# Enhanced Convolutional Neural Networks with Hilbert Transform for Harmful Brain Activity Classification

**Master of Technology**
**in**
**Computer Science and Engineering**

*by*

**Saurabh Roy**
**23CS60R76**

Under the guidance of

**Prof. Debasis Samanta**

COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Department of Computer Science and
Engineering
Indian Institute of Technology,
Kharagpur
India - 721302

# CERTIFICATE

This is to certify that we have examined the thesis entitled **Enhanced Convolutional Neural Networks with Hilbert Transform for Harmful Brain Activity Classification**, submitted by **Saurabh Roy**(Roll Number: *23CS60R76*) a postgraduate student of **Department of Computer Science and Engineering** in partial fulfillment for the award of degree of Master of Technology. We hereby accord our approval of it as a study carried out and presented in a manner required for its acceptance in partial fulfillment for the Post Graduate Degree for which it has been submitted. The thesis has fulfilled all the requirements as per the regulations of the Institute and has reached the standard needed for submission.

**Supervisor**

**Department of Computer
Science and Engineering**
Indian Institute of Technology,
Kharagpur

**Place: Kharagpur**
**Date:**

# ACKNOWLEDGEMENTS

I extend my heartfelt gratitude and appreciation to my dedicated supervisor, **Prof. Debasis Samanta**, for his invaluable guidance and support throughout my Master's Thesis Project. Prof. Debasis Samanta graciously provided me with the opportunity to undertake this research under his mentorship. His unwavering support, from the early stages of my journey as a novice to my growth and learning in advanced research, has been instrumental. I am deeply thankful to him for believing in my potential and for nurturing my academic development.

I would also like to express my thanks to the **Computer Science and Engineering at IIT Kharagpur** for providing the platform for my research.

My profound gratitude extends to my parents, teachers, and friends for their unwavering encouragement and support throughout my academic career.

I would like to acknowledge **Prof. Arobinda Gupta**, the Head of the Department of Computer Science and Engineering, for his leadership and for fostering an environment conducive to academic growth.

This acknowledgment is a small token of appreciation for the significant contributions made by these individuals and institutions that have shaped my academic journey.

**Saurabh Roy**
Computer Science and Engineering
IIT Kharagpur
Date: 10/11/2024

# ABSTRACT

Electroencephalogram (EEG) signal classification is critical for diagnosing and monitoring neurological disorders, yet the interpretation of these signals poses significant challenges due to their inherent complexity and variability. This gap in reliable EEG analysis is addressed in our study, which introduces an advanced convolutional neural network (CNN) model incorporating a Hilbert Transform layer for enhanced feature extraction. This method is tailored specifically for the nuanced demands of EEG signal classification, drawing from insights gained through the Havard Medical Science-Harmful Brain Activity Classification Dataset. Our proposed approach leverages the spatial-temporal characteristics of EEG data, enriching the feature set with amplitude and phase information obtained via the Hilbert Transform—a technique that captures the dynamic properties of EEG signals more comprehensively. This enriched feature extraction allows our model to identify and categorize critical brain activity patterns, such as seizures and periodic discharges, from extensive EEG recordings. The model's architecture efficiently learns complex patterns, which is demonstrated through training with both cross-entropy and KL-Divergence loss functions to finely align the model's output with expert annotations. Empirical results indicate that the proposed model achieves a classification accuracy of 95%, significantly outperforming traditional methods. The Hilbert Transform layer in particular not only accelerated convergence but also improved the model's generalization capabilities across various brain activity states. This research makes a substantial contribution to biomedical signal processing by enhancing the accuracy and reliability of EEG signal classification. The advancements reported in this paper could potentially revolutionize EEG-based diagnostics, providing clinicians with more precise and dependable tools, and pave the way for future innovations in EEG analysis methodologies.

**Keywords**: Electroencephalography(EEG), Spectogram, Harmful Brain Activity, Seizures, Periodic Discharges, ,Convolutional Neural Networks(CNN), Hilbert Transform, Deep Learning in Medicine

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Electroencephalography (EEG) is a foundational tool in neurocritical care and epilepsy management, providing essential insights into brain activity that aid in diagnosing and treating neurological disorders. However, EEG interpretation remains challenging due to the subtle and often variable nature of neural signals. Traditionally, EEG data is reviewed manually by specialized neurologists. While valuable, this process is both time-consuming and subject to variation between reviewers, which limits its scalability in high-stakes settings where timely intervention is crucial. Automated systems that can improve the speed, accuracy, and reliability of EEG classification are needed, particularly to detect harmful brain activity such as seizures and other irregular patterns.

Manual EEG analysis, though essential, is labor-intensive and prone to inconsistencies from human error or fatigue. This can lead to delays in diagnosis and treatment, especially in critical care, where prompt decisions make a significant difference in patient outcomes. Current automated methods provide some support, but they often struggle with the inherent complexity of EEG data, particularly in distinguishing between similar types of brain activity. My research seeks to address these challenges by developing an improved approach to EEG classification through deep learning, enhanced by signal processing techniques.

The primary goal of this project is to develop an accurate, automated EEG classification model that integrates the Hilbert Transform with Convolutional Neural Networks (CNNs) to create a specialized model, which I refer to as **HTCNN**. This **HTCNN** model introduces the Hilbert Transform layer to capture both amplitude and phase information from EEG signals, adding a new dimension to feature extraction that traditional CNNs lack. Specifically, this project focuses on classifying harmful brain activities into distinct patterns that are clinically significant. The objectives of this research include:

1. Developing an HTCNN model that uses deep learning to classify harmful brain activities in EEG data, distinguishing critical patterns like seizures and periodic discharges.

2. Validating the effectiveness of this model using a comprehensive dataset of EEG signals from critically ill patients, ensuring robustness and adaptability in real-world clinical scenarios.

3. Demonstrating the practical benefits of HTCNN in clinical and research settings, including streamlining EEG analysis, reducing diagnostic delays, and enhancing the accuracy of neurological assessments.

In this research, HTCNN was trained and validated on the HMS dataset, which includes EEG segments labeled by experts. The model employs cross-validation techniques to ensure robustness, with metrics like accuracy, precision, recall, and F1-score used as benchmarks for comparison with existing approaches. The addition of the Hilbert Transform layer allows HTCNN to extract both spatial and temporal features critical for differentiating complex patterns in brain activity. By automating this analysis process, HTCNN holds the potential to reduce dependency on expert review, streamline the diagnostic workflow, and provide faster and more reliable EEG analysis in critical care.

In summary, this project contributes to biomedical signal processing by presenting a new approach to EEG classification that incorporates the Hilbert Transform within a CNN architecture. This advancement improves classification accuracy and reliability and paves the way for future research into more sophisticated methods for automated brain signal analysis, benefiting both clinical and research applications.

# Chapter 2

# Literature Survey

[1] showcases the development of SPaRCNet, a dense-CNN architecture based algorithm crafted to match or surpass the diagnostic accuracy of expert neurophysiologists in classifying complex EEG patterns. Employing a robust dataset of 6,095 scalp EEGs, segmented into 50,697 parts and meticulously annotated by 20 trained neurophysiologists, the algorithm was rigorously trained and evaluated. SPaRCNet demonstrated outstanding performance across several metrics, achieving high sensitivity, specificity, and precision in identifying seizures (SZs), lateralized and generalized periodic discharges (LPD, GPD), and lateralized and generalized rhythmic delta activity (LRDA, GRDA). Notably, the algorithm excelled in calibration and discrimination metrics, often outperforming human experts. SPaRCNet exceeds the following percentages of 20 experts-ROC: 45%, 20%, 50%, 75%, 55%, and 40%; PRC: 50%, 35%, 50%, 90%, 70%, and 45%; and calibration: 95%, 100%, 95%, 100%, 100%, and 80%, respectively. These results highlight SPaRCNet's potential as a transformative tool for EEG interpretation, capable of enhancing diagnostic accuracy and efficiency in neurocritical care and epilepsy management, offering a substantial leap forward in automated EEG analysis.

[2] investigates the reliability among EEG experts in diagnosing seizures and related patterns, revealing considerable variability in interpretations. Analyzing 2,711 EEGs reviewed by 30 experts, the study highlights moderate pairwise interrater reliability with an average percent agreement of 52% and a kappa value of 42%. Majority agreement metrics were notably higher, suggesting that group consensus might reflect more reliable interpretations. The discrepancies primarily stemmed from differences in decision thresholds among the experts. These findings underscore the necessity for standardized training and calibration among professionals to enhance diagnostic

consistency in clinical EEG assessments, pointing to significant implications for both clinical practice and the development of automated diagnostic systems.

[3] investigates the relationship between epileptiform activity (EA) burden and clinical outcomes in acutely ill patients. Using a convolutional neural network (CNN) model, the study analyzes over 11 terabytes of continuous EEG (cEEG) data from 2,000 patients to quantify EA, including seizures, lateralized periodic discharges (LPDs), generalized periodic discharges (GPDs), and lateralized rhythmic delta activity (LRDA). The model measured EA burden in three ways: first 24-hour burden, peak burden in a 12-hour window, and cumulative burden over 72 hours. Results indicated a significant association between peak EA burden and poor neurological outcomes (modified Rankin Scale 5-6), with a 35% increase in the probability of a poor outcome as peak burden rose from 0% to 100%. Calibration errors remained low across the different discharge periods, supporting the model's accuracy. The strongest link between EA burden and poor outcomes was observed in patients with hypoxic ischemic encephalopathy (HIE), followed by those with acute seizures or status epilepticus. This study highlights the effectiveness of automated EA annotation and its potential role in guiding interventions and improving prognostic models for severe brain injuries and seizure disorders.

[4] introduces ProtoPMed-EEG, a deep learning-based model aimed at assisting clinicians in classifying EEG patterns such as seizures, lateralized periodic discharges (LPD), generalized periodic discharges (GPD), lateralized rhythmic delta activity (LRDA), and generalized rhythmic delta activity (GRDA) along the ictal–interictal injury continuum (IIIC). This model emphasizes interpretability, providing explanations for its predictions using case-based reasoning with prototypical EEG examples, in contrast to black-box models like SPaRCNet. ProtoPMed-EEG was trained on a large dataset of 50,697 EEG segments from 2,711 ICU patients, and its performance was validated through a user study involving eight non-expert clinicians. The results demonstrated a significant improvement in diagnostic accuracy, increasing from 47% to 71% with AI assistance, and an enhancement in interrater reliability (IRR), suggesting the model's potential to improve clinical decision-making in the absence of expert neurologists.ProtoPMed-EEG achieved high area under the receiver operating characteristic curve (AUROC) scores for multiple EEG patterns, including 0.87 for seizures, 0.93 for LPD, 0.96 for GPD, 0.92 for LRDA, and 0.93 for GRDA, demonstrating superior performance compared to existing models. Moreover, when tested

on an external dataset from a different hospital, ProtoPMed-EEG maintained a strong AUROC of 0.85, highlighting its generalizability across different clinical settings. The model's ability to provide interpretable outputs not only enhances diagnostic accuracy but also serves as an educational tool for clinicians, aiding in the learning and understanding of complex EEG patterns. The study concludes that ProtoPMed-EEG has the potential to significantly improve both the accuracy and confidence of non-expert clinicians in diagnosing harmful brain activity, while also contributing to training efforts in EEG interpretation.

[5] explores the advancements and challenges in seizure detection using EEG data, emphasizing the evolution from traditional signal processing to more sophisticated machine learning and deep learning approaches. The review highlights several studies that have contributed to the field by employing various algorithms to improve the sensitivity and specificity of seizure detection. It discusses the shift towards deep learning methods, particularly the use of recurrent neural networks like LSTM, which are adept at handling time-series data inherent in EEG signals. The paper underscores the potential of these advanced computational techniques to overcome the limitations of manual interpretation and earlier automated methods, providing a comprehensive background that sets the stage for introducing their approach to integrating machine learning and deep learning for enhanced diagnostic accuracy in epilepsy monitoring.

