Machine Learning Methods for Seizure Detection

Meenakshi Khosla
Cornell University
Ithaca, NY
mk2299@cornell.edu

James Redd
Cornell University
Ithaca, NY
jsr283@cornell.edu

***Abstract*—Machine learning techniques hold significant potential to detect seizures from EEG data. This is consequential, since timely seizure detection is essential for the success of responsive neurostimulation therapies in patients with drug-resistant epilepsy. In this paper, we briefly review the literature for seizure detection using EEG data and recent advances in the field. We further describe the techniques implemented by our team for the ECE5040 Kaggle Challenge and discuss the results obtained using different combinations of features and algorithms on the dataset.**

***Index Terms*—Machine learning, seizures, epilepsy, neurostimulation, EEG**

I. Introduction

Patients with epilespy frequently develop intuitions as to when a seizure is imminent. This allows them to prepare, rendering their environment safer for them and others around them. For example, it is incredibly dangerous for a patient to have a seizure while driving on the highway. Warning that a seizure is going to occur would give them time to pull over and come to a stop. However, much of the time this intution fails to emerge, and when it does it is innacurate or provides insufficient warning time. A portable device that uses readily accessable signals could be a path to early warning as well as a way for a stimulator to inject current and stop the seizure form ever occuring.

EEG signals require only the placement of non-invasive recording electrodes on the scalp. Many publicly available EEG devices are emerging that do this in a low energy, subtle form factor that would increase patient compliance. Therefore, there is a clear path to deployment of a device that classifies seizure and pre-seizure activity. The first step in development of such a device would be the software that can recieve EEG signals and determine when a seizure will occur. The data and labels provided by the ECE5040 Kaggle Competition was a first step towards creating that capability.

In the ECE5040 Kaggle Competition, we are provided with EEG clips of 7 patients during ictal and non-ictal periods. Our goal is to classify the two categories, with submissions assessed on the AUC of the Receiver Operating Characteristic (ROC) curve. We extract features previously reported to be indicative of seizures for our prediction framework.

We adopt a patient-specific approach, wherein the classifier is trained separately for each subject. The parameters of the classifier are, however, kept the same across subjects to prevent overfitting and to gain a true assessment of the generalizability of our model to previously unseen subject data.

II. Data

A. Data description

Each clip contains 1s of EEG recording across differing number of electrodes for each patient, in the range of .

2 of the 7 subjects had the EEG recordings sampled at 5kHz, while the rest had data sampled at 500Hz.

B. Data Visualization

As a preliminary step, we visualized the data to gain a better understanding of how EEG activity is altered during seizures.

Figure 1 depicts a general classification routine. The rest of the paper is outlined as follows: Section III discusses the data preprocessing steps and the relevant features considered in our model. Section IV described the various classification algorithms implemented. Section V presents the results obtained using the different combinations of features and algorithms. Finally, in section VI, we discuss the results obtained using different methods and why we believe certain techniques failed to yield good results.

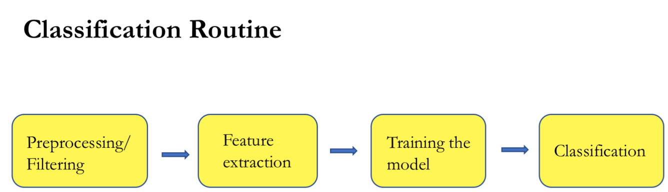

Fig. 1. Classification Routine

III. Preprocessing and Feature extraction

A crucial step in seizure detection is the selection or extraction of appropriate features. Traditionally, the features extracted from EEG signals are classified into two broad categories: Time-domain features and Frequency-domain features.

A. Time-domain features

Several time domain features have been proposed in literature for seizure detection. Commonly used features include line-length, signal variation or standard deviation, energy etc.

Previous studies suggest that epileptic seizures are characterized by hypersynchronous states. Correlations provide a significantly rich feature space for machine learning models to learn the discriminating data patterns across categories. Thus, we included the correlation structure of multi-channel intracranial EEG into our model. For n channels, this gives *n(n-1)/2* correlation values after excluding the diagonal entries and extracting the upper triangle of the symmetric correlation matrix.

Eigenvalues of the correlation matrix further reflect additional metric compactly representing some aspects of the correlation structure [7]. Typically, the magnitude of these eigenvalues represents the correlation values in uncorrelated directions known as eigenvectors. We concatenated the n eigenvalues of the *nxn* correlation matrix with the existing features, hoping that it captures another quantity predictive of seizures.

For time domain features, we also experimented with several preprocessing steps like downsampling the data or applying dimensionality reduction on the correlation values using Principal Comopnent Analysis (PCA).

