Classification of Schizophrenia patients based on Stimuli of Speech Perception using Deep Learning and Machine Learning

MSc Research Project
Data Analytics

Darshan Vetal
Student ID: 23215194

School of Computing
National College of Ireland

Supervisor: Dr. Teerath Kumar

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Darshan Vetal
Student ID:	23215194
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr. Teerath Kumar
Submission Due Date:	12/12/2024
Project Title:	Classification of Schizophrenia patients based on Stimuli of Speech Perception using Deep Learning and Machine Learning
Word Count:	XXX
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Darshan Vetal
Date:	8th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Schizophrenia patients based on Stimuli of Speech Perception using Deep Learning and Machine Learning

Darshan Vetal
23215194

Abstract

This research focuses on classifying individuals as schizophrenic or healthy, and further distinguishes schizophrenic patients with and without auditory hallucinations using MRI scans stimulated by speech perception tasks. Schizophrenia, a mental disorder characterized by symptoms such as delusions, memory loss, disorganized thinking, and auditory hallucinations, was studied using deep learning and machine learning models. The analysis targeted Broca's area, associated with auditory processing, to understand neural differences. For Research Question 1, distinguishing hallucinating from non-hallucinating schizophrenic patients, a Convolutional Neural Network (CNN) achieved 48% accuracy initially. Applying ADASYN for oversampling improved