The paper [6] provides an in-depth exploration of deep learning techniques applied to the detection of epileptic seizures using intracranial EEG (iEEG) signals. The authors present a comparative analysis of multiple convolutional neural network (CNN)-based models, such as S-CNN, Modif-CNN, CNN-SVM, and Comb-2CNN, which are evaluated on the American Epilepsy Society database. Their study demonstrates the effectiveness of CNN models in classifying epileptic states with a high degree of accuracy, with the Modif-CNN model achieving the best performance with an accuracy of 97.96%. The paper highlights the importance of real-time seizure detection and prediction, which could significantly improve patient outcomes in clinical settings. Moreover, the research underscores the potential of deep learning to overcome limitations faced by traditional machine learning approaches in handling large, complex iEEG datasets. This study contributes to the growing body of work focused on leveraging neural networks for improved accuracy and efficiency in epileptic seizure detection.

The paper [7] introduces a smart neurocare framework combining cloud and fog computing for efficient epileptic seizure detection using temporal analysis of EEG signals. The study focuses on single-channel EEG data, which undergoes preprocessing through a maximum variance-based channel selection method. After filtering and segmentation into temporal windows, the EEG data is processed by three deep learning models: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Stacked Autoencoders (SAEs). The CNN-based approach demonstrates superior performance, achieving an accuracy of 96.43%, sensitivity of 100%, and specificity of 93.33% on the CHB-MIT dataset, and 100% across all metrics on the Bonn dataset. This framework is designed to operate efficiently in real-time at the fog layer, minimizing latency and bandwidth usage, making it a promising solution for computationally efficient epileptic seizure detection in neurocare applications.

The paper titled [8] proposes an innovative system architecture for detecting and classifying epileptic seizures using EEG data. The authors employed deep learning techniques, focusing on binary and multigroup classification tasks, tested on the Temple University Hospital Seizure Corpus (TUSZ). The study emphasizes optimizing EEG acquisition settings, such as sampling frequency and the number of electrodes, to achieve high performance while maintaining interpretability. The system showed significant success with 87.7% sensitivity and 91.16% specificity in seizure detection, and accuracy ranging from 95-100% in classifying seizure types. The use of interpretable machine learning techniques, such as activation maximization, highlighted specific EEG patterns corresponding to different seizure types, improving model transparency. Additionally, source reconstruction techniques were used to distinguish between focal and generalized seizures. Overall, the study provides a framework for optimizing EEG settings while ensuring reliable and interpretable seizure classification.

[9] presented a novel approach in their paper by addressing the interpretability issue of deep learning models in the medical domain. They developed a CNN-based architecture aimed at detecting epileptic seizures from EEG signals with a focus on interpretability. The model utilized a minimal pre-processing approach and explored various kernel sizes in the first convolutional layer to assess its impact on learning frequency patterns, relevant to seizure detection. The REPO2MSE dataset, consisting of multi-channel EEG recordings from 568 patients, was used, and the methodology was tested both at the segment and seizure level. The model achieved a high F1 score of 0.873 and detected 90% of seizures, demonstrating good generalizability to

unseen patients. Additionally, the study introduced two visualization techniques, including SHAP values, to highlight significant ictal features on the input EEG signals, further enhancing model interpretability. The findings underscore the importance of amplitude and high-frequency components in seizure classification, contributing to the advancement of interpretable AI in clinical practice.

[10] introduces a novel deep active learning framework to detect and classify EEG patterns along the Interictal Ictal Injury Continuum (IIIC), which includes seizures and various types of discharges (LPDs, GPDs, LRDA, GRDA). The authors leverage active learning to minimize the amount of labeled data required by selecting the most informative samples during training. Using a convolutional neural network (CNN) with expert annotations, the model was able to classify IIIC patterns with high accuracy. The dataset, composed of EEG recordings from multiple hospitals, was preprocessed and labeled by expert electroencephalographers. The proposed deep active learning model achieved a significant reduction in required labeled data while maintaining a high classification performance, with an overall sensitivity of 85% and specificity of 91%. The study highlights the potential of deep active learning to reduce annotation costs and improve diagnostic accuracy in clinical applications.

Table 2.1: Comparative analysis of various EEG-based brain activity detection and classification techniques that employs machine learning and deep learning paradigm.

| **Paper** | **Method** | **Dataset** | **Contribution** | **Remarks** |
|---|---|---|---|---|
| [1] | SPaRCNet (Dense-Net CNN architecture) | Own EEG Dataset with User Ratings and Expert Annotations | CNN-based model to identify and classify seizures | SPaRCNet fails to identify some EEG patterns of clinical relevance. |
| [2] | Pairwise IRR and Majority IRR for class comparison | Massachusetts General Hospital Patient Dataset with 2711 EEG recordings | Shows moderate reliability between experts | Lack of expert scores for some types of IIIC patterns. |
| [3] | CNN model for IIIC classification and Multivariable logistic regression model for EA burden prediction | Own Dataset with 2000 EEG recordings | Increasing EA burden associated with worse neurologic outcomes | Focuses only on outcomes at hospital discharge rather than long-term. |
| [4] | CNN-based feature extractor with prototype layer and angular distance-based linear layer | Massachusetts General Hospital Dataset | Interpretable DL algorithm classifies six clinically relevant EEG patterns | Accuracy & Sensitivity can still be improved. |
| [5] | Logistic Regression, KNN, SVM, ANN, LSTM | UCI-Epileptic Seizure Recognition Dataset | Comparative analysis of ML/DL models including LSTM for seizure detection | Demonstrates effectiveness of LSTM in seizure detection but lacks detailed implementation. |

| Paper | Method | Dataset | Contribution | Remarks |
|---|---|---|---|---|
| [6] | S-CNN, Modif-CNN, CNN-SVM and Comb-2CNN models for seizure detection | American Epilepsy Society Database | CNN-based models for classifying epilepsy states with highest accuracy from Modif-CNN | High classification accuracy with Modif-CNN achieving 97.96% success rate. |
| [7] | CNN, RNN, SAE models | CHB-MIT, Bonn EEG Database | CNN-based approach achieved 96.43% accuracy, 100% sensitivity on CHB-MIT | Uses cloud and fog computing for temporal analysis. |
| [8] | CNN, Multi-group RNN for Seizure Detection and Classification | Temple University Hospital Seizure Corpus (TUSZ) | Optimization of EEG acquisition settings with interpretable ML for seizure detection | Achieved 95-100% accuracy, with 87.7% sensitivity and 91.16% specificity. |
| [9] | CNN-based architecture, SHAP values for interpretability | REPO2MSE Dataset (multichannel EEG from 568 patients) | CNN-based seizure detection with interpretability, using SHAP | Significant contribution to interpretability in deep learning models for EEG. |
| [10] | CNN with active learning | EEG recordings from multiple hospitals | Developed an active learning framework to reduce labeled data for accurate IIIC pattern classification | Achieved 85% sensitivity and 91% specificity and reduced annotation costs significantly. |

# Chapter 3

# Methodology

The following sections outline the detailed steps taken for the research methodology, as represented in the flowchart (Fig. 3.1).

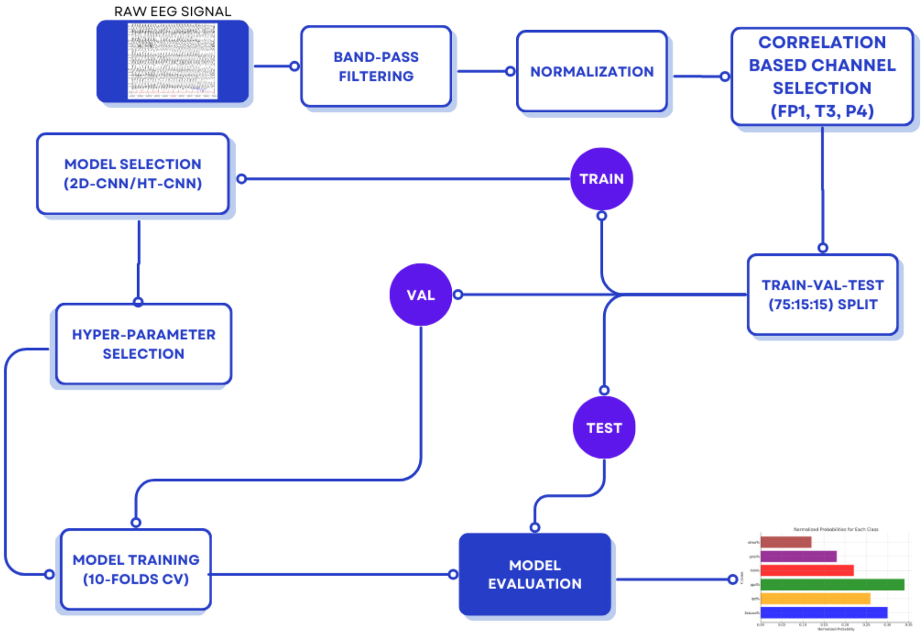

RAW EEG SIGNAL
BAND-PASS FILTERING
NORMALIZATION
CORRELATION BASED CHANNEL SELECTION (FP1, T3, P4)
MODEL SELECTION (2D-CNN/HT-CNN)
TRAIN
TRAIN-VAL-TEST (75:15:15) SPLIT
VAL
HYPER-PARAMETER SELECTION
TEST
MODEL TRAINING (10-FOLDS CV)
MODEL EVALUATION
Normalized Probabilities for Each Class
other%
grda%
lrda%
gpd%
lpd%
Seizure%
Class
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Normalized Probability

Figure 3.1: Research Methodology Flowchart.

### 3.0.1 Raw EEG Signal

The raw EEG signals serve as the primary input data for deep learning models in the classification of harmful brain activities. Electroencephalography (EEG) is a non-invasive technique used to record electrical activity of the brain through electrodes placed on the scalp. These signals are captured as time-series data, with each electrode recording the voltage fluctuations at a specific location on the scalp over time. The raw EEG data consists of multiple channels, each representing the activity of a different region of the brain. In our study, EEG signals were recorded from 19 electrodes (e.g., 'Fp1', 'F3', 'C3', 'P3', etc.) and an additional electrocardiogram (EKG) channel, sampled at 200 Hz. These raw signals typically exhibit a high level of noise and variability, requiring preprocessing steps such as filtering and normalization before being fed into the deep learning models. The multi-channel nature of EEG data provides a rich temporal and spatial representation of brain activity, which deep learning models can leverage to identify complex patterns associated with harmful brain conditions. By learning from these raw signals, models such as 2D-CNNs and HT-CNNs can capture both temporal dependencies and spatial correlations, allowing for the effective classification of seizure and other brain activity patterns.

### 3.0.2 Band-Pass Filtering

To prepare the raw EEG signals for analysis, band-pass filtering is applied to remove unwanted noise and retain the frequencies of interest. The filter allows frequencies between 0.5 Hz and 40 Hz to pass through, which are known to contain significant components related to brain activity, including seizure events. This range was chosen because it encompasses the delta, theta, alpha, beta, and low gamma bands, which are known to be relevant in neurological studies. The filtering process was conducted using a second-order Butterworth filter, which is known for its stability and effectiveness in preserving signal integrity. The band-pass filter helps eliminate low-frequency artifacts such as movement or electrode drift and high-frequency noise from environmental interference. Let $x(t)$ represent the raw EEG signal and $y(t)$ represent the filtered signal:

$$y(t) = x(t) * h(t)$$

where $*$ denotes the convolution operation, and $h(t)$ is the impulse response of the band-pass filter. By preserving only the relevant frequency components, band-pass filtering ensures that the data is clean and suitable for feature extraction and model training.