We also observed a clear distinction between correlation patterns during the ictal and non-ictal periods visually, as shown in Figure 2

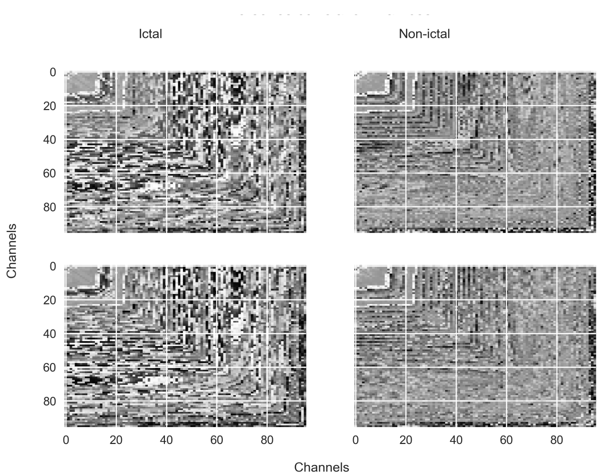

Fig. 2. Correlation structure of time series across channels

B. Frequency domain features

Several frequency domain features have been previously shown to predict seizures. These include power spectral density, ratios of power spectral density, mean phase coherence etc. Of these, the power spectral density (PSD) features are the most frequently used predictors of seizures. Previous research has shown that significant changes occur in the spectrogram during seizures.

FFT is commonly used to extract the content in different frequency bands of the signal. Since the signal duration is 1s, we can extract frequency features at a resolution of 1 Hz. In accordance with previous research, we expected the frequency content in alpha, beta and lower gamma bands to be useful for seizure prediction. Omerhodzic et. al. [4] showed that energy distribution in delta (0.5-4Hz) and theta (4-8 Hz) bands increases during seizures, whereas the alpha (8-12 Hz), beta (12-25 Hz) and gamma (typically 25-100 Hz) bands experience a lower energy distribution. We confirmed a similar trend using our own visualizations of spectral power, as shown in Figure 3. Most studies focus on the 1-60 Hz frequency range. Since the optimal frequency range for our classification task is unknown, we experimented with 4 frequency ranges: (1-30) Hz, (1-45) Hz, (1-60) Hz and (1-100) Hz. Frequency-domain features extracted from all the channels are concatenated together with other features to test the final input for classification.

Rather than feature selection, we proceed with the strategy of feature elimination. We first include all the proposed features in our model and then remove each of these sequentially to estimate the importance of feature subsets.

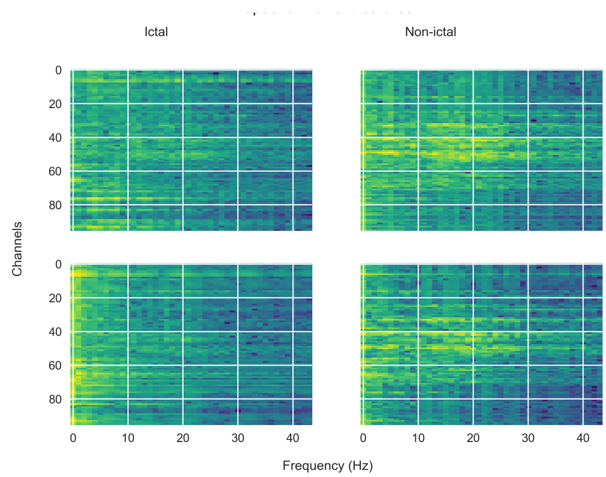

Fig. 3. Spectral Power calculated using FFT

IV. Classification Algorithms

A. Support Vector Machines

SVMs optimize for a hyperplane with maximum margin between the output classes. It is one of the most commonly used linear classifiers for seizure detection. Mathematically, SVMs minimize the following loss function to get the optimal weights θ^*=$\{\mathbf{w}^*, b^*\}$.

$$\mathcal{L}_{\mathcal{S}} = \sum_{i=1}^{n} \max(0, 1 - y_i(w^T x + b))^2 + \beta \left\| w \right\|^2$$

There are several variants to the above loss function. For example, the squared hinge-loss (i.e. the first term) can be replaced a hinge-loss or the regularization term (i.e. the second term) can be an *L1* penalty instead of *L2*. We use the above loss function to compute our predictions.

