
# Report on Harmful Brain Activity Classification

from the course of studies Mathematical Foundations for Data Science

at the Beijing University of Posts and Telecommunications

by

**Hao Chen**

23.06.2024

| | |
|---|---|
| **Student ID, Course:** | 2022213763, Final Report |
| **Company:** | Queen Mary School Hainan, 100080, Beijing, China |
| **Supervisor in the Company:** | Professor Yonggang Qi, Professor Yang Yang |

# Table of Contents

# List of Acronyms

| | |
|---|---|
| **CNN** | Convolutional Neural Network, a type of neural network designed for processing data with grid-like topology, such as images and videos, widely used in computer vision. |
| **EEG** | Electroencephalogram, a method of recording brain electrical activity used to diagnose neurological |
| **PCKLS** | Parallel Convolutional Kernels for Long and Short-term features, a method for extracting features simultaneously in time and frequency domains, suitable for capturing dynamic signal changes and complex feature extraction. |
| **ViT** | Vision Transformer, a visual processing model based on the Transformer architecture, suitable for tasks like image classification and processing. |
| **iAFF** | integrated Adaptive Feature Fusion, a method for integrating information from different data sources to enhance model accuracy and robustness. |

# Abstract

This study addresses the critical need for accurate detection and classification of abnormal brain activities using EEG signals. We introduce a multimodal approach combining EEG and spectrogram data, employing the iAFF method for effective feature fusion. Our PCKLS feature extraction method enhances sensitivity to signal dynamics, while EfficientNet B0 proves optimal for processing spectrograms. Leveraging two-stage training and intra-group cross-validation, we achieve robust model performance, advancing neurocritical care and epilepsy treatment. These contributions offer new tools for improving treatment outcomes and enhancing understanding in neuroscience.

# 1. Introduction

Detecting abnormal brain activity in patients and diagnosing signs of potential brain injury are among the most pressing issues in the medical field today. Accurately analyzing electroencephalogram (EEG) signals and precisely classifying seizures and other harmful brain activities hold transformative benefits for neurocritical care, epilepsy treatment, and drug development. Globally, approximately 5% of the population suffers from epilepsy, with around 150,000 deaths annually attributed to epilepsy-related causes, including accidents and other complications resulting from seizures. Currently, EEG monitoring relies entirely on manual analysis by specialized neurologists. This process is not only time-consuming and expensive but also prone to fatigue-related errors and inconsistencies in assessment.

Therefore, developing a model capable of accurately detecting and classifying seizures and other harmful brain activities is crucial. In our project, we use 106,800 rows of EEG data paired with corresponding 2D spectrograms as inputs to develop a model that predicts the true disease labels of samples. On the test set, the closer the predicted distribution is to the real distribution, the smaller the Kullback-Leibler (KL) divergence [1], indicating more accurate predictions.

