. Backend Seizure Prediction

The backend workflow consists of three main stages: data processing, feature extraction, and classification.

1) Data Processing: Data used in the study was extracted from the International Epilepsy Electrophysiology Portal (IIEG.org), an open source repository containing various EEG datasets. Each dataset is identified by a unique ID. The sampling frequency for our chosen datasets is 5000 Hz. We selected three patient datasets that contain regions of seizures and non-seizures and a neurologist’s annotation of the start and end times of the seizures activities. Each dataset contains discrete voltage readings over time from a varied
number of channels attached to unique regions of the scalp. Each channel is filtered using MATLAB’s built in high pass and low pass filter function to eliminate noise. The filters attenuate signals outside of a predetermined frequency range to remove noise of extreme frequencies that can be associated with the patient’s interaction with other electronics or energy sources[11].

2) Feature Extraction: After acquiring and preparing our data for evaluation, we implemented feature extraction using three common mathematical measurements of EEG signals: line length sum, energy, and the correlation coefficient across each channel pair. When extracting features from the dataset, we divided the data into ten second windows. For each given window, we calculated values for our determined features for input into the classifier.

- Line Length - The line length sum of a signal is a measure of relative amplitude, which is quantified by the total length of the sum of differences between successive points within range of size N[12]. In our case, size corresponds to the number of data points included in each ten second window. Line length sum is defined as

\[ \text{LineLength} = \sum_{i=1}^{N} |x_{i-1} - x_i| \quad (1) \]

where \( x_{i-1} \) and \( x_i \) are two discrete successive voltage readings of a channel. Each line length calculation represented the sum of the absolute value of the difference in amplitudes between 5000*10 - 1 data point pairs. This is because our sample rate of 5000 implies that each data set contains 5000 data points per second over a 10 second window. High values of line length correspond to larger, more drastic fluctuations of amplitude (moving quickly from very low voltage to very high voltage).

- Energy - Energy of a signal is a measure of absolute amplitude over time, which is quantified by the total amplitude squared. Energy is defined as the following:

\[ \text{Energy} = \sum_{i=1}^{N} |x_i|^2 \quad (2) \]

For our calculations, energy is the sum of 5000 * 10 squared values of amplitude. High values of energy correspond to consistently high values of voltage. Unlike line length, which denotes high levels of fluctuation, energy indicates areas of the data that are comprised of consistently high voltages.

- Correlation Coefficient - Correlation coefficient (Pearson product correlation coefficient) is a measure of the linear dependence between readings from two channels. Correlation coefficient is defined as

\[ \rho(A, B) = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{A_i - \mu_A}{\sigma_A} \right) \left( \frac{B_i - \mu_B}{\sigma_B} \right) \quad (3) \]

where \( \mu_A \) and \( \sigma_A \) are the mean and standard deviation of A. In our data, if the recording used \( n \) channels, our feature extraction would return \( \frac{n(n-1)}{2} \) features corresponding to a correlation for each pair of channels. For each ten second window, the feature extractor would return values that are representative of the average correlation across each pair of channels over the ten second interval.

B. Classification

Once feature extraction was complete, three types of classification algorithms were used: linear discriminant analysis (LDA), support vector machine (SVM), and RUSBoost. RUSBoost is a built-in MATLAB classifier that is specialized for classifying imbalanced data[13]. The algorithm uses a combination of random undersampling (RUS) and boosting. Random undersampling removes examples at random from the majority class until the desired balance is achieved. Boosting improves the performance of any weak classifier by improving the accuracy of any given learning algorithm. Boosting in RUSBoost is based on AdaBoost, which iteratively constructs an ensemble of models[14]. During each iteration in AdaBoost, example weights are adjusted with the purpose of correctly classifying examples in the next iteration which were incorrectly classified in the current iteration. Afterwards, unlabeled examples are classified based on weighted voting of all constructed models. Our training data contains on average greater than 99% non-seizure activities and fewer than 1% seizure activities. The algorithm takes \( N \), the number of members in the class with the fewest members in the training data, as the basic unit for sampling. Classes with more members are under sampled by taking only \( N \) observations of each class[15].

IV. USER INTERFACE

Our user interface brings all aspects of the project together. The user uploads a binary file of raw data in which the first section of the data is annotated. Our web server calls MATLAB code that filters the data, extracts features, trains
the classifier on given annotations, and runs the classifier for results on the remaining data. The results of our classifier are displayed over raw data on our user interface.

