HMS – Harmful Brain Activity

Harmful Brain Activity

Seizure Classification and Prediction

- Karan Kalra

HMS – Harmful Brain Activity

WHAT IS THE PURPOSE AND VISION?

EEG?

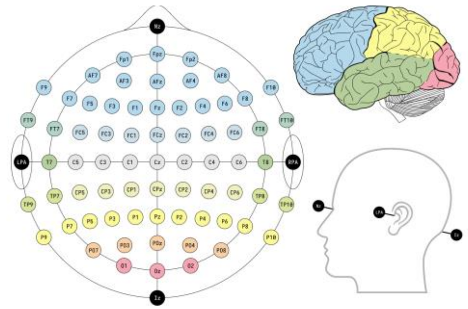
An electroencephalogram (EEG) is a test that detects electrical activity in the brain by attaching tiny metal discs (electrodes) to the scalp.

PURPOSE

to identify and categorize different types of brain seizures

VISION

My work will significantly increase electroencephalography **seizure categorization accuracy**, resulting in dramatic improvements for **neurocritical care, epilepsy, sleep disorders, medication discovery**, and in some cases brain tumor.



HMS – Harmful Brain Activity

Dataset and it's approach

FEATURES –

eeg_id - A unique identifier for the entire EEG recording.
eeg_sub_id - An ID for the specific 50 second long subsample this row's labels apply to.
eeg_label_offset_seconds - The time between the beginning of the consolidated EEG and this subsample.
spectrogram_id - A unique identifier for the entire EEG recording.
spectrogram_sub_id - An ID for the specific 10 minute subsample this row's labels apply to.
spectogram_label_offset_seconds - The time between the beginning of the consolidated spectrogram and this subsample.
label_id - An ID for this set of labels.
patient_id - An ID for the patient who donated the data.
expert_consensus - The consensus annotator label. Provided for convenience only.

TARGET –

[seizure/lpd/gpd/lrda/grda/other]_vote - The count of annotator votes for a given brain activity class. The full names of the activity classes are as follows: lpd: lateralized periodic discharges, gpd: generalized periodic discharges, lrd: lateralized rhythmic delta activity, and grda: generalized rhythmic delta activity .

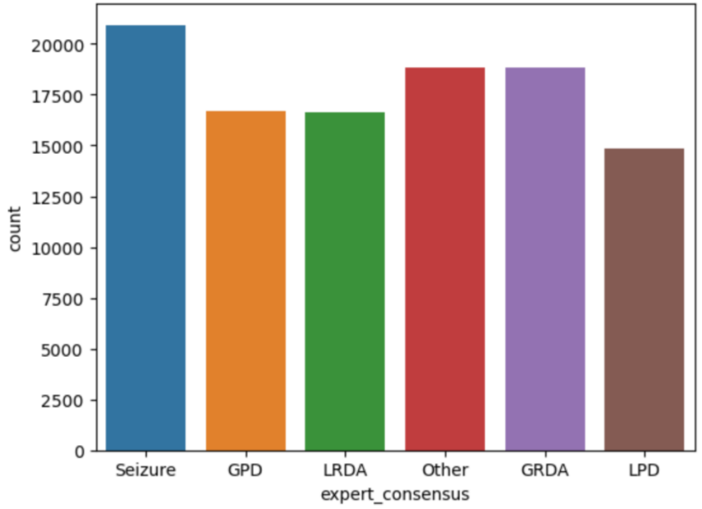
After understanding the data, it was important to recognize the data and how the parameters work in the eeg with their samples. The samples were read with .parquet file.