B. Logistic Regression

The logistic regression classifier is used to estimate the probability of a binary outcome by fitting a linear regression model on the input features. In this model, the log-odds of the output prediction are assumed to be a linear function of the inputs. Mathematically, logistic regression minimizes the following loss function problem to find the optimal weight parameters θ^*={$\mathbf{w}^*$, b^*},

$$\mathcal{L}_{\mathcal{LR}} = \sum_{i=1}^{n} \log(\exp(-y_i(w^T x_i + b)) + 1) + \alpha \|w\|^2$$

C. Random Forests

Random Forest Classifiers are ensembles of decision trees, with each tree trained on a random subset of the training data. Decision tree classifiers build a flowchart of decision rules based on feature attribute values. For each test sample, the tree is traversed following these rules at each internal node and the final classification label is assigned at the leaf node. Each subset for training individual decision trees is selected using sampling with replacement from the training data. This model averaging strategy helps in preventing overfitting, and yields reasonable probability estimates for the classes. Significant tuning parameters for random forests include the number of individual estimators (decision trees), maximum number of features to use during split searching, maximum depth of the tree, minimum samples required for internal node splitting etc. We speculated that number of individual estimators to have the most significant effect on AUC, and hence focused on tweaking this parameter.

D. Gaussian Naive Bayes

The Gaussian Naive Bayes algorithm seeks to find an optimal decision boundary between the binary classes under the assumptions that the data is normally distributed and the covariance matrix of the features is diagonal. This algorithm assumes conditional independence between features, which might be a strong assumption in our case. However, since it is a probabilistic classifier, we decided to test the model on our dataset.

E. Majority Voting

We also employed another strategy by building a majority voting ensemble of these individual models. This scheme weighs the binary outputs of individual classifiers equally to generate a final probability for the classes.

Implementation toolkit: scikit-learn **Coding Language:** Python

V. Results

We consider Support Vector Machines and Logistic Regression as a classical machine learning baseline. Although the SVMs can be tweaked to yield probabilities for the classes instead of categorical predictions, the interpretation of this continuous metric as probability is not straightforward. Further, we observed that logistic regression yielded probabilities very close to the binary values of probability, indicating that the classifier becomes very confident in its prediction. This could result in negative effects on the AUC when the categorical predictions are not very accurate. Subsequently, we proceeded with other classifiers like Random Forests, which give more realistic and conservation probabilistic predictions.

A. Table

The following table briefly summarizes results obtained using the main models discussed in this paper.

Prediction Results	
Method	AUC
FFT+Random Forest (4000)	0.89028
FFT+Correlation+Eigenvalues (unscaled time series)+Random Forest	0.9199
FFT+Correlation+Eigenvalues+Logistic Regression	0.82642
FFT+Correlation+Eigenvalues+Random Forest (500)	0.93351
FFT+Correlation+Eigenvalues+Random Forest (4000)	**0.93478**
FFT+Correlation+Eigenvalues+Voting	0.81156
FFT (1-100Hz)+Correlation+Eigenvalues+ Random Forest (4000)	0.84829

B. Results using Validation Data

In order to assess our algorithms and features without having to submit our predictions to Kaggle, we split the given training data into a training and validation set. Particularly, we randomly select 20% of the data for each subject and use it for training, while computing performance metrics on the validation set. Figure 4 depicts the ROC curve for all subjects using our best model, i.e. Random Forests trained using all the aforementioned features. Across all subjects, we achieved

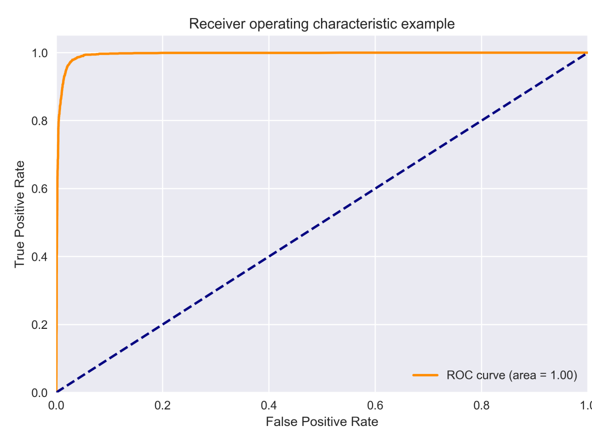

Fig. 4. Receiver Operating Characteristic (ROC) Curve

a sensitivity of **100%** and specificity of **99.93%** using our validation data. We observed that the performance obtained using validation splits was significantly higher compared to the independent test set on Kaggle and are still trying to understand the gap in performance.

