Detecting Harmful Brain Activity with Ensemble Learning

Ajay Pillay (apillay@umich.edu), Finn Moore (fdm@umich.edu),
Nikhil Patel (nikhilap@umich.edu), Nilay Gautam (gnilay@umich.edu),
Soham Das (sohamdas@umich.edu)

June 25, 2024

Abstract

This project seeks to address the challenge of accurately detecting and classifying forms of harmful brain activity using electroencephalography (EEG) signals from critically ill patients. Focusing on six distinct patterns, we aim to automate the analysis process, overcoming the limitations of manual review by specialized neurologists. The significance lies in advancing neurocritical care and epilepsy treatment through quicker and more accurate diagnoses. We propose a weighted ensemble model composed of three successful models: EfficientNet, ResNet, and WaveNet, significantly outperforming each model individually. Our model evaluation aligns with the competition's metric, utilizing Kullback-Leibler divergence. This project strives to contribute to accessible, accurate, and cost-effective neurocritical care solutions with transformative implications for medical treatments and diagnoses.

1 Introduction

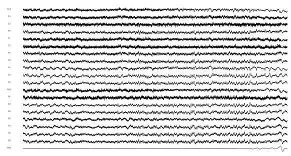
The capacity to accurately detect and classify harmful brain activity, such as seizures and various periodic discharges, is vital to the field of neurocritical care. Utilizing EEG signals from hospitalized patients, this project aims to develop a machine learning model to accurately classify six specific patterns of brain activity that indicate potential neurological issues: seizures, generalized periodic discharges, lateralized periodic discharges, lateralized rhythmic delta activity, generalized rhythmic delta activity, and "other".

Our motivation for solving this problem is driven by the need for timely and precise diagnoses in critical care settings where rapid responses are paramount. The current manual analysis of EEG signals, while accurate, is often slow, resource-intensive, and subject to human error. By automating the process through machine learning, we anticipate not only accelerating the diagnostic process but also mitigating the inconsistency of human interpretation.

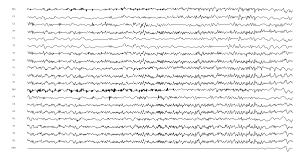
Currently, global access to neurologists is limited. An algorithm would be very easy to distribute globally without much need for human resources, granting the developing world substantially better neurocritical care. This project could also enhance patient outcomes by enabling early intervention, reduce the workload on healthcare professionals, and possibly augment our understanding of seizure patterns, leading to novel treatments. Overall, solving this problem would lead to cheaper, more accurate, and more available treatments and diagnoses.

2 Data Preprocessing

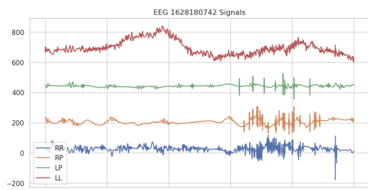
The EEG data used in this project was sourced from critically ill patients, with the aim of detecting harmful brain activity. The dataset comprised of raw EEG signals and their corresponding 11,138 spectrograms collected from 1,950 patients. Initial data handling involved cleaning and preprocessing the raw EEG data to ensure consistency and usability for machine learning applications.



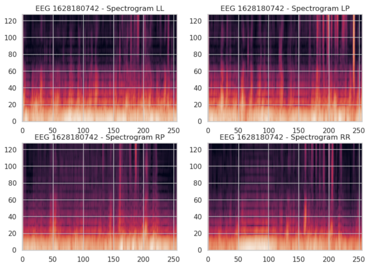
(a) EEG signals before denoising



(b) EEG signals after denoising



(c) EEG signals



(d) Mel-spectrograms

Figure 1: Data Preprocessing

Spectrogram Transformation: Raw EEG signals were denoised using Daubechies 8(db8) and transformed into spectrograms using the Short-Time Fourier Transform (STFT). This conversion aids the application of CNNs by representing the data in the time-frequency domain. We use the following formulas to calculate spectrograms.[3]

