Beyond Brainstorms: Predicting Epilepsy Surgery Outcome with Functional Connectivity Centrality

Beyond Brainstorms: Predicting Epilepsy Surgery Outcome with Functional Connectivity Centrality

A DISSERTATION PRESENTED
BY
SIYE ANNIE ZHU
TO
THE DEPARTMENT OF COMPUTER SCIENCE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR'S
IN THE SUBJECT OF
COMPUTER SCIENCE AND MATHEMATICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

MARCH 2022

Thesis advisor: Dr. Eleonora Tamilia Siye Annie Zhu

Beyond Brainstorms: Predicting Epilepsy Surgery Outcome with Functional Connectivity Centrality

Abstract

For patients with medically refractory epilepsy (MRE) who do not respond to anti-seizure medicine, invasive intracranial EEG (icEEG) and resective surgery are used to respectively identify and remove epileptogenic zones (EZ). Advances in signal processing and network analysis have sparked interest in the processing and interpretation of icEEG data for more accurate identification of the EZ. The aim of this study is to use functional connectivity centrality measures extracted from interictal icEEG recording to localize seizure onset zones (SOZ) and predict post-surgical outcomes. The SOZ is the most reliable estimator for EZ with observable ground truth, and the current standard practice relies on clinicians to visually identify individual electrodes as SOZ during every ictal recording, which is an inaccurate and inefficient method. This paper analyzed five minutes of interictal icEEG data from 40 MRE patients who underwent surgery at Boston Children's Hospital. First, classification models using centrality measures were able to robustly identify SOZ regions (AUC 0.70). Then, using likelihood-based diagnostic models and proportion of resection, we were able to predict post-surgical outcome with AUC 0.74. Our results suggest that functional connectivity centrality measures are efficient and useful features for clinicians in the treatment of MRE.

Contents

1 Introduction 1
- 1 Background and definitions . . . 1
- 2 Functional connectivity and related work . . . 3
- 3 Overview of our contributions . . . 5

2 Models and approaches 7
- 1 Logistic regression (LR) . . . 8
- 2 Gradient-boosted trees (XGB) . . . 9
- 3 Principal component analysis (PCA) . . . 10
- 4 Kernel density estimation (KDE) . . . 11
- 5 Variational autoencoder (VAE) . . . 12

3 Dataset and methodology 15
- 1 Patients and feature extraction . . . 15
- 2 SOZ prediction . . . 19
- 3 Outcome prediction . . . 23

4 Results 26
- 1 Data exploration . . . 26
- 2 SOZ predictions . . . 27
- 3 Supervised classification of SOZ electrodes . . . 28
- 4 Dimensional reduction for unsupervised clustering . . . 31
- 5 Identification of patient outcome . . . 34

5 Conclusion 40

Appendix A Centrality measures correlations 43

Appendix B Performance metrics 47
- 1 AUC from SOZ predictions . . . 48
- 2 AUC from outcome predictions . . . 49

References 51

For Mom, Dad, and David

Acknowledgments

First and foremost, I would like to express my deep gratitude for my advisor, Dr. Eleonora Tamilia, for being the most knowledgeable and patient mentor. Her guidance and support allowed me to find the common ground between my theoretical statistics and computer science background and the applied field of epilepsy research, which I had known little about prior to the project. I would also like to thank members of the DtAK lab, in particular Weiwei and Yaniv, for providing useful insights whenever I hit a roadblock in my research. I also thank Prof. Finale Doshi-Velez for offering me advice on the research and for being one of my readers and taking the time to provide feedback on this thesis.

Second, I would like to thank the departments of math, computer science, and statistics for inspiring me in the intellectual pursuit of knowledge. The past four years would not have been as eye-opening for me without any one of you, and I am so fortunate to have such a wonderful college experience.

Last but not least, I would like to thank my family and friends for their unending love support, both throughout the process of this thesis and in helping me become who I am today. I would especially like to thank David, without whose infinite patience and encouragement I would have never managed to undertake this challenge. To all of you, I am eternally grateful.

1

Introduction

1 Background and definitions

Epilepsy is a chronic electrophysiological disorder of the brain that currently affects over 50 million people worldwide, with around 5 million new diagnosed cases every year.[1] Untreated epilepsy can increase the risk of premature death by up to three times. 50% of epilepsy cases begin in childhood

or adolescence. Epilepsy is characterized by recurrent seizures, which are episodes of excessive neural firing that can lead the patient into prolonged convulsions. The cause of epilepsy is unknown for about 50% of the cases.*

Many treatment methods have been developed for epilepsy. While up to 70% of epilepsy patients can become seizure free after taking anti-seizure medicine, 30% of the patients do not respond to medical therapy and are considered as affected by medically refractory epilepsy (MRE).[11] The most effective treatment for MRE is surgical resection of the *epileptogenic zone* (EZ), which is defined as the brain tissue that is responsible for generating seizures in the patient.

Clinicians diagnose epilepsy patients and develop surgical plans by referring to electroencephalography (EEG) recordings of neural activity. Non-invasive EEG recording involves placing small metal discs on the scalp of the patient. For some patients, invasive intracranial EEG (icEEG) recording is also required, which consists of either placing subdural grids of electrodes directly on the surface of the brain (electrocorticography, or ECoG) or depth electrodes inside the target brain areas (stereo-electroencephalography, or sEEG). The goal of EEG recordings is to localize the *seizure onset zone* (SOZ), which as the name suggests describe the neural region of earliest onset of epileptic activity during at least one seizure observed during EEG monitoring. Since the EZ cannot be directly observed, the SOZ is considered as one of the most reliable guides for determining surgical resection areas. However, it is also widely understood that the observed SOZ often does not completely coincide with the EZ for a variety of reasons, and as a result, only 27%-67% of the patient become

*Statistics taken from https://www.who.int/news-room/fact-sheets/detail/epilepsy and collected by the World Health Organization (WHO).

seizure-free post surgery. [16]

An EEG recording is *ictal* if it is recorded while the patient is undergoing a seizure, whereas a recording is *interictal* if it is made between seizures. While ictal recordings contain more information of neural pathology, oftentimes they can only be captured by prolonged continuous EEG monitoring. Although such an approach often captures many seizures, it requires heavy maintenance and processing and is therefore undesirable. [5]

2 Functional connectivity and related work

To this day, visual inspection of EEG recordings remain the standard practice for determining SOZ. However, recent advances in computer-assisted data analysis methods has revolutionized the study of brain function and presents many new possibilities for the analysis of EEG recordings in determining the SOZ, which can improve the success rate of epilepsy surgery. The technological advances are manifold. To start, while the icEEG data can be interpreted by visual inspection, the process can be inaccurate time-consuming, whereas automated methods have been developed to efficiently capture high frequency oscillations charactieristic to epileptogenic zone. [7,1] Moreover, such methods can be applied to ictal and interictal periods alike, where the latter is much more abundant yet not as meaningful to a human inspector. [21] Most importantly, graph-based approaches have proven to be important for understanding the the functional organization of the brain in terms of functional connectivity.

Functional connectivity between neurons is defined as the "temporal coincidence of spatially dis-

tant neurophysiological events."[17] In recent year, the EZ is increasingly considered as a network of neurons rather than a localized region of the brain, so understanding the functional network of the brain has become important in predicting the EZ.[19] A region of the brain has high *centrality* if it is significant in the brain connectivity network on a local and global scale. Locally, *mean connectivity* is a centrality measure that captures the average functional connectivity between a given electrode and all other electrodes. In addition to showing increased connectivity in certain frequency bands, epileptic brains do not have the healthy small-world topology of healthy brains and rather exhibit organizational patterns that allow for fast propagation of seizures.[20] Since mean connectivity is only able to capture network features on a local scale, we extract the minimum spanning tree (MST) on the functional connectivity graph. The MST retains the most important paths on which neural signals travel, thereby giving a more global picture.

Four kinds of centrality measures can be defined for a graph $G = (V, E)$ with adjacency matrix representation of edges A in order to characterize its topology:

- *Degree*: the number of electrodes coming from a vertex,[14] i.e.

$$d_i = \sum_{j \in V} a_{ij}.$$

- *Betweenness*: the percentage of shortest pathways that pass through the vertex,[14] i.e.

$$b_i = \sum_{j,k \in V, j \neq k} \frac{n_{jk}(i)}{n_{jk}},$$

  where n_{jk} denotes the number of shortest paths from j to k, and $n_{jk}(i)$ denotes the number of shortest paths from j to k passing through i. In an MST, $n_{jk} = 1$ so an vertex is either *leaf* if $b_i = 0$ or *central* if $b_i = 1$.

- *Closeness*: the reciprocal of average distance to other electrodes,[3] measured as

$$c_i = \frac{N}{\sum_{j \in V} d(i,j)},$$

where $N = |V|$ and $d(i,j)$ is the distance between vertices i and j.[†]

- *Eigenvector*: the eigenvector corresponding to the greatest eigenvalue of the adjacency matrix,[6] i.e. the vector e corresponding to the largest λ such as the following system has a solution

$$e_i = \lambda^{-1} \sum_{j \in V} a_{i,j} e_j \Leftrightarrow Ae = \lambda e$$

By the Perron-Frobenius theorem, for any nonnegative, irreducible matrix, there exists an eigenvector with nonnegative entries such that its corresponding eigenvalue is positive and greater than or equal to the absolute value of any other eigenvalue,[18] so this definition is almost always unique on empirical data.

Studies have shown that high betweenness centrality on the gamma frequency band (30-50 Hz) is correlated with electrode resection among patients with good post-surgical outcome[21]. Hubs, i.e. regions of dense connections, have also been shown to characterize with epileptogenic zone and are more likely to stand out with high centrality measure in a MST[21,12]. In our study, we measure the functional connectivity centrality of measures, which we expect to be predictive on

3 Overview of our contributions

This paper studies functional connectivity centrality measures that can be used to localize the SOZ and predict post-surgical epilepsy-freedom for MRE patients. Firstly, SOZ electrodes are demonstrated to have higher centrality measures than non-SOZ electrodes. Secondly, centrality measures

[†]By the original definition, the denominator of c_i is $N-1$ to account for there being $N-1$ vertices other than N, but for large N these two definitions give very similar measures.

are used to predict SOZ electrodes. Finally, based on the difference between the distribution of centrality measures of SOZ and non-SOZ electrodes, diagnostic models are proposed for measuring the epileptogenicity of individual electrodes, and the proportion of resection among epileptogenic electrodes is used to predict the post-surgical outcome of MRE patients. This paper analyzes the performance of our methodology on a retrospective profile of 40 patients who underwent epilepsy resection surgery at the Boston Children's Hospital.

The paper is divided into five chapters. Chapter 1 provides the background information on the definitions of epilepsy, functional connectivity, and centrality, as well as a literature review on previous studies showing that high centrality characterize the EZ. Chapter 2 gives an overview of the major models used in this study. Chapter 3 elaborates on the feature extraction process for our dataset and how the models are applied to the data. Chapter 4 gives the results from our analysis. Chapter 5 concludes the contributions of the paper and presents directions for future work.

