Generative Modeling and Prediction of Spontaneous Epileptic Seizures

Christine Liu, Peter Maldonado, Keegan Mehall
{cliu99, phm, kmehall}@stanford.edu

Motivation

- Epilepsy is a neurological disorder that affects more than 65 million people.
- Absence epilepsy is a type of epilepsy characterized by short seizures that tend to occur in clusters.
- Seizures have traditionally been described as stochastic events.
- Research studies have suggested that there exists a slower shift in network state that influences seizure probability, however, the exact relationship between the seizure and pre-seizure shape is unknown.
- Currently no seizure prediction device exists, mostly due to the lack of strong confidence in predictive algorithms and the high stochasticity of the data.
- A seizure prediction system would be clinically invaluable and improve the well-being of epilepsy patients.
- This project seeks to predict epileptic seizures in two ways: first with binary classification, and second with a generative approach

Seizure vs Non-Seizure Signals

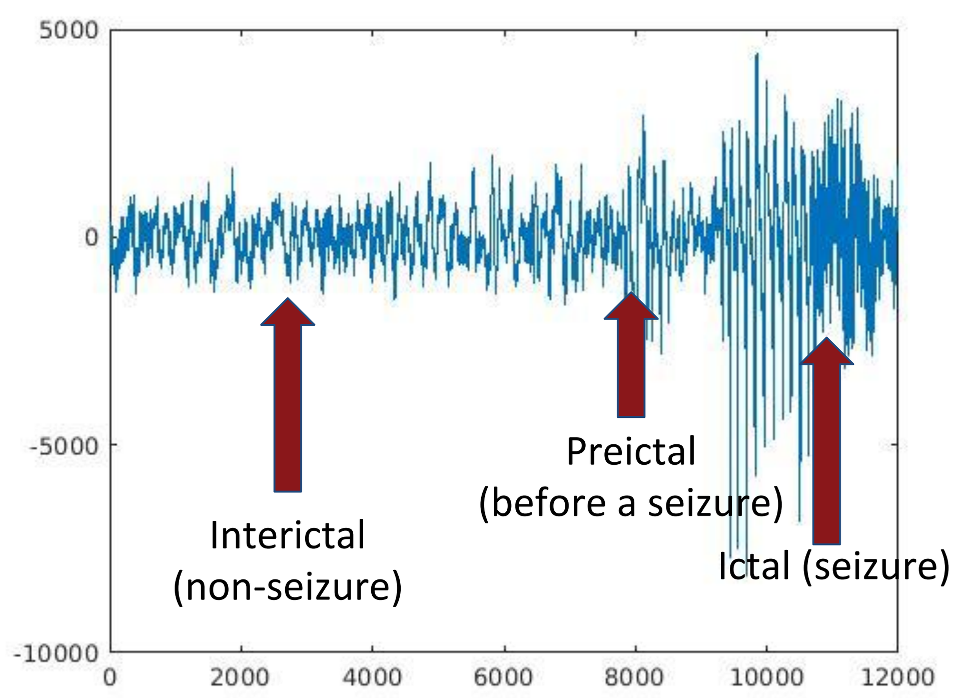


Fig 1: Electroencephalogram (EEG) showing 12 second data segment

Dataset

We are working with 2 datasets. First, there is an electroencephalogram (EEG) dataset of absence epileptic seizures from the Huguenard Lab sampled at 1000 hz. This dataset has seizure data from 9 rodents with genetic absence epilepsy (WAGRij). This dataset consists of hours of extra-cellular recordings of individual cortical and thalamic neurons from rodents with genetic absence epilepsy. The data is segmented into 12 second windows spanning each seizure.

The second dataset is hosted on Kaggle and contains data from intracranial EEGs from both dogs and humans. The EEG is sampled from 16 electrodes at 400 Hz and 5000 Hz, and contains the recorded voltages. The Kaggle dataset was mostly used to validate our models and check for robustness. The kaggle dataset consits of tonic-clonic seizures

Models and Method

This project focused on predicting epileptic seizures using a multi-step approach:

Step 1. Data Collection

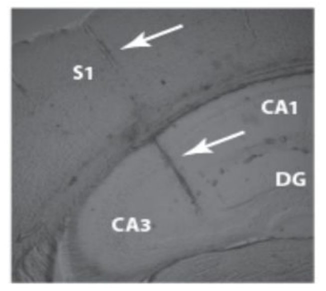


Fig 2: Multi-electrode silicon probe implanted through somatosensory cortex and thalamus

Step 2. Data pre-processing

Data is filtered to eliminate noise from recordings so that the signal is cleaner. One problem with raw EEG data from rodents with spontaneously occurring seizures is the highly stochastic nature of the data.

Step 3. Preictal vs Interictal Binary Prediction Task

We train two models for a prediction task between preictal and interictal segments in order to distinguish between seizure and non-seizure classifications. We also chose to implement a version of our model with fourier transform and without.
Logistic Regression w/ a=0.01

- This decision treats each signal as a stochastic process, where we maximize the likelihood that a given signal is properly classified. This is a reasonable assumption, as the EEG data is collected across a number of noisy channels during each time step.

Support Vector Machines w/ radial basis function kernel

- The SVM learns the maximum margin decision boundary, which in the context of our problem means we focus on dividing the mass of seizure signals from that of the non-seizure signals. The RBF kernel is useful for a projection into an infinite dimensional feature space, which is helpful as our EEG data is representing significantly higher-dimensional neural activity in the subjects.

Step 4. Generative Modeling of seizure signal using Long Short-Term Memory Networks

- We trained long short-term memory (LSTM) models in order to try and predict the shape of a seizure based off the preictal shape of the EEG data.
- LSTMs are empirically well-adapted for longer time series datasets such as continuous EEG data.
- We chose to implement this task both on the raw data, and with filtering for high frequencies.
- This task has not been attempted in seizure literature, as the highly stochastic nature of the data makes it incredibly difficult.