LL Spec = (spec(Fp1 - F7) + spec(F7 - T3) + spec(T3 - T5) + spec(T5 - O1))/4

LP Spec = (spec(Fp1 - F3) + spec(F3 - C3) + spec(C3 - P3) + spec(P3 - O1))/4

RP Spec = (spec(Fp2 - F4) + spec(F4 - C4) + spec(C4 - P4) + spec(P4 - O2))/4

RR Spec = (spec(Fp2 - F8) + spec(F8 - T4) + spec(T4 - T6) + spec(T6 - O2))/4

Data Augmentation: To increase the robustness of our model against overfitting and to enhance its ability to generalize across different EEG patterns, we applied data augmentation techniques such as horizontal flipping.

Filtering and Downsampling: A Butterworth low-pass filter was applied to the raw EEG signals to attenuate high-frequency noise, enhancing the signal quality relevant for brain activity analysis. Raw EEG signals were also downsampled to reduce data dimensionality, making the computational processing more efficient without losing critical information.

Both EEG signals and spectrograms were standardized and Log transformations were employed on spectrograms to improve the visibility of less prominent features.

3 Method

In order to leverage the many types of data and features given to us, we are employing an Ensemble method composed mainly of three models: EfficientNet, ResNet, and WaveNet. Building upon established models and configurations developed by other researchers, our approach focuses on refining hyperparameters and strategically integrating these models so that we can achieve better performance than using any single model alone.

3.1 Preprocessing Ablation Study

In order to test the contribution of different steps in our preprocessing pipeline, we performed an ablation study where we systematically disabled different steps. All tests were conducted using a hybrid model composed of EfficientNetB2 and ResNet trained on Kaggle Spectrograms, EEG Spectrograms, and Raw EEG Signals. The results can be seen in 1, and we can observe that removing the spectrogram standardization has the biggest effect on model performance by a wide margin.

Steps Removed	Score
None	0.593062
Spectrogram Log Transform	0.922007
Spectrogram Standardization	2.050984
EEG Standardization	0.778278
Spectrogram and EEG Standardization	2.156891

Table 1: Ablation Study Results

3.2 Evaluator Analysis

When examining the distribution of data points based on number of expert evaluators, an intriguing pattern emerges. This pattern seems to suggest the data may have been aggregated from two distinct studies, one with less than seven expert evaluators, and another with greater than ten expert evaluators per data point, as can be seen in 2. We hypothesize the data points with greater number of expert votes are of higher quality and experiment with the EfficientNetB3 model by limiting our training data to those points. This key insight allows us to achieve our best performing model.

3.3 EfficientNet

EfficientNet is a CNN that is commonly trained to classify images. It possesses a unique compound coefficient, which is used to systematically scale up the network's depth, width, and resolution in a way that achieves better accuracy and efficiency than scaling any singular dimension.

Adapting the process employed by Danial Zakaria [2]. We employ transfer learning – so we use a version of EfficientNet pre-trained on ImageNet and tune it to fit our brain activity classification task. We start with EfficientNetB2 and add a global average pooling 2D layer and a dense layer.

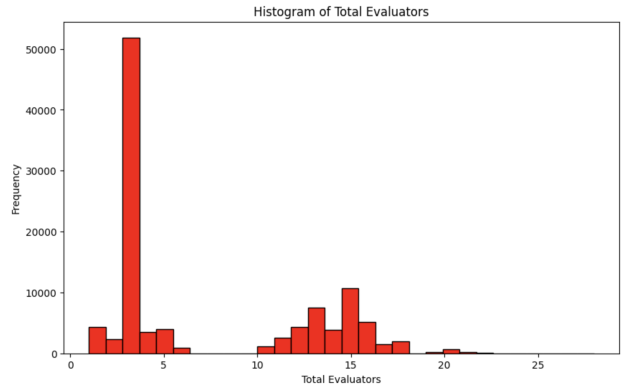


Figure 2: Histogram of Total Evaluators

The pooling layer reduces each feature map to a single value, which helps to minimize the number of parameters and avoids overfitting. The dense layer maps the pooled features to the six class probabilities we are predicting. All original layers are frozen, and then we can proceed to train just the last 2 layers on our dataset.