2

Models and approaches

We give brief overview of the motivations and mathematical foundations of the models we employ in Chapter 3.

1 Logistic regression (LR)

Logistic regression is a generalized linear model often used for binary classification.[2] Given feature vector $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip}) \in \mathbb{R}^p$ for data points $i = 1, \ldots, n$, a logistic regression model hypothesizes that $Y_i \sim \text{Bern}(\pi_i)$ with

$$\pi_i = \frac{\exp\left(\sum_{j=1}^p \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^p \beta_j x_{ij}\right)}, \text{ or } \text{logit}\left(\pi_i\right) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j=1}^p \beta_j x_{ij}$$

for some coefficient vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}} \in \mathbb{R}^p$. The choice of logit as the link function is motivated by the fact that logit symmetrically maps $[0, 1]$ to the real line. In GLM terminology, we observe that $\pi_i = F(\sum_{j=1}^p \beta_j x_{ij})$ where $F(z) = e^z/(1+e^z)$ is the CDF of the logisitic distribution, which is an exponential family distribution. The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is the solution to the likelihood equations $\partial L(\boldsymbol{\beta})/\partial \beta_j = 0$, which in this case simplifies to

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n n_i \left(y_i - \pi_i\right) x_{ij} = 0.$$

While there is no closed form solution to this equation, a numerical solution can be approximated using the Newton-Raphson algorithm, which iteratively computes

$$\pi_i^{(t)} = \frac{\exp\left(\sum_{j=1}^p \beta_j^{(t)} x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^p \beta_j^{(t)} x_{ij}\right)},$$

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left\{\boldsymbol{X}^{\mathrm{T}} \text{Diag}\left[n_i \pi_i^{(t)}\left(1 - \pi_i^{(t)}\right)\right] \boldsymbol{X}\right\}^{-1} \boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{s} - \boldsymbol{\mu}^{(t)}\right).$$

In this study, logistic regressions are fitted using the `LogisticRegression` function from the scikit-learn package (version 0.24.2) in Python.

2 Gradient-boosted trees (XGB)

Gradient tree boosting is the product of a long line of tree-based ensemble models. Efficient, flexible (non-linear), and robust to overfitting, it has become one of the most popular methods in the data science community, producing state-of-the-art performance in all types of challenges ranging from Kaggle to industry benchmarks.[4]

Given observations $x_1, \ldots, x_n$, a tree ensemble model predicts

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i),$$

where $f_k(\boldsymbol{x}) = w^{(k)}_{q(\boldsymbol{x})}$ is a regression tree, where $q(\boldsymbol{x})$ reflects the tree strucure and $w^{(k)}$ is the vector of predictions at the leaf nodes. Training the regression involves optimizing a regularized objective

$$\mathcal{L}(\phi) = \ell(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w^{(k)}\|^2,$$

where Ω is a penalty function on the complexity of the regression tree. Such a model is trained additively, i.e. in the t-th iteration, if the current predictions are $\hat{y}_i^{(t)}$, then we seek to add a regression

tree f_t to our current tree in order to minimize the objective

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \ell\left(y_i, \hat{y}_i^{(t-1)} + f_t\left(\mathbf{x}_i\right)\right) + \Omega\left(f_t\right).$$

We find the optimal f_t in a two-step process. Fixing the tree structure q and using a second order approximation for the loss function, we can compute the optimal leaf nodes. Fixing the leaf node predictions, we can greedily add branches to the tree (call this a *split*) and keep the one that introduces the decrease in the loss. Additional methods such as column subsampling, which forces each regression tree to only use a subsample of the features, can also help prevent overfitting.

Similar to the extension from linear model to binary classification, gradient tree boosting models can be extended to make binary predictions, which fits our use case. In this study, gradient boosted trees are fitted using the `XGBClassifier` function from the xgboost package (version 1.5.0) in Python.

3 Principal component analysis (PCA)

Principal component analysis is a dimensional reduction method. [10] Given column-wise zero-meaned data $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, the top principal component can be found by finding the variance-maximizing direction of projection

$$\boldsymbol{w}_{(1)} = \arg \max_{\|\boldsymbol{w}\|_2=1, \boldsymbol{w} \in \mathbb{R}^p} \sum_{i=1}^{n} (\boldsymbol{x}_i \boldsymbol{w})^2 = \arg \max_{\|w\|_2=1} \|\boldsymbol{X}\boldsymbol{w}\|^2.$$

When $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ is full rank, we know that $\boldsymbol{w}_{(1)}$ is the eigenvector corresponding to the largest eigenvalue λ_1 of $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. Then, the first principal component of all datasets is $\boldsymbol{X}\boldsymbol{w}_{(1)} \in \mathbb{R}^n$, where $\boldsymbol{x}_i\boldsymbol{w}_{(1)}$ is the dimensional reduced representation of $\boldsymbol{x}_i$. To compute further principal components, assume that we have already computed the first k components. Then, we can remove their effect by computing $\hat{\boldsymbol{X}}_k = \boldsymbol{X} - \sum_{s=1}^{k-1} \boldsymbol{X}\boldsymbol{w}_{(s)}\boldsymbol{w}_{(s)}^{\mathrm{T}}$ and proceed as before to extract the next component.

For high-dimensional data, extracting the first few principal components gives us meaningful latent space representations because it preserve as much of the data variation as possible. The choice of how many components to extract is usually determined heuristically based on our understanding of the data set and the cumulative explained variances, i.e. $\sum_{i=1}^{k} \lambda_k$ for the first k components.

4 Kernel density estimation (KDE)

Kernel density estimation is a non-parametric model for estimating probability distribution. Given observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the kernel density estimator gives the distribution

$$\hat{f}_h(\boldsymbol{x}) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_i\|}{h}\right),$$

where h is a smoothing parameter and K is the kernel. A common choice for the kernel is the pdf of the multivariate normal distribution with identity covariance. Due to the localized nature of the construction of KDE, it does not have good performance when the data is sparse and/or in high-dimensions.

5 Variational autoencoder (VAE)

A variational autoencoder is a probabilistic graphical model that models the distribution of a dataset on the high-dimensional input space using a parmetrized distribution and a low-dimensional latent space.[8] Denote the observed, input data by x and the unobserved, latent representation by z, and let the joint distribution be $p_\theta(x, z)$. $p_\theta(x|z)$ is typically taken to be some Gaussian distribution $\mathcal{N}(f(z), \sigma^2 I)$ for some function f, where the *output variance* σ^2 is hyperparameter of our model. $p_\theta(z|x) = p_\theta(x, z)/p_\theta(x)$ is hard to compute because $p_\theta(x) = \int_z p_\theta(x|z)p_\theta(z)\mathrm{d}z$ is intractable. The goal is to find a function $q_\phi(z|x)$ that best approximates $p_\theta(z|x)$ with respect to Kullback-Leibler (KL) divergence. The naive architecture of a VAE, depicted in Fig. 2.1a, contains a probabilistic encoder and probabilistic decoder that sample from $q_\phi(z|x)$ (conditioning on input) and $p_\theta(x|z)$ (conditioning on latent representation), respectively.

The KL divergence can be further expanded as

$$
\begin{aligned}
&\mathrm{KL}(q_\phi(z|x)||p_\theta(z|x)) \\
=&\mathbb{E}_q[\log(q_\phi(z|x))] - \mathbb{E}_q[\log(p_\theta(z|x))] \\
=&\mathbb{E}_q[\log(q_\phi(z|x))] - \mathbb{E}_q\left[\log \frac{p_\theta(x, z)}{p_\theta(x)}\right] \\
=&\mathbb{E}_q[\log(q_\phi(z|x)) - \log p_\theta(x, z)] + \log p_\theta(x) \\
=&\mathbb{E}_q[\log(q_\phi(z|x)) - \log p_\theta(z) - \log p_\theta(x|z)] + \log p_\theta(x) \\
=&\mathrm{KL}(q_\phi(z|x)||p_\theta(z)) - \mathbb{E}_q[\log p_\theta(x|z)] + \log p_\theta(x),
\end{aligned}
$$

which we can rewrite as

$$\begin{aligned}\log p_\theta(x) &= \mathrm{KL}(q_\phi(z|x)||p_\theta(z|x)) + (\mathbb{E}_q[\log p_\theta(x|z)] - \mathrm{KL}(q_\phi(z|x)||p_\theta(z))) \\ &= \mathrm{KL}(q_\phi(z|x)||p_\theta(z|x)) + \mathbb{E}_q\left[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\right],\end{aligned}$$

where the second term on the right is the evidence lower bound (ELBO) of $\log p_\theta(x)$. Therefore, minimizing

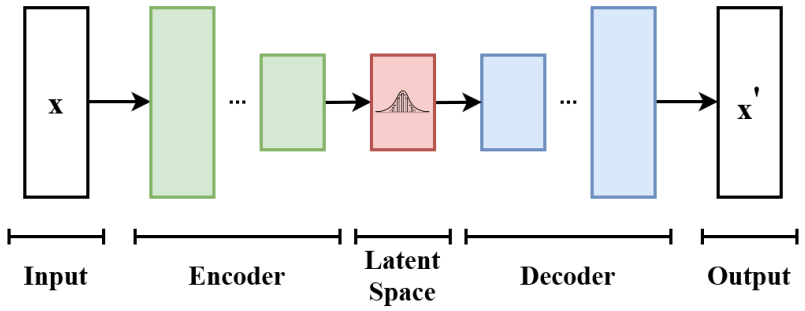
$$L_{\theta,\phi} = -\log p_\theta(x) + \mathrm{KL}(q_\phi(z|x)||p_\theta(z|x)) = -ELBO$$

simultaneously maximizes ELBO (and therefore the likelihood on observed data $p_\theta(x)$) and minimizes the KL divergence between $q_\phi(z|x)$ and $p_\theta(z|x)$.

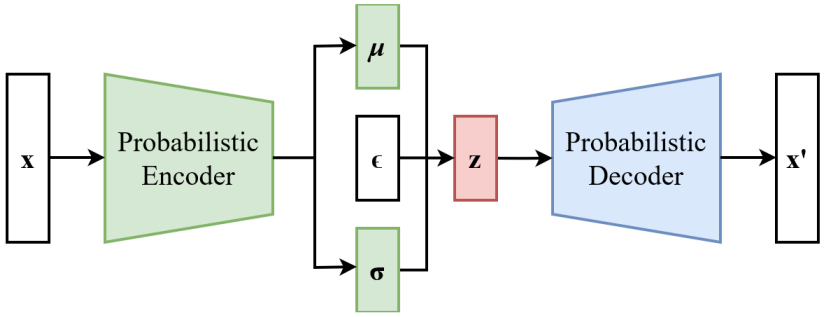
Finally, we reparametrize q_θ so that it can our loss can be minimized using backpropagation and stochastic gradient descent. We assume that $q_\phi(z|x)$ is reallly $\mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are output by the encoder. The modified VAE architecture is given in Fig. 2.1b. This completes the construction of the model.