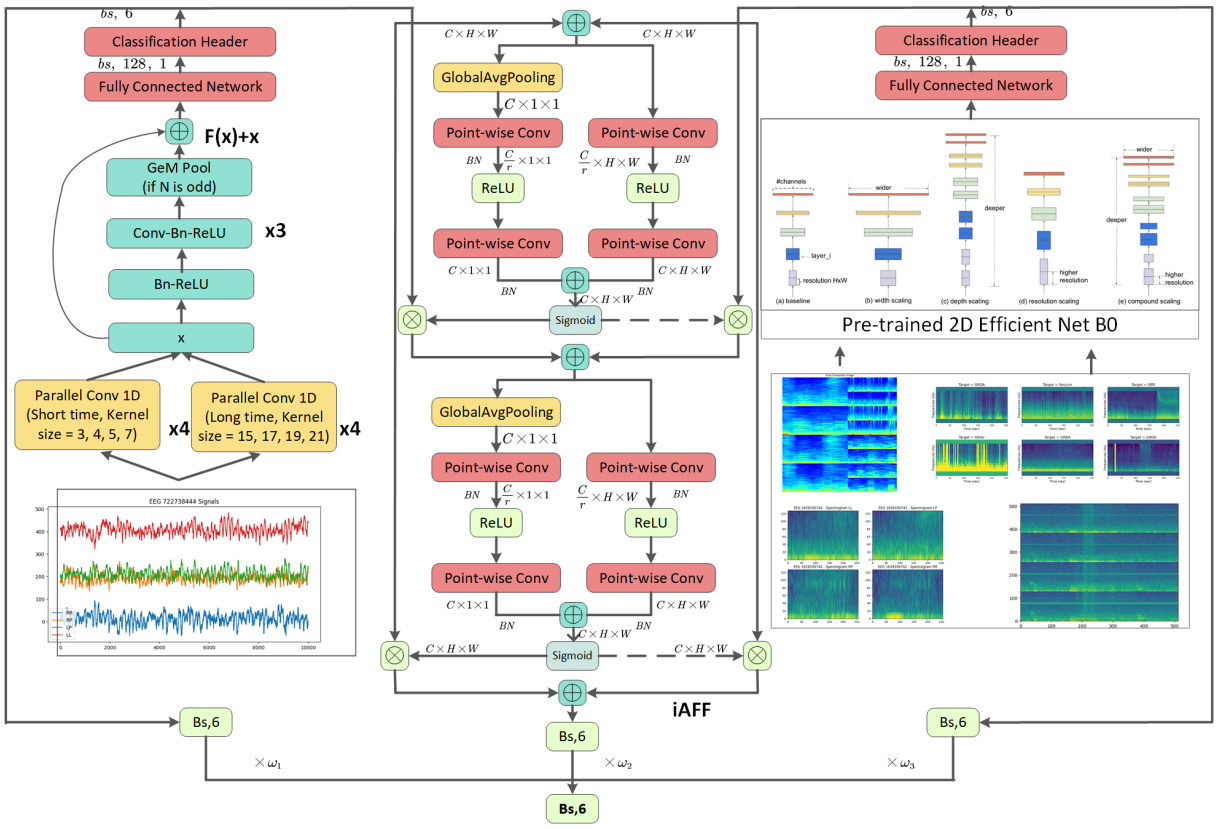


Figure 1 — Flow chart

Given that EEG signals and their corresponding 600-second spectrograms are provided in our project, we introduce a feature fusion method. Conventional feature fusion methods typically integrate information from different sources or modalities into a unified feature representation, using approaches such as simple weighted averaging, concatenation, or parallel processing. However, for complex multimodal tasks, these traditional methods may face challenges due to differences between modalities, insufficient complementarity of information, or mismatched feature dimensions, limiting their effectiveness.

Hence, for integrating one-dimensional EEG signals and two-dimensional spectrograms, we utilize the iAFF method [2] , which addresses these challenges more effectively. By combining EEG and spectrogram features, iAFF leverages complementary information within multimodal data, thereby enhancing task accuracy and robustness. This method not only considers feature integration but also effectively models the structural characteristics of different data types, demonstrating significant advantages in multimodal contexts.

In traditional EEG signal processing, common feature extraction methods include techniques based on time-domain or frequency-domain analysis. Time-domain methods typically use filters, statistical metrics, or temporal analyses to capture dynamic changes and waveform characteristics of signals (e.g., mean, variance, slope). While straightforward and efficient for capturing basic signal features, these methods may be limited for complex EEG signals. Frequency-domain methods provide information on signal distribution in the frequency domain but may lack direct insights into dynamic changes in the time domain, limiting their utility for tasks requiring integrated time-frequency information.

In contrast, our approach, the Parallel Convolutional Kernels for Long and Short-term features (PCKLS) [3], effectively overcomes these limitations. By simultaneously considering long and short-term convolutional kernels, PCKLS extracts multi-level feature representations in the time domain, capturing temporal patterns and frequency characteristics across different time scales. This parallel structure enhances sensitivity to signal dynamics, improves feature extraction complexity and expressiveness, and proves highly applicable and advantageous in EEG feature extraction.

Spectrograms corresponding to EEG signals provide more semantically rich features to models. While the spectrograms provided in our project correspond to a 600-second duration, clinical focus typically centers on a central 10-second segment of each EEG sample. Hence, we reverse-engineer the transformation between EEG and spectrogram data, generating spectrogram data corresponding to 50 seconds and 10 seconds. By employing gradient concatenation and two other concatenation methods, we derive four types of spectrogram data: **original spectrograms**, **spectrograms concatenated with gradients by size**, **spectrograms corresponding to 50 seconds** and **vertically stacked spectrograms**, serving as inputs for our 2D model (as shown in Figure 1). After comparing results with 2D CNN and Vision Transformer (ViT) models, we select EfficientNet B0 for its superior prediction accuracy.

In initial attempts with single models and single-class data, we achieved prediction accuracy close to random guessing. Given varying total vote counts ranging from 3 to 20, we adjusted model depth and training methods. Notably, we adopted a two-stage training approach: first training on all data and then retraining on data with total votes exceeding 10, emphasizing these data points more. Moreover, total vote count serves as a confidence measure for label accuracy, mitigating noise from low vote counts.

Finally, we employ intra-group cross-validation among the 1D model, 2D model, and fused model. As EEG samples from the same individuals exhibit high similarity, preventing overfitting and inflated accuracy, we utilize GroupKFold cross-validation. Coefficients derived from cross-validation results of the three models are weighted based on performance, yielding a final prediction as the weighted sum of the three results, concluding our comprehensive task approach.

In summary, our contributions are as followed:

1. **Development of a multimodal EEG analysis model:** We developed a model that integrates EEG signals and spectrograms to accurately detect and classify epileptic seizures and other harmful brain activities, enhancing neurocritical care and epilepsy treatment outcomes.
2. **Introduction of the iAFF feature fusion method:** We introduced the iAFF method, effectively integrating EEG signals and spectrograms to improve the accuracy and robustness of the classification model.
3. **Inference of EEG-to-spectrogram conversion formula and generation of diverse spectrogram data:** We reverse-engineered the conversion between EEG and spectrogram data, generating diverse spectrogram data across different time segments to enhance inputs for 2D models.
4. **Innovative PCKLS EEG feature extraction method:** We developed the PCKLS method, extracting EEG features simultaneously in the time and frequency domains, enhancing sensitivity to dynamic signal changes and complex feature extraction.
5. **Preference of EfficientNet as the optimal model:** Through comparative analysis, we identified EfficientNet B0 as the optimal choice for processing spectrogram data. We utilized two-stage training and intra-group cross-validation to ensure the model's robustness and accuracy.

These efforts advance the field of neuroscience, providing new tools and methods to improve treatment outcomes for epilepsy patients and enhance neurocritical care.

## 2. Related Work

### Convolutional Neural Networks (CNNs) in EEG Signal Analysis

Convolutional Neural Networks[4] (CNNs) have been extensively applied in EEG signal processing due to their ability to automatically learn hierarchical representations of data. In the context of disease prediction from EEG signals, CNNs have shown promise in capturing both spatial and temporal dependencies within 1D EEG data. Typically, CNN architectures involve convolutional layers for feature extraction followed by fully connected layers for classification. Recent advancements include the integration of parallel convolutional layers to capture multi-scale features effectively, which is crucial for analyzing EEG signals that exhibit varying frequencies and complex temporal patterns.

