

Final Project Report – Harmful Brain Activity

Group Members: Saisruthi Kotagiri, Terry Griffin, Mia Tsivitse, Nishi Surana, Will Abbott

Kaggle Competition:

https://www.kaggle.com/competitions/hms-harmful-brain-activity-classification/data?select=train_eegs

Project Scope:

This Kaggle competition was hosted by Harvard Medical School, and our goal was to detect and classify seizures and other types of harmful brain activity. We developed models trained on electroencephalography (EEG) signals recorded from critically ill hospital patients.

Currently, EEG monitoring relies on manual analysis by specialized neurologists. Current methods are labor-intensive, expensive, and prone to errors. In addition to its time-consuming nature, the process results in reliability concerns amongst reviewers, regardless of their expertise. These issues create an important problem to solve, as detecting seizures and harmful brain activity is crucial for timely treatment and preventing brain damage.

Data Description:

A team of experts categorized the harmful brain activity data from the competition. Their annotations span from clear patterns with unanimous expert consensus to ambiguous cases where the experts have differing opinions. Consequently, the classification faces challenges in terms of reliability among expert reviewers. Presented below are six discernible EEG patterns agreed upon by multiple experts to enhance reliability, each providing insight into the internal functioning of the brain. These patterns could arise from diverse conditions like epilepsy, brain trauma, infections, or other neurological disorders.

Six EEG Patterns of Interest:

- Seizure (SZ): Sudden, uncontrolled electrical disturbance in the brain. High variation in duration and severity.
- Generalized Periodic Discharges (GPD): Abnormal electrical discharges in the brain that occur synchronously and involve large areas of the brain simultaneously.
- Lateralized Periodic Discharges (LPD): Like GPD, but involves electrical discharges that are confined to one hemisphere or one side of the brain.
- Lateralized Rhythmic Delta Activity (LRDA): Rhythmic delta waves (slow brain waves) that are predominantly seen on one side of the brain.

- Generalized Rhythmic Delta Activity (GRDA): Rhythmic delta waves that are generalized and involve both hemispheres of the brain simultaneously.
- "Other": Abnormal brain activity that does not fit into the aforementioned patterns.

Within the 6 EEG patterns used for classification, there are “Idealized Patterns” with high agreement amongst experts, and cases where $\sim 1/2$ of experts gave a label as “other” and the other $\sim 1/2$ gave one of the remaining five labels, called “proto patterns”. Cases where experts were approximately split between 2 of the 5 named patterns, are called “edge cases”.

In this competition, Kaggle provided both spectrograms and raw EEG waveforms. The spectrograms provided by Kaggle span 10 minutes, while the EEG waveforms cover 50 seconds. Interestingly, the middle 50 seconds of both datasets depict the same time window, capturing identical events. In essence, these datasets present the same information in two different formats, with spectrograms serving as visual representations of the raw waveforms.

Exploratory Data Analysis:

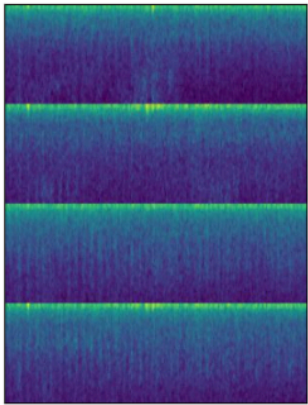
The data exploration and visuals revealed insights into brain activity classifications and EEG sample distribution. The distribution of brain activity classifications plot indicated that most classifications in the training data lean towards seizures, followed by “other” and Generalized Rhythmic Delta Activity (GRDA). Lateralized Periodic Discharges (LPD) were the least common classified type of harmful brain activity in the training data.

Next, a histogram of EEG sample offsets indicated higher concentration within the 0-1000 seconds interval. From this information, it is possible to infer that potential artifacts are present during this period, such as eye blinks or muscle movements, particularly at the start of recording sessions.

