Harmful Brain Activity Classification

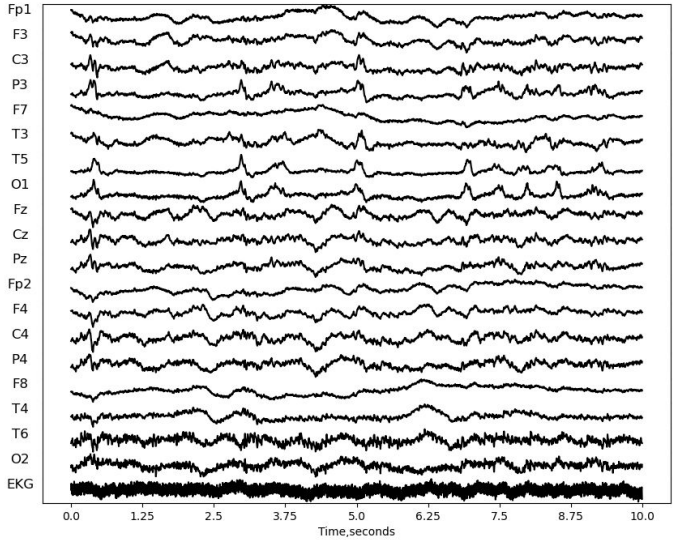
Evgeniya Lagoda, Kshitiz Parihar, Souparna Purohit, Jianing Yang

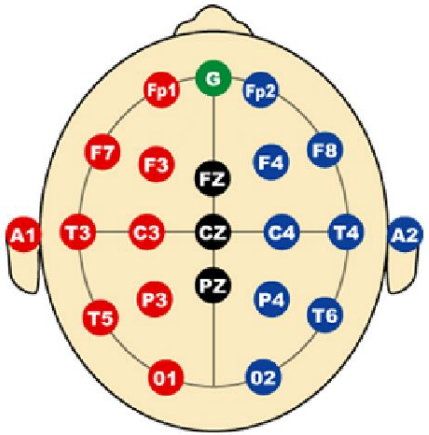
Guiding research question

The goal of this project is to detect and classify harmful brain activities such as seizures with EEG signals.

Electroencephalography (EEG) is a method to record the spontaneous electrical activity of the brain. 19 electrodes are placed on the scalp to detect electrical signals from four regions of the brain (LL, LP, RP, RL).

A good brain activity classifier will improve EEG pattern recognition, and help doctors and brain researchers diagnose harmful brain activities faster and more accurately.





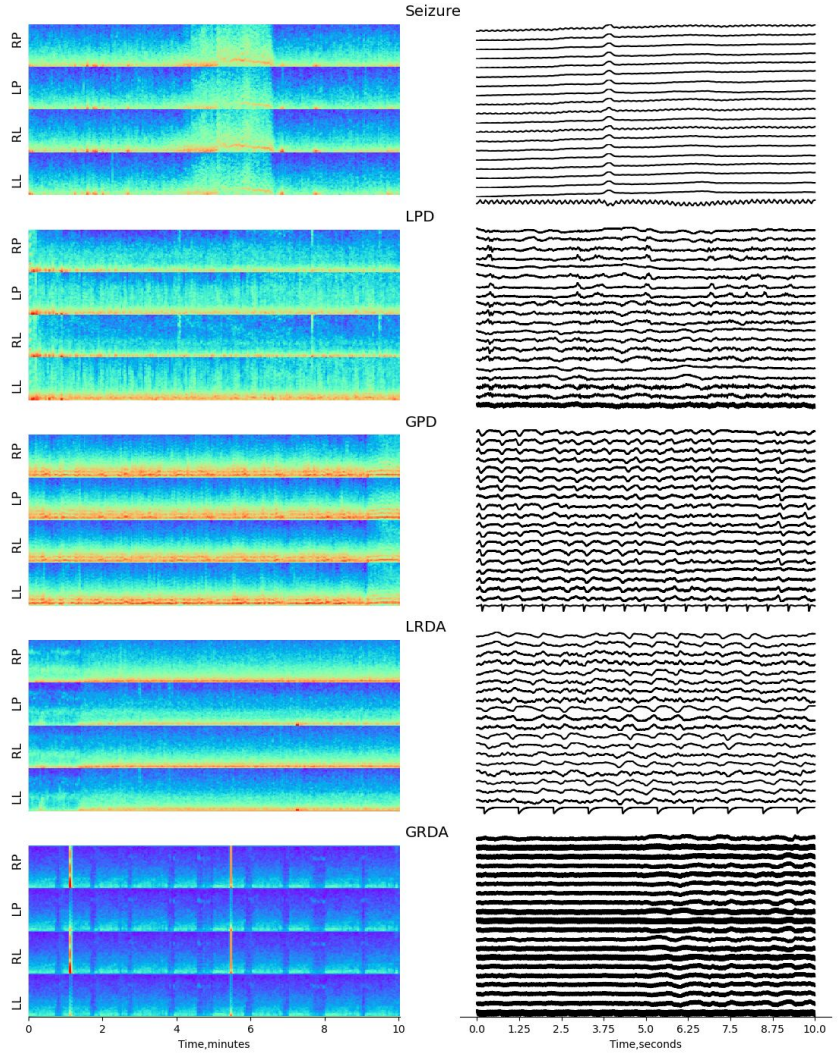
The letters used are:

F - Frontal lobe
T - Temporal lobe
C - Central lobe
P - Parietal lobe
O - Occipital lobe

"Z" refers to an electrode placed on the mid-line.

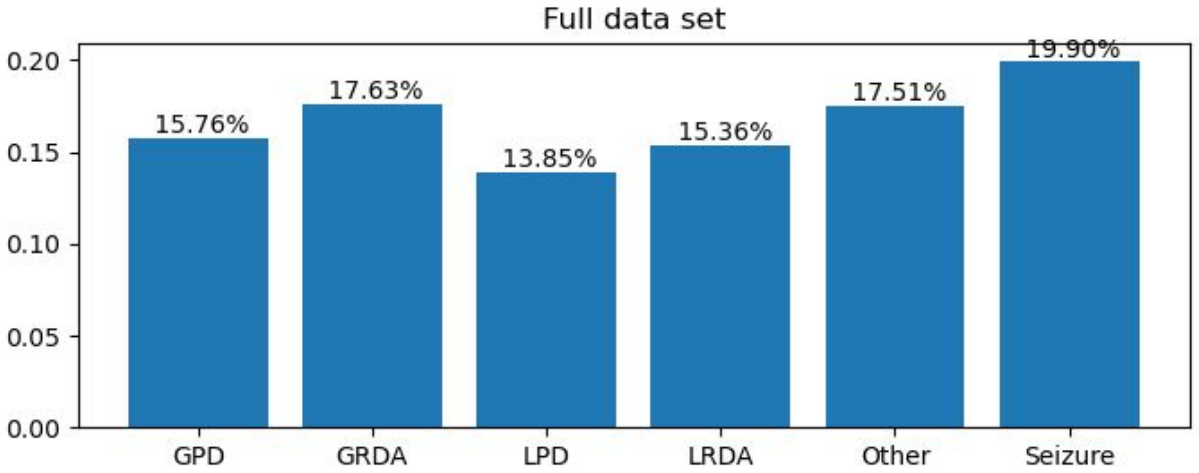
Categories of seizure-like brain activities