### Vision Transformer (ViT) for Image-Based EEG Spectrogram Analysis

Vision Transformer[5] (ViT) has revolutionized image classification tasks by leveraging self-attention mechanisms to capture global dependencies across image patches. Applied to EEG spectrogram analysis, ViT treats the 2D spectrogram data as sequences of tokens, enabling it to model interdependencies among different frequency bands and temporal segments effectively. While ViT excels in handling 2D data representations like spectrograms, its adaptation to EEG signal analysis requires careful consideration of tokenization strategies and the interpretation of learned representations in the context of disease prediction tasks.

### Two-Stage Learning Approaches

Two-stage learning approaches[6] are gaining popularity in medical signal analysis, particularly for integrating information from diverse data modalities such as 1D EEG signals and 2D spectrograms. In the first stage, specialized models like CNNs are employed to extract discriminative features from each modality independently. This stage focuses on capturing modality-specific characteristics crucial for disease identification. In the second stage, these learned features are fused at a higher level to enhance predictive performance. Such approaches leverage the strengths of individual modalities while mitigating their respective limitations, thereby improving overall diagnostic accuracy.

### Feature Fusion and Multimodal Integration

Feature fusion plays a pivotal role in combining complementary information extracted from different modalities to enhance disease prediction accuracy. In the context of EEG analysis, integrating features derived from CNNs processing 1D data with those from ViT processing 2D spectrograms can capture both temporal dynamics and frequency distributions relevant to disease states. Effective fusion strategies include concatenation, attention mechanisms, or multi-layer perceptrons, which aim to preserve informative aspects from each modality while suppressing noise and irrelevant features.

# 3. Methodology

We will structure the Methodology section into four parts: **Problem**, **Dataset**, **Data Engineering** and **Algorithms**. Significant data will be mentioned, while detailed data and their significance will be referenced in the `README.md` file.

## 3.1. Problem (Definition)

The main goal of this task is to develop a model that trains on long 50-second electroencephalogram (**EEG**) signals recorded from critically ill hospital patients, along with long 600-second **spectrograms**. The model's objective is to accurately classify epileptic seizures and other harmful brain activities based on EEG data. Ultimately, it will predict on undisclosed annotated data and evaluate performance using the ***Kullback-Liebler*** divergence between predicted probabilities and observed targets.

While the task appears not highly challenging, it will face issues such as **slow training due to large dataset sizes**, **inefficient prediction due to the long timespan of spectrograms**, and **varying confidence levels due to fluctuating vote counts**. Furthermore, single-model and single-data prediction accuracies are notably low, demanding higher requirements for feature engineering and model fusion.

## 3.2. Dataset

The data is primarily divided into two categories. **One category** consists of one-dimensional electroencephalogram (EEG) data from one or multiple overlapping samples, containing the names of electrode positions of the EEG leads. They are collected at a rate of 200 samples per second.

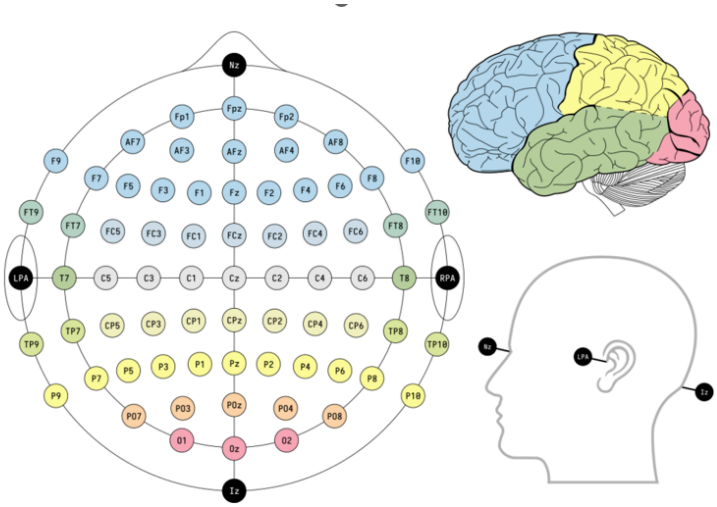


Figure 2 — Electrode positions correspond to regions of the brain

The names of electrode positions correspond to regions of the brain, including the frontal lobe (F), temporal lobe (T), central sulcus (C), parietal lobe (P), and occipital lobe (O). These letters may be accompanied by numbers, where odd and even numbers indicate electrodes on the left and right sides of the midline of the head, respectively. “z” indicates electrodes on the midline (as shown in Figure 2). In this project, they are sampled in simplified form (as shown in Figure 3).

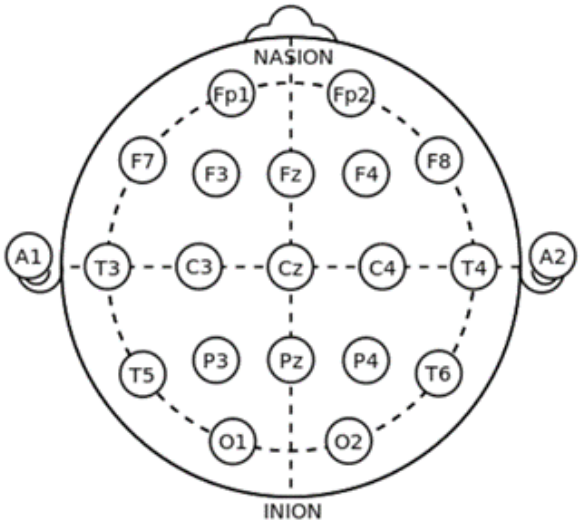


Figure 3 — Simplified positions

Abbreviations corresponding to anatomical positions are as follows: LL = left lateral; RL = right lateral; LP = left parasagittal; RP = right parasagittal. EEG curves are derived by subtracting adjacent electrodes within each region (as shown in Figure 4).