Results

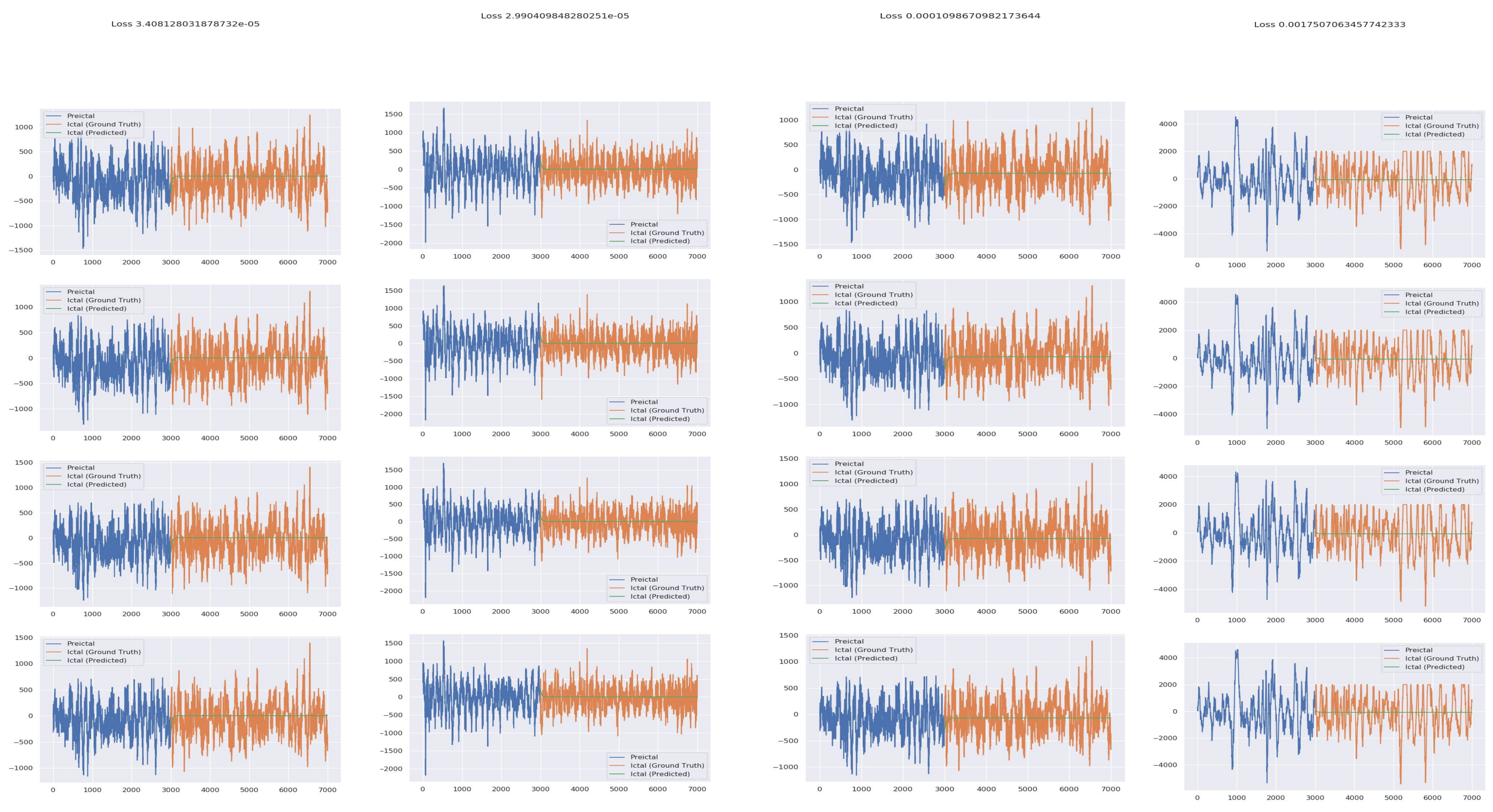
Binary Prediction Task:

	Dataset 1 Accuracy	Dataset 2 Accuracy
Logistic Regression	0.552	0.765
SVM	0.633	0.878
Logistic w/ Fourier	0.575	x
SVM w/ Fourier	0.620	x

The accuracies for the binary prediction task were better than random chance. Absence epilepsy has been noted to be highly stochastic and the seizures are more difficult to predict than typical grand-mal tonic clonic which attributes for the lower accuracies from Dataset 1 as opposed to Dataset 2.

Generative Modeling Task:

Non-Filtered **Filtered**



As noted in the resulting graphs, the LSTM was not able accurately generate the signal. This is mostly due to the high dimensionality and stochasticity of the data. This could be improved by trying to reduce dimensionality through PCA, or increasing the training time.

References

1. Sorokin, J., Ganguli, S., Huguenard, J. "Identification of unique pre-ictal states via non-negative tensor decomposition of single unit recordings." COSYNE Conference 2015.

2. Florian Mormann, Ralph G. Andrzejak, Christian E. Elger, Klaus Lehnertz, Seizure prediction: the long and winding road, Brain, vol. 130, Issue 2, February 2007

3. Carney, Paul R et al. "Seizure prediction: methods." Epilepsy & behavior : E&B vol. 22, Suppl 1 2011

4. Syed Muhammad Usman, Shehzad Khalid, Rizwan Akhtar, Zuner Bortolotto, Zafar Bashir, Haiyang Qiu, "Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: Review of available methodologies," Seizure, Volume 71, 2019, Pages 258-269