### 3.0.3 Normalization

After band-pass filtering, normalization is performed on the EEG signals to bring the data values into a standardized range. Normalization ensures that all EEG channels have similar scales, which improves the model's ability to learn effectively without being influenced by variations in magnitude across different channels. In this study, Z-score normalization was used, which involves standardizing each data point by subtracting the mean and dividing by the standard deviation of the channel. For an EEG signal $x_i$, the normalized value $x'_i$ is computed as:

$$x'_i = \frac{x_i - \mu}{\sigma}$$

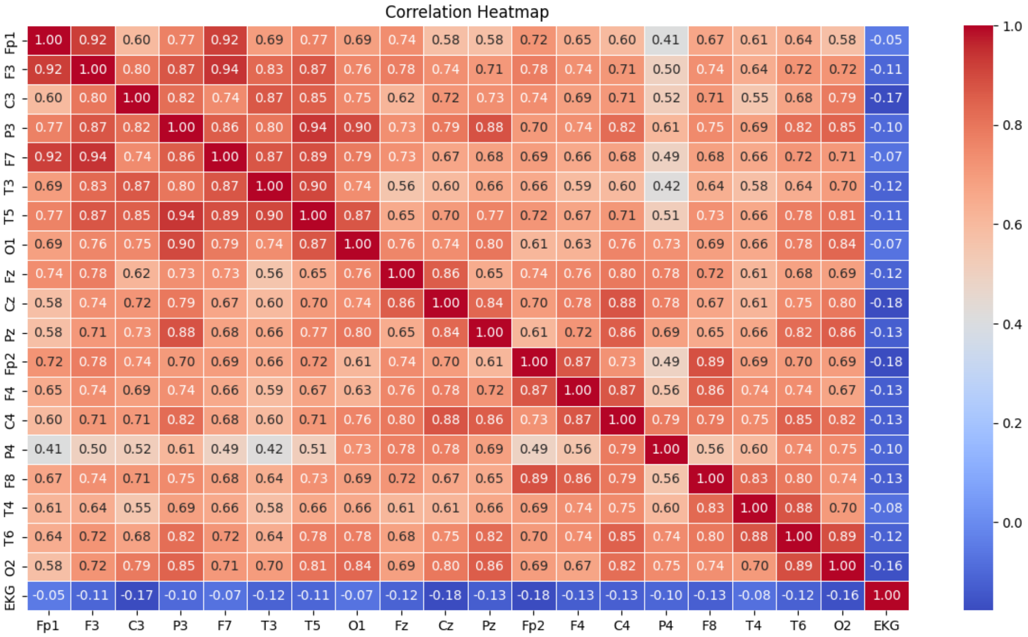
where $\mu$ is the mean value of the channel, and $\sigma$ is the standard deviation. By standardizing each channel, the model can focus on the relative differences in the signal, which is crucial for identifying patterns like seizures, discharges, and other brain activities. Normalization is particularly important in EEG analysis as it addresses variations due to different recording conditions, patient-specific characteristics, or electrode impedance. This ensures that the input data fed to the model is consistent and prevents the model from converging towards suboptimal solutions due to large-scale differences in the input features.

### 3.0.4 Correlation-Based Channel Selection

To reduce computational complexity and focus on the most informative features, correlation analysis was used to select a subset of EEG channels. EEG signals often exhibit redundancy across channels due to the spatial proximity of the electrodes, which can lead to overlapping information. To address this, a correlation-based feature selection method was employed to identify the channels that contain the most discriminative information for harmful brain activity classification. In this study, the channels Fp1, T3, and P4 were selected based on their high correlation with seizure activity and other relevant brain patterns. The Pearson correlation coefficient $\rho_{xy}$ between two channels $X$ and $Y$ is calculated as follows:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

where $X_i$ and $Y_i$ are individual data points from channels $X$ and $Y$, $\bar{X}$ and $\bar{Y}$ are the mean values of $X$ and $Y$, respectively, and $n$ is the number of data points. A high correlation value indicates that a particular channel shares significant information with seizure events or other important EEG patterns. By selecting the channels Fp1, T3, and P4, we ensure that the most relevant features are retained for model training, while minimizing computational overhead. This selection not only improves the efficiency of the model but also enhances its interpretability, as fewer channels simplify the analysis and interpretation of the results (Fig. 3.2).



| | Fp1 | F3 | C3 | P3 | F7 | T3 | T5 | O1 | Fz | Cz | Pz | Fp2 | F4 | C4 | P4 | F8 | T4 | T6 | O2 | EKG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fp1 | 1.00 | 0.92 | 0.60 | 0.77 | 0.92 | 0.69 | 0.77 | 0.69 | 0.74 | 0.58 | 0.58 | 0.72 | 0.65 | 0.60 | 0.41 | 0.67 | 0.61 | 0.64 | 0.58 | -0.05 |
| F3 | 0.92 | 1.00 | 0.80 | 0.87 | 0.94 | 0.83 | 0.87 | 0.76 | 0.78 | 0.74 | 0.71 | 0.78 | 0.74 | 0.71 | 0.50 | 0.74 | 0.64 | 0.72 | 0.72 | -0.11 |
| C3 | 0.60 | 0.80 | 1.00 | 0.82 | 0.74 | 0.87 | 0.85 | 0.75 | 0.62 | 0.72 | 0.73 | 0.74 | 0.69 | 0.71 | 0.52 | 0.71 | 0.55 | 0.68 | 0.79 | -0.17 |
| P3 | 0.77 | 0.87 | 0.82 | 1.00 | 0.86 | 0.80 | 0.94 | 0.90 | 0.73 | 0.79 | 0.88 | 0.70 | 0.74 | 0.82 | 0.61 | 0.75 | 0.69 | 0.82 | 0.85 | -0.10 |
| F7 | 0.92 | 0.94 | 0.74 | 0.86 | 1.00 | 0.87 | 0.89 | 0.79 | 0.73 | 0.67 | 0.68 | 0.69 | 0.66 | 0.68 | 0.49 | 0.68 | 0.66 | 0.72 | 0.71 | -0.07 |
| T3 | 0.69 | 0.83 | 0.87 | 0.80 | 0.87 | 1.00 | 0.90 | 0.74 | 0.56 | 0.60 | 0.66 | 0.66 | 0.59 | 0.60 | 0.42 | 0.64 | 0.58 | 0.64 | 0.70 | -0.12 |
| T5 | 0.77 | 0.87 | 0.85 | 0.94 | 0.89 | 0.90 | 1.00 | 0.87 | 0.65 | 0.70 | 0.77 | 0.72 | 0.67 | 0.71 | 0.51 | 0.73 | 0.66 | 0.78 | 0.81 | -0.11 |
| O1 | 0.69 | 0.76 | 0.75 | 0.90 | 0.79 | 0.74 | 0.87 | 1.00 | 0.76 | 0.74 | 0.80 | 0.61 | 0.63 | 0.76 | 0.73 | 0.69 | 0.66 | 0.78 | 0.84 | -0.07 |
| Fz | 0.74 | 0.78 | 0.62 | 0.73 | 0.73 | 0.56 | 0.65 | 0.76 | 1.00 | 0.86 | 0.65 | 0.74 | 0.76 | 0.80 | 0.78 | 0.72 | 0.61 | 0.68 | 0.69 | -0.12 |
| Cz | 0.58 | 0.74 | 0.72 | 0.79 | 0.67 | 0.60 | 0.70 | 0.74 | 0.86 | 1.00 | 0.84 | 0.70 | 0.78 | 0.88 | 0.78 | 0.67 | 0.61 | 0.75 | 0.80 | -0.18 |
| Pz | 0.58 | 0.71 | 0.73 | 0.88 | 0.68 | 0.66 | 0.77 | 0.80 | 0.65 | 0.84 | 1.00 | 0.61 | 0.72 | 0.86 | 0.69 | 0.65 | 0.66 | 0.82 | 0.86 | -0.13 |
| Fp2 | 0.72 | 0.78 | 0.74 | 0.70 | 0.69 | 0.66 | 0.72 | 0.61 | 0.74 | 0.70 | 0.61 | 1.00 | 0.87 | 0.73 | 0.49 | 0.89 | 0.69 | 0.70 | 0.69 | -0.18 |
| F4 | 0.65 | 0.74 | 0.69 | 0.74 | 0.66 | 0.59 | 0.67 | 0.63 | 0.76 | 0.78 | 0.72 | 0.87 | 1.00 | 0.87 | 0.56 | 0.86 | 0.74 | 0.74 | 0.67 | -0.13 |
| C4 | 0.60 | 0.71 | 0.71 | 0.82 | 0.68 | 0.60 | 0.71 | 0.76 | 0.80 | 0.88 | 0.86 | 0.73 | 0.87 | 1.00 | 0.79 | 0.79 | 0.75 | 0.85 | 0.82 | -0.13 |
| P4 | 0.41 | 0.50 | 0.52 | 0.61 | 0.49 | 0.42 | 0.51 | 0.73 | 0.78 | 0.78 | 0.69 | 0.49 | 0.56 | 0.79 | 1.00 | 0.56 | 0.60 | 0.74 | 0.75 | -0.10 |
| F8 | 0.67 | 0.74 | 0.71 | 0.75 | 0.68 | 0.64 | 0.73 | 0.69 | 0.72 | 0.67 | 0.65 | 0.89 | 0.86 | 0.79 | 0.56 | 1.00 | 0.83 | 0.80 | 0.74 | -0.13 |
| T4 | 0.61 | 0.64 | 0.55 | 0.69 | 0.66 | 0.58 | 0.66 | 0.66 | 0.61 | 0.61 | 0.66 | 0.69 | 0.74 | 0.75 | 0.60 | 0.83 | 1.00 | 0.88 | 0.70 | -0.08 |
| T6 | 0.64 | 0.72 | 0.68 | 0.82 | 0.72 | 0.64 | 0.78 | 0.78 | 0.68 | 0.75 | 0.82 | 0.70 | 0.74 | 0.85 | 0.74 | 0.80 | 0.88 | 1.00 | 0.89 | -0.12 |
| O2 | 0.58 | 0.72 | 0.79 | 0.85 | 0.71 | 0.70 | 0.81 | 0.84 | 0.69 | 0.80 | 0.86 | 0.69 | 0.67 | 0.82 | 0.75 | 0.74 | 0.70 | 0.89 | 1.00 | -0.16 |
| EKG | -0.05 | -0.11 | -0.17 | -0.10 | -0.07 | -0.12 | -0.11 | -0.07 | -0.12 | -0.18 | -0.13 | -0.18 | -0.13 | -0.13 | -0.10 | -0.13 | -0.08 | -0.12 | -0.16 | 1.00 |

Figure 3.2: EEG Channel Correlations

### 3.0.5 Train-Validation-Test Split

To ensure the model's robustness and to avoid overfitting, the EEG dataset was divided into three distinct subsets: training, validation, and testing. The dataset was split in the ratio of 75:15:15, as described below:

**Training Set (75%)** Used to train the model and learn the optimal parameters through backpropagation.

**Validation Set (15%)** Used during training to tune hyperparameters and monitor the model's performance, allowing us to identify and prevent overfitting. Early stopping was employed based on the performance on this set.

**Test Set (15%)** Used for the final evaluation of the trained model to assess its generalizability on unseen data.

The data split was done randomly while ensuring that all classes were well represented in each subset. This stratified approach ensured that the distribution of the target classes—seizure (SZ), generalized periodic discharges (GPD), lateralized periodic discharges (LPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), and ”other”—was consistent across the three subsets. The goal of the train-validation-test split is to provide a fair estimate of the model's performance on new data, ensuring that the model has not overfitted to specific data samples. The validation set helps in fine-tuning hyperparameters, while the test set evaluates the final performance in a real-world setting.