There are 8 different versions of EfficientNet, B0-B7, based on the compound coefficient. Referring to various discussion posts on effective models, we noted the performance of EfficientNet with respect to its compound coefficient plateaus after B3. Due to the increased training time on the deeper models, we decided to experiment with EfficientNetB3 and train it using different parameters.

3.4 ResNet

ResNet is a CNN that overcomes the vanishing gradient problem – an issue where the gradient becomes too small for effective learning as the depth of the network increases. By adding residual connections that allow the gradient to skip layers, we can construct models with many layers. Thus, since we often need to process around 10,000 time steps for each sample, ResNet is well suited for time-series data. Additionally, we are able to include both convolutional layers, which can learn local spatial patterns, and recurrent layers, which can model long-range temporal dependencies.

We are adapting a modified version of ResNet – called EEGNet – from Med Ali Bouchhioua [1]. We combine 9 ResNet_1D_Blocks in a sequential manner with different kernel sizes, strides, and padding. Each ResNet_1D_Block consists of two 1D convolutional layers with batch normalization and ReLU activation. We can concatenate the output of these blocks with the output of a recurrent layer and pass it through a fully connected layer to produce the final predictions.

3.5 WaveNet

When administering an EEG, the electrodes are placed in a formation such that they form 4 montage chains that track the signal across different paths in the brain. Instead of processing the data from all electrodes, it is sometimes useful to group the data by chain and process them in parallel – a process adapted from Danial Zakaria [2].

WaveNet has an exceptional ability to model long-range dependencies in waveform data. Using dilated convolutions, the kernel is spaced out, which allows it to have larger coverage with the same number of parameters and computational cost. This also preserves the temporal resolution, which would be lost with pooled or strided convolutions. Additionally, the network can vary the dilation rate to integrate features from multiple temporal scales – which is important since EEG data contains both short-term and long-term patterns.

3.6 Vision Transformer

Vision Transformers (ViTs) are a relatively new architecture that can leverage the power of transformers to classify images. CNNs traditionally use spatial convolutions to extract features. However, ViTs divide the input image into patches and process them as a sequence. This allows them to capture global dependencies in the data very effectively. This power makes them especially well suited for processing EEGs. We use a ViT pretrained on ImageNet and fine tune it using our HMS data. Unfortunately, this method did not yield promising results, achieving a public score no better than 4.9, which is considered quite poor. However, we also experimented with a hybrid vision transformer that performed two convolutions during the patch embedding process, with the hope that this would help filter for frequency bands of interest. This modification led to a score of 1.37, an improvement that was adequate but still fell short of our expectations, leading us to deprioritize this architecture in favor of other alternatives.

3.7 Ensemble Pipeline

We have multiple types of data: Kaggle's spectrograms (K), EEG spectrograms (E), and Raw EEG signals (R). Spectrograms are images while EEGs are time series waveform data. We currently build one ResNet model using just EEG Data, as well as 7 models using either EfficientNet, WaveNet, or both using some combination of all types of available data. This yields 8 models which can be seen in table 2. In instances where EfficientNet and WaveNet are used together, a hybrid model is created by concatenating their output layers and passing them through a dense layer.

During inference, we run the input through all 8 models and average the results, weighting the prediction of each model using their individual performance when run on the test set. However, there are multiple interesting ways to optimize the weight function in the future. A visualization of the inference process can be seen in Figure 3.