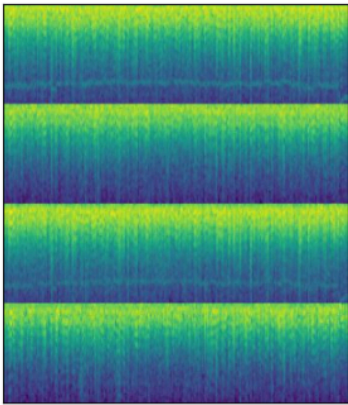
The distribution of EEG agreement as consensus plot solidified the finding that some eeg_ids have poor consensus, with less than 20% agreement. This insight validated the premise of the project, as classification suffers from reliability issues, even between expert reviewers. Between the six patterns of harmful brain activity, the visualization indicated that Seizure_vote had the highest rates of consensus amongst experts and gpd_vote had the lowest rates of consensus amongst experts.

Finally, we analyzed sample EEG structure to support identification of patterns or anomalies in the EEG signals and plotted visuals of example EEGs from the directory.

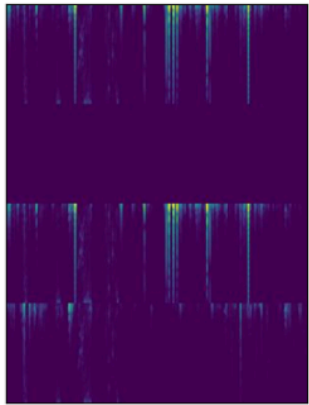
Train_Images



Other.1423489515.jpeg

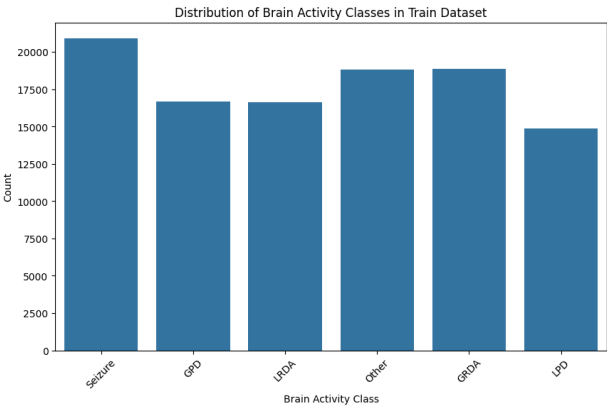
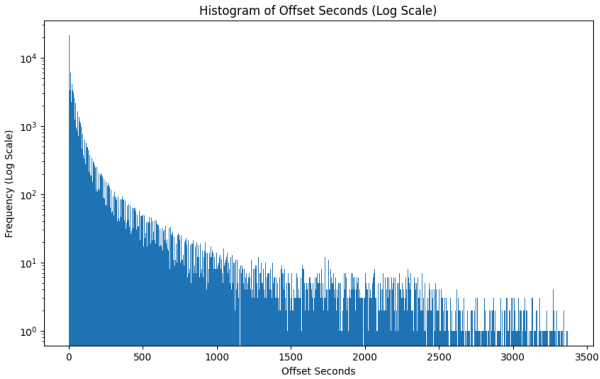
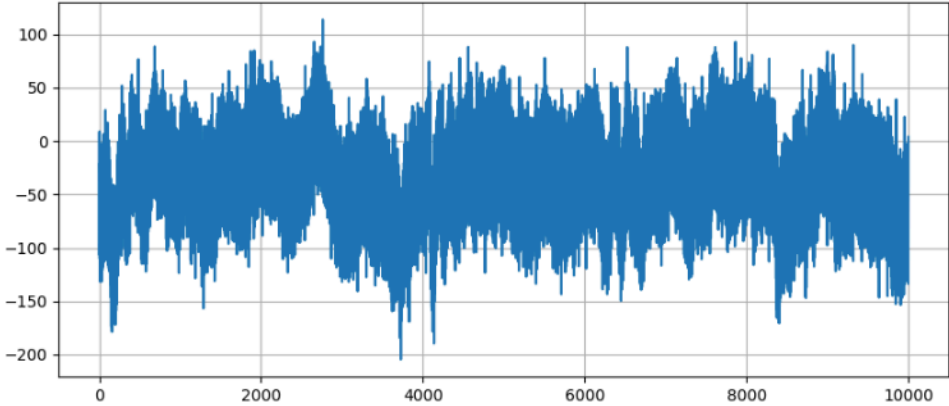