### 3.0.6 Model Selection

In this study, two types of Convolutional Neural Networks (CNNs) were evaluated for classifying harmful brain activity: the 2D-CNN and the Hilbert Transform-based CNN (HT-CNN). Each model's architecture and specific layers are detailed below, along with the rationale for selecting these architectures and how they contribute to effective learning.

#### 3.0.6.1 2D-CNN

The 2D-CNN architecture was chosen for its ability to capture spatial dependencies in the EEG data. The EEG signals were transformed into a 2D representation, with one axis representing the different EEG channels and the other axis representing time. Convolutional layers were used to extract local patterns and relationships across multiple channels over time. The detailed architecture, with all parameter and output shapes is depicted in Table 3.1.

- **Input Layer**: The input to the model is a 2D matrix representation of the EEG signal, where the rows correspond to EEG channels and columns correspond to time points.

- **Convolutional Layers**: Several convolutional layers are employed to detect spatial features across channels. Each convolutional layer applies a set of filters (kernels) that slide across the input matrix to extract important local features.

Table 3.1: 2DCNN Architecture for EEG Classification Model

| **Layer (type)** | **Output Shape** | **Param #** |
|---|---|---|
| conv2d | (None, 3, 10000, 64) | 576 |
| conv2d_1 | (None, 3, 5000, 64) | 32,832 |
| max_pooling2d | (None, 3, 2500, 64) | 0 |
| conv2d_2 | (None, 3, 2500, 128) | 65,664 |
| conv2d_3 | (None, 3, 1250, 128) | 131,200 |
| max_pooling2d_1 | (None, 1, 625, 128) | 0 |
| conv2d_4 | (None, 1, 625, 256) | 524,544 |
| conv2d_5 | (None, 1, 313, 256) | 1,048,832 |
| max_pooling2d_2 | (None, 1, 156, 256) | 0 |
| global_average_pooling2d | (None, 256) | 0 |
| dense | (None, 256) | 65,792 |
| dropout | (None, 256) | 0 |
| dense_1 | (None, 128) | 32,896 |
| dense_2 | (None, 64) | 8,256 |
| dropout_1 | (None, 64) | 0 |
| dense_3 | (None, 6) | 390 |
| **Total params:** | 1,910,982 | |
| **Trainable params:** | 1,910,982 | |
| **Non-trainable params:** | 0 | |

Mathematically, the output feature map $y_{i,j,k}$ of a convolutional layer can be represented as:

$$y_{i,j,k} = f\left(\sum_{m}\sum_{n} x_{i+m,j+n,k} \cdot w_{m,n} + b_k\right)$$

where $x_{i,j,k}$ represents the input feature map, $w_{m,n}$ is the filter, $b_k$ is the bias term, and $f$ is the activation function (typically ReLU). The convolution operation allows the model to learn spatial patterns in the EEG signals, such as correlations between different channels, which is crucial for identifying seizure activity.

- **Pooling Layers**: Pooling layers are used to reduce the spatial dimensions of the feature maps while retaining the most important information. Max pooling is often applied, which selects the maximum value within each pooling window. This helps in reducing computational complexity and controlling overfitting by providing a form of translational invariance.

- **Fully Connected Layers**: After the convolutional and pooling layers, the output feature maps are flattened and fed into fully connected layers. These layers perform high-level reasoning based on the extracted features, enabling the classification of the EEG signals into one of the six target classes: seizure (SZ), generalized periodic discharges (GPD), lateralized periodic discharges (LPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), or "other."

- **Output Layer**: The output layer consists of six neurons, each corresponding to one of the target classes, with a softmax activation function to generate a probability distribution over the classes.

The choice of a 2D-CNN was motivated by the spatial relationships present in EEG data, where different brain regions are represented by distinct channels. Convolutional layers are particularly effective in learning local patterns, such as the synchronization between neighboring brain regions, which is critical for detecting seizure-like activity. By stacking multiple convolutional layers, the model can learn more complex features, which contributes to a deeper understanding of the EEG signals.

#### 3.0.6.2 Hilbert Transform-based CNN (HT-CNN)

In addition to the 2D-CNN, a Hilbert Transform-based CNN (HT-CNN) was implemented. The Hilbert Transform was used to enhance feature extraction by providing both amplitude and phase information, which are important for distinguishing different types of brain activity.

- **Hilbert Transform Layer**: The first layer of the HT-CNN applies the Hilbert Transform to the input EEG signals. The Hilbert Transform generates an analytic signal that contains both the original amplitude and the instantaneous phase information. The output $H(t)$ of the Hilbert Transform for a signal $x(t)$ is given by:

  $$H(t) = \frac{1}{\pi}\,\text{P.V.}\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau}\,d\tau$$

  where P.V. denotes the Cauchy principal value. This transformation enhances the model's ability to learn phase-related features, which are known to be important for detecting rhythmic brain activities.

Table 3.2: HTCNN Architecture for EEG Classification Model

| **Layer (type)** | **Output Shape** | **Param #** |
|---|---|---|
| conv2d | (None, 3, 10000, 64) | 576 |
| conv2d_1 | (None, 3, 5000, 64) | 32,832 |
| max_pooling2d | (None, 3, 2500, 64) | 0 |
| hilbert_layer | (None, 3, 5000, 64) | 0 |
| conv2d_2 | (None, 3, 2500, 128) | 65,664 |
| conv2d_3 | (None, 3, 1250, 128) | 131,200 |
| max_pooling2d_1 | (None, 1, 625, 128) | 0 |
| conv2d_4 | (None, 1, 625, 256) | 524,544 |
| conv2d_5 | (None, 1, 313, 256) | 1,048,832 |
| max_pooling2d_2 | (None, 1, 156, 256) | 0 |
| global_average_pooling2d | (None, 256) | 0 |
| dense | (None, 256) | 65,792 |
| dropout | (None, 256) | 0 |
| dense_1 | (None, 128) | 32,896 |
| dense_2 | (None, 64) | 8,256 |
| dropout_1 | (None, 64) | 0 |
| dense_3 | (None, 6) | 390 |
| **Total params:** | 1,910,982 | |
| **Trainable params:** | 1,910,982 | |
| **Non-trainable params:** | 0 | |

- **Convolutional Layers**: Similar to the 2D-CNN, convolutional layers are employed to extract local features from the transformed EEG data. By incorporating the Hilbert Transform, these layers can learn features that are more sensitive to both amplitude and phase changes in the EEG signals.

- **Pooling Layers**: Pooling layers are used to downsample the feature maps, reducing the dimensionality and allowing the model to focus on the most relevant features. This also helps in preventing overfitting.

- **Fully Connected Layers and Output Layer**: The flattened feature maps from the convolutional layers are passed through fully connected layers, which perform high-level classification based on the extracted features. The output layer, with a softmax activation function, generates the probability distribution for each of the six target classes.

The HT-CNN was chosen to enhance the feature extraction process by incorporating both amplitude and phase information, which is essential for distinguishing complex brain activities. The Hilbert Transform provides an additional perspective on the EEG data, allowing the model to capture subtle differences in phase and frequency components that might be indicative of specific brain states. This additional information improves the model's ability to differentiate between different harmful brain activities.

#### 3.0.6.3 Comparison and Selection

Both the 2D-CNN and HT-CNN were trained and evaluated to determine the most effective model for classifying harmful brain activities. The HT-CNN demonstrated improved classification accuracy and faster convergence due to the additional features derived from the Hilbert Transform. The final selection of the model was based on achieving the highest accuracy while maintaining interpretability and computational efficiency. The inclusion of convolutional layers allows both models to learn spatial dependencies across EEG channels, while the Hilbert Transform in the HT-CNN helps capture phase information, leading to a more nuanced understanding of brain activity. This combination of features contributes to the effective learning and classification of harmful brain patterns.

# Chapter 4

# Experimental Setup

### 4.0.1 Objectives of the Experiment

The primary objective of this experiment is to develop and evaluate deep learning models for the classification of harmful brain activities from EEG data. EEG signals, due to their non-stationary and complex nature, present significant challenges in detecting harmful activities like seizures and other abnormal brain patterns. Deep learning models, particularly Convolutional Neural Networks (CNNs), are well-suited for capturing spatial and temporal patterns from raw EEG data. This experiment aims to explore the effectiveness of different deep learning architectures, specifically 2D-CNN and HT-CNN, in improving the classification accuracy of harmful brain activities in the real-time settings.

The experiment further seeks to optimize key aspects of model training, including the selection of loss functions, training strategies, and evaluation metrics, ensuring robust performance across different patterns of harmful brain activities.

#### 4.0.1.1 Loss Functions

A critical component of deep learning model training is the choice of the loss function, as it guides the model in learning by minimizing the error during backpropagation. In this experiment, two loss functions were employed for model training:

- **Categorical Cross-Entropy (CCE)**: CCE is a widely used loss function for multi-class classification tasks. It computes the logarithmic loss between the predicted probability distribution and the true labels. The cross-entropy loss is defined as:

$$L_{\mathrm{CCE}} = -\sum_{i=1}^{N} y_i \log(p_i)$$

where $y_i$ is the true label and $p_i$ is the predicted probability for the $i$-th class. CCE was chosen because it penalizes incorrect predictions more heavily, pushing the model to learn better class separation.

- **Kullback-Leibler Divergence (KL-Divergence)**: To compare the predicted class distribution against the expert-annotated votes for each EEG segment, KL-Divergence was also used as a loss function. It measures how one probability distribution diverges from a reference distribution, which in this case is the true label distribution provided by multiple expert annotators. KL-Divergence is given by:

$$L_{\mathrm{KL}} = \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right)$$

where $p_i$ is the true distribution and $q_i$ is the predicted distribution. This loss function is particularly useful in handling cases where the true labels are distributed across multiple classes (as seen in ambiguous patterns).

#### 4.0.1.2 Performance and Evaluation

The evaluation of the trained models was carried out to assess their ability to classify harmful brain activities effectively. Several metrics were used to evaluate the performance, including accuracy, sensitivity, specificity, precision, F1-score, confusion matrix, ROC, AUROC, and Cohen's Kappa. This subsection provides details on these evaluation metrics and how they were used to interpret the model's effectiveness.

- **Accuracy**: Accuracy measures the proportion of correctly classified instances out of the total instances. It is given by:

$$\text{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Accuracy provides an overall measure of how well the model performs but may not be sufficient for imbalanced datasets.

- **Sensitivity (Recall)**: Sensitivity, also known as recall or true positive rate, measures the model's ability to correctly identify positive instances. It is defined as:

  $$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

  Sensitivity is crucial for detecting harmful brain activities such as seizures, where false negatives must be minimized.

- **Specificity**: Specificity measures the model's ability to correctly identify negative instances. It is given by:

  $$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

  Specificity is important for ensuring that normal brain activity is not misclassified as harmful.

- **Precision**: Precision measures the proportion of true positives among all instances classified as positive. It is defined as:

  $$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

  Precision is important in evaluating the reliability of the model when it predicts harmful activity.

- **F1-Score**: The F1-score is the harmonic mean of precision and recall, providing a balanced measure when the class distribution is imbalanced. It is given by:

  $$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

  A high F1-score indicates a good balance between precision and recall.

- **Confusion Matrix**: The confusion matrix is used to provide a detailed breakdown of the classification performance across different classes. It shows the number of true positives, true negatives, false positives, and false negatives for each class, allowing for an in-depth understanding of the model's strengths and weaknesses.