It is empirically known that a well-trained VAEs can identify clusters from the input space in the latent space even if the input space is high-dimensional and there is high noise in the data. This is because a VAE model can distinguish between high and low relative density. These good properties make VAE a suitable candidate for dimensional reduction and likelihood estimation.

The VAE we use in this study is modified from an implementation in PyTorch from Yaniv Jacoby.



(a) Naive VAE



(b) Reparameterized VAE

Figure 2.1: Architecture of VAE. Images taken from https://en.wikipedia.org/wiki/Variational_autoencoder

3

Dataset and methodology

1 Patients and feature extraction

In this study, we retrospectively analyze 40 pediatric patients (age: 2-21 years) with medical refractory epilepsy who underwent treatment at Boston Children's Hospital. All patients included meet the following criteria:

- Availability of high-quality pre-surgical and post-surgical MRI recordings, which allow us to build a 3D model of the brain and measure the Euclidean distance between each electrode and the closest margin of the resected area.* Adjustments are made to account for brain shifts and other factors that could introduce inaccuracies in our measurements.
- Availability of five continuous minutes of high-quality interictal epileptiform icEEG recordings, which allow us to model the functional connectivity of the brain as a graph of the electrode network;
- Labelling of whether electrodes are in seizure onset zone (SOZ) as determined by clinicians by ictal recordings. SOZ labels are generated independently from our study; they are observed from seizures that are not included in the five minutes of icEEG recording that we use to measure functional connectivity. Moreover, since the total number of recorded seizures vary greatly between patients and are not necessarily recorded, the number and relative frequency of seizure onsets is not used in our study. For future studies, this information can become useful if we develop ways to standardize it properly between different patients.
- Engel surgical outcome scale as determined by clinicians during post-surgical follow-ups, where class 1 (completely seizure free) are considered as good outcome and all other classes are considered as poor outcome. Among the patients in our dataset, 25 had good outcome and 15 had poor outcome.

The five minutes of recording is further divided into 100 disjoint three-second epochs, and feature extractions are performed on a per-epoch basis.

1.1 Spike, hybrid, and silent data

An experienced reader is tasked with identifying spikes on every three-second epoch. For a visualization of spikes as captured by EEG, see Figure 3.1. The presence of spikes is almost always indicative of pathological behavior in the brain [13]. Epochs are classified as spike and silent depending on the

*For the purpose of our study, we consider electrodes with Euclidean distance $<$ 5mm to the resection zone as resected.

presence and absence of spikes. Using this classification, we obtain three different datasets: 1) spike data, which contains only spike epochs, 2) silent data, which contains only silent epochs, and 3) hybrid data, which contains both. The proportion of spike and silent epochs vary greatly between the patients, two of which have no spike epochs and one of which have no silent epochs. As a result, when performing analysis on spike data and silent data, we are considering one or two fewer patients.

1.2 Functional connectivity and centrality

Within each epoch, we perform Fourier transform and divide our coefficients into four frequency bands: high-delta (3-4 Hz), spike (4-35 Hz), ripple (80-250 Hz), and fast-ripple (250-500 Hz). Ripple and fast-ripple frequency oscillations are collectively known as high-frequency oscillations and considered as an important biomarker for epileptogenic tissue.[23,15] We compute the squared correlation of the spectral density between every pair of electrodes to measure functional connectivity. Mathematically, if we denote the signal of electrode i at time t as $x_i(t)$, then for each three-second epoch $[t_\ell, t_{\ell+1}]$, we can compute

$$\begin{aligned}
\hat{x}_i^{(\ell)}(f) &= \int_{t_\ell}^{t_{\ell+1}} e^{-i2\pi ft} x_i(t) \mathrm{d}t \\
S_{i,j}^{(\ell)}(b) &= \int_{l_b}^{u_b} x_i^{(\ell)}(f) x_j^{(\ell)}(f) \mathrm{d}f \\
C_{i,j}^{(\ell)}(b) &= \frac{S_{i,j}^{(\ell)}(b)^2}{S_{i,i}^{(\ell)}(b) S_{j,j}^{(\ell)}(b)},
\end{aligned}$$

where l_b and u_b denote the lower and upper bound of frequency band b, and $C_{i,j}^{(\ell)}(b)$ is our estimate of the functional connectivity between electrodes i and j during $[t_\ell, t_{\ell+1}]$ within the frequency band b. The measure is then averaged over all time epochs in the dataset for a given patient.

For each frequency band, we first compute the mean connectivity of every electrode of the patient. Then, the functional connectivity matrix of the patient is converted into a graph consisting of the electrodes as vertices and the inverse of functional connectivity measures as edge weights following the approach described by [19]. Under this set-up, edges with the smallest weight correspond to higher synchronization between electrodes, so that edge weights model costs of signals to travel and shortest paths in the graph correspond to more likely routes through which the neurons communicate. The MST of the graph is computed using Kruskal's algorithm,[9] and we generate a subgraph using all edges of the MST with weight set to 1. Four cenrtality measures–degree, betweenness, closeness, and eigenvector–are computed on the MST. Across all frequency bands, this gives us $(5 \text{ measures}) \times (4 \text{ bands}) = 20$ features.

As another variant to the features, we de-noise the connectivity graph via fuzzy thresholding. More concretely, we create a subgraph using only the edges that have the 25%, 50% or 75% most functional connectivity, a.k.a. the lightest edge weights by our construction, and we set the weight of all kept edges as 1. We then similarly compute the four centrality measures by extracting a MST from the subgraph. This gives us $(4 \text{ measures}) \times (4 \text{ bands}) \times (3 \text{ thresholds}) = 48$ more features, for a total of 68 features.

Finally, we normalize each feature of all electrodes within the patient to have mean 0 and standard

deviation 1. This step is especially necessary because the number of electrodes varies greatly for the patients, so measures of centrality that scale with the size of the graph, such as mean connectivity and degree, need to be properly standardized in order to be meaningfully compared between the patients.

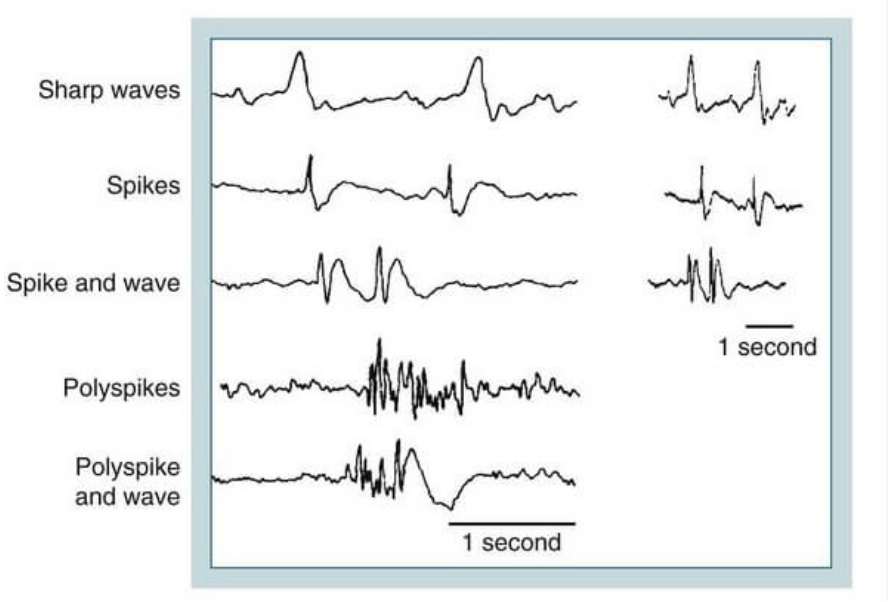


Figure 3.1: Common spikes as observed in EEG. Image taken from https://eastneurology.com.au/eeg-electroencephalogram-brain-wave-tests/.

2 SOZ prediction

The first goal of our study is to understand whether the features we have extracted are predictive of whether an electrode is in the SOZ.

2.1 Single feature analysis

For a simple sanity check on whether our features contain information about whether electrodes are in SOZ, we verify that the SOZ and non-SOZ electrodes exhibit significant differences in the derived features. In particular, many studies suggest that high-centrality electrodes are associated with the epileptogenic zone.[21,12] Since such measures cannot be compared between patients, all comparisons must be made with paired data, i.e., between the SOZ and non-SOZ electrodes for each patient. As such, we compute the mean for each of our 68 features and run hypothesis tests with one-sided alternative of SOZ electrode feature means being greater than the means of non-SOZ electrodes.

Differences between SOZ and non-SOZ electrodes could also depend on the patient outcome. To assess this, we compute for each patient the difference between the mean of the feature value for SOZ electrodes and the same mean for non-SOZ electrodes. The distribution of this statistic among good- and poor-outcome patients is then compared.

2.2 Classification

Since single functional connectivity measures can predict whether an electrode is SOZ, the next step is combine these measures to build a model for classification. As previously discussed, identification of SOZ electrodes helps clinicians decide which sections of the brain to resect for the patient to become seizure-free. As such, accurate classification of SOZ electrodes from short time periods of EEG recording is clinically significant.

For all of the experiments below, we take a leave-one-out approach on the patients: for each patient i, we train our model on the electrodes of all patients except i, and use the trained model to make predictions on all electrodes of patient i. To prevent overfitting, we also train iterations of our model using subsets of the features. Specifically, we compare the results of a logistic regression model using:

- All features.
- Only the functional connectivity from the raw graph and not any of the features derived from the de-noised graph. Denote this by `lr_fc`. We can similarly do this for any of the variations below, in which case we use `_thd` to denote that we are also considering the threshold graph and the absence thereof to denote that we are only considering the raw graph features.
- Use only features from a particular frequency band, i.e. one of high-delta, spike, ripple and fast ripple. Denote this by `lr_BAND`.
- Use only features from a particular functional connectivity measure, i.e. one of mean connectivity, betweenness, closeness, degree, and eigenvalue. Denote this by `lr_FCMEASURE`.
- Use the top four principal components. The choice of four components is due to evidence from Section 1 and Section 4.1 in Chapter 4. Denote this by `lr_pca`.
- Use extreme gradient-boosted trees instead of logistic regression. Denote this by `xgb`.

AUC values from different models are compared to determine the relative predictive power from each of the approaches. To evaluate the performance of actual predictions rather than predictde probabilities, we choose the threshold that maximizes the Youden's index[22], a common heuristic that measures

$$J = \text{sensitivity} + \text{specificity} - 1.$$

We remark that many other criteria exist for the choice of thresholds, and using the one index is hardly sufficient for summarizing the performance of a diagnostic model. In this paper we will use AUC as the metric for comparing the performance between different models rather than dwelling on statistics derived from the confusion matrix produced by a specific threshold.