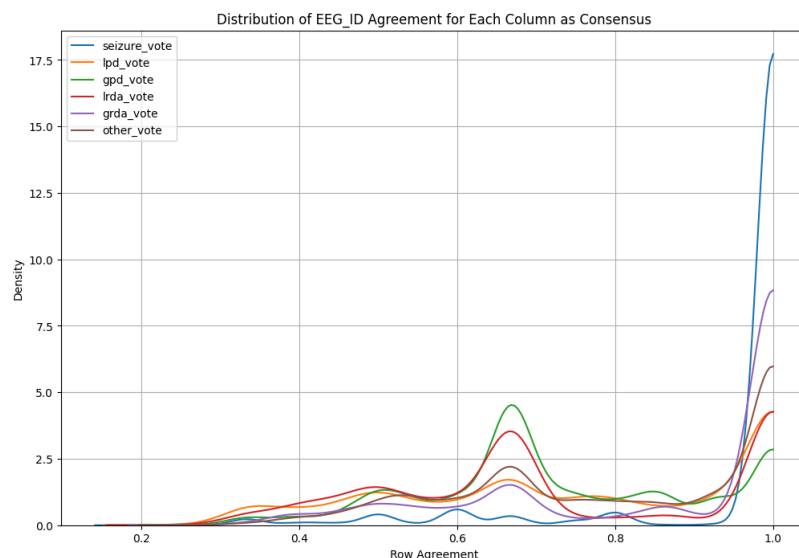
GPD.1505414177.jpeg



Other.453548281.jpeg

Fp1





Data Pre-Processing:

Deep learning models have shown remarkable success in various computer vision tasks, including image classification, object detection, and segmentation. However, achieving optimal performance requires careful preprocessing of input data. Preprocessing techniques aim to enhance data quality, reduce noise, and ensure consistency, thereby enabling models to learn robust representations. In this paper, we focus on preprocessing methods for image data, with a particular emphasis on spectrogram images used in harmful brain activity classification.

The first step in preprocessing image data involves loading the dataset and mapping image IDs to their corresponding file paths. We utilize Python libraries such as Pandas and Glob to efficiently handle data loading tasks. By creating ID mappings for EEG and spectrogram images, we establish associations between image IDs and their respective file paths.

Image preprocessing encompasses several essential steps aimed at standardizing and enhancing image data before feeding it into deep learning models. We introduce a comprehensive preprocessing pipeline tailored to spectrogram images, including normalization, resizing, and log transformation. Two instances of ImageDataGenerator are initialized for training and testing data. The images are resized to the target size of (299, 299) and normalized. These techniques ensure that pixel values are within a standardized range, enhance image contrast, and reduce the effect of outliers.

Data augmentation is a crucial technique for expanding the training dataset's diversity and improving model generalization. We discuss common augmentation techniques such as rotation, translation, and flipping, which introduce variations into the training data without altering the underlying labels. By augmenting spectrogram images, we increase the model's exposure to diverse input variations, leading to better performance on unseen data.

Splitting the dataset into training and validation sets is essential for evaluating model performance and preventing overfitting. We employ stratified splitting to ensure that each class is adequately represented in both the training and validation sets. This approach helps maintain class balance and ensures that the model learns to generalize across different classes effectively.

As mentioned, preprocessing plays a critical role in preparing image data for deep learning tasks. By employing techniques such as data loading, augmentation, normalization, and dataset splitting, we can enhance the quality of input data and improve model performance. Through practical implementation examples, we demonstrate the effectiveness of these preprocessing techniques in facilitating model training and achieving better classification results of harmful brain activity.