3.8 Evaluation

All submissions to the competition are evaluated on the Kullback-Leibler (KL) divergence between the predicted probability and the observed target probability. KL divergence is a measure of how different two probability distributions are. In this case, it quantifies the difference between the predicted probability distribution (by our model) and the observed target probability distribution (determined by human experts). Formally, for discrete probability distributions P and Q on the same sample space $\mathcal{X}$, the KL divergence D_{KL} is defined as follows:

$$D_{KL}(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

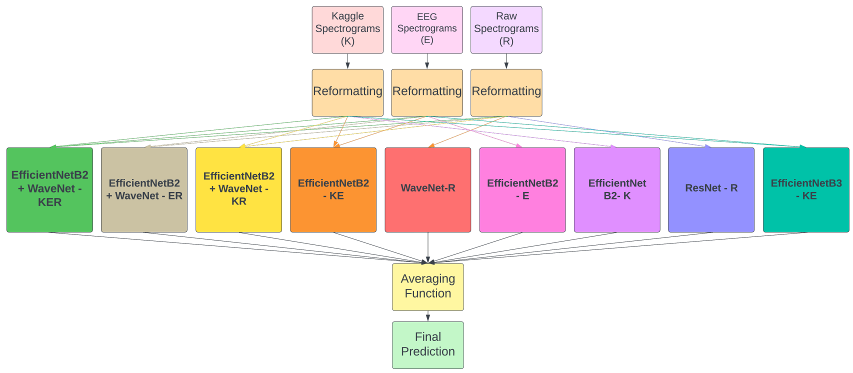


Figure 3: Block Diagram of Inference Pipeline

KL divergence is undefined when $P(x) = 0$ and $Q(x) > 0$, so the convention adopted in the competition is to explicitly set the contribution to 0 in these cases.

4 Related Work

The detection and classification of harmful brain activity using EEG signals have seen considerable advancements over the years, with a number of existing methods and state-of-the-art approaches being developed. Traditional methods primarily relied on manual analysis by specialized neurologists, which, despite its accuracy, is labor-intensive and requires specialized skills. With the evolution of machine learning and deep learning, automated methods have gained prominence. Notably, architectures such as EEGNet, EfficientNet, WaveNet and Long Short-Term Memory (LSTM) have emerged as powerful tools for processing and interpreting EEG signals, capitalizing on their ability to capture spatial and temporal features inherent in the data.

Our approach builds upon these existing methodologies by integrating ensemble techniques that leverage the strengths of EfficientNet, ResNet, and WaveNet, thus facilitating a comprehensive analysis of EEG signals from various aspects. Similar to other works, we utilize CNNs and deep learning models capable of handling the complexity of EEG data to accurately classify harmful brain activity patterns. This alignment with contemporary research underscores our commitment to employing proven, effective techniques in the field.

However, our methodology distinguishes itself through the use of an ensemble model that combines the predictions from multiple deep learning architectures, each tailored to specific types of data within the EEG spectrum. This innovative approach aims to harness the diverse characteristics of EEG signals with the aim to achieve classification performance beyond what individual models can do on their own. By blending the predictive power of EfficientNet, ResNet, and WaveNet, our model seeks to mitigate the weaknesses of any single approach, thereby providing a more robust and accurate classification system. This ensemble strategy represents a novel contribution to the field, demonstrating an advanced application of machine learning techniques to the critical task of detecting harmful brain activity.

5 Results

Kaggle evaluates submissions against a hidden test set and a leaderboard (LB) score is published. The lower the score, the more closely the predicted distribution matches the true distribution. Hence, the objective is to obtain as low of a score as possible. There are two types of leaderboard: public and private. The public leaderboard is based on evaluation over a smaller dataset available before the conclusion of the competition. The private leaderboard is the final evaluation of the submitted models based on a larger dataset. The final ranking is done based on this private LB scores and it is available after the competition ended. We shall present the quality of model performances based on their LB scores. Note that the public and private LB scores might be different for the model.