VI. Discussion

The best model was obtained using a Random Forest with large number of estimators (4000 Decision Trees) and the combination of all aforementioned features, which include spectral power components in the frequency range 1-45 Hz, the time-series correlation values across all possible pairs of channels and finally the eigenvalues of this correlation matrix. We observed that inclusion of spectral power components beyond 45 Hz actually decreased our prediction accuracy (and AUC). We speculate this might be occurring because of possible confounds introduced by the broadband gamma components.

We also assessed the impact of dimensionality reduction before supervised learning on the final prediction precision. However, in our case, dimensionality reduction using PCA (with 500 components) before training the Random Forest Classifier actually decreased the AUC from **93.35%** to **81%**.

We also observed the impact of preprocessing techniques on the features and resultant classification performance. Particularly, we observed that resampling the time series of each channel to fewer samples (500-1000) actually shows better performance. This might help get rid of spurious time series correlations to yield more reasonable estimates of correlations and eigenvalues. Further, we also observed that scaling the time series data

Another innovative approach would have been to identify the channels localized to parts of the brain most affected by seizures. Channel selection can help build a robust prediction model insensitive to noise in EEG channels unaffected by seizures.

Another limitation of our approach is that we discarded high frequency spectral features. For patients sampled at 5kHz, inclusion of very high frequency components ($\geq$ 1000 Hz) could lead to better predictions. Several studies have shown the prevalence of high-frequency oscillations in epileptic patients during seizures.

Furthermore, in order to capture changes in signal frequency content over time, we could have used sliding windows and short-time Fourier Transform. This would yield dynamic spectral features rather than static features from the 1s clip, which might help improve the prediction accuracy. The dynamics of the correlation structure itself have previously shown to be predictive of seizures ([7])

In the development of this machine learning system, it was notable the degree to which the addition and subtraction of particular extracted features were impactful in our final AUC result. For example, using FFT and a Random Forest had an AUC of .89, while adding correlation and Eigenvalues increased that to .91. This is a common feature of relatively simple machine learning systems, where the selection of extracted features is sometimes more important than the ML algorithim chosen. This means that the features we were selecting for were specific to this dataset and could potentially not generalize well to other patients, though we did use a robust validation scheme.

In order to improve the generalizability of seizure detection, we are interested in future work in using Deep Learning. Though feature extraction is used in DL, it is notable in that it can recieve an entire unfiltered dataset and create accurate results. This would allow a system to observe many possible features and select the ones most valuable for classification for each patient. There are many different ways we could reconfigure the data to be fed into a DL classifier. We would use a Recurrent Neural Network or a Long Short Term Memory network and directly feed in the signals themselves, as well as the FFT or the eigenvalues. There has also been work on manipulating EEG signals into a 2 dimensional image and applying a Convolutional Neural Network with high fidelty results. [8] The downside of a the Deep Learning approach would be the high likelihood that the Neural Net itself would have to be hosted on a server and not on chip, increasing latency.

Furthermore, the use of Deep Learning on all of the available signals may potentially inform new physiological discoveries in Epilespy. Interrogating the network may provide information about patterns we were not previously aware of, which could inform future devices. In classical machine learning this would not be possible as we a-priori select the features and connections the system is using to classify seizures.

Acknowledgment

We would like to thank Professor Mahsa Shoaran and Hanchen Jin for their continued guidance and support during the course of this project.

References

[1] Ulate-Campos A., Coughlin F., Ganza-Lein M. et. al, "Automated seizure detection systems and their effectiveness for each type of seizure," Seizure, vol. 40, pp. 88-101, 2016.

[2] Ramgopal Sriram, Thome-Souza, Sigride Jackson et. al, Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy, Epilepsy & Behavior, vol. 37. 2014, pp.291-307.

[3] Shoeb Ali, Guttag John, "Application of Machine Learning To Epileptic Seizure Detection," in International Conference on Machine Learning, 2010.

[4] Omerhodzic Ibrahim, Avdakovic Samir, Nuhanovic Amir "Energy Distribution of EEG Signals: EEG Signal Wavelet-Neural Network Classifier," Arxiv Id: 1307.7897, July 2013.

[5] Zijlmans Maeike, Jiruska Premysl, Zelmann Rina, "High-frequency oscillations as a new biomarker in epilepsy" Annals of Neurology, vol.71, 2012, pp. 169-178.

[6] Staba Richard J, Stead Matt, Worrell Gregory A, "Electrophysiological Biomarkers of Epilepsy," Neurotherapeutics, vol. 2, pp. 334–346, 2014

[7] Schindler Kaspar, Leung Howan, Elger Christian E, "Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG," Brain, vol. 130, pp. 65–77, 2007.

[8] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In 23rd International Conference on Intelligent User Interfaces (IUI '18). ACM, New York, NY, USA, 43-53.