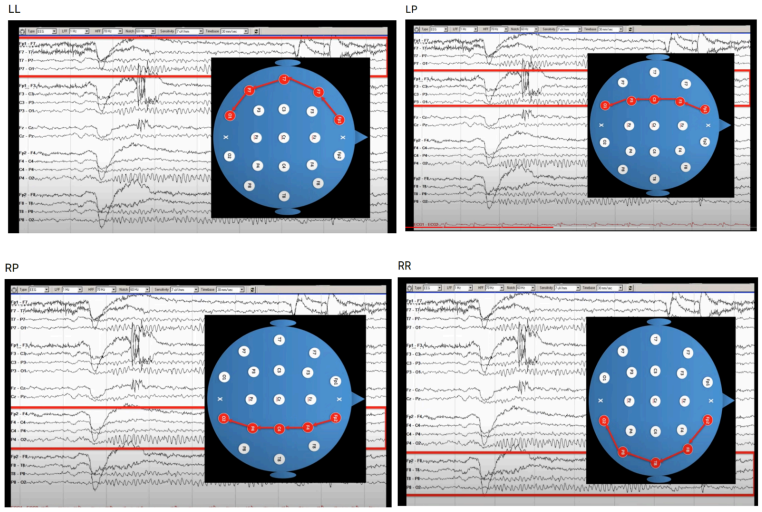


Figure 4 — Abbreviations corresponding to anatomical positions

The other category consists of two-dimensional spectrograms collected from EEG data, containing information such as frequency and the recording areas of EEG electrodes (as shown in Figure 5). Spectrogram data are transformed from EEG using signal processing methods like Fourier transform. They are included in the dataset and can be directly accessed, though specific transformation methods for these spectrograms are not provided.

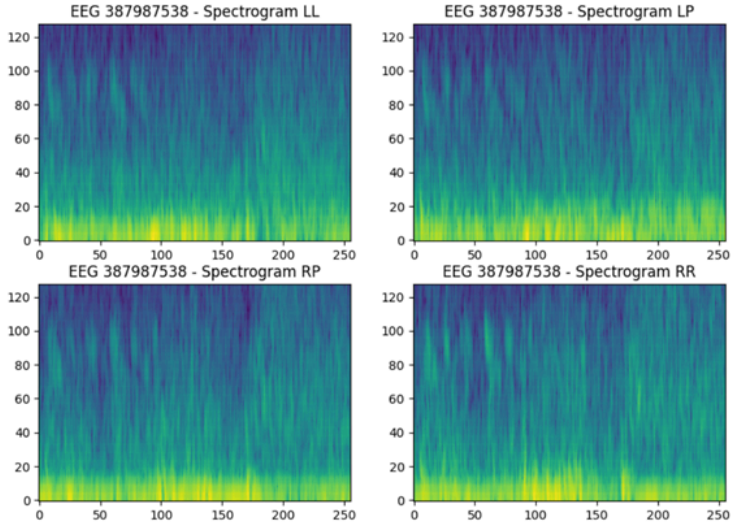


Figure 5 — Spectrograms collected from EEG data with a duration of 600s

Each sample from these two types of data undergoes simultaneous review and voting by 3-20 experts to preliminarily determine its correct label (6 classes), including seizure activity (SZ), generalized periodic discharges (GPD), lateralized periodic discharges (LPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), or "other." In some cases where experts unanimously agree on the correct label, it is termed "idealized patterns." Approximately half of the experts label samples as "other," while the other half assign

one of the remaining five labels, referred to as “proto patterns.” Experts generally fall into two of the five naming patterns, which are termed “edge cases” (as shown in Figure 6).