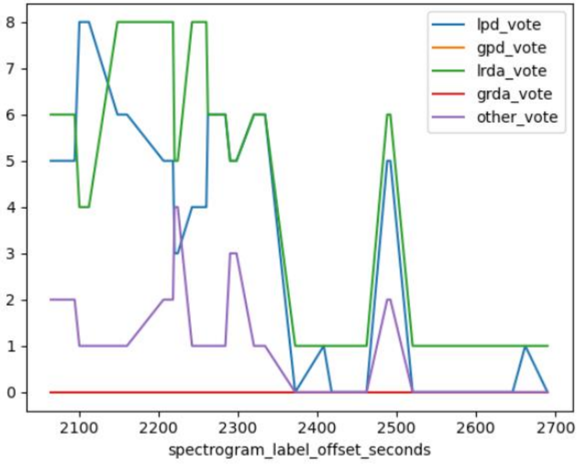
HMS – Harmful Brain Activity

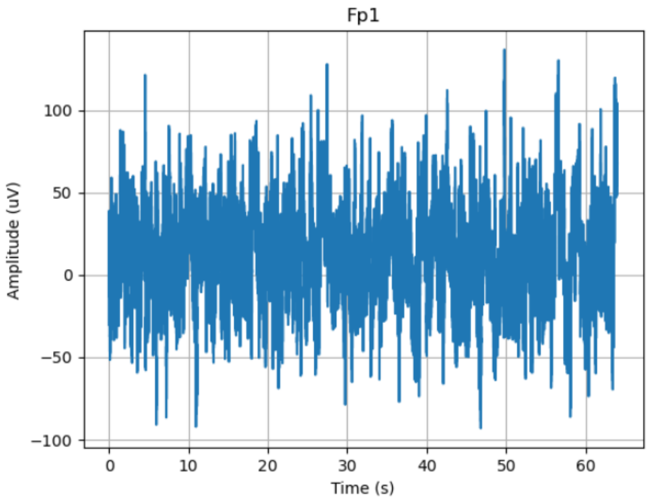
EDA and results

The dataset contains both the train and test data for EEG patterns. EEG data came from some overlapping samples that were in .parquet files.

Understanding the data with expert_consensus , electrodes, while dividing the different types of reading and seizures diverted into votes.



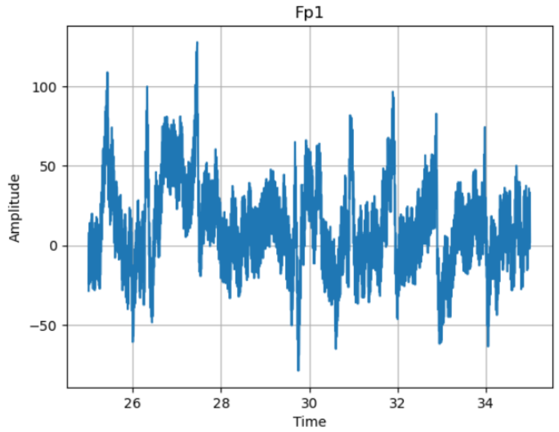


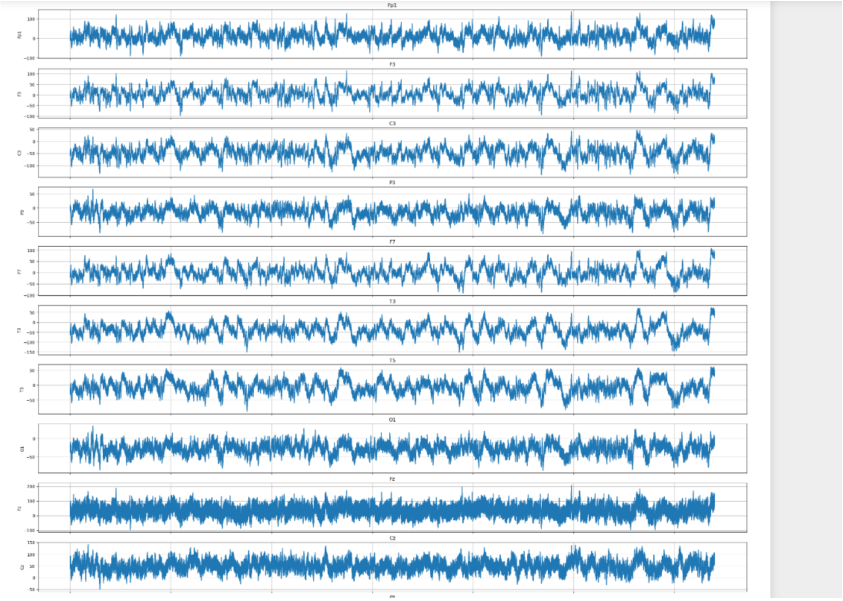


HMS – Harmful Brain Activity

Focusing and Zooming out the Samples

First, we take a sample of the one sample window and zoom in to understand the variability in middle seconds window to get accuracy. For e.g., **even if subsample is 50 seconds long, the annotation is done by looking at the central 10 seconds.** Then, we plot the whole set of electrodes based on different designations having various second's window.





HMS – Harmful Brain Activity

Potential -ve impact of EDA

Signals are quite noisy and do not provide much insight. One probable reason of the noise in our signal is the presence of an interfering signal that adds to the actual signal.

Merging the .parquet with the train csv was required as the next step of exploration and modelling.

HMS – Harmful Brain Activity

Fitting the models

Initially, I split the data into training and validation sets and used 80% of the data for training and 20% for validation.