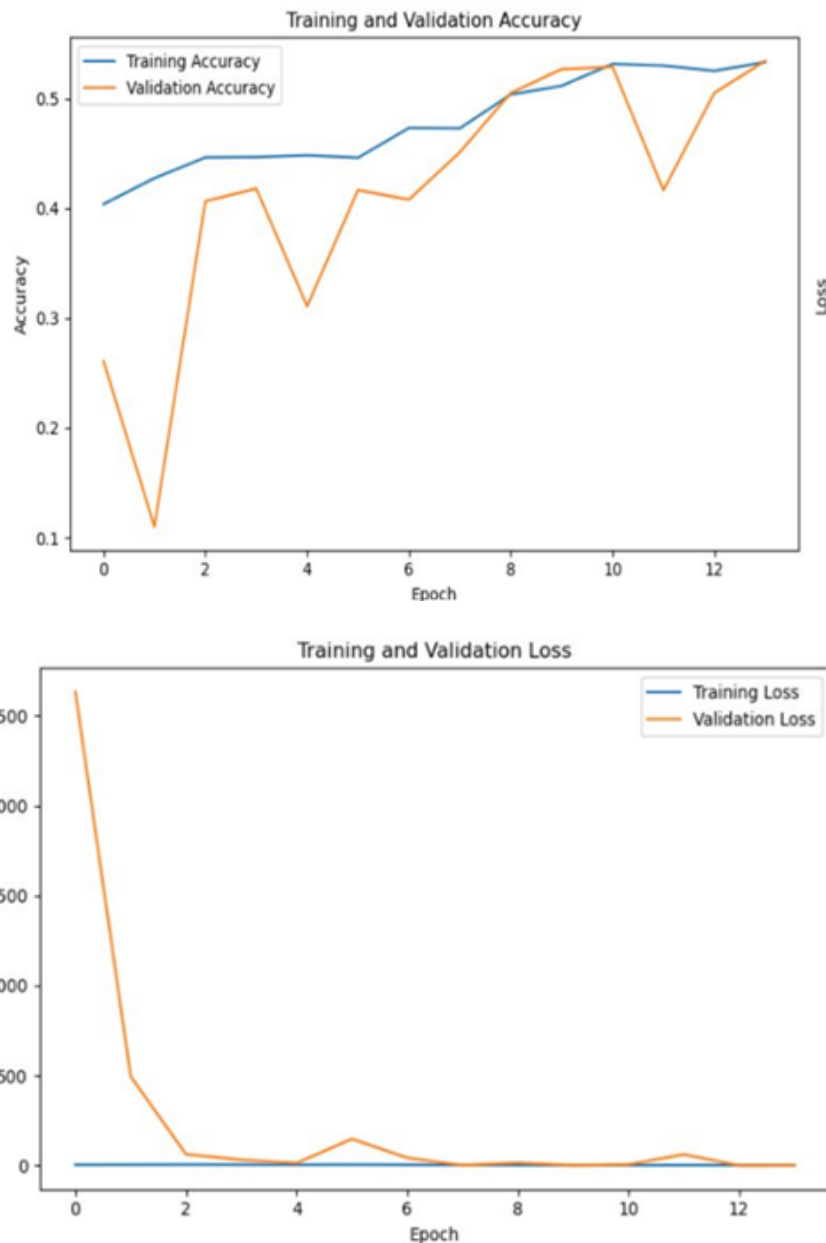
Models:

Through observing examples of successful networks and architectures in the Kaggle competition, we chose to leverage ResNet and EfficientNet models and experiment with new architectures and modeling techniques to improve performance.

1. Transfer Learning

For this experiment, we looked to deploy transfer learning with two renowned image classification models, ResNet50V2 and InceptionResNetV2. Both are popular choices for image classification tasks due to their effective architectures and strong performance in various computer vision challenges.