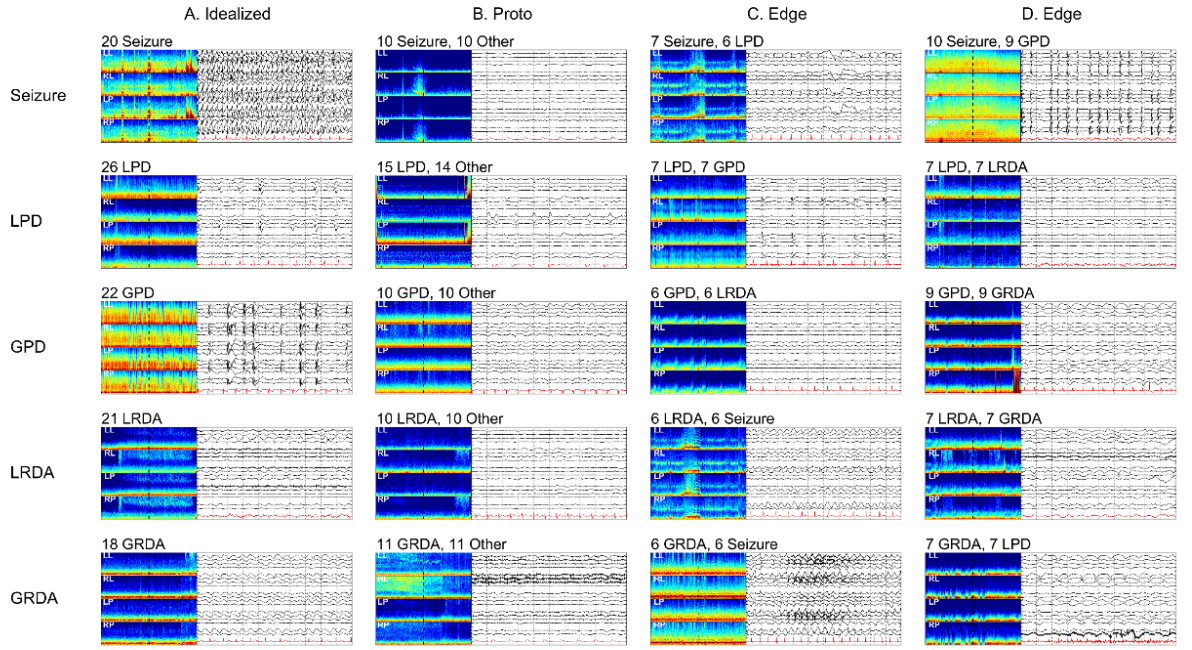


Figure 6 — Four types of voting distributions

Annotators reviewed 50-second EEG samples and corresponding spectrograms covering a 10-minute window, which were simultaneously centered and labeled for the central 10 seconds (as shown in Figure 7). Additionally, separate test sets for exactly 50 seconds and exactly 10 minutes of data were provided. The dataset comprises 106,800 rows of data, involving 17,089 unique EEG records (eeg_ids), 11,138 unique spectrogram records (spectrogram_ids), and 1,950 unique patients (patient_ids). This is because some EEG records and spectrograms were sampled multiple times or covered by different time windows.

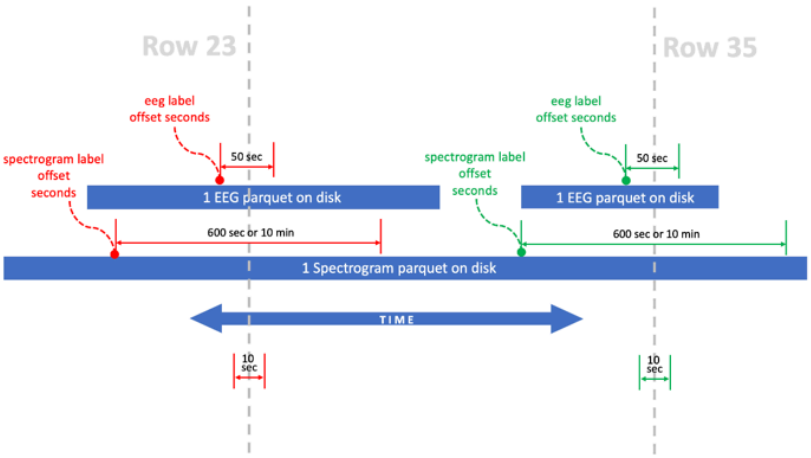


Figure 7 — Time period diagram

## 3.3. Data Engineering

**Data Filtering** There are many samples with the same EEGid in the data (as shown in Figure 8). To prevent deterioration in model prediction caused by noise from duplicate data, the data needs to undergo a filtering step. In Experiment 1, we discussed various filtering rules and ultimately chose to keep only the first row of data for each identical EEGid, which significantly improved model predictions.

| eeg_id | eeg_label_offset_seconds | expert_consensus | seizure_vote | lpd_vote | gpd_vote | lrda_vote | grda_vote | other_vote |
|---|---|---|---|---|---|---|---|---|
| 2578018731 | 0.0 | Other | 0 | 0 | 0 | 0 | 0 | 1 |
| 2578018731 | 20.0 | GRDA | 0 | 0 | 2 | 0 | 9 | 2 |
| 2578018731 | 26.0 | GRDA | 0 | 0 | 2 | 0 | 9 | 2 |
| 2578018731 | 32.0 | GRDA | 0 | 0 | 2 | 0 | 9 | 2 |

Figure 8 — Time period diagram

**Confidence Level and Normalization** Because some data only have votes from 3 experts, such results are not entirely convincing. We first calculate the total number of votes for each sample, then normalize all voting results by dividing by the total votes to obtain the proportion of each voting result. At the same time, we treat the total number of votes as a symbol of confidence level; the higher the total votes, the more trustworthy the voting results.

**Spectrum Transformation** Spectrum images are effective additional data for improving model prediction accuracy. However, the competition provided only one spectrum image with a total duration of 600 seconds. Nevertheless, data corresponding to 50 seconds of EEG data are more valuable than those corresponding to 600 seconds. We exported the given spectrum image transformation method using computational tools and obtained a spectrum image corresponding to 50 seconds of EEG data based on this transformation method. We mapped it into a matrix of the same shape as the 128*256*4 dimensions of the 600-second spectrum image.

**Gradient Concatenation of Spectrum Images** We concatenate spectrum images with size gradients, including the central 10-second spectrum image, the spectrum image corresponding to EEG 50 seconds, and the provided 600-second spectrum image. We map and concatenate them in descending order of size, corresponding to their relative importance.

## 3.4. Algorithm

The task primarily utilized a multimodal feature fusion approach. For the 1D EEG signal, features capturing long-term and short-term characteristics were extracted by feeding the signal into parallel one-dimensional convolutional layers with different kernel sizes. These features were then processed through multiple residual 1D convolutional layers and a recurrent neural network (**RNN**) classification head, resulting in a 6-dimensional vector. Simultaneously, four types of 2D spectrograms were inputted into a pre-trained **EfficientNet B0** model. After passing through an RNN classification head, these spectrograms were also transformed into 6-dimensional vectors.

The outputs from the 1D and 2D models were further fused using the **iAFF** model for multimodal feature integration, yielding a final 6-dimensional vector. These three 6-dimensional vectors underwent **GroupKFold** cross-validation internally, where the quality of the cross-validation results determined the corresponding weights $\omega_i (i = 1, 2, 3)$. These weights were used to aggregate the results based on their respective performance evaluated using the **Kullback-Leibler** divergence.

$$D_{\mathbb{KL}}(P \parallel Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right)$$

It is a method for measuring the difference between two probability distributions. Here, $P(i)$ and $Q(i)$ are the probability density or probability mass functions of distributions $P$ and $Q$, respectively, for event or variable $i$. KL divergence is non-negative and equals zero if and only if $P = Q$, indicating identical distributions.