In the main dashboard, the user views the summary of the most recently uploaded patient EEG data sorted by date of upload. The display of graphs is supported by Dygraphs, an open source JavaScript charting library. The summary provides a preview of annotated data with the minimized view displaying the beginning minutes of the EEG recording. The highlighted regions indicate possible seizures. When the user clicks on a particular ‘View Uploaded Data’ button, the page is redirected to display the uploaded EEG data with possible seizure regions highlighted in red. The user is able to zoom in and zoom out by adjusting the range selector below the main dygraphs view and review the data in chronological order from start to end by sliding the range selector. In addition, the user may click and drag for more zooming functionality. Dygraphs also allow panning to view more of the zoomed-in data by holding down the shift key and clicking and dragging the mouse sideways.

V. RESULTS

A. Metrics

Metrics are defined as follows:

\[
\text{Accuracy(\text{non - seizures})} = \frac{TP + TN}{TN + FP + TP + FN} \tag{4}
\]

\[
\text{Precision(\text{non - seizures})} = \frac{TP}{TP + FP} \tag{5}
\]

\[
\text{Baseline Precision(\text{non - seizures})} = \frac{\#\text{\text{non - seizures}}}{\text{total \#datapoints}} \tag{6}
\]

B. Results Analysis

A true positive (TP) is a non-seizure period determined by both the algorithm and the neurologist; a false negative (FN) is a period labeled by the neurologist as a non-seizure period but classified as a seizure period by the learning algorithm; a true negative (TN) is a seizure epoch determined by both algorithm and neurologist; a false positive (FP) is a seizure epoch incorrectly classified as a non-seizure period by the algorithm only.

Accuracy presents the percentage of correctly predicted non-seizures and possible seizures over the total data points. Precision, defined as the number of correctly labeled non-seizure activities divided by the total number of labeled non-seizure activities, is the main metric for evaluating classifier performance. In the context of the clinical application, false positives, or seizures that are incorrectly classified as non-seizures are more undesirable than false negatives, or non-seizures that are labeled as possible seizures. False positives might lead to critical error in diagnosis in which the neurologist overlooks an occurrence of a seizure, whereas false negative would lead to more EEG data needed to be reviewed by a neurologist. Therefore, our primary focus was to minimize the amount of false positives produced by our classifier.

Based on the results from the ROC plot from Figure 4, we decided to use the RUSBoost classifier as our primary model for the web application. The closer the curve is to the top left corner, the better the classifier performs under different thresholds of data. This performance is also highlighted by the Area Under the Curve (AUC). An AUC of 1 represents a perfect test while an AUC of 0.5 represents a test that
performs no better than random guessing \cite{16}. This would be shown by a straight line bisecting the X and Y axis at a 45 degree angle starting from the origin to the top right corner. The RUSBoost classifier performed the best due to its ability to handle the large data imbalance to prevent it from producing many false positives as well as using boosting to increase its accuracy.

The RUSBoost classifier performed the best due to its ability to handle the large data imbalance to prevent it from producing many false positives as well as using boosting to increase its accuracy.

Fig. 5. The table displays the performances of the SVM and LDA classifiers based on each individual feature used separately for training data.

<table>
<thead>
<tr>
<th>Feature</th>
<th>SVM</th>
<th>LDA</th>
<th>SVM</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Length Sum</td>
<td>98%</td>
<td>93%</td>
<td>88%</td>
<td>80%</td>
</tr>
<tr>
<td>Energy</td>
<td>98%</td>
<td>93%</td>
<td>92%</td>
<td>95%</td>
</tr>
<tr>
<td>Correlation Coefficient</td>
<td>98%</td>
<td>93%</td>
<td>88%</td>
<td>91%</td>
</tr>
</tbody>
</table>

The results from Figure 5 indicate that correlation coefficient is a poor feature to include in the classifier. This can be attributed to the fact that correlation is susceptible to noise. Because we have used clinical data throughout this project, the noise included in the recordings might have contributed to the low performance of our classifier when incorporating correlation coefficient. Thus, we decided to only include line length sum and energy as the only features in the final classifier.

Fig. 6. A learning curve showing the comparison of misclassified seizures vs correctly identified seizures for the first patient. For each training data split, a percentage of seizure regions was added to the training data. The remaining seizures and non-seizure regions were left as testing data.

One notable characteristic of our EEG discrete time-series data is that the non-seizure regions represent greater than 99% of the data points, leaving fewer than 1% of the data containing seizure activities. In fact, most seizures last from a few seconds to a few minutes, with the average seizure lasting fewer than 11 minutes \cite{17}. The data used is representative of typical data in our use case, where the patient is being monitored for long periods of time - ranging from a few hours to weeks. Infrequent occurrences of seizure and relatively short periods of seizure activity in comparison to the long hours of non-seizure data regions result in a large imbalance of data. Figure 6 shows that increasing percentages of seizure data used for training correlates to fewer mispredictions of seizure as non-seizure.

If the classifier predicts the whole duration of the data as being non-seizure, the classifier will still yield high baseline precision (99.8%) because of the imbalanced data of mostly non-seizure regions. To account for high baseline precision, precision of the classifier is compared to the baseline precision. As shown in Figure 7, the possible improvement percentage increase of precision over baseline precision rises with the increased percentage of seizures in training data.