- **Receiver Operating Characteristic (ROC) and AUROC**: The ROC curve is a graphical representation of the true positive rate (sensitivity) versus the false positive rate (1-specificity) for various threshold values. The Area Under the ROC Curve (AUROC) provides a single scalar value summarizing the overall performance of the model:

  $$\text{AUROC} = \int_0^1 \text{TPR}(t)\, d(\text{FPR}(t))$$

  where TPR represents the true positive rate and FPR represents the false positive rate. A higher AUROC value indicates better model discrimination capability between positive and negative classes.

- **Cohen's Kappa**: Cohen's Kappa is used to measure the level of agreement between the predicted and true labels, adjusted for the agreement that occurs by chance. It is given by:

  $$\kappa = \frac{P_o - P_e}{1 - P_e}$$

  where $P_o$ is the observed agreement and $P_e$ is the expected agreement by chance. Cohen's Kappa helps evaluate the model's performance considering the class imbalance and is particularly useful when comparing multiple models.

These performance metrics were computed after each training session and averaged over multiple cross-validation runs to ensure reliable and robust model evaluation. The models were also compared in terms of convergence speed and stability during training, providing a complete view of their performance.

#### 4.0.1.3 Model Comparison

To compare the performance of the two models (2D-CNN and HT-CNN), the following aspects were considered:

- **Convergence Speed**: The convergence speed of the models was analyzed to determine how quickly each model reached optimal performance. The HT-CNN model showed faster convergence due to the additional features extracted by the Hilbert Transform layer.

- **Classification Performance**: Classification metrics such as accuracy, sensitivity, specificity, precision, F1-score, AUROC, and Cohen's Kappa were used to compare the effectiveness of the 2D-CNN and HT-CNN models.

- **Error Analysis**: The error analysis focused on the distribution of false positives and false negatives for each class, which provided insights into the strengths and weaknesses of each model.

### 4.0.2 Dataset Description

The dataset used for this work is titled *"HMS - Harmful Brain Activity Classification"*, which is hosted on Kaggle by the Sunstella Foundation in collaboration with Harvard Medical School, Persyst, Jazz Pharmaceuticals, and the Clinical Data Animation Center (CDAC). The goal of this dataset is to help researchers preserve and enhance brain health by providing annotated EEG data for harmful brain activity classification. The dataset contains EEG recordings collected at a frequency of 200 samples per second, with expert annotators labeling 50-second-long EEG samples along with matched spectrograms covering a 10-minute window centered at the same time. The central 10 seconds of each segment were labeled by experts, and overlapping samples have been consolidated. The dataset includes metadata in a file named *"data.csv"*, which provides information about the individual subsets annotated by raters. The dataset contains multiple channels representing EEG electrode locations, with one additional channel for electrocardiogram (EKG) data that records heart activity. The six patterns of interest are seizure (SZ), generalized periodic discharges (GPD), lateralized periodic discharges (LPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), and "other." The EEG segments have been annotated by experts, and in some cases, disagreements occurred. Segments with high levels of agreement are termed as *"idealized patterns"*, whereas segments with mixed opinions are classified as *"proto patterns"* or *"edge cases"* based on the level of expert disagreement. The dataset provides a robust basis for training and evaluating models for harmful brain activity classification.

Table 4.1: Dataset Overview

| **Query** | **Value** |
|---|---|
| Number of Patients | 17,089 |
| Number of EEG Samples | 106,000 |
| Sampling Frequency | 200 Hz |
| Sample Window Length | 50s |
| Target Classes | sz, lpd, gpd, lrda, grda, oth |
| EEG Channels | Fp1, F3, C3, P3, F7, T3, T5, O1, Fz, Cz, Pz, Fp2, F4, C4, P4, F8, T4, T6, O2, EKG |

### 4.0.3 Experimental Setup

#### 4.0.3.1 Hyper-Parameter Selection

Selecting appropriate hyper-parameters is crucial for optimizing model performance and ensuring effective learning. In this study, several hyper-parameters were tuned for both the 2D-CNN and HT-CNN models, including learning rate, batch size, number of filters, kernel size, and dropout rate. Each of these hyper-parameters and the rationale behind their selection are described below.

- **Learning Rate**: The learning rate controls how much the model's parameters are adjusted with respect to the loss gradient during each iteration of training. The weight update rule for a parameter $w$ can be expressed as:

  $$w \leftarrow w - \eta \cdot \frac{\partial L}{\partial w}$$

  where $\eta$ is the learning rate and $\frac{\partial L}{\partial w}$ is the gradient of the loss function $L$ with respect to the parameter $w$. A learning rate of $1 \times 10^{-5}$ was chosen to provide a balance between stable convergence and avoiding overshooting the optimal solution. A lower learning rate ensured that the model could gradually minimize the loss function without oscillating.

- **Batch Size**: The batch size was set to 64, meaning that 64 samples were used to estimate the gradient during each update of the model parameters. This batch size was selected to maintain a balance between computational efficiency and the stability of gradient estimation. Smaller batch sizes can lead to noisier updates, while larger batch sizes require more memory and may slow down training.

- **Number of Filters and Kernel Size**: The number of filters in each convolutional layer and the kernel size were tuned based on the complexity of the features to be extracted. The output of a convolutional layer can be represented as:

  $$y_{i,j,k} = f\left(\sum_m \sum_n x_{i+m,j+n,k} \cdot w_{m,n} + b_k\right)$$

  where $x_{i,j,k}$ represents the input feature map, $w_{m,n}$ is the filter, $b_k$ is the bias term, and $f$ is the activation function (typically ReLU). Smaller kernel sizes (e.g., $3 \times 3$) were used in the initial layers to capture fine-grained details, while deeper layers used larger kernels to extract more abstract features. The number of filters was increased in subsequent layers (e.g., 32, 64, 128) to enable the model to learn more complex patterns.

- **Dropout Rate**: To prevent overfitting, a dropout layer was used after the fully connected layers, with a dropout rate of 0.5. Dropout is a regularization technique that randomly drops a fraction of the neurons during training. Mathematically, the output $y_i$ of a neuron after dropout can be represented as:

  $$y_i = \begin{cases} 0 & \text{with probability } p \\ \frac{h_i}{1-p} & \text{with probability } 1-p \end{cases}$$

  where $h_i$ is the original output of the neuron, and $p$ is the dropout rate. This technique prevents the model from relying too heavily on specific neurons and improves its ability to generalize to unseen data.

- **Activation Function**: The ReLU (Rectified Linear Unit) activation function was used in the convolutional layers to introduce non-linearity. The ReLU function is defined as:

  $$f(x) = \max(0, x)$$

  ReLU is computationally efficient and helps mitigate the vanishing gradient problem, enabling the model to learn effectively by allowing only positive values to pass through.

- **Early Stopping**: Early stopping was employed to monitor the validation loss and prevent overfitting. Training was stopped if the validation loss did not improve for 50 consecutive epochs, and the model with the best validation performance was saved.

- **Optimizer**: The Adam optimizer was chosen for its adaptive learning rate capabilities, which combine the advantages of both RMSprop and SGD optimizers. The parameter update rule in Adam can be expressed as:

  $$w_{t+1} = w_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon}$$

  where $m_t$ and $v_t$ are the first and second moment estimates at time step $t$, respectively, $\eta$ is the learning rate, and $\epsilon$ is a small constant to avoid division by zero. Adam helps in faster convergence by adapting the learning rate for each parameter, making it well-suited for training deep networks.

The combination of these hyper-parameters allowed the models to converge efficiently while minimizing overfitting. By fine-tuning the learning rate, batch size, and other hyper-parameters, the models were able to achieve optimal performance in classifying harmful brain activities from EEG data.

#### 4.0.3.2 Model Training

The training process of the model was designed to ensure its generalizability and robustness. To achieve this, the model was trained multiple times using different random splits, along with various other techniques to prevent overfitting and enhance performance. Each component of the training process is explained in detail below:

- **Repeated Training with Random Splits**: To evaluate the consistency and robustness of the model, it was trained 10 times using different random splits of the dataset in the ratio of 75% for training, 15% for validation, and 15% for testing. Each time, a different random split was used, and the model's performance was averaged over all 10 trials to obtain a reliable estimate. This approach helps assess how well the model generalizes across different subsets of the data and reduces the risk of the model overfitting to a particular split.

- **Early Stopping**: To prevent overfitting, early stopping was used during training. If the validation loss did not improve for 50 consecutive epochs, the training process was halted. The model with the best validation performance was saved, ensuring that the model did not overfit to the training set.

- **Regularization Techniques**: To further prevent overfitting, regularization techniques such as dropout and batch normalization were employed. Dropout, with a rate of 0.5, was applied to fully connected layers to randomly deactivate neurons during training, thereby improving generalizability. Batch normalization was also used to standardize the inputs to each layer, ensuring that the model remains stable throughout the training process. The output of batch normalization is given by:

  $$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

  where $x_i$ is the input, $\mu$ is the batch mean, $\sigma^2$ is the batch variance, and $\epsilon$ is a small constant added for numerical stability.

- **Training Infrastructure**: The model training was carried out using GPU acceleration to reduce the training time. The experiments were run on a machine equipped with an NVIDIA RTX 3060 GPU with 24GB of VRAM memory, allowing for efficient matrix operations, which are fundamental to deep learning training.

- **Training Summary**: The combination of repeated training with different random splits, appropriate loss functions, adaptive learning rate scheduling, and regularization techniques ensured that the model learned effectively and generalized well to unseen data. The final trained model demonstrated good classification accuracy and stability, validated by the metrics recorded during training.

# Chapter 5

# Results Observed

In this chapter, we present the performance evaluation of the two primary models(2D-CNN model and the HT-CNN model) along with other SOTA models used for EEG signal classification. We evaluate these models using several metrics, including accuracy, precision, recall, F1-score, AUROC, and Cohen's Kappa, to provide a comprehensive analysis of each model's strengths and weaknesses. The evaluation metrics are based on the performance of the models in classifying six EEG patterns: seizure (sz), lateralized periodic discharges (lpd), generalized periodic discharges (gpd), lateralized rhythmic delta activity (lrda), generalized rhythmic delta activity (grda), and "other" (oth).

## 5.1 Accuracy and Loss Curves

The training process of the 2D-CNN and HT-CNN models using categorical cross-entropy loss and kl divergence loss was monitored through the accuracy and loss plots, which are displayed in the below Fig. 5.1, Fig. 5.2, Fig. 5.4 and Fig. 5.5 respectively. The training and validation accuracy and loss plot of the HT-i-CNN (sequential HT layer with CNN layer) is also depicted in Fig. 5.3. These plots depict the model's behaviour over the course of 500 epochs, 250 epochs and 300 epochs, providing insight into both the training and validation phases.