2.3 Dimensional reduction and clustering

In addition to conventional supervised classification methods, we are interested in asking whether unsupervised clustering methods can give us useful insights via dimensional reduction. There are two possible outcomes from this approach. In the first scenario, the latent space representation of the data shows clear, visual clusters that we can attribute to either SOZ or non-SOZ. In the second scenario, we discover that latent space representation of SOZ and non-SOZ electrodes show distinct yet overlapping distributions. In this case, clean clustering is not possible, but the latent space representation remain useful in predicting the relative epileptogenicity of an electrode. See Section 3.

To check for clustering, PCA, TSNE, and VAE are trained on all electrodes, and the latent space is visualized. To test for differences between the distribution of SOZ and non-SOZ electrodes, the models are also fitted only on SOZ electrodes, and the resulting latent space representation of SOZ and non-SOZ electrodes are compared. In particular, after fitting a PCA on the non-SOZ electrodes and then transforming all electrodes, we can consider the distribution of the Mahalanobis distance of SOZ and non-SOZ electrodes on the latent space. Given observations that fol-

low a distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, the Mahalanobis distance is then defined as $d(\boldsymbol{x}) = \sqrt{\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})}$.[10] Mahalanobis distance can be considered as a generalization of the z-score. In particular, it is most interpretable when $\boldsymbol{x}$ is normally distributed, so that $p(\boldsymbol{x}) \propto \exp\left(-d(\boldsymbol{x})^2/2\right)$ and data points with larger Mahalanobis distance land in lower-density regions. Therefore, if SOZ electrodes have significantly larger Mahalanobis distances than non-SOZ electrodes in the latent space derived from PCA fitted to non-SOZ electrodes, then we can conclude that the distribution of SOZ and non-SOZ electrodes are quite different. Similarly, for a VAE model trained on non-SOZ electrodes, we can use Monte Carlo methods to estimate the likelihood of all electrodes and then compare the likelihood of SOZ and non-SOZ electrodes.

3 Outcome prediction

Since the primary application of our study is to provide guidance for determining the resection zone, one may ask whether the models we have derived thus far can help achieve this goal. Indeed, both the predicted SOZ probability from classification models and the distribution of centrality features on the latent space can be used to estimate epileptogenicity of the inidividual electrode. More concretely, denote the set of all electrodes for patient i by S_i and denote the set of resected electrodes by R_i. Assume that we can use fuzzy inference to establish a criterion for whether an electrode is considered epileptogenic based on our features above; call this set E_i. Then, good-outcome patients are expected to have higher proportion of resection among the epileptogenic electrodes, which we denote by $re_i = |E_i \cap R_i|/|E_i|$, whereas the proportion of non-resection among the nonepilep-

togenic electrodes, denote by $rn_i = |E_i^C \cap R_i|/|E_i^C|$, should not be different between good- and poor-outcome patients.

The choice of (re_i, rn_i) as the features to study here is motivated by a variety of factors. On one hand, they are very interpretable and align with clinical understanding of epilepsy treatment: SOZ electrodes are more likely to be epileptogenic, and the removal or most epileptogenic electrodes increase the likelihood of the patient becoming epilepsy-free post surgery. On the other hand, the predictive power of single feature is easy to measure, whereas involving multiple features can easily lead to overfitting and unstable predictions given the size of the dataset. An alternative feature we have considered is to take the mean of the epileptogenicity feature within or outside the resection zone, and taking their difference. However, there is not significant evidence that any feature derived along this route has a different distribution between good- and poor-outcome patients. In the future, with the inclusion of more patient profiles, there is room for developing more complex interpretable features and models for combining the features.

The following features of epileptogenicity are explored:

- SOZ labels provided by the dataset;
- predicted SOZ probability from classification models trained on SOZ labels;
- Mahalanobis distance of latent space representation from PCA trained on non-SOZ electrodes, where larger Mahalanobis distance are associated with out-of-distribution data points, higher SOZ probability, and higher epileptogenicity;[†]

[†]The results from training the PCA model on SOZ electrodes only is not included because the resulting latent space representation does not approximately follow a multivariate normal distribution in shape, and Mahalanobis distance measured on such datasets is not very meaningful.

- Estimated likelihood from VAE and KDE trained on non-SOZ electrodes, where lower likelihood is associated with higher SOZ probability and higher epileptogenicity;
- Estimated likelihood from VAE and KDE trained on SOZ electrodes, where higher likelihood is associated with higher SOZ probability, and higher epileptogenicity.

We remark that the inclusion of KDE alongside VAE is due to the fact that both provide likelihood estimates for the data. However, VAE is expected to have better performance because KDE does not perform as well in high dimensions. We also observe that in when SOZ labels are only used to select our training set, our epileptogenicity predictions are more robust to noise in the SOZ labels that can distract the model from the true goal.

The predictive power of re_i and rn_i for patient outcome are measured in AUC. When the feature of epileptogenicity is continuous, we establish a concrete threshold for epileptogenicity by grid searching over the 20-quantiles of the distribution of the feature.

4

Results

1 Data exploration

A Kolmogorov-Smirnov test is applied to all 68 features in Table 4.1 and shows that the features are not normally distributed. Therefore, nonparametric hypothesis tests such as Wilcoxon signed rank and Mann-Whitney U rank are used in place of t-tests, as the latter has less power when the data

does not follow a normal distributed.

Correlation is measured across frequency bands, centrality measures, and de-noising thresholds for all patient's electrodes in hybrid datain an effort to understand the relationship between features. Select results are given in Appendix A and they reflect the general trend. Centrality measures are highly correlated within the same frequency band, whereas centrality measures from different frequency bands are largely uncorrelated. Different de-noising thresholds produce similar centrality measures, which are quite different from those produced without de-noising.

2 SOZ predictions

2.1 Single feature analysis

For spike data, 60 out of 68 features show significant difference (p-value<0.05), and 46 show highly significant difference (p-value<0.001), as is recorded in Section 2.1. For hybrid data 58 out of 68 features are significantly different, and for silent data include number out of 51 features are significantly different. As expected, spike data has the highest centrality measure difference between SOZ and non-SOZ electrodes, though hybrid and silent data still have many significantly different features. This suggests that SOZ electrodes exhibit differences even in the absence of epileptic activity. Such findings are valuable for most patients for whom silent and hybrid epochs are more frequent.

As seen in Section 2.1, most features do not exhibit a significant difference in the centrality measure difference between SOZ and non-SOZ electrode between good- and poor-outcome patients.

	$p < 0.05$	$p < 0.01$	$p < 0.001$
spike	66	63	61
hybrid	64	63	62
silent	63	62	62

Table 4.1: Counts of significant p-value among 68 features from Kolmogorov-Smirnov test for feature normality. Small p-value indicates that the features are not normal.

	$p < 0.05$	$p < 0.01$	$p < 0.001$
spike	60	55	49
hybrid	58	57	42
silent	51	42	32

(a) Counts of significant p-value among 68 features from one-sided Wilcoxon signed-rank test between the mean feature values of SOZ and non-SOZ electrodes. Small p-value indicates that SOZ electrodes have higher value than non-SOZ electrodes.

	$p < 0.05$	$p < 0.01$	$p < 0.001$
hybrid	2	0	0
spike	3	1	0
silent	3	1	0

(b) Counts of significant p-value among 68 features from two-sided Mann-Whitney U rank test between good- and poor-outcome patients. Small p-value indicates that the distinction between SOZ and non-SOZ electrodes is different between good- and poor-outcome patients.

Table 4.2: Counts from single-feature tests on SOZ versus non-SOZ electrodes and good- versus poor-outcome patients.

The few that do are likely due to p-value having the uniform distribution under the null hypothesis. Therefore, we have reason to believe that for the purpose of identifying SOZ electrodes, we do not need to consider the distinction between good- and poor-outcome patients.

3 Supervised classification of SOZ electrodes

The logistic regression model is trained for up to 500 iterations until convergence. For XGB, the parameters are determined using a grid-search with five-fold cross-validation on all patient electrodes with area under ROC as the evaluation metric. Since this process is computationally expensive, we only do it once. While this is not standard practice, it is the best we can do given that we do not have

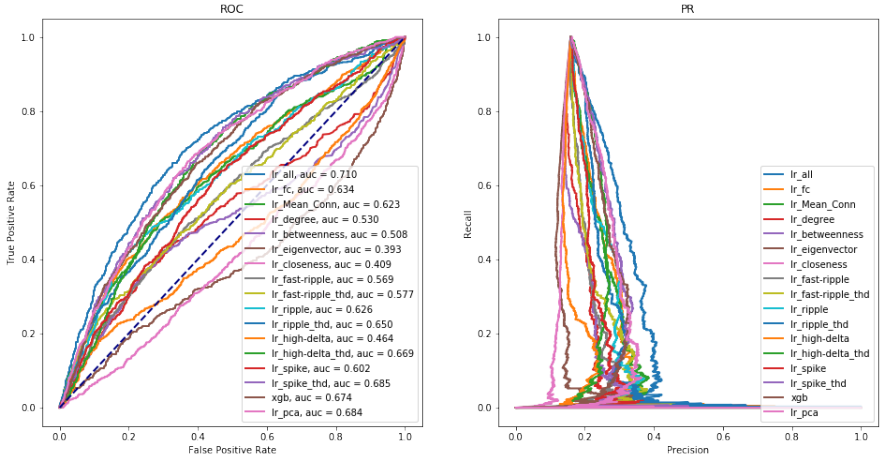
Act. \ Pred.	non-SOZ	SOZ
non-SOZ	2157	1193
SOZ	197	438

Table 4.3: Confusion matrix of SOZ electrode predictions on spike data using logistic regression model on all features.

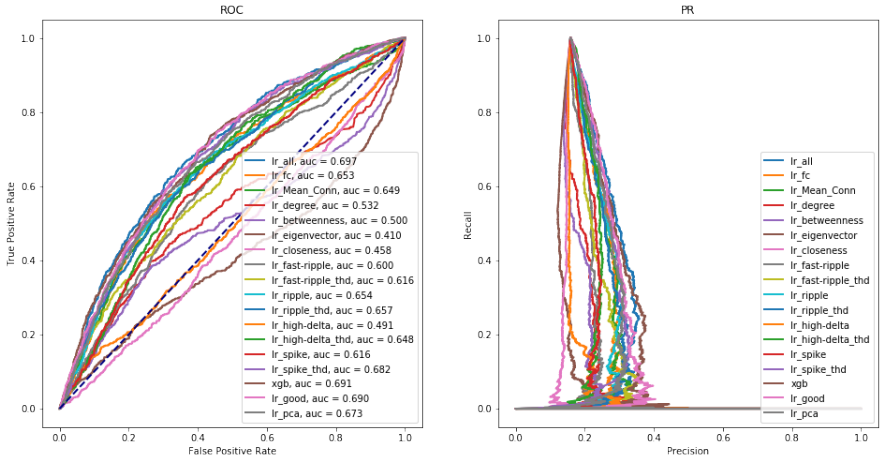
enough data to split into training, validation, and testing sets for this purpose.