VI. DISCUSSION

The largest strength of our project is the percent growth over the baseline that we have achieved in our classifier. Our user interface also serves as a strength - not only have we achieved impactful high precision and accuracy in our classifier, we have also created a means through which users can upload data and utilize a visualization of the results to make more informed diagnoses in a fraction of the time. In addition, our approach utilizes a simple algorithm and requires low computational cost associated with feature extraction.

Because of the scarcity of long EEG recordings annotated by neurologists (also further exemplifies the problem of interest), we were only able to test our algorithm on three patients. In the future, we would like to test this on a larger patient set. While incorporating two simple features in RUSBoost classifier have produced high precision and accuracy, we would like to experiment with additional features to fine tune the classifier. In particular, we are interested in understanding how the performance of the non-seizure prediction algorithm differs across different types, severity, duration, frequency of occurrence of seizures as well as the demographics and medical conditions of the patients.

There exists a trade-off between precision and application of the trained classifier to predict non-seizures across different patients. Our approach involves training on a portion of a patient’s continuously monitored EEG recordings and running the classifier on the rest of the given dataset to maximize precision. We chose this method because medical
conditions are highly personal and vary on an individual basis. The threshold of abnormal voltage spikes for seizures differs for each patient. In addition, the number of channels attached to a given patient's scalp and the physical placement of electrodes on the scalp vary. Seizures are classified into two groups: generalized seizures that affect both sides of the brain and focal seizures that affect only a certain area of the brain. Therefore, comparing the voltage readings taken from different parts of the brain of different patients and seizure types yields low precision and accuracy.

Lastly, our user-interface has limited ability to load and display huge volumes of data using Dygraphs. The current user-interface is noticeably slow when all channels are graphed at once. In the future, we would like to implement incremental loading with AJAX to display only the amount of data needed for the given selected window to improve performance. Currently, the reduced time neurologists needed to review EEG data is estimated based on the percentage of data identified as non-seizure. Yet, neurologists will most likely spend more time reviewing EEG data than estimated because of additional time spent learning and changing habits to make use of our user interface. Given more time, we would conduct user case studies to make improvements to our user interface based on user feedback.

VII. CONCLUSION

Seizure prediction based solely on EEG recordings has long been a field of study with unsolved challenges. Our novel approach builds upon existing work with seizure prediction and shifts the paradigm from predicting seizures to predicting non-seizures and labeling the remaining data as possible seizure for neurologists to review. With two simple, yet robust extracted features of line length sum and energy for the RUSBoost classifier, we have produced an average of 99.9% precision of non-seizure prediction for all three data sets. We have also developed a web application that focuses on optimizing the user experience tailored to the needs and preferences of the neurologists. Our study has demonstrated the feasibility of building a non-seizure prediction tool that can expedite neurologists' review of EEG data. We have developed a more consistent and objective approach in predicting non-seizures that complements the conventional, manual approach that often results in highly variant and inconsistent diagnoses.

VIII. ETHICAL / PRIVACY CONSIDERATIONS

Data used in the study was extracted from the International Epilepsy Electrophysiology Portal (IEEG.org), a platform for sharing data, tools, and expertise between researchers. Each dataset is labeled by a unique identifier, in which the patient information is de-identified. IEEG.org is HIPAA compliant and protects patient privacy. There is minimal risk or harm associated with the misuse of the data. We have built a prototype for a non-seizure prediction tool that assists the neurologist in interpreting EEG data of patients and is not intended to be a replacement for reading EEG data.

Moving forward, additional ethical/privacy considerations should be made pertaining to user settings of the user interface that displays our results. Although no login functionality has been implemented because of exclusive use of data from the IEEG portal, future use of the tool to evaluate patient data will warrant additional security features to be added as follows:

- Include security level in user registration to be confirmed upon registration.
- When creating a new patient account or uploading data associated with patient account, data access settings will be specified to indicate which users have access to the data that user is uploading.
- Different levels of security to correspond to aspects of the patient’s account information and uploaded data the user can view and change.

IX. ACKNOWLEDGMENT

We thank Dr. Brian Litt and Dr. Christian Murphy for their support and mentorship. We also thank Steven Baldassano, Dr. Jason Moyer, and John Frommeyer for their guidance in data acquisition, analysis and development of the user interface.

REFERENCES

[9] Rosana; Echauz, Javier; D’Alessandro, Maryann; Shor, Rachel; Henry; Thomas; Pennell, Page; Epstein, Charles; Bakay, Roy; Dichter, Marc; Vachtsevanos, George. In: Neuron, Vol. 30, No. 1, 04.2001, p. 51-64.