Both ResNet50V2 and InceptionResNetV2 have been pre-trained on large-scale datasets like ImageNet and serve as a good basis for our model. For each base model, the script loads the pre-trained weights without the top layers and freezes the initial layers. It then extracts the output features, flattens them, and concatenates them vertically. The output features from the two base models are concatenated, followed by dropout regularization to prevent overfitting. Finally, a dense layer with softmax activation is added to generate predictions for the six classes in the dataset. The combined model is compiled using the Adam optimizer and categorical cross-entropy loss function. The script trains the model on the training set while validating its performance on the validation set. ReduceLROnPlateau and EarlyStopping callbacks are used to adjust the learning rate and prevent overfitting. This model received a validation accuracy of 0.53 and a validation loss of 1.93.



2. EfficientNetB0 Model

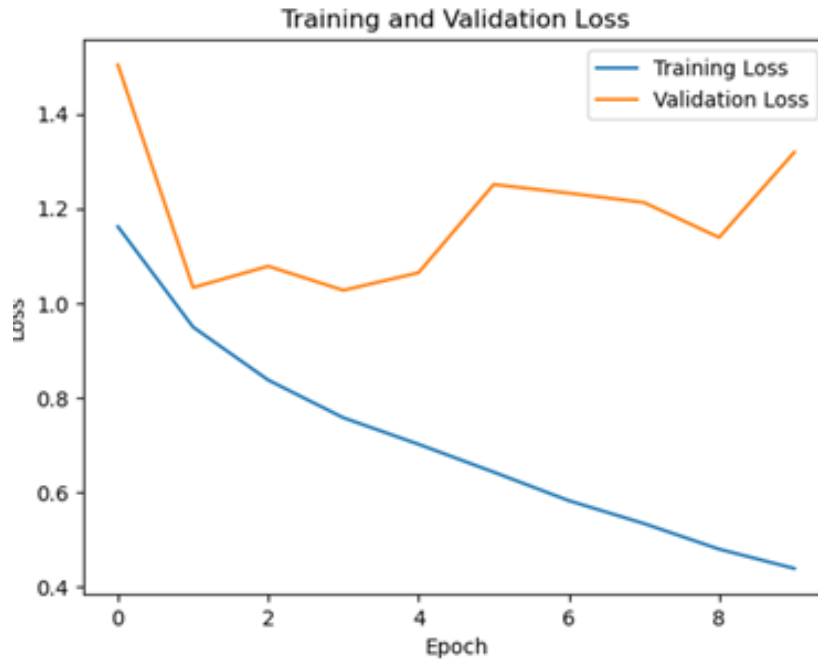
For this experiment, we created two CNN EfficientNetB0 models. We chose to explore the EfficientNetB0 model as it is well known for its effectiveness in image classification.

Model 1 achieved classification by learning features from the images through the layers of EfficientNetB0 and then predicting the class probabilities using the added classification layers. We began by loading the EfficientNetB0 model pre-trained on ImageNet without its top classification layer. To begin modifications, we added a GlobalAveragePooling2D layer to reduce the spatial dimensions of the feature maps followed by a Dense layer with softmax activation for multi-class classification, producing outputs for six classes. The model was compiled using the KL

(Kullback-Leibler) Divergence loss function and the Adam optimizer. During training, the model was fitted to the training data for 10 epochs and a batch size of 64. The training progress was monitored using both training and validation accuracy and loss metrics. To evaluate the performance, we created visualizations of the accuracy and loss for both training and validation data. This model received a validation accuracy of 0.611 and a validation loss of 1.21.

Model 2 also employed the EfficientNetB0 architecture for multi-class image classification but with a different approach. We began by loading the pre-trained EfficientNetB0 model without the top classification layer. The pre-trained layers were frozen to prevent further training. In addition to the frozen layers, we added a GlobalAveragePooling2D layer and a Dense output layer with SoftMax activation. Like the first model, this model was also compiled using the KL Divergence loss function and the Adam optimizer. This model was also trained for 10 epochs with a batch size of 64, and the training progress was visualized in accuracy and loss plots. This model benefited from the pre-trained layers, which learned valuable features from ImageNet data. By freezing those layers, computational resources were conserved, and the risk of overfitting was reduced. This model received a validation accuracy of 0.64 and a validation loss of 1.19, which improved from the first model.