Performance metrics of different models for SOZ classification are given in Table B.1 and visualized in Figure 4.1. Using spike data generally gives us the best performance with AUC 71.0% and AUC-PR 27.7%. The logisitic regression model using all features shows better performance across all datasets than almost other models by these metrics. Lastly, the non-linear XGB model does not show significant performance gain over the logistic regression.

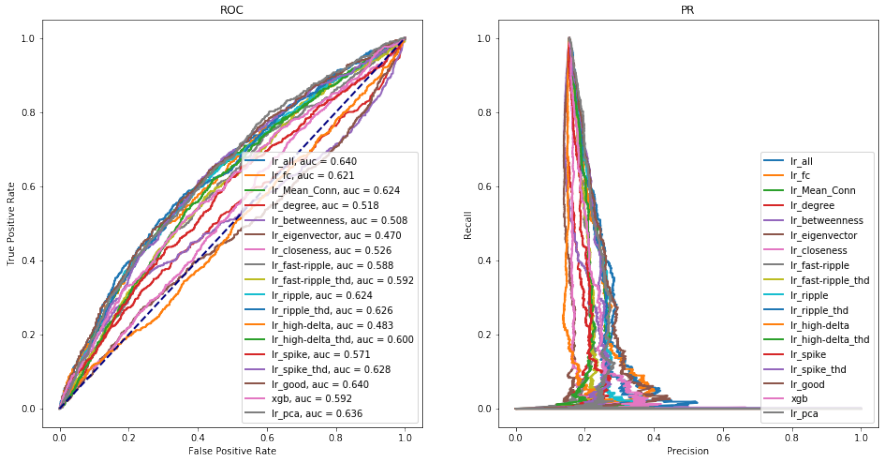
Table 4.3 gives the confusion matrix from the all-feature logistic regression model using the Youden's index-maximizing threshold on hybrid data. This model gives sensitivity 0.271, specificity 0.910, positive predictive value 0.620, negative predictive value 0.686, and accuracy 0.675. To put these values in context, we remark that only 16% of the electrodes are actually SOZ, so we are inflating the proportion of epileptogenic electrodes. However, Section 5.1 show that obsessing over a good fit of the SOZ electrodes is in fact not in line with our ultimate goal of predicting whether a patient has good post-surgical outcome. Therefore, our predictions here, though imperfect with respect to the ground truths that they are trained on, can potentially remain meaningful.



(a) spike data



(b) hybrid data



(c) silent data

Figure 4.1: AUC and PR curves across many different models for spike data, silent data, and hybrid data.

4 Dimensional reduction for unsupervised clustering

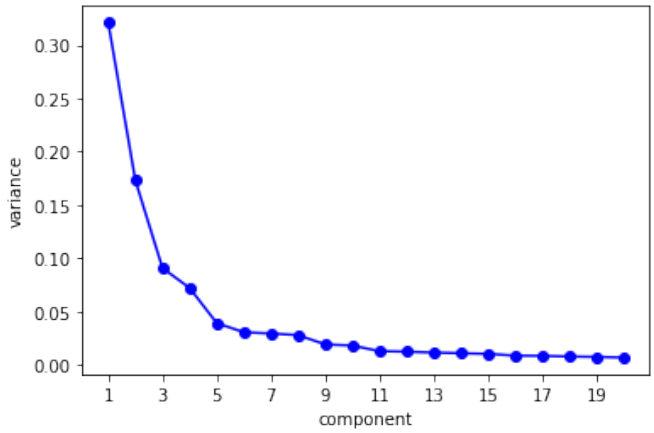
4.1 Using top principal components

We start off with the linear PCA model. Graphing the explained variance ratio for hybrid data, we get Figure 4.2a, which along with the interpretability of the data coming from four frequency bandwidths prompts us to consider the first four components, which explain 66% of the variance in our data. A pair plot of these four components is given in Figure 4.2b. We applied the same methodology to spike data and silent data, which gave similar results that we omit here.

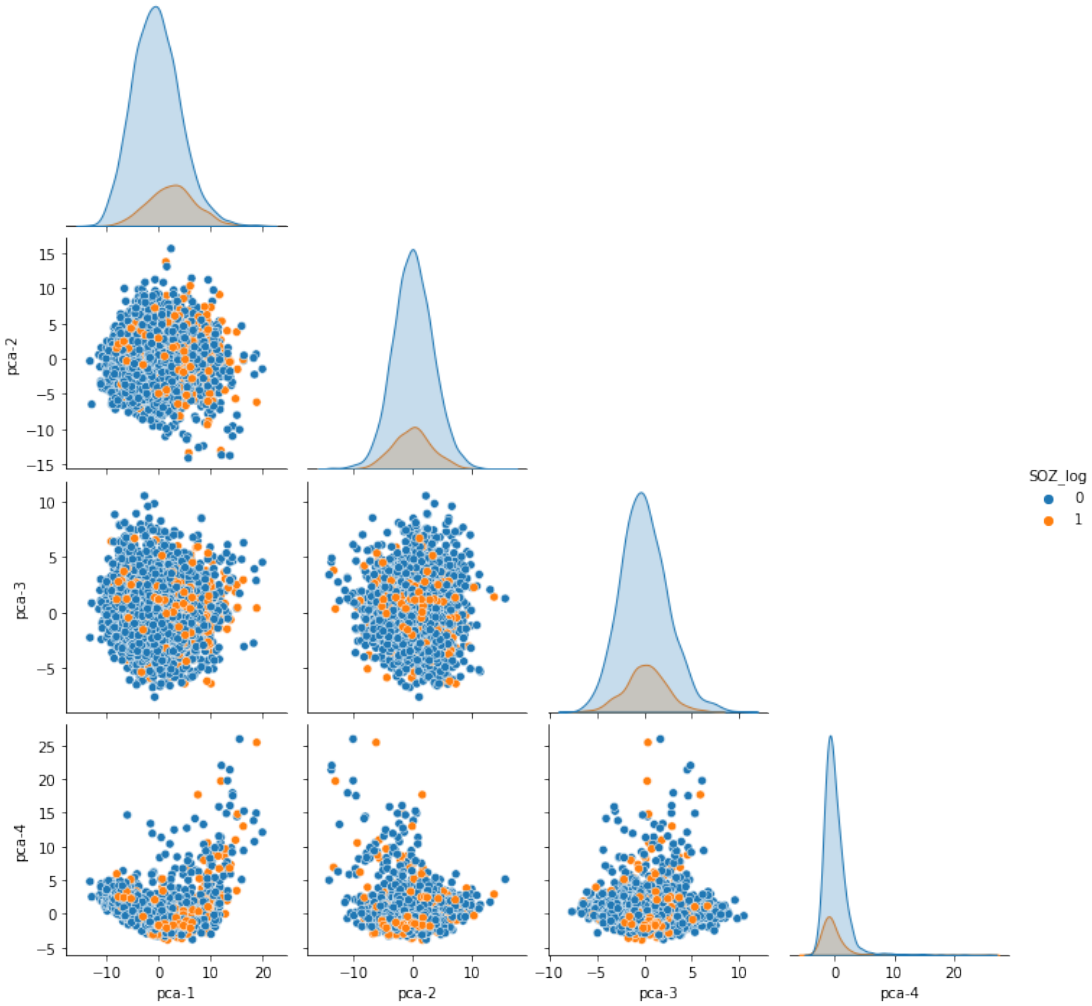
Nevertheless, there is significant evidence of a difference between the distribution of SOZ and non-SOZ electrodes. The Mahalanobis distance of the latent representations of all electrodes from a PCA trained on non-SOZ electrodes is plotted in Figure 4.3. Applying two-sided Mann-Whitney U rank test shows highly significant difference (p-value<0.001) for spike dataand significant difference for (p-value<0.05) for hybrid data and silent data. Therefore, larger Mahalanobis distance can imply a higher probability for an electrode to be in the seizure onset zone.

4.2 Using nonparametric methods

A TSNE model with perplexity 30 on hybrid data gives the following fit, which again fails to show clear clustering. Applying TSNE to silent data and hybrid data similarly shows no clustering.



(a) Explained variance of top 20 components.



(b) Pair plot of the top 4 components.

Figure 4.2: Results from applying PCA to hybrid data.

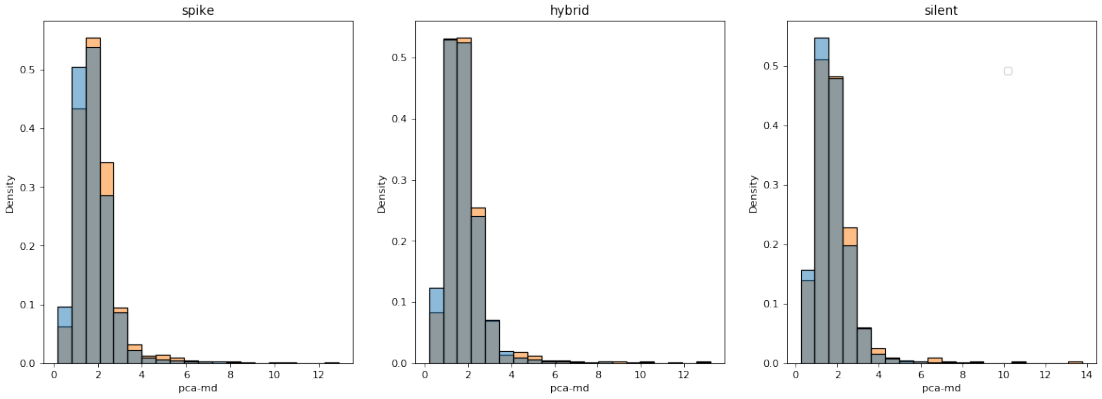


Figure 4.3: Mahalanobis distance on PCA four-component latent space for SOZ (orange) and non-SOZ (blue) electrodes.

4.3 Using the latent space from VAE

Finally, we consider a VAE model. We use a decoder and an encoder with three hidden fully connected layers of 50 electrodes and leaky ReLU activation. We use a grid search to tune hyperparameters: we split all our data into a training set (roughly two-thirds of the patients) and testing set (roughly one-third of the patients), and then we select the model that gives the best estimated likelihood on the test set while having the smallest latent dimension. This ensures that we achieve a good fit while minimizing overfitting. We end up with the optimal latent dimension of 4, output variance of 0.25.

Training it on all electrodes in hybrid datagives latent representation in Figure 4.5, which again shows no meaningful clustering. Nevertheless, similar to Section 4.1, if we graph the estimated likelihood of all electrodes from a VAE trained only on non-SOZ we again have evidence of a difference in distribution between SOZ and non-SOZ electrodes, as demonstrated in Figure 4.6.