#### 3.4.1. Parallel one-dimensional convolution

Parallel one-dimensional convolution is a method used for time series prediction and analysis, particularly effective in handling sequential data. It combines both long-term and short-term information to better capture patterns and trends within time series data. This approach involves using two independent convolutional neural networks simultaneously.

The long-term convolution utilizes larger kernels to capture long-range dependencies and trends in the time series, aiding in identifying patterns and periodic variations over extended periods. Suppose the time series is $x = [x_1, x_2, ..., x_T]$ where $T$ is the length of the time series, the output of the long-term convolution can be represented by:

$$y_t^{long} = f^{long}\left(x_{t-k_{long}+1}, x_{t-k_{long}+2}, ..., x_t\right)$$

where $k_{long}$ is the size of the long-term kernel and $f^{long}$ is the activate function of the long-term kernel.

However, the short-term convolution employs smaller kernels to focus on capturing short-term fluctuations and local patterns within the time series, facilitating the extraction of transient features and rapid changes in the data. The output of the short-term convolution can be represented by:

$$y_t^{short} = f^{short}\left(x_{t-k_{short}+1}, x_{t-k_{short}+2}, ..., x_t\right)$$

where $k_{short}$ is the size of the short-term kernel and $f^{short}$ is the activate function of the short-term kernel.

This dual approach allows for a comprehensive analysis of time series data by simultaneously considering both overarching trends and immediate fluctuations, enhancing the model's ability to understand and predict complex temporal patterns.

On one hand, such a model exhibits enhanced feature extraction capabilities by integrating both long-term and short-term information, enabling a comprehensive capture of patterns and structures across various time scales within time series data. On the other hand, it achieves higher prediction accuracy by leveraging convolutional neural networks to learn complex patterns in the data, thereby improving predictive accuracy and robustness.

#### 3.4.2. EfficientNet B0

EfficientNet B0 is an efficient and powerful convolutional neural network structure designed to strike a balance between computational efficiency and model performance. EfficientNet employs a technique called Compound Scaling, which simultaneously scales the network's depth, width, and resolution to improve model performance while maintaining computational efficiency. The model constructs different versions of EfficientNet (from B0 to B7) by adjusting the width multiplier, depth multiplier, and input resolution. This approach enhances the overall efficiency and effectiveness of the model architecture. The formulas of compound scaling are:

$$Depth = \alpha^{\Phi}, Width = \beta^{\Phi}, Resolution = \gamma^{\Phi}$$

where $\alpha, \beta, \gamma$ are hyperparameters, $\Phi$ is the overall scaling factor. The selection of these parameters can optimize the model's performance and computational efficiency through a compound scaling method.

EfficientNet B0 is fundamentally composed of blocks based on Depthwise Separable Convolution, augmented with Residual Connections, to enhance the model's representational power and training efficiency. This design optimizes the network structure and parameters, maximizing accuracy and generalization under constrained computational resources. Depthwise Separable Convolution reduces computational complexity while maintaining model effectiveness.

This efficient model effectively extracts features from four types of spectrogram data, ultimately producing six-dimensional vectors of the same shape as those from the 1D model. Empirical evidence demonstrates that this approach is more effective compared to other models such as **ViT** (Vision Transformer).

#### 3.4.3. Multimodal feature fusion

Multimodal feature fusion involves combining features from different layers or branches, which is ubiquitous in modern network architectures. Typically achieved through simple linear operations such as summation or concatenation, these methods may not always be optimal. We employ Attention Feature Fusion (AFF) as a unified and versatile solution suitable for most common scenarios, including those involving short and long skip connections and feature fusion within Inception layers. To better integrate features with semantic and scale disparities, we introduce the Multi-Scale Channel Attention Module (MS-CAM), which addresses issues arising from merging features of different scales. However, initial feature fusion can

become a bottleneck, so we mitigate this challenge by using the Iterative Attention Feature Fusion Module (**iAFF**).