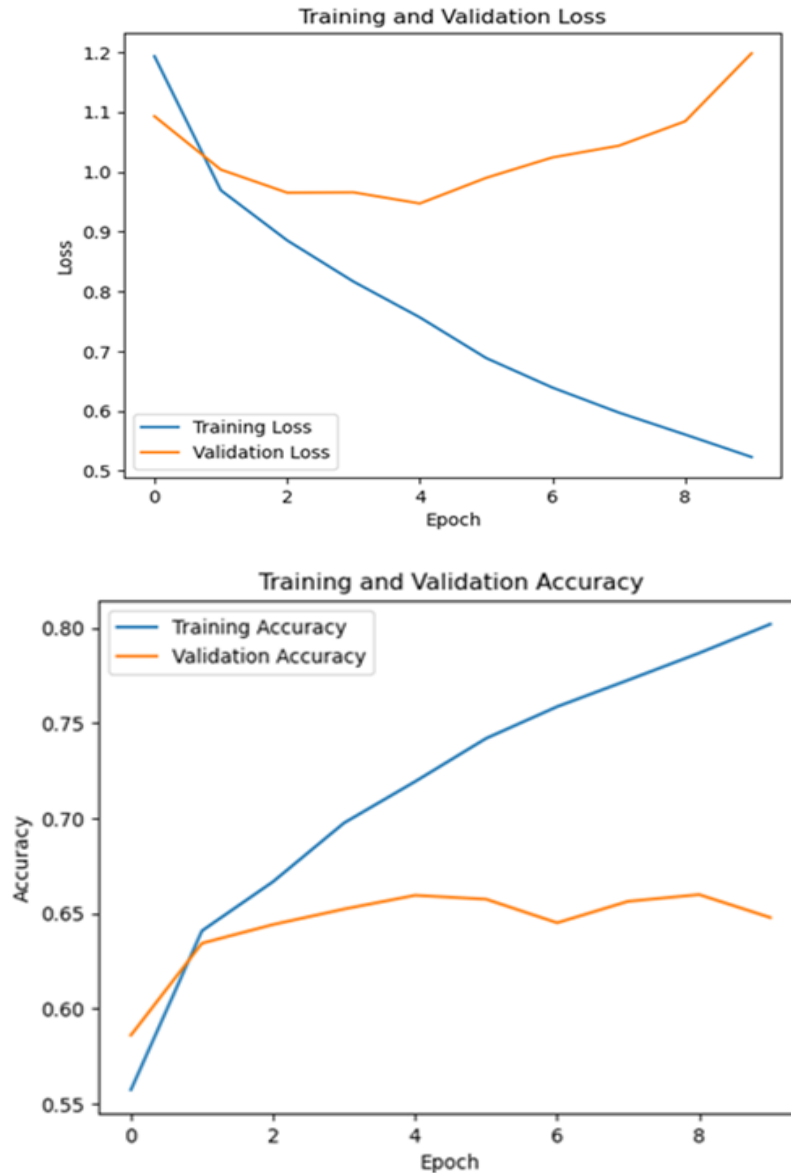
3. EfficientNetV2B2 Model

This is a pre-trained model of ImageClassifier using the EfficientNet B2 architecture. It is an upgraded version of previous model, with 6 convolutional blocks. The number of filters in each convolutional block is scaled by $\text{width_coefficient}=1.1$ and $\text{depth_coefficient}=1.2$. We ran it with 10 epochs and 32 batches.

The train accuracy is 0.8003 and validation accuracy 0.651.

We even tried running this model with hyperparameter tuning learning rate, number of epochs and batch size but there was no significant change in accuracy or loss.