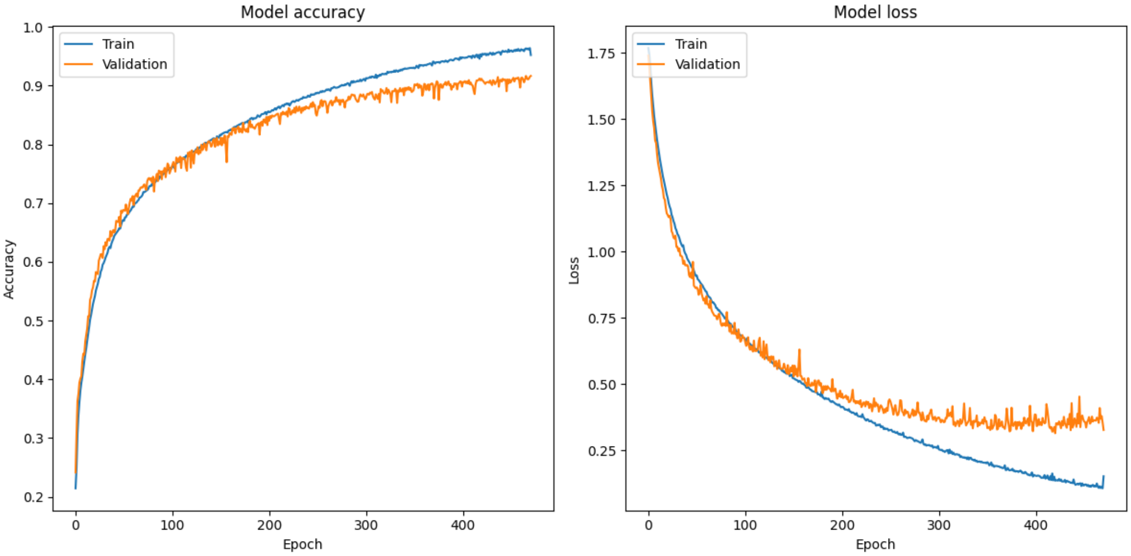


Figure 5.1: Training and Validation Accuracy and Loss curves for 2D-CNN model with Cross-Entropy Loss

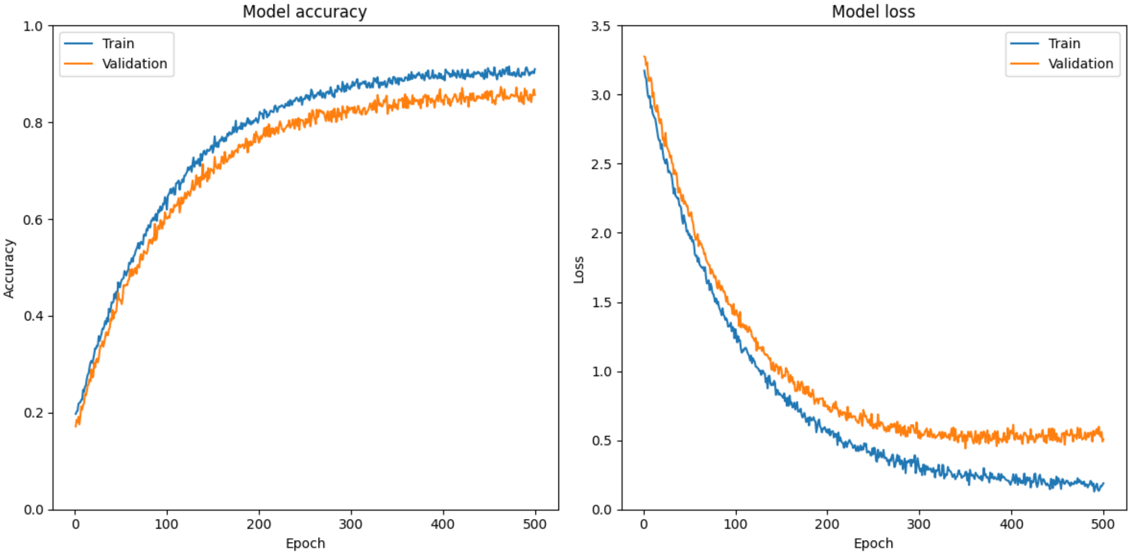


Figure 5.2: Training and Validation Accuracy and Loss curves for 2D-CNN model with Kullback-Leibler Divergence Loss

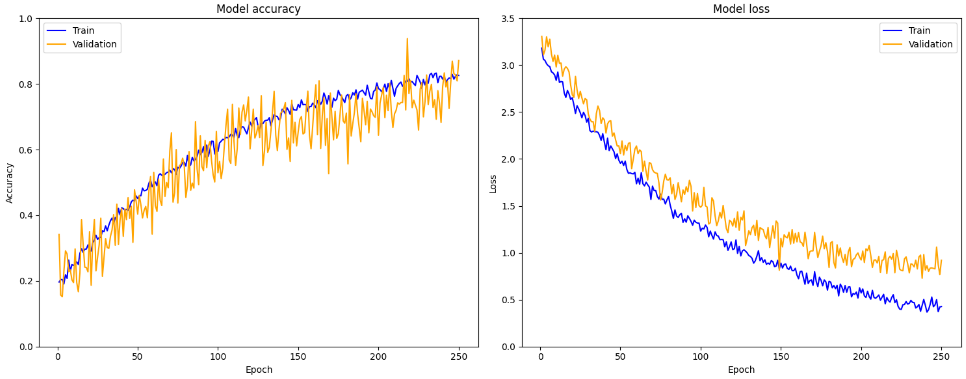


Figure 5.3: Training and Validation Accuracy and Loss curves for HT-i-CNN model with Cross-Entropy Loss

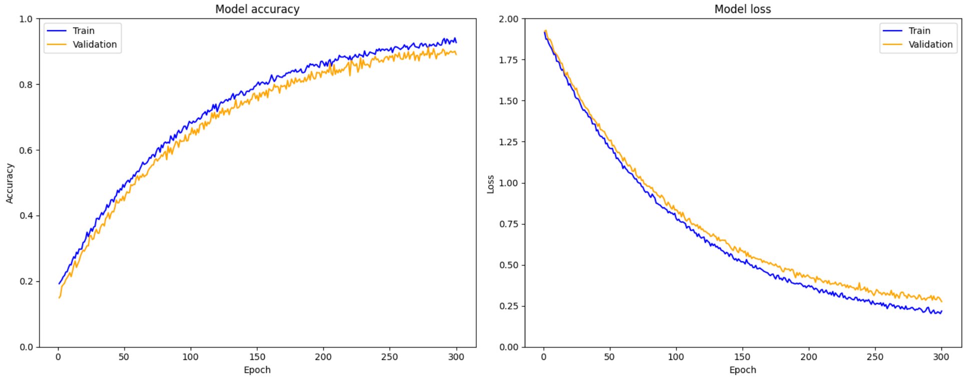


Figure 5.4: Training and Validation Accuracy and Loss curves for HT-CNN model with Cross-Entropy Loss

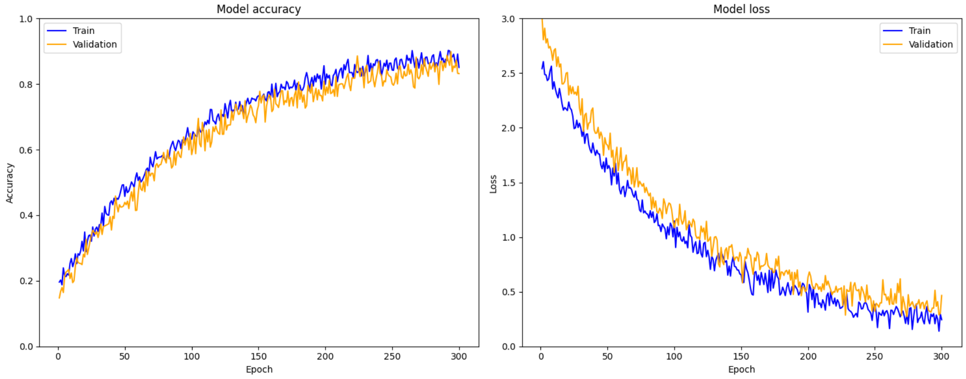
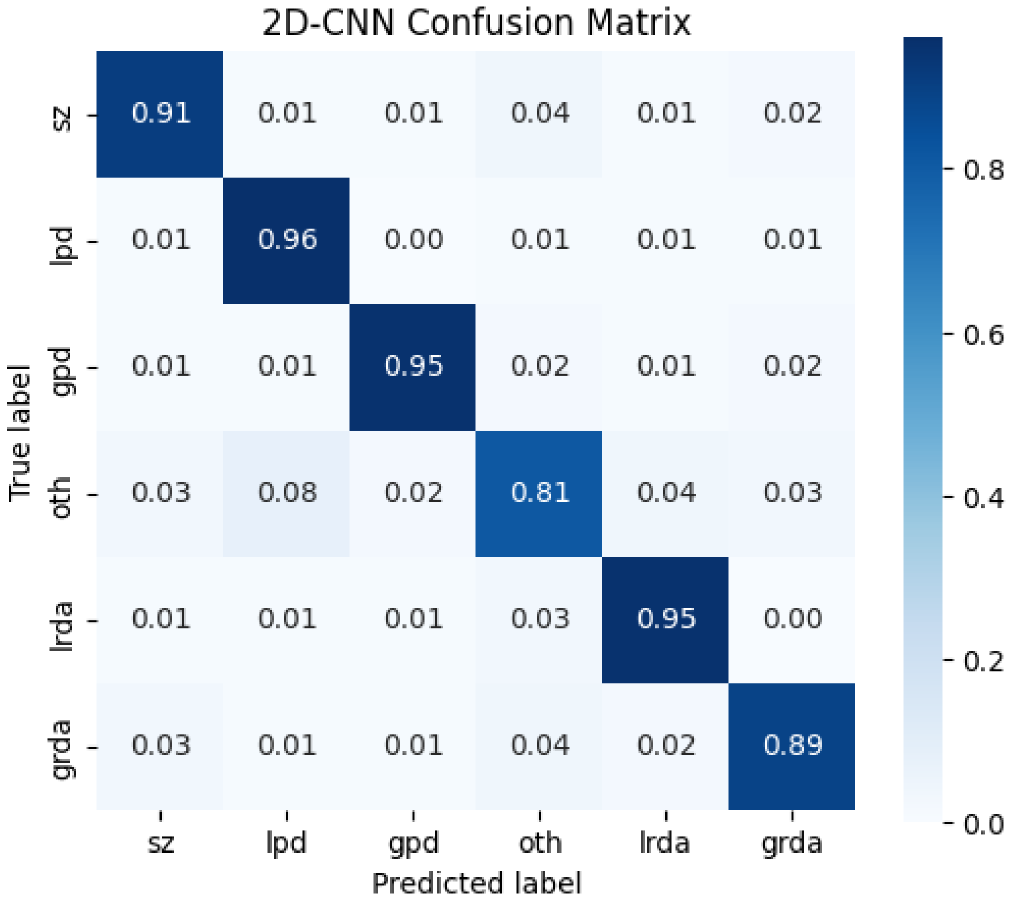


Figure 5.5: Training and Validation Accuracy and Loss curves for HT-CNN model with Kullback-Leibler Divergence Loss

The accuracy and loss curves indicate that the HT-CNN model generally outperforms the 2D-CNN model, achieving a higher accuracy and lower loss in both training and validation sets. The HT-CNN model shows faster convergence, particularly with the Cross-Entropy loss.