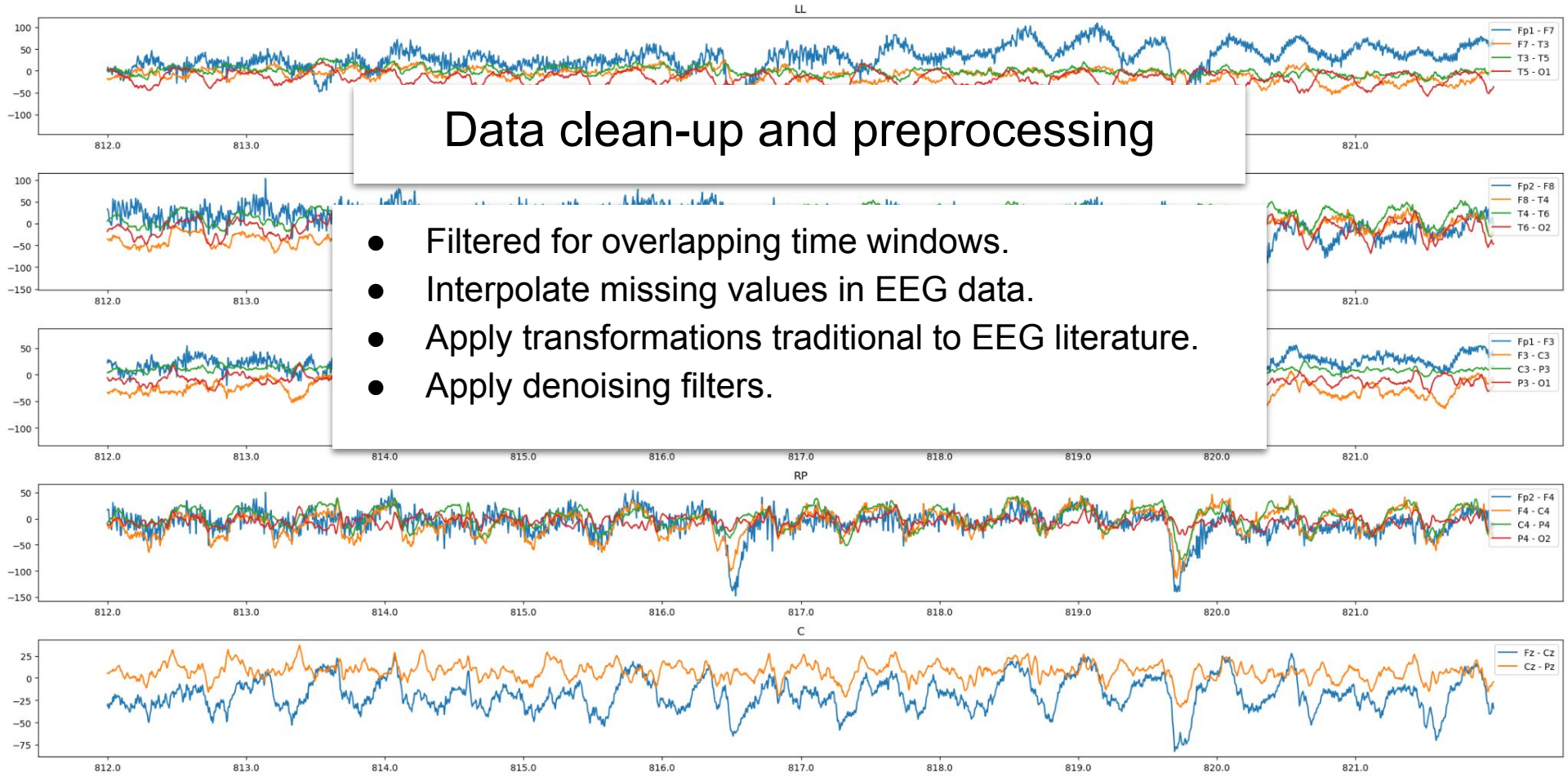
- Seizure
- Generalized Periodic Discharges (GPD)
- Lateralized Periodic Discharges (LPD)
- Lateralized Rhythmic Delta Activity (LRDA)
- Generalized Rhythmic Delta Activity (GRDA)
- “Other”



Structure of the data

- The dataset for the project is publicly available at Kaggle Competition “HMS – Brain Activity Classification”. The data consists of EEG and spectrogram files from 1950 patients.
- The middle 10sec of each EEG with corresponding spectrogram segments have been annotated by a group of experts into the six categories:

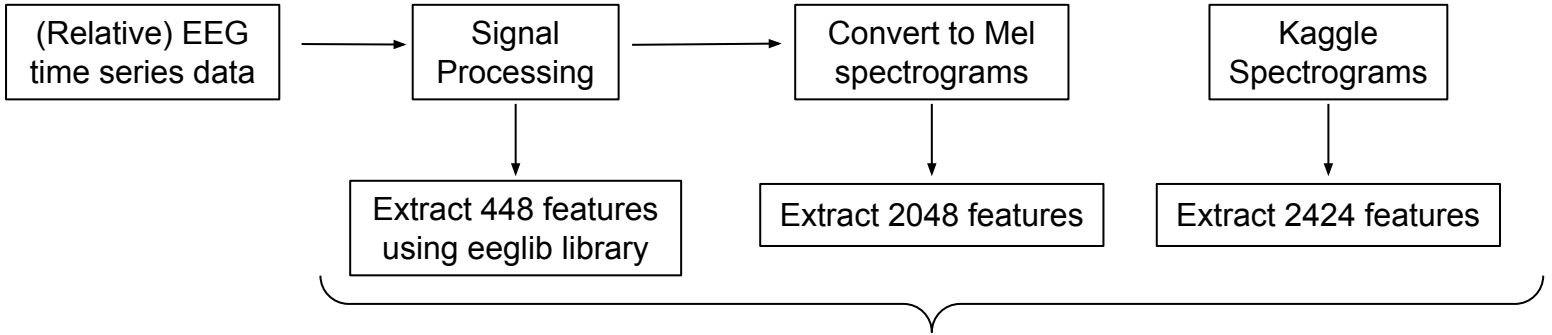


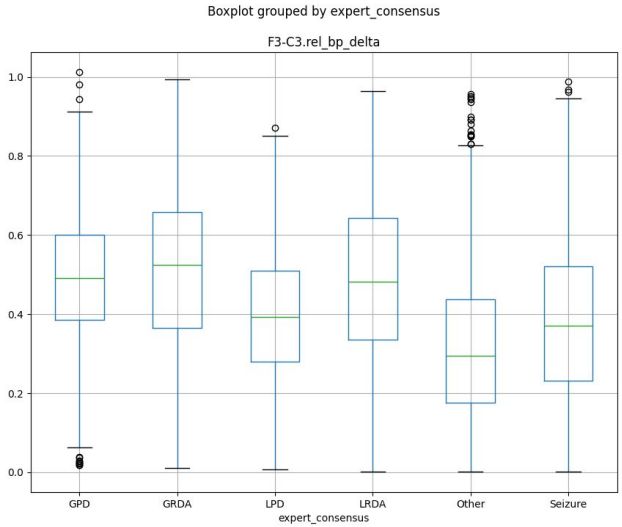


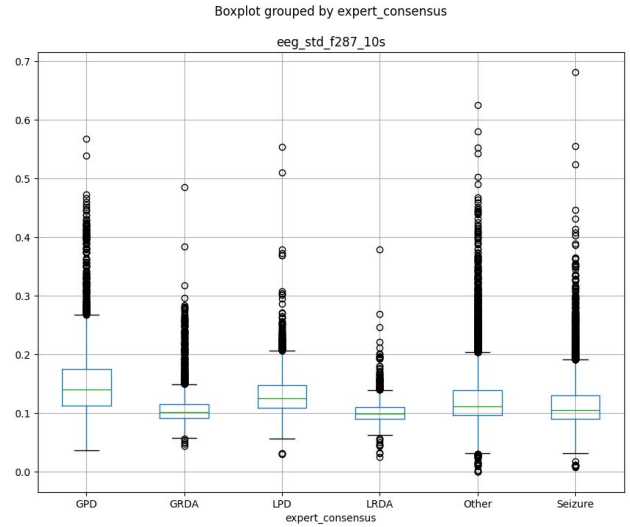
Data clean-up and preprocessing

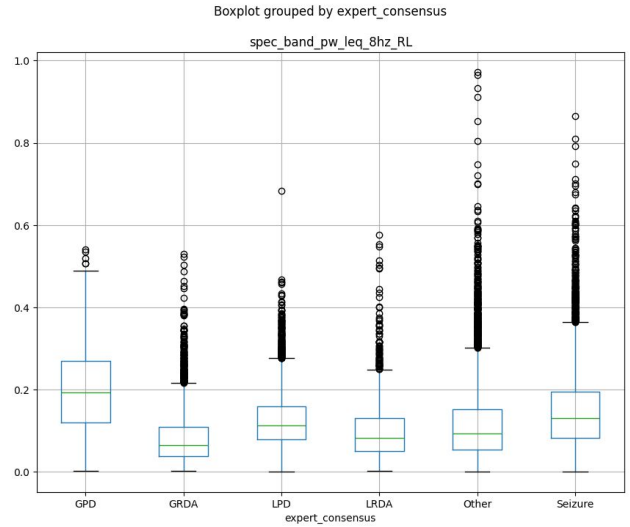
- Filtered for overlapping time windows.
- Interpolate missing values in EEG data.
- Apply transformations traditional to EEG literature.
- Apply denoising filters.