- **Accuracy score**
- **Random forest classifier**
- **Classification report**
- **Confusion matrix**

```
The accuracy score is: 0.9995318352059925
The random classification score is:              precision    recall  f1-s
core   support

           0       1.00      1.00      1.00      3444
           1       1.00      1.00      1.00      3709
           2       1.00      1.00      1.00      2910
           3       1.00      1.00      1.00      3375
           4       1.00      1.00      1.00      3757
           5       1.00      1.00      1.00      4165

    accuracy                           1.00     21360
   macro avg       1.00      1.00      1.00     21360
weighted avg       1.00      1.00      1.00     21360
```

```
The selected accuracy score is: 0.9317415730337079
The selected random classification score is:
precision    recall  f1-score   support

           0       0.96      0.96      0.96      3444
           1       0.97      0.96      0.97      3709
           2       0.88      0.91      0.90      2910
           3       0.94      0.97      0.95      3375
           4       0.91      0.86      0.88      3757
           5       0.93      0.93      0.93      4165

    accuracy                           0.93     21360
   macro avg       0.93      0.93      0.93     21360
weighted avg       0.93      0.93      0.93     21360

The selected confusion matrix is:
[[3300   11   25    4   71   33]
 [  30 3579   11   16   54   19]
 [  30    8 2659   63   88   62]
 [   1   10   50 3265   28   21]
 [  63   73  144   75 3242  160]
 [  25   23  119   42   99 3857]]
```

Also calculated the Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2) for error metrics, Decision tree, respectively for each model.

HMS – Harmful Brain Activity

Parameter tuning

COMBINATION OF THE FREQUENCIES WITH EVERY EEG AND SPECTROGRAM ID

Accuracy score

Random forest classifier

F1 score

```
Model for seizure_vote: Accuracy = 0.8078, F1 Score = 0.7870
Model for lpd_vote: Accuracy = 0.8617, F1 Score = 0.8435
Model for gpd_vote: Accuracy = 0.8897, F1 Score = 0.8748
Model for lrda_vote: Accuracy = 0.8570, F1 Score = 0.8377
Model for grda_vote: Accuracy = 0.8412, F1 Score = 0.8213
Model for other_vote: Accuracy = 0.7162, F1 Score = 0.6767
```

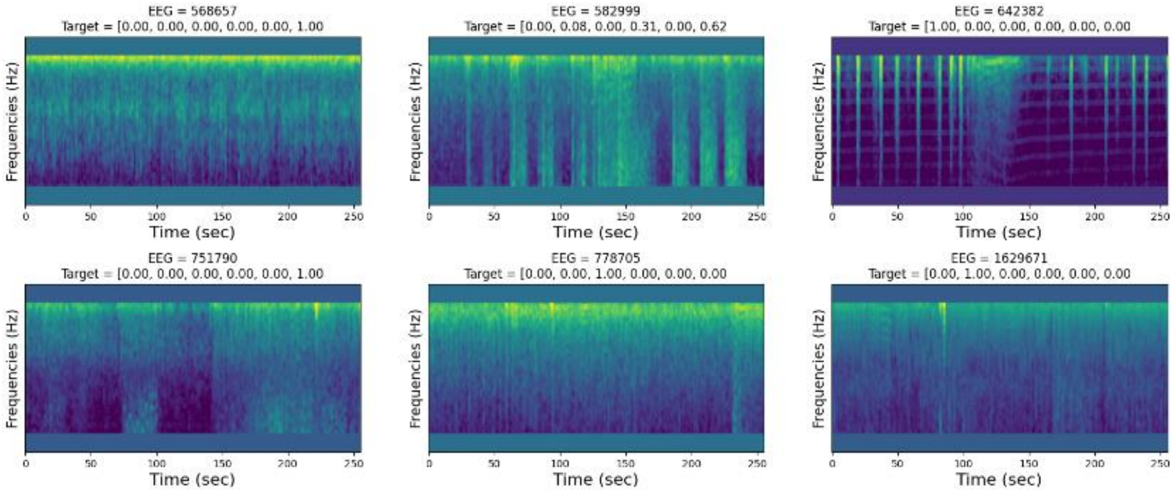
Improved accuracy up to 83%. Although, what else can we do?

Can we implement something else?

WAVENET Neural Network model

- It is a TensorFlow implementation of DeepMind's WaveNet

It is a fully convolutional neural network, where the convolutional layers have various dilation factors that allow its receptive field to grow exponentially with depth and cover thousands of timesteps with different connections.



HMS – Harmful Brain Activity

Final Conclusion

The notebook effectively sets up the data pipeline for the WaveNet neural network model, including data loading, augmentation, and visualization.

By iterating through batches, it displays spectrograms and their corresponding target labels, providing a clear understanding of the dataset's structure and preparation steps. This visualization aids in verifying the correctness of the data preprocessing pipeline and ensures the model receives accurately processed inputs.

The approach ensures a robust preparation of EEG data for subsequent training and evaluation phases.

The iteration through the DataLoader, plotting the spectrogram images for each sample in a grid layout. The images are normalized, and each subplot is labeled with the corresponding target values and EEG ID.