## 5.2 Confusion Matrices

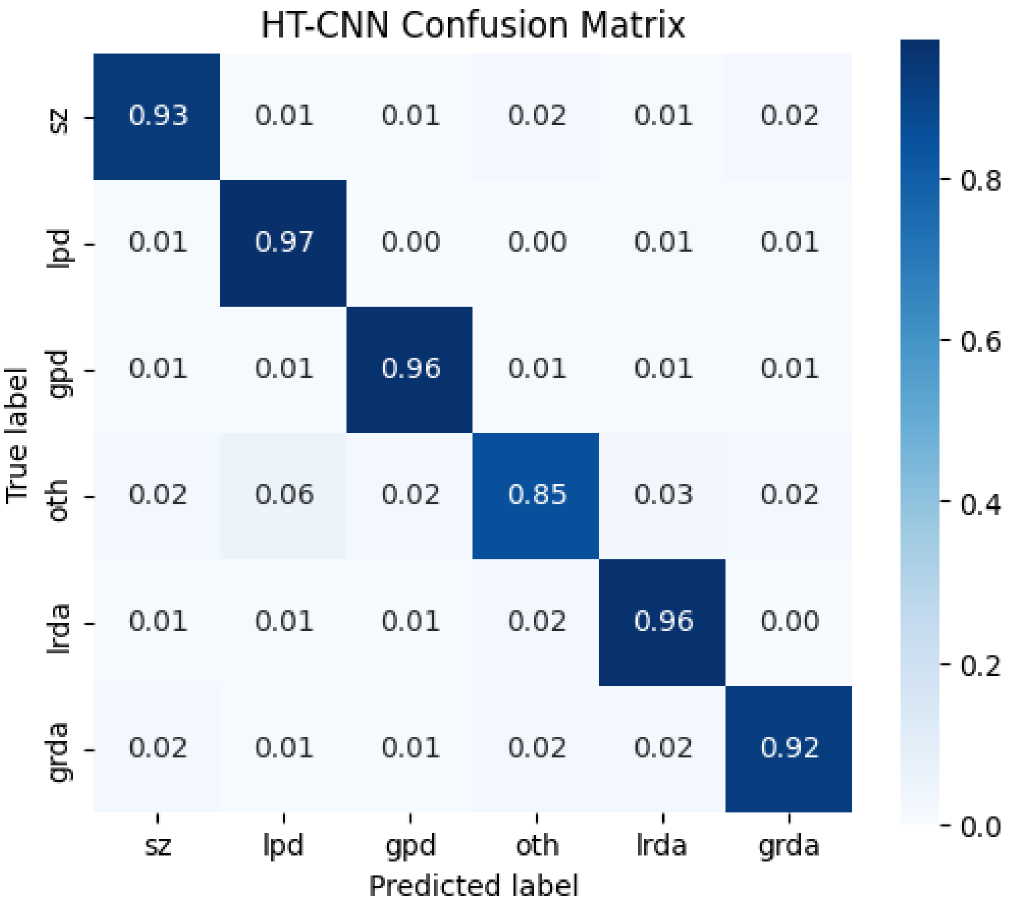
The confusion matrices for each model provide insight into the classification performance across each EEG pattern class.



2D-CNN Confusion Matrix

| True label \ Predicted label | sz | lpd | gpd | oth | lrda | grda |
|---|---|---|---|---|---|---|
| sz | 0.91 | 0.01 | 0.01 | 0.04 | 0.01 | 0.02 |
| lpd | 0.01 | 0.96 | 0.00 | 0.01 | 0.01 | 0.01 |
| gpd | 0.01 | 0.01 | 0.95 | 0.02 | 0.01 | 0.02 |
| oth | 0.03 | 0.08 | 0.02 | 0.81 | 0.04 | 0.03 |
| lrda | 0.01 | 0.01 | 0.01 | 0.03 | 0.95 | 0.00 |
| grda | 0.03 | 0.01 | 0.01 | 0.04 | 0.02 | 0.89 |

Figure 5.6: Normalized Confusion Matrix for 2D-CNN model (Cross-Entropy Loss)



| | sz | lpd | gpd | oth | lrda | grda |
|---|---|---|---|---|---|---|
| sz | 0.93 | 0.01 | 0.01 | 0.02 | 0.01 | 0.02 |
| lpd | 0.01 | 0.97 | 0.00 | 0.00 | 0.01 | 0.01 |
| gpd | 0.01 | 0.01 | 0.96 | 0.01 | 0.01 | 0.01 |
| oth | 0.02 | 0.06 | 0.02 | 0.85 | 0.03 | 0.02 |
| lrda | 0.01 | 0.01 | 0.01 | 0.02 | 0.96 | 0.00 |
| grda | 0.02 | 0.01 | 0.01 | 0.02 | 0.02 | 0.92 |

Figure 5.7: Normalized Confusion Matrix for HT-CNN model (Cross-Entropy Loss)

The confusion matrices highlight that the HT-CNN model shows improved classification accuracy across the classes compared to the 2D-CNN model. For example, in the HT-CNN model, the seizure (sz) and lateralized periodic discharges (lpd) classes show a higher rate of correct classification with fewer misclassifications.

## 5.3 ROC Curves

The Receiver Operating Characteristic (ROC) curves for each class in both models are shown below. The Area Under the Curve (AUC) values provide an additional measure of classification quality, with higher values indicating better model performance.

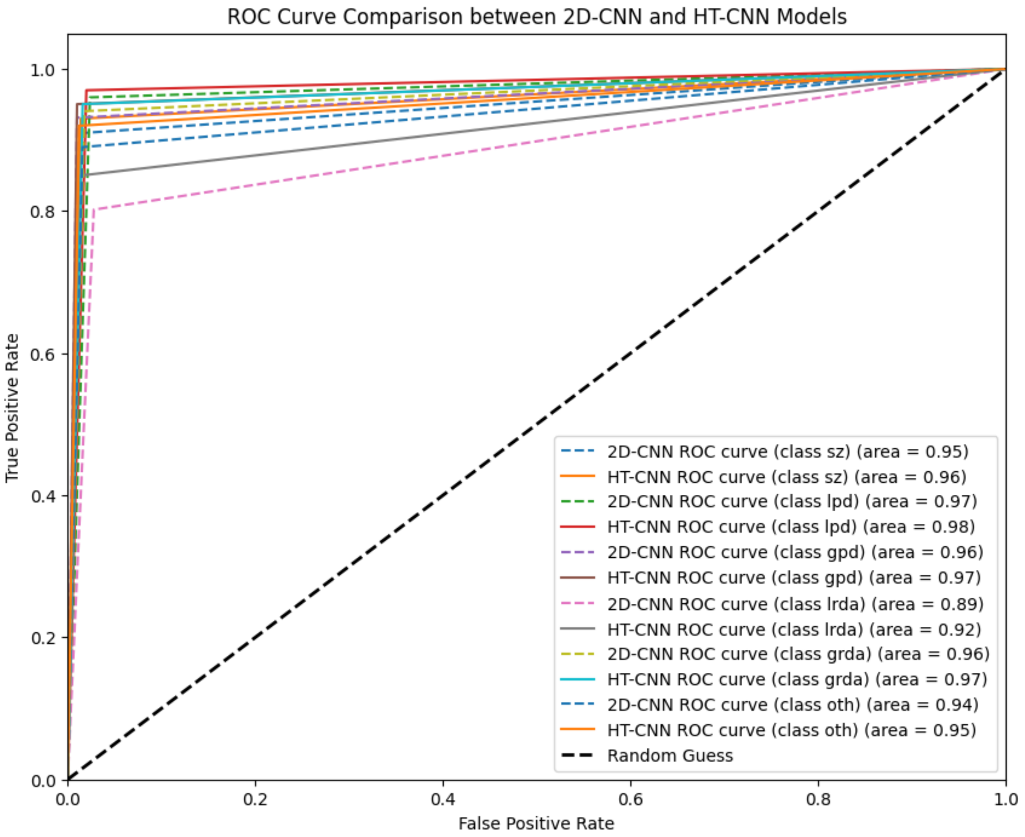


Figure 5.8: ROC Curve Comparison between 2D-CNN and HT-CNN Models

The ROC curves indicate that the HT-CNN model achieves a higher AUROC across most classes compared to the 2D-CNN model, further validating the superior performance of the HT-CNN model in distinguishing between different EEG patterns.

## 5.4 Performance Metrics

Table 5.1 summarizes the performance metrics for both models, providing a comprehensive view of their classification quality.

Table 5.1: Performance Metrics for 2D-CNN, HT-CNN and other SOTA Models

| **Model** | **Acc** | **Prec** | **Rec** | **F1** | **AUROC** | **Kappa** |
|---|---|---|---|---|---|---|
| 2D-CNN | 90.56% | 0.9055 | 0.9056 | 0.9051 | 0.9434 | 0.8868 |
| HT-i-CNN | 83.98% | 0.8037 | 0.8273 | 0.8153 | 0.8533 | 0.8145 |
| HT-CNN | **92.86%** | **0.9288** | **0.9286** | **0.9282** | **0.9571** | **0.9143** |
| Bi-LSTM | 59.77% | 0.6000 | 0.5977 | 0.5988 | 0.6150 | 0.4970 |
| EEGNET-V4 | 43.34% | 0.4400 | 0.4334 | 0.4350 | 0.5100 | 0.3580 |
| DeepConvNet | 49.12% | 0.4950 | 0.4912 | 0.4930 | 0.5300 | 0.4100 |
| ShallowFBCSPNet | 50.64% | 0.5100 | 0.5064 | 0.5080 | 0.5400 | 0.4250 |
| ATCNet-V2 | 44.72% | 0.4500 | 0.4472 | 0.4485 | 0.5150 | 0.3650 |

The HT-CNN model consistently outperforms the 2D-CNN model across all metrics. Notably, the HT-CNN model achieves a higher F1 Score and AUROC, which are critical metrics for evaluating the model's performance in classifying imbalanced classes. Cohen's Kappa also shows a notable improvement, suggesting that the HT-CNN model achieves a higher level of agreement between predictions and true labels compared to the 2D-CNN model.

## 5.5 Visualization of HMS Layers

To gain insights into the internal representations learned by our Harmful Brain Activity Classification model, we visualized the Representational Dissimilarity Matrices (RDM) for each layer. These RDMs offer a detailed view of the similarity between neural activations across different data samples, providing a clear picture of how the model differentiates between brain activity patterns.

### 5.5.1 Layer-wise Analysis

Each layer's RDM is constructed to understand the progression of learned representations through the network hierarchy. The RDMs for layers 1 to 11 are shown in Figures 5.9 to 5.19. We observe that as we move deeper into the layers, the patterns become more distinctive, indicating a progression in the model's ability to abstract and encode complex features.