Features of Interest

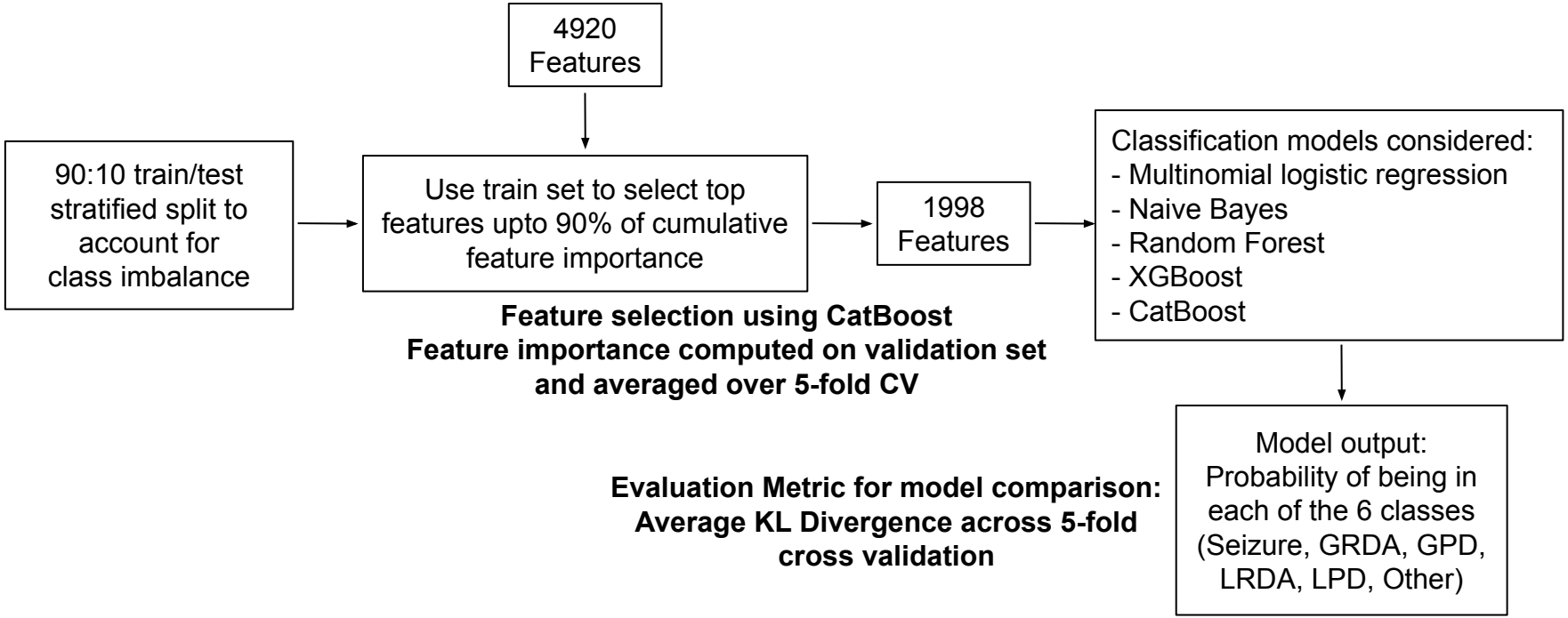








Modeling approach



Model Performance

Model	Average KL Divergence across 5-fold CV
Baseline	1.36
Logistic Regression	1.17
Naive Bayes	16.72
Random Forest	0.81
XGBoost	**0.78**
CatBoost	0.79

Baseline model: Probability of being in each of the 6 classes is the proportion of the samples in the training set with that class

On Test Set

KL Divergence: 0.71 Balanced Accuracy: 0.53

Predicted \ True	Seizure	LPD	GPD	LRDA	GRDA	Other
Seizure	433	57	97	32	35	117
LPD	8	276	16	41	7	60
GPD	36	13	248	2	2	26
LRDA	3	10	0	41	8	26
GRDA	30	18	26	55	285	66
Other	125	152	48	56	105	415

Confusion matrix

	Precision	Recall
Seizure	0.56	0.68
LPD	0.68	0.52
GPD	0.76	0.57
LRDA	0.47	0.18
GRDA	0.59	0.64
Other	0.46	0.58

Weighted Precision: 0.58
Weighted Recall: 0.57

Weighted (# of true instances of each class) average of individual precision and recall values to account for class imbalance

Future research

- Model is reasonable at detecting Seizures, LPD, GPD, and GRDA, but very poor on LRDA
- Run deep learning models such as CNNs (unsuccessful on Mel spectrogram data)
- Develop more sophisticated data preparation pipeline following state of the art EEG literature for such models

Thank You