EfficientNetV2B2 Model had faster training speed and better efficiency than previous models, giving us higher validation accuracy and lower loss. Hence, this turned out our best model.



The winning competition team was identified, but the specific model and results were not publicly posted. Depending on the model being run, there are varying results for the test data provided. Hence, we looked at how well our final model performed as compared to other Kaggle teams that used similar model techniques. We did this by comparing the probabilities our model predicted with the probabilities from other models using a score called KL divergence.

By comparing the predicted probabilities of our model with those from other models on Kaggle on test data, we are provided with valuable insights into our model's performance.

Classes	Kaggle Model	Kaggle Model 2	Our Model
other_vote	0.498040	0.409436	0.667508
lrda_vote	0.338065	0.406408	0.089399
lpd_vote	0.073867	0.066091	0.121445
seizure_vote	0.044966	0.094089	0.118242
grda_vote	0.041618	0.022012	0.002312
gpd_vote	0.003442	0.001964	0.001093

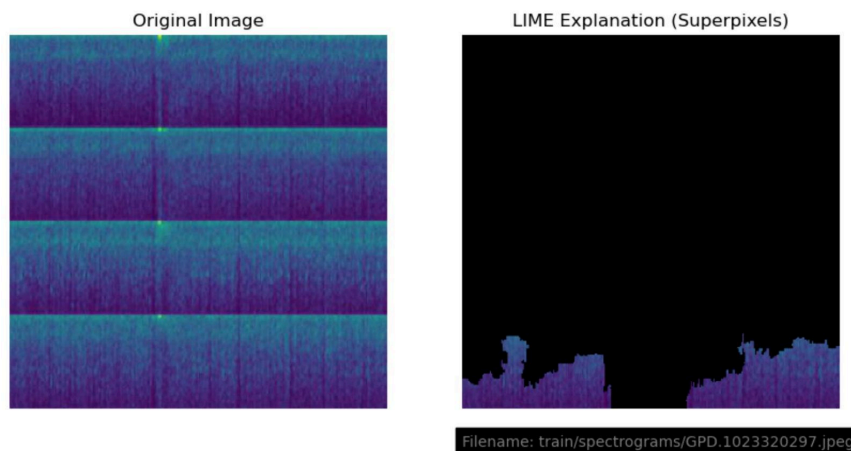
As per all the three models, the highest probability is for other_vote class. Our model could improve on lrda_vote class as it has the second highest probability after others, but our model differed on that. The least probability was for grda and gpd class, aligned between all three model results on test data (EEG ID - 3911565283). We got a KL divergence score of 1.46 suggests that there's a moderate level of difference between the two sets of probabilities for Kaggle model 1 and our final model.

This helps us see where our model is doing well and where it needs improvement compared to others.

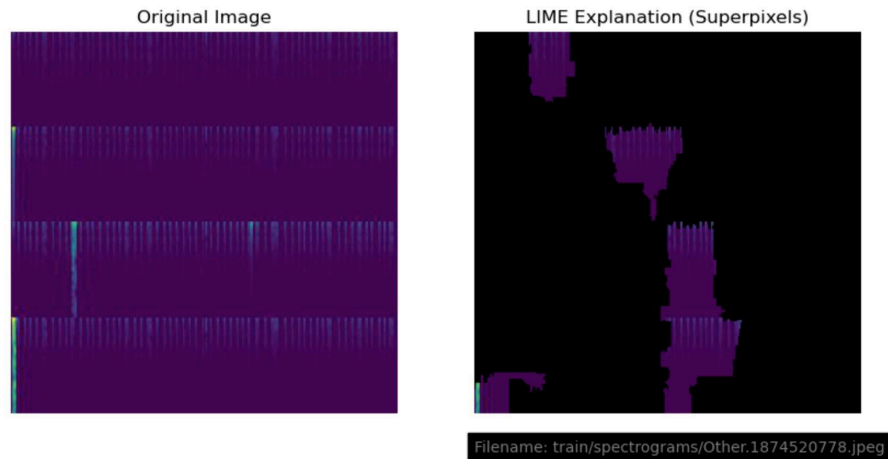
Explainable AI: Local Interpretable Model-agnostic Explanations (LIME)