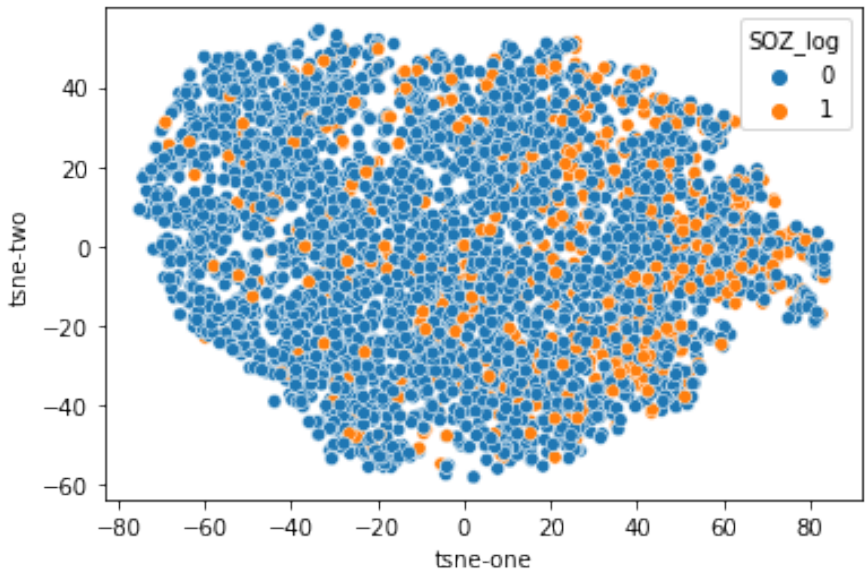


Figure 4.4: Reduction of hybrid data to two dimensions using TSNE.

5 Identification of patient outcome

For simplicity, we will only present results from hybrid data, as it exists for all 40 patients.

As a sanity check, we first observe that among the 40 patients, the proportion of resected electrodes $r_i = |R_i|/|S_i|$ is predictive of post-surgical outcome with AUC 70.4%; a histogram is given in Figure 4.7. In particular, all patients with $r_i > 40\%$ have good outcome, whereas all patients with $r_i < 10\%$ are significantly more likely to have poor outcome. When r_i is very large (small), both re_i and rn_i are inflated/deflated, which confounds the performance of our approach. Therefore, we truncate our dataset and only consider patients with $10\% \leq r_i \leq 40\%$, as within this range, there is an even mixture of patients with good and poor post-surgical outcome, and we expect prediction based on re_i to be most interesting. Among these 28 patients, r_i only has 52.8% AUC.

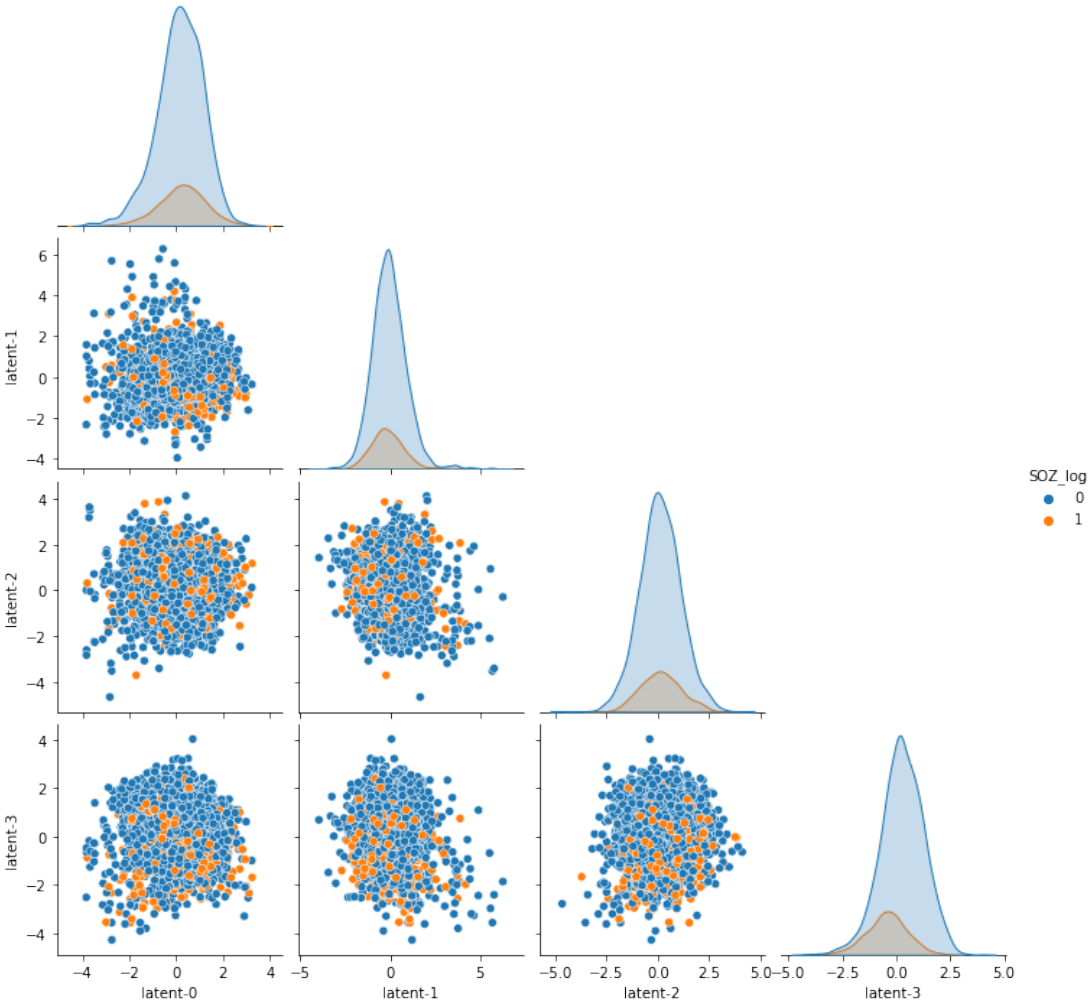


Figure 4.5: Four-dimensional latent space from VAE trained on hybrid data.

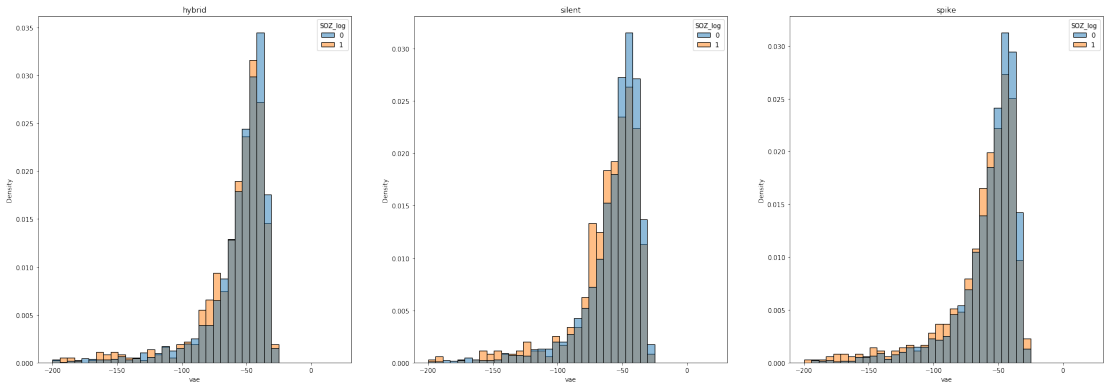


Figure 4.6: Estimated log likelihood of all electrodes from VAE trained on non-SOZ elecrodes hybrid data.

The summarized results from the following different criteria for epileptogenicity are given in Figure 4.8 and Table B.2.

5.1 Supervised classification

When SOZ labels are used to determine epileptogenic electrodes, the AUC from re_i and rn_i are 53.9% and 56.1%, respectively, which shows us that our SOZ labels do not give very meaningful predictions of patient outcome. We then consider using the probability of being SOZ as predicted by logistic regression and extreme gradient-boosted trees. Since SOZ electrodes are more likely to be epileptogenic, we consider an electrode as epileptogenic if its predicted SOZ probability is above a certain threshold. This approach does do not have significant prediction power because the AUC of re lingers around 0.5 and is often less than the AUC of rn.

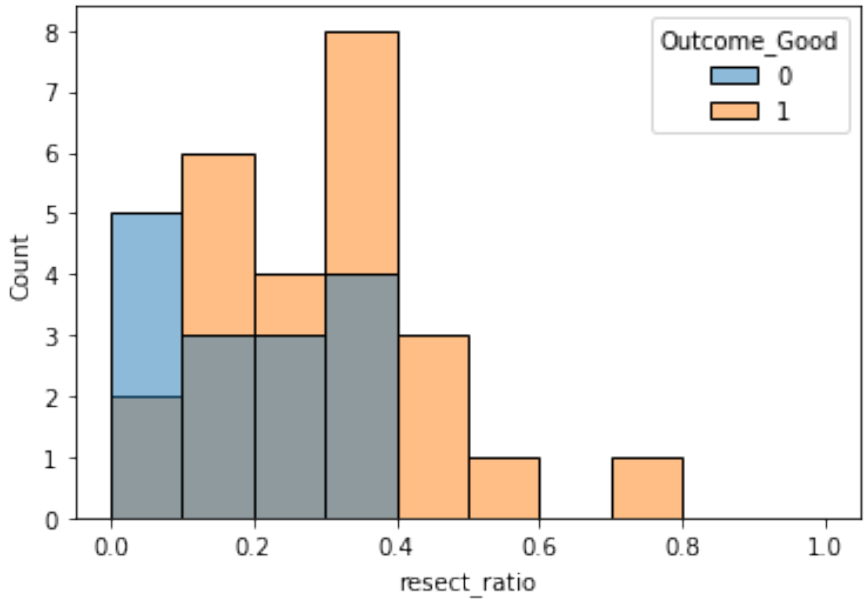


Figure 4.7: Proportion of resected electrodes $|R_i|/|S_i|$ for all 40 patients in hybrid data.

5.2 Dimensionality reduction and probabilistic methods

An optimal KDE bandwidth is chosen using 20-fold grid search with the objective of maximizing the estimated likelihood on the data. The hyperparameters for the remaining models are as described previously.

The results are visualized in Figure 4.8 and detailed in Table B.2. The AUC of re is 0.744 while the AUC of rn is 0.500 when we consider the top 10% most epileptogenic electrodes as measured by a PCA. Also, the AUC of re is 0.656 while the AUC of rn is 0.533 when we consider the top 25% most epileptogenic electrodes as measured by a VAE trained on non-SOZ electrodes. Using Youden's index as a heuristic for choosing the threshold, the VAE model gives sensitivity 0.611, specificity 0.8, positive predictive value 0.846, negative predictive value 0.533, and accuracy 0.679.