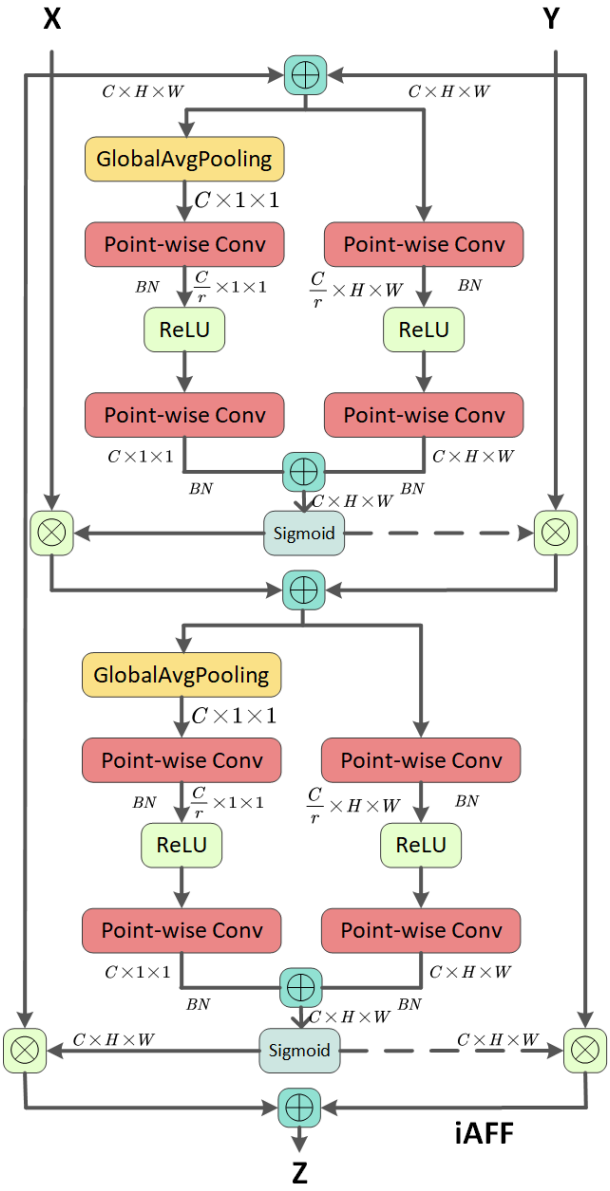


Figure 9 — iAFF

The approach involves initial feature fusion of two input features, $X$ and $Y$, followed by a sigmoid activation function that outputs values between 0 and 1. The authors aim to achieve soft selection by using 1 minus this fusion weight, allowing for weighted averaging of $X$ and $Y$. Through training, the network determines the weights for each feature.

$$X \uplus Y = M(X + Y) \otimes X + (1 - M(X + Y))$$

This method enables multi-level feature fusion, where iterative modules fuse features across different layers. By iterating multiple times, the module can integrate features at various lev-

els and dynamically learn the importance of different features. This dynamic learning process enhances the integration and utilization of each feature's information, thereby improving feature representation and minimizing information loss over iterations.

By adopting this multi-level fusion approach, we effectively integrate features extracted from 1D and 2D data, resulting in meaningful outcomes. Moreover, during internal cross-validation of the three outputs, the fused result obtains the highest weight, leading to more accurate predictions.

#### 3.4.4. GroupKFold

GroupKFold is used because multiple samples from the same patient share similarities in both their true labels and predictions. In standard k-fold cross-validation, if these similar samples are distributed into different folds, it can lead to erroneously high scores or severe overfitting.

Therefore, we employ GroupKFold to ensure that samples from the same patient are grouped together in the same fold, thereby enhancing the reliability of cross-validation.

#### 3.4.5. Two-Stage Training

Based on a research from Harvard Medical School [7] Figure 10, samples with a higher total number of votes ($\geq 10$) are more confidently accurate, whereas those with fewer votes ($\geq 3$) may contain more noise. Hence, we adopt a two-stage training approach. In the first stage, all samples are trained together. In the second stage, we specifically train on samples with a total vote count of at least 10 to emphasize the weights of more confident data in the model. Importantly, our two-stage training method has shown excellent results in final predictions. Additionally, we experimented with using data from samples with higher total votes to assist in correcting labels for samples with fewer votes, thereby improving the accuracy of predictions for samples with fewer votes.

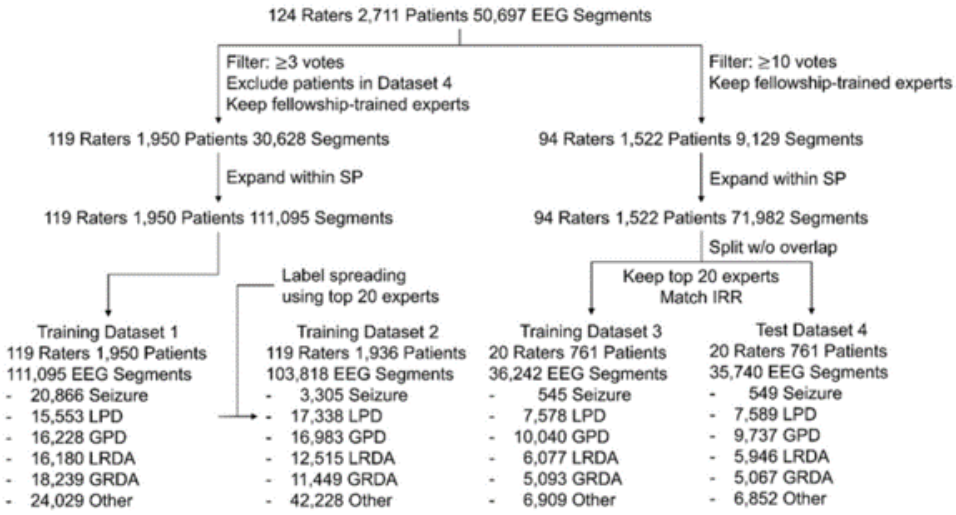


Figure 10 — Classifying by 3 and 10

# 4. Experiments & Results

Our model comprises three modules: a 1D module, a 2D module, and a feature fusion module. We have two types of input data: one type consists of 1D EEG data, and the second type consists of 2D spectrogram data. After our analysis and processing, the spectrogram data were reprocessed and concatenated, resulting in four variations of the second type of data and one variation of the first type. The first type has a shape of (10000*8), while the second type includes spectrogram data with shapes as follows: spectrograms corresponding to 600 seconds (128*256*4), spectrograms corresponding to 50 seconds (128*256*4), gradient-concatenated spectrograms (512*512*1), and spectrograms vertically stacked in four regions (512*512*1). Therefore, we will experiment with various combinations and permutations of these inputs. Additionally, we will discuss initial feature fusion strategies for our model and also evaluate the processing and generation of data.

During the data processing stage, we use the total vote count to represent confidence, where a higher total indicates greater confidence, and convert these vote counts into probabilities.