After comparing model performance with respect to accuracy, we integrated explainable AI in the form of Local Interpretable Model-agnostic Explanations (LIME). LIME provides further insights into what the model is identifying as key characteristics in classification of harmful brain activity within the images. This allows domain experts to validate the model's decisions, enhancing trust in the model's outputs to make critical treatment decisions based on those results.

In the following example, LIME is showing that the EEG signal readings in the lower right and left section of the image are contributing the most to the classification of "GPD". This may differ greatly from the idealized patterns of the entire image that domain experts could interpret as meaningful in the diagnosis.



In the next LIME explanation, 4 specific areas are identified as contributing most to model classification of the “Other” classification. Only the bottom left section has any lit up readings that stand out visually from the other brain scans highlighted in the EEG. These details could be shared with specialized neurologists to further understand why certain regions are contributing to the classification or others are being ignored for further model tuning.



Conclusion:

Detecting and classifying seizures and harmful brain activity through automated EEG analysis of critical patients with deep learning models presents many challenges. Traditional EEG monitoring issues associated with manual analysis include inefficiency, subjectivity, and reliability of the individual observers. To address these points and ensure that subjective biases were not carried into the models, meticulous data preprocessing and model selection were used. Leveraging pretrained models to enhance classification accuracy and reduce the need for manual interventions minimizes the burden on healthcare workers by providing quick and accurate diagnosis of harmful activity.

EfficientNetV2B2 proved to be the most promising model tested resulting in the highest validation accuracy score. Accuracy is an important measure for the models due to the fact that consensus of manual EEG readings is below 20% in some of the cases with the imaging data. This exemplifies the significance of model architecture and tuning to achieve optimal image classification for all types of activity presented in the EEG results.

The EfficientNet models may have outperformed the ResNet models for several reasons. One reason is that EfficientNet has historically high performance on image classification since it is trained with ImageNet, which allows the model to learn features from diverse image data. Additionally, EfficientNet is highly parameter-efficient in comparison to ResNet, often requiring less computational resources. Finally, EfficientNet models are known for effectively capturing both local and global features because of their balanced scaling. The balanced scaling likely supported improved generalization in the model compared to that of the ResNet model, as the EEG patterns include a high amount of variability and complexity.

The successful application of deep learning models for classification of harmful brain activity seen in EEG imaging offers a promising path forward for automating EEG analysis. Implementing advanced machine learning techniques should be a priority for improving critical care responsiveness and reliability of harmful brain activity diagnoses, leading to better outcomes for patients and medical professionals.

Data Files:

Kaggle Competition Data Files:

<https://www.kaggle.com/competitions/hms-harmful-brain-activity-classification/data>

Files: 28463 files | Size 26.4 GB

train.csv: Metadata for the training set containing EEG samples and matched spectrograms. Includes IDs, time offsets, labels, and patient information.

test.csv: Metadata for the test set with EEG and spectrogram IDs, and patient information. No overlapping samples.

train_eegs/: EEG data from overlapping samples for training. Columns represent electrode locations and EKG data, sampled at 200 Hz.

test_eegs/: EEG data for testing, exactly 50 seconds long.

train_spectrograms/: Spectrograms assembled from EEG data for training. Columns indicate frequency and recording regions of EEG electrodes.

test_spectrograms/: Spectrograms assembled from 10 minutes of EEG data for testing.

Sources:

Jin Jing, Zhen Lin, Chaoqi Yang, Ashley Chow, Sohier Dane, Jimeng Sun, M. Brandon Westover. (2024). HMS - Harmful Brain Activity Classification. Kaggle.

<https://kaggle.com/competitions/hms-harmful-brain-activity-classification>