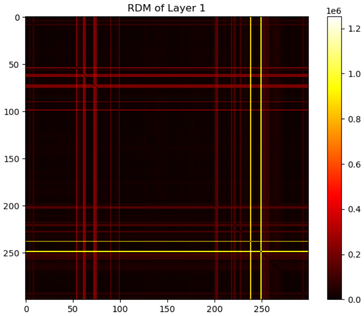


Figure 5.9: RDM of Layer 1

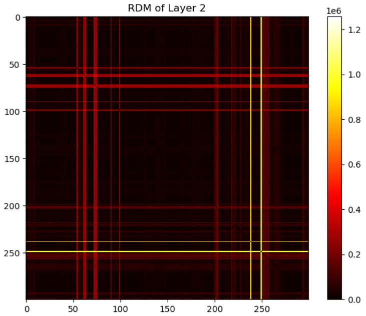


Figure 5.10: RDM of Layer 2

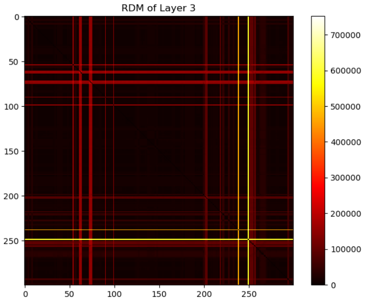


Figure 5.11: RDM of Layer 3

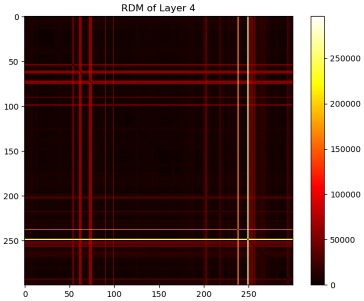


Figure 5.12: RDM of Layer 4

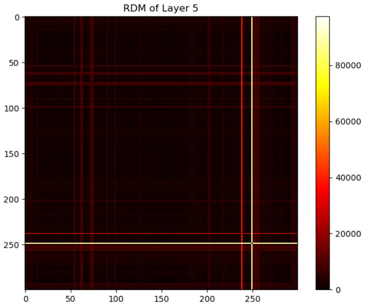


Figure 5.13: RDM of Layer 5

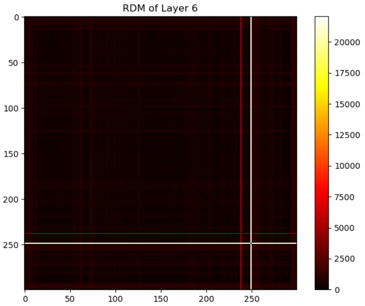


Figure 5.14: RDM of Layer 6

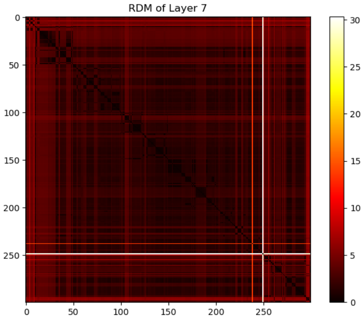


Figure 5.15: RDM of Layer 7

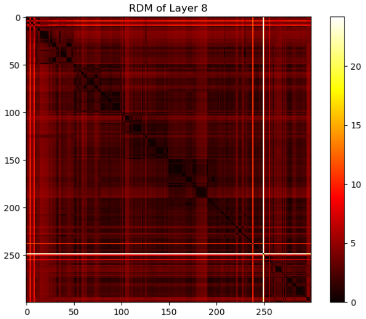


Figure 5.16: RDM of Layer 8

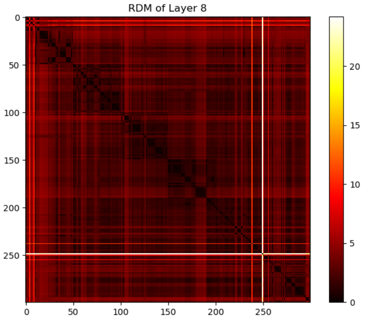


Figure 5.17: RDM of Layer 9

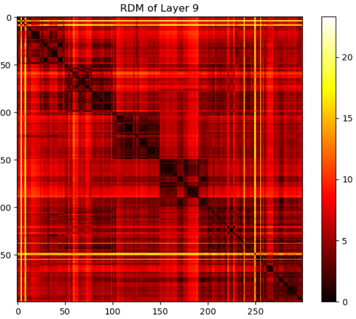


Figure 5.18: RDM of Layer 10

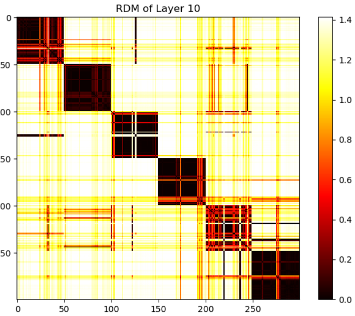


Figure 5.19: RDM of Layer 11

### 5.5.2 Observations

The RDMs reveal a transformation in feature representation as the data moves through the network's layers. Early layers exhibit higher similarity across samples, indicating low-level feature extraction. As we progress to deeper layers, the dissimilarity between different brain activity patterns becomes more pronounced, reflecting the model's capability to separate complex features associated with various types of brain activities.

This visualization supports the interpretability of our model, providing insights into how different layers contribute to differentiating harmful brain activities. Such analysis not only validates the model's learned representations but also serves as a foundation for further improvements and fine-tuning of the model architecture.

## 5.6 Summary of Results

In summary, the results indicate that the HT-CNN model with a Hilbert Transform layer significantly enhances classification performance over the 2D-CNN model. This improvement is evident across various performance metrics, including accuracy, AU-ROC, and Cohen's Kappa. The integration of the Hilbert Transform layer in the HT-CNN model allows for better feature extraction from EEG data, thereby achieving a higher degree of separation between different EEG patterns. The findings suggest that the HT-CNN model is a promising approach for EEG signal classification in clinical settings, potentially aiding in faster and more accurate diagnosis of brain activities.

# Chapter 6

# Future Work

This research demonstrates promising results in automated EEG classification using the HT-CNN model. However, there is ample scope to further improve and extend this work by incorporating emerging techniques in deep learning and artificial intelligence. Specifically, three advanced methodologies—Vision Transformer models, Large Language Models (LLMs), and Meta-Learning—present new avenues for enhancing the accuracy, adaptability, and interpretability of EEG classification systems. This chapter explores how each of these methods could contribute to the future development of this research.

## 6.1 Vision Transformer Models

Transformers have recently revolutionized various fields in machine learning, showing remarkable success in computer vision tasks through Vision Transformers (ViTs). Vision Transformers use self-attention mechanisms to capture complex spatial dependencies within input data, making them particularly effective for tasks requiring detailed pattern recognition. Applying Vision Transformers to EEG classification could enhance the model's ability to identify intricate and non-linear relationships within EEG signals, which are inherently high-dimensional and time-dependent.

In the context of this research, Vision Transformers could be adapted to handle 2D or 3D EEG spectrograms, which are visual representations of EEG data over time. The self-attention mechanism in ViTs would allow the model to focus on the most relevant temporal and spatial features in the EEG signal, potentially leading to more accurate classification of different brain states. Integrating Vision Transformers into

EEG classification models could also help in achieving greater cross-subject generalization, as the model learns to extract robust features from varied EEG patterns. The success of Vision Transformers in vision tasks suggests that this model could provide a new level of interpretability and performance, making it a valuable direction for future research.

## 6.2 LLMs for EEG Classification

Large Language Models (LLMs), such as GPT and BERT, have shown significant potential in a variety of domains beyond natural language processing. In the field of EEG analysis, LLMs could be used to analyze EEG patterns by treating them as a sequence of tokens or events. By encoding EEG signals into a form compatible with LLMs, the model could leverage sequential dependencies within EEG data to better classify complex brain states.

LLMs could contribute to EEG classification in multiple ways. Firstly, they can capture long-range dependencies in time-series data, which is crucial for EEG, as brain activity patterns are often dependent on context across multiple time points. Secondly, by fine-tuning LLMs on EEG data, the model could learn more nuanced representations of brain activity, potentially improving classification accuracy for complex patterns. Lastly, LLMs could provide interpretability by allowing researchers to visualize and understand the ”attention” placed on different time segments of the EEG signal, highlighting which parts of the EEG sequence were most influential in the model’s predictions. This approach could open up innovative ways of understanding EEG data, aligning with recent trends in explainable AI.

## 6.3 Meta-Learning for Enhanced Cross-Subject Generalization

A critical challenge in EEG analysis is achieving reliable performance across different subjects, as individual brain patterns can vary significantly. Meta-learning, or ”learning to learn,” aims to train models that can quickly adapt to new tasks or subjects with minimal data. By using meta-learning techniques, this research could develop models capable of adapting to the specific EEG patterns of new subjects, thereby improving cross-subject generalization.

Meta-learning techniques such as Model-Agnostic Meta-Learning (MAML) could be applied to fine-tune the model on a small subset of new subjects before deployment. This would enable the model to rapidly adjust to individual differences in EEG patterns without extensive retraining. Meta-learning could also help address data scarcity by maximizing the utility of existing labeled EEG data, reducing the need for large-scale annotated datasets for each new subject. Integrating meta-learning into EEG classification could therefore result in a more flexible and robust model, capable of achieving high performance even on previously unseen subjects, which is essential for practical clinical applications.

# Chapter 7

# Conclusion

This research presents a significant advancement in automated EEG classification by introducing and evaluating convolutional neural network architectures, specifically the 2D-CNN and the enhanced HT-CNN models, for classifying harmful brain activities. Through the integration of the Hilbert Transform layer, the HT-CNN model demonstrated superior performance over conventional approaches, achieving higher accuracy and interpretability in distinguishing complex EEG patterns. This improvement underscores the potential of advanced signal processing techniques in deep learning models to capture the subtle nuances of brain activity, facilitating timely and precise diagnosis in neurocritical care settings.

One of the core strengths of this research lies in its attempt to ensure **person invariability**, where the model generalizes across subjects by cross-subject training and testing, effectively minimizing person-specific variability. This focus enhances the robustness of the model, making it suitable for real-world applications where subjects vary significantly. Furthermore, this work marks a pioneering effort in utilizing the HMS-Harmful Brain Activity Classification dataset, as no previously published research has explored this dataset. By establishing a baseline with promising results, this study lays a foundation for future research in this domain. Additionally, we aimed to **minimize the number of EEG channels** required for accurate classification, thus making the model more resource-efficient without significantly compromising accuracy. This approach ensures that the model remains practical and adaptable, especially in resource-constrained environments where fewer EEG electrodes are preferable.

The success of this approach not only validates the effectiveness of deep learning in analyzing complex biomedical signals but also opens new possibilities for real-time

EEG interpretation. By automating the classification of EEG patterns, the proposed model could substantially reduce the dependence on manual analysis, alleviating the burden on neurologists and enabling faster intervention in clinical settings.

Looking ahead, the potential for further development is vast. Emerging techniques such as Vision Transformers, Large Language Models, and Meta-Learning offer exciting pathways for enhancing the model's adaptability, accuracy, and interpretability. Vision Transformers could deepen spatial feature extraction in EEG data, Large Language Models might reveal temporal dependencies within sequences, and Meta-Learning could ensure cross-subject generalization, addressing the variability of EEG patterns between individuals. The significance of this research lies not only in its contribution to automated EEG analysis but also in its broader implications for advancing brain health diagnostics.

As technology progresses, the integration of these models into clinical workflows could bring about a transformative shift in neurocritical care, enabling early detection, accurate classification, and personalized treatment strategies for neurological conditions. This work lays a foundation for future advancements, propelling EEG analysis toward becoming a pivotal tool in modern healthcare, with far-reaching benefits for patients and practitioners alike.

# Bibliography

[1] S. H. e. a. Jin Jing, Wendong Ge, "Development of expert-level classification of seizures and rhythmic and periodic patterns during eeg interpretation," *PubMed*, vol. Volume Number, p. Page Numbers, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10136013/

[2] A. F. S. e. a. Jin Jing, Wendong Ge, "Interrater reliability of expert electroencephalographers identifying seizures and rhythmic and periodic patterns in eegs," *PubMed*, vol. Volume Number, p. Page Numbers, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10136018/

[3] R. E. e. a. Zafar SF, "Automated annotation of epileptiform burden and its association with outcomes," *PubMed*, vol. Volume Number, p. Page Numbers, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/34231244/

[4] G. Z. e. a. Barnett AJ, "Improving clinician performance in classifying eeg patterns on the ictal–interictal injury continuum using interpretable machine learning," *PubMed*, vol. Volume Number, p. Page Numbers, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38872809/

[5] G. M. . G. P. Kunekar P., "Detection of epileptic seizure in eeg signals using machine learning and deep learning techniques," *Journal of Engineering and Applied Science*, vol. Volume Number, p. Page Numbers, 2023. [Online]. Available: https://jeas.springeropen.com/articles/10.1186/s44147-023-00353-y

[6] A. N. e. a. Echtioui A., "Epileptic seizures detection using ieeg signals and deep learning models," *Circuits, Systems, and Signal Processing*, vol. Volume Number, p. Page Numbers, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00034-023-02527-8

[7] M. J. Singh K., "Smart neurocare approach for detection of epileptic seizures using deep learning based temporal analysis of eeg patterns," *Multimedia Tools and Applications*, vol. Volume Number, p. Page Numbers, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11042-022-12512-z

[8] B. V. e. a. Statsenko Y, "Automatic detection and classification of epileptic seizures from eeg data: Finding optimal acquisition settings and testing interpretable machine learning approach," *Biomedicines*, vol. 11, p. 2370, 2023. [Online]. Available: https://www.mdpi.com/2227-9059/11/9/2370

[9] T. T. Valentin Gabeff, "Interpreting deep learning models for epileptic seizure detection on eeg signals," *Artificial Intelligence In Medicine*, vol. 117, p. 102084, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365721000774

[10] J. J. Wendong Ge, "Deep active learning for interictal ictal injury continuum eeg patterns," *Journal of Neuroscience Methods*, vol. 351, p. 108966, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027020303897