The PCA model gives sensitivity 0.556, specificity 1.0, positive predictive value 1.0, negative predictive value 0.556, and accuracy 0.714. See the confusion matrix in Table 4.4.

act. \ pred.	VAE		PCA	
	poor	good	poor	good
poor	8	2	10	0
good	7	11	8	10

Table 4.4: Confusion matrix for VAE- and PCA-derived outcome predictions.

Figure 4.8: AUC of re_i and rn_i in predicting patient outcome under different metric and threshold of epileptogenicity.

5

Conclusion

There are two major contributions of this paper. First, it demonstrates that functional connectivity centrality measures extracted from five minutes of interictal icEEG recording contain valuable information (AUC 0.70) for identifying the SOZs. The current standard practice relies on clinicians to visually identify individual electrodes as SOZ during every ictal recording. This is both inaccurate due to the human factor and inefficient because ictal recordings are sparse and not every represen-

tative pattern of seizure onset is guaranteed to be recorded by the icEEGs. Our approach provides a huge improvement over the standard practice and can be used to advise future clinicians in epilepsy surgery. Second, this paper reveals that likelihood-based diagnostic models on functional connectivity centrality can predict patient outcome. In particular, our model scores the epileptogenicity of individual electrodes, and the proportion of resection among predicted epileptogenic electrodes is predictive of patient post-surgical outcome (AUC 0.74) when the naive proportion of resection among all electrodes or only SOZ electrodes is not (AUC 0.53, 0.54). This suggests that our model can overcome the insufficiency of clinically determined SOZ labels and offer a better understanding of the EZ. Our diagnostic model can provide useful insights for the success of resective surgery, thereby opening new doors for MRE treatment.

For future work, many interesting approaches can also be tried with the inclusion of more data. For instance if we can collect more patient profiles, deep neural networks can potentially give us even better performance on SOZ classification. With our current dataset size of 4000 data points, deep neural networks are likely to overfit, whereas simpler networks are unlikely to perform better than the models we have already attempted. Meanwhile, our current analysis of the functional connectivity network produces centrality measures for each individual electrode. While the centrality measures that we use do incorporate information about the network on a less local scale, a full holistic understanding of the impact of resective surgery on the patient's functional connectivity network remains amiss. Recent advances in machine learning such as graph neural networks, which process data that is best represented using the graph data structure, is suitable for our use case. However,

at the current dataset size of 40 patients, training such complex models is largely intractable. In general, the field of epilepsy research is limited by the lack of standardized guidelines for data collection. Researchers often use patient data from the university and/or hospital with which they are affiliated. Such datasets are usually not public, whereas public datasets often do not fit the criteria for inclusion in order to be used for specific studies. Nonetheless, functional connectivity network analysis in MRE treatment is a burgeoning field of research, and the continued development of new approaches for analyzing the network hold promises for finding the cure for epilepsy.

A

Centrality measures correlations

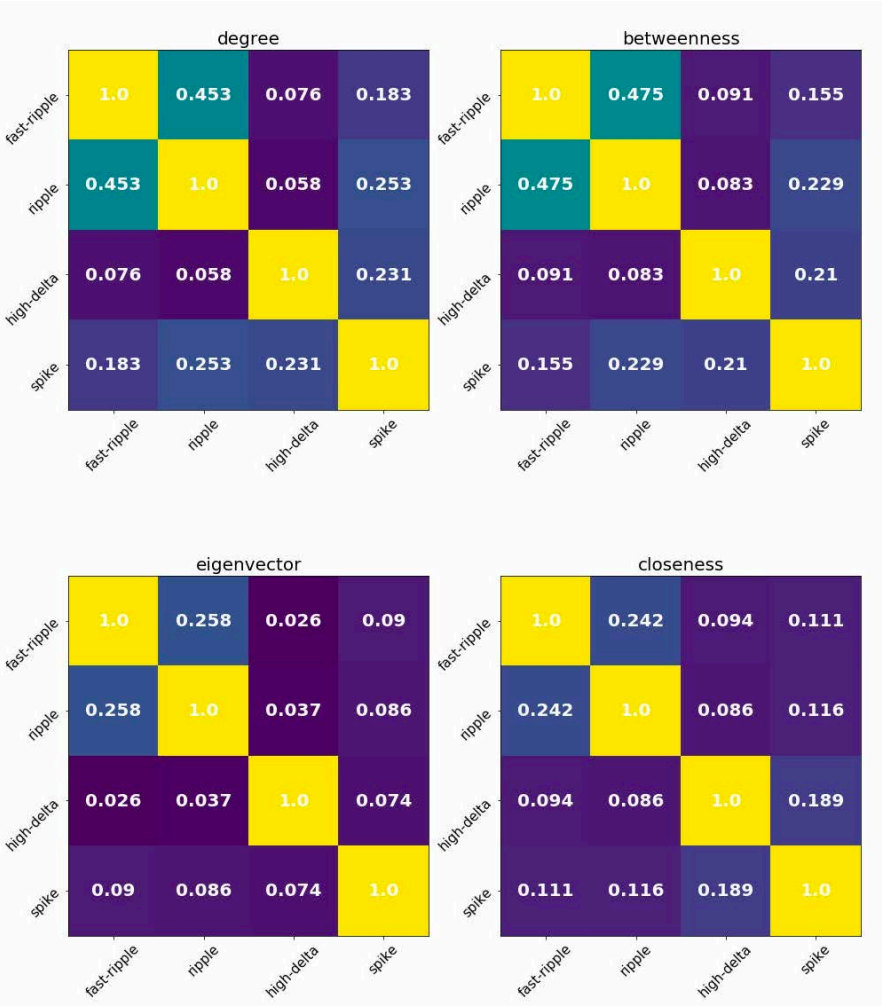


Figure A.1: Correlation between frequency bands.

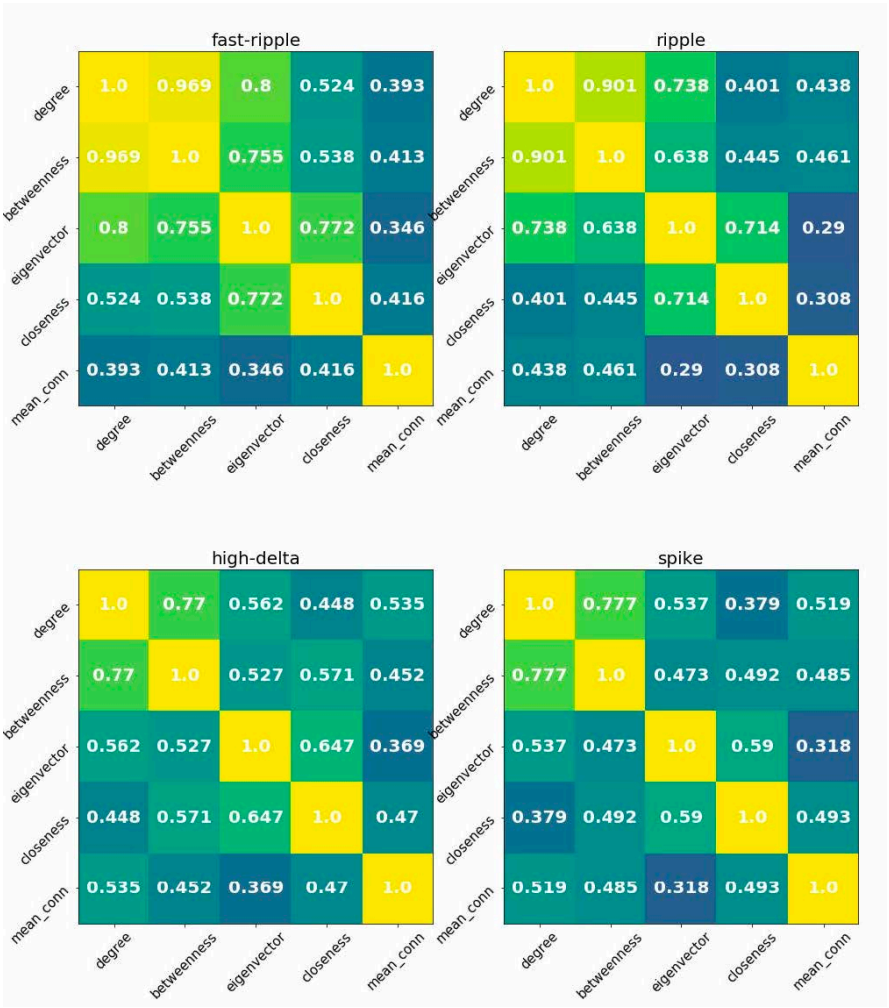


Figure A.2: Correlation between centrality measures.

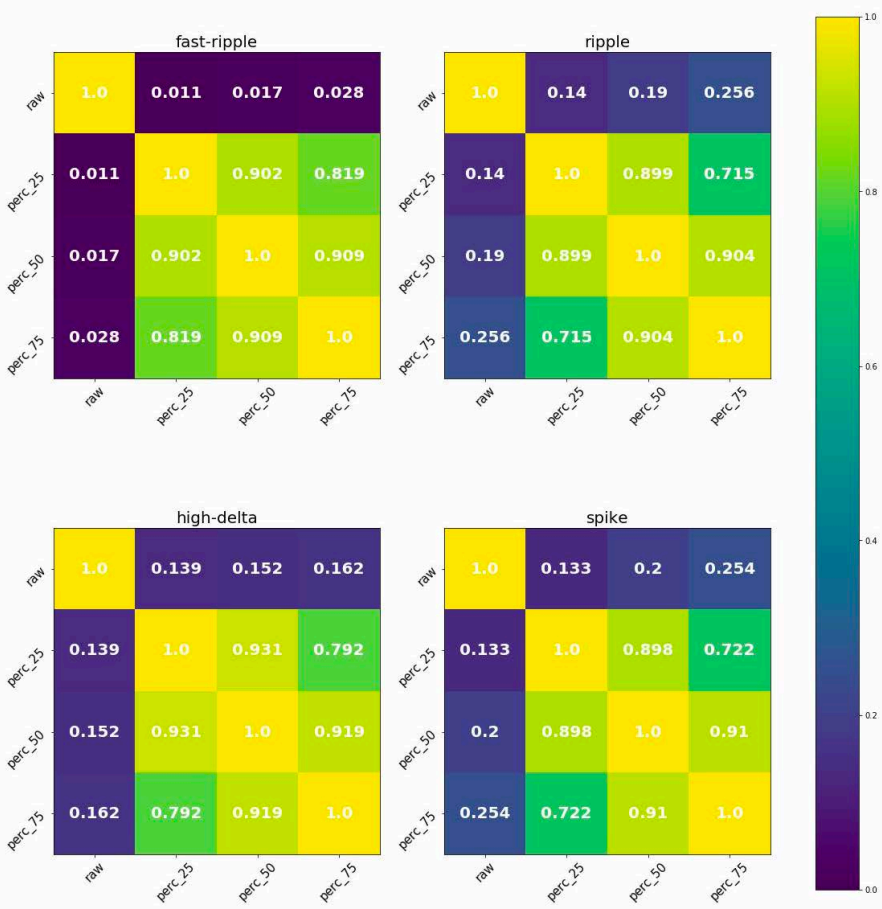


Figure A.3: Correlation between de-noising thresholds using degree as centrality measure.