The ensemble model involved training each individual model on different folds of the training data, and then combining the prediction results from the weights obtained from each individual fold. The ensemble model made use of 5-fold cross-validation, which we adapted the ResNet model to follow as well. The models, when evaluated individually, gave the following range of public LB scores as outlined in Table 2.

MODEL	DATA TYPE	PUBLIC LB SCORE
ResNet	R	0.43
EfficientNetB2	K	0.41
EfficientNetB2	E	0.39
WaveNet	R	0.41
EfficientNetB2	KE	0.37
EfficientNetB2 + WaveNet	KR	0.39
EfficientNetB2 + WaveNet	ER	0.38
EfficientNetB2 + WaveNet	KER	0.36

Table 2: Initial Models against LB Scores

The ensemble model combines the predictions in the form of a weighted average based on each model's LB score. Models that perform better contribute more to the predicted outcome and vice versa for models that perform worse. Interestingly enough, the ensemble model performs better than any of the individual models, obtaining an LB score of 0.34, with or without ResNet. We aim to incorporate other models into this ensemble as well to bring down the LB score further. The following table shows the final ensemble models that we submitted and their public and private LB scores. LB scores of our ensemble models are provided in Table 3.

MODEL	LB SCORE	
	Public	**Private**
Initial Ensemble Inference	0.345	0.417
ResNet + Transformer	0.353	0.429
EfficientNetB3 Inference	0.350	0.421
Ensemble + ResNet + EfficientNetB3	0.335	0.408
KE + ER + KER + EfficientNetB3	0.321	0.390
KER + EfficientNetB3	**0.318**	**0.386**

Table 3: Ensemble models against LB Scores

The final private LB score for our ensemble method was 0.389 ranking 912 (top 33%), compared to the lowest LB score of 0.272. Post-deadline, we got this score down to 0.386. Our results confirm the effectiveness of aggregating predictions across diverse models to enhance overall performance.

6 Conclusion

In conclusion, we have successfully navigated the initial stages of data pre-processing and understanding existing models, experimenting with a method of weighted ensemble to achieve a ranking in the top 33% of all competitors. Through the development of an ensemble model comprising EfficientNet, ResNet, and WaveNet, we have demonstrated significant improvements in classification performance compared to individual models.

Furthermore, our ensemble pipeline incorporates multiple types of data, including Kaggle's spectrograms, EEG spectrograms, and raw EEG signals, leveraging the strengths of each data type to achieve better classification results. We identified key insights such as the importance of expert evaluator analysis and the effectiveness of ensemble methods in improving model performance.

7 Author Contributions

Soham and Nilay explored data preprocessing and some initial models involving LSTM. Ajay, Nikhil, and Finn explored publicly available models to understand the landscape at the time. All co-authors were involved in the training and testing pipelines upon converging on the ensemble model approach. All co-authors were involved in writing this report and contributed equally to this project. This information can also be seen in Table 4.

Contribution	**Soham**	**Nilay**	**Ajay**	**Nikhil**	**Finn**
Data Pre-processing		✓			
Explored Initial LSTM Models	✓	✓			
Explored Public Models			✓	✓	✓
Training and Testing Pipelines	✓	✓	✓	✓	✓
Writing the Report	✓	✓	✓	✓	✓

Table 4: Summary of author contributions to the project

8 References

[1] Med Ali Bouchhioua. HMS Resnet1D-GRU Inference — kaggle.com. https://www.kaggle.com/code/medali1992/hms-resnet1d-gru-inference, 2024. [Accessed 03-10-2024].

[2] Danial Zakaria. Features+Head Starter — kaggle.com. https://www.kaggle.com/code/nartaa/features-head-starter/notebook, 2024. [Accessed 03-10-2024].

[3] Danial Zakaria. HMS EEG — kaggle.com. https://www.kaggle.com/datasets/nartaa/hms-eeg, 2024. [Accessed 03-10-2024].