1. **Handling 1D EEG samples with the same EEG ID (see Figure 8)**

   The dataset includes multiple samples with the same EEG ID, and we employed various methods to effectively handle them to minimize noise. We observed that for 10-second samples with the same EEG ID, there is significant overlap between adjacent samples (e.g., 20-30s and 26-36s). Therefore, we implemented several approaches to address this issue:
   1. Take the first entry.
   2. Remove samples with identical vote counts to improve training efficiency.
   3. Sum the vote values for identical EEG IDs.
   4. Data from the start to the midpoint.
   5. Select samples with a time difference >10s.

   These methods are advantageous for avoiding excessive duplicate data, selecting samples from different time stages, reducing data volume, shortening training time, and preventing model overfitting. These data processing methods are all reasonable. However, training time remains relatively long for the method of selecting data with time differences greater than 10 seconds and removing samples with identical vote counts, which also shows relatively poorer results. Additionally, summing the vote values for identical EEG IDs may cause the training focus to deviate from samples with fewer votes. Based on the volume of data, prediction accuracy, and time required, we rank and list the following table:

| Num | Metric | | Rows of Samples |
|---|---|---|---|
| | Accuracy | Order of Time-Costs | |
| 1 | 78.1% | No.5 | 37,601 |
| 2 | 77.6% | NO.1 | 53,475 |
| 3 | 57.0% | NO.2 | 37,601 |
| 4 | 69.6% | NO.4 | 37,601 |
| 5 | 69.1% | NO.3 | 57.977 |

According to these results, we will use the first method in all subsequent experiments, which is to retain the first entry. This approach reduces data volume, enhances training efficiency, and ensures the robustness and accuracy of the model.

2. **Feature Fusion**

   There are various ways to perform feature fusion, such as addition, multiplication, or weighted combination. Two simple methods include: merging the data first and then inputting it into the same model, and having data from two classes go through their respective model networks before merging the trained features into a classification header. In the first method, we reshape 1D data into 2D data, and then add or concatenate the newly generated EEG data with spectrograms. Regardless of whether it is addition or concatenation, the predictive results are very poor, approaching accuracy levels close to random assignment. Therefore, we focus on the second method of feature fusion, using the iAAF model for multimodal feature fusion, maximizing the retention of different information from different modalities.

3. **Handling 2D Spectrogram Data**

   In this task, we reverse-engineered the relationship between EEG signals and spectrograms to generate four different types of spectrogram data corresponding to four time periods: original spectrograms, 50-second spectrograms, 10-second spectrograms, and gradient-connected spectrograms. These spectrogram data variations were designed to be stacked and combined as inputs to our model to assess their impact on detecting harmful brain activity in neuro-monitoring and epilepsy treatment.

   During the experiments, we employed a pre-trained EfficientNet B0 model as our baseline model, feeding it with combinations of each spectrogram data type. We compared the effects on model prediction accuracy when using each spectrogram data type alone versus their stacked combinations. We found that both the original spectrograms and 50-second spectrograms have a shape of 128x256x4, which can be reshaped into 512x512x1, consistent with the shapes of the other two spectrogram types. Concatenating these three input data sets resulted in a shape of 1536x512x1, which was validated as optimal.

## 5. Conclusion

We have made significant progress in addressing the critical challenges of detecting abnormal brain activity and enhancing neurocritical care through advanced EEG analysis. By integrating innovative methods such as the iAFF feature fusion technique and the PCKLS EEG feature extraction method, we developed a comprehensive model capable of accurately classifying epileptic seizures and other harmful brain activities. Our approach not only leverages the complementary strengths of EEG signals and spectrograms but also enhances the robustness and accuracy of classification models crucial for epilepsy treatment and neurocritical care.

Through meticulous experimentation and validation, we have demonstrated the effectiveness of these methods in real-world applications, particularly in leveraging diverse spectrogram data transformations and selecting optimal neural network architectures like EfficientNet B0. Our contributions also include advancements in data feature extraction, where the development of the PCKLS method enhances sensitivity to dynamic signal changes and complex feature extraction by simultaneously considering time and frequency domains.

Furthermore, by reverse-engineering the EEG-to-spectrogram data transformation formula, we generated diverse spectrogram data across different time segments to enhance inputs for 2D models. Through gradient concatenation and two other concatenation methods, these innovative approaches effectively optimized model inputs, thereby improving classification accuracy and robustness.

Looking ahead, our interdisciplinary efforts represent a significant step forward in bridging the gap between traditional EEG analysis and modern machine learning techniques, promising enhanced capabilities in neurocritical care and epilepsy treatment. These efforts not only contribute new insights and methodologies to neuroimaging and computational neuroscience but also provide new tools and strategies to improve diagnosis, treatment, and overall prognosis for patients with epilepsy and other neurological disorders.

# 6. Contribution of Group Members

**I have completed this report all by myself without any outer assistance.**

# References

[1] A. Malinin and M. Gales, "Reverse kl-divergence training of prior networks: Improved uncertainty and adversarial robustness," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[2] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 3560–3569.

[3] P. Jiang, H. Fu, H. Tao, P. Lei, and L. Zhao, "Parallelized convolutional recurrent neural network with spectral features for speech emotion recognition," *IEEE access*, vol. 7, pp. 90368–90377, 2019.

[4] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999–7019, 2021.

[5] L. Yuan *et al.*, "Tokens-to-token vit: Training vision transformers from scratch on imagenet," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 558–567.

[6] V. Suárez-Paniagua, R. M. R. Zavala, I. Segura-Bedmar, and P. Mart\inez, "A two-stage deep learning approach for extracting entities and relationships from medical texts," *Journal of biomedical informatics*, vol. 99, p. 103285–103286, 2019.

[7] J. Jing *et al.*, "Development of expert-level classification of seizures and rhythmic and periodic patterns during eeg interpretation," *Neurology*, vol. 100, no. 17, p. e1750–e1762, 2023.