B

Performance metrics

epoch model	roc hybrid	 silent	 spike	pr hybrid	 silent	 spike
lr_all	0.696509	0.639746	0.710392	0.276770	0.243700	0.303476
lr_mean_conn	0.648716	0.623944	0.623485	0.245934	0.220346	0.246927
lr_betweenness	0.500349	0.508362	0.508020	0.180441	0.184184	0.192873
lr_betweenness_thd	0.645562	0.621908	0.652452	0.254231	0.223734	0.261836
lr_closeness	0.458478	0.526260	0.409452	0.143725	0.167978	0.130809
lr_closeness_thd	0.685354	0.627606	0.697428	0.252322	0.218047	0.273960
lr_degree	0.532299	0.517873	0.529877	0.192450	0.175009	0.194284
lr_degree_thd	0.683940	0.634267	0.694821	0.259754	0.222260	0.278531
lr_eigenvector	0.410077	0.469889	0.392998	0.144474	0.152003	0.140940
lr_eigenvector_thd	0.672573	0.632542	0.684020	0.255690	0.229391	0.272253
lr_fast-ripple	0.600338	0.588345	0.568696	0.208429	0.193478	0.203627
lr_fast-ripple_thd	0.616017	0.591855	0.577069	0.231883	0.199183	0.223145
lr_fc	0.652596	0.621077	0.633501	0.244182	0.232723	0.250817
lr_good	0.690377	0.640188	0.690368	0.275493	0.242655	0.275487
lr_high-delta	0.490998	0.483207	0.464266	0.168033	0.156533	0.163057
lr_high-delta_thd	0.648179	0.600130	0.668664	0.221482	0.197867	0.235556
lr_pca	0.673497	0.636124	0.684465	0.255178	0.225327	0.268663
lr_ripple	0.654129	0.623562	0.626420	0.247475	0.226218	0.241778
lr_ripple_thd	0.656594	0.626379	0.649645	0.249188	0.231735	0.251089
lr_spike	0.616016	0.571107	0.601803	0.212819	0.194315	0.217304
lr_spike_thd	0.682202	0.628115	0.684652	0.258819	0.219549	0.265314
xgb	0.691471	0.591837	0.674136	0.284699	0.215016	0.255578

Table B.1: Area under the ROC and the PR curve across many different models for spike data, silent data, and hybrid data.

	lr		pca		vae,nonsoz		vae,soz		kde,nonsoz		kde,soz	
quantile	re	rn	re	rn	re	rn	re	rn	re	rn	re	rn
0.000	0.528	0.500	0.528	0.500	0.528	0.500	0.528	0.500	0.528	0.500	0.528	0.500
0.050	0.556	0.403	0.578	0.394	0.533	0.564	0.533	0.558	0.544	0.414	0.528	0.500
0.100	0.550	0.511	0.556	0.411	0.544	0.531	0.556	0.536	0.558	0.525	0.528	0.500
0.150	0.528	0.539	0.567	0.519	0.544	0.539	0.511	0.622	0.550	0.536	0.528	0.500
0.200	0.511	0.461	0.567	0.478	0.550	0.558	0.528	0.664	0.558	0.533	0.511	0.683
0.250	0.544	0.467	0.556	0.467	0.550	0.525	0.483	0.644	0.567	0.514	0.514	0.708
0.300	0.531	0.511	0.572	0.494	0.558	0.511	0.494	0.611	0.544	0.497	0.522	0.708
0.350	0.500	0.517	0.583	0.522	0.528	0.539	0.522	0.617	0.536	0.519	0.539	0.653
0.400	0.517	0.539	0.644	0.458	0.556	0.519	0.506	0.675	0.550	0.522	0.542	0.628
0.450	0.531	0.531	0.628	0.483	0.550	0.511	0.522	0.650	0.561	0.525	0.544	0.589
0.500	0.500	0.544	0.686	0.481	0.519	0.536	0.506	0.639	0.556	0.572	0.544	0.561
0.550	0.500	0.528	0.647	0.483	0.553	0.597	0.506	0.633	0.542	0.589	0.558	0.553
0.600	0.514	0.511	0.572	0.500	0.583	0.583	0.514	0.578	0.528	0.561	0.528	0.561
0.650	0.506	0.506	0.528	0.528	0.600	0.558	0.525	0.578	0.550	0.575	0.558	0.556
0.700	0.506	0.539	0.586	0.506	0.617	0.561	0.481	0.594	0.614	0.522	0.575	0.517
0.750	0.486	0.553	0.564	0.517	0.656	0.533	0.489	0.572	0.614	0.522	0.589	0.550
0.800	0.494	0.572	0.619	0.506	0.556	0.578	0.492	0.586	0.614	0.522	0.508	0.572
0.850	0.519	0.594	0.667	0.494	0.597	0.550	0.494	0.567	0.614	0.522	0.522	0.550
0.900	0.453	0.572	0.744	0.500	0.556	0.558	0.478	0.556	0.614	0.522	0.539	0.561
0.950	0.553	0.536	0.694	0.528	0.539	0.539	0.356	0.567	0.614	0.522	0.417	0.544
1.000	0.528	0.528	0.500	0.528	0.500	0.528	0.500	0.528	0.614	0.522	0.500	0.528

Table B.2: AUC of re_i and rn_i in predicting patient outcome under different metric and threshold of epileptogenicity.

References

[1] Acharya, U. R., Vinitha Sree, S., Swapna, G., Martis, R. J., & Suri, J. S. (2013). Automated eeg analysis of epilepsy: A review. *Knowledge-Based Systems*, 45, 147–165.

[2] Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley Series in Probability and Statistics. Nashville, TN: John Wiley & Sons.

[3] Beauchamp, M. A. (1965). An improved index of centrality. *Behavioral Science*, 10(2), 161–163.

[4] Chen, T. & Guestrin, C. (2016). XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: ACM.

[5] Chu, C. J. (2015). High density eeg—what do we have to lose? *Clinical Neurophysiology*, 126(3), 433–434.

[6] Fletcher, J. M. & Wennekers, T. (2018). From structure to activity: Using centrality measures to predict neuronal activity. *International Journal of Neural Systems*, 28(02), 1750013. PMID: 28076982.

[7] Gliske, S. V., Irwin, Z. T., Davis, K. A., Sahaya, K., Chestek, C., & Stacey, W. C. (2016). Universal automated high frequency oscillation detector for real-time, long term eeg. *Clinical Neurophysiology*, 127(2), 1057–1066.

[8] Kingma, D. P. & Welling, M. (2013). Auto-encoding variational bayes.

[9] Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1), 48–50.

[10] Mardia, K. V., Kent, J. t., & Bibby, J. M. (1979). *Multivariate Analysis*. San Diego, CA: Academic Press.

[11] Mohanraj, R. & Brodie, M. J. (2006). Diagnosing refractory epilepsy: response to sequential treatment schedules. *European Journal of Neurology*, 13(3), 277–282.

[12] Ortega, G. J., Sola, R. G., & Pastor, J. (2008). Complex network analysis of human ecog data. *Neuroscience Letters*, 447(2), 129–133.

[13] Rodin, E., Constantino, T., Rampp, S., & Wong, P. (2009). Spikes and epilepsy. *Clinical EEG and Neuroscience*, 40(4), 288–299. PMID: 19780350.

[14] Sanz-Garcia, A., de Sola, R. G., Vega-Zelaya, L., Pastor, J., & Ortega, G. J. (2016). Network theoretical approach to describe epileptic processes. In C. Hintermüller (Ed.), *Advanced Biosignal Processing and Diagnostic Methods* chapter 4. Rijeka: IntechOpen.

[15] Tamilia, E., Dirodi, M., Alhilani, M., Grant, P. E., Madsen, J. R., Stufflebeam, S. M., Pearl, P. L., & Papadelis, C. (2020). Scalp ripples as prognostic biomarkers of epileptogenicity in pediatric surgery. *Annals of Clinical and Translational Neurology*, 7(3), 329–342.

[16] Téllez-Zenteno, J. F., Dhar, R., & Wiebe, S. (2005). Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain*, 128(5), 1188–1198.

[17] Toga, A. W. (2015). *Brain Mapping*. Elsevier Science.

[18] Upton, G. & Cook, I. (2008). *Perron–Frobenius theorem*. Oxford University Press.

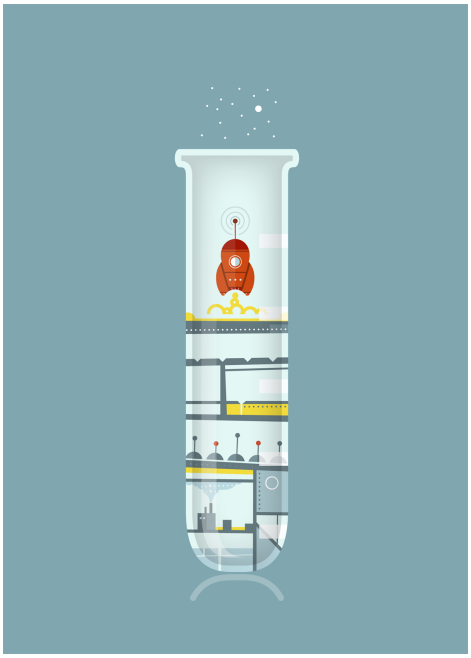
[19] van Dellen, E., Douw, L., Hillebrand, A., de Witt Hamer, P. C., Baayen, J. C., Heimans, J. J., Reijneveld, J. C., & Stam, C. J. (2014). Epilepsy surgery outcome and functional network alterations in longitudinal meg: A minimum spanning tree analysis. *NeuroImage*, 86, 354–363.

[20] Varotto, G., Tassi, L., Franceschetti, S., Spreafico, R., & Panzica, F. (2012). Epileptogenic networks of type ii focal cortical dysplasia: A stereo-eeg study. *NeuroImage*, 61(3), 591–598.

[21] Wilke, C., Worrell, G., & He, B. (2011). Graph analysis of epileptogenic networks in human partial epilepsy. *Epilepsia*, 52(1), 84–93.

[22] Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35.

[23] Zijlmans, M., Jiruska, P., Zelmann, R., Leijten, F. S., Jefferys, J. G., & Gotman, J. (2012). High-frequency oscillations as a new biomarker in epilepsy. *Annals of Neurology*, 71(2), 169–178.



This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, *Science Experiment 02*, was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD dissertation with this look & feel has been released under the permissive AGPL license, and can be found online at github.com/suchow/Dissertate or from its lead author, Jordan Suchow, at suchow@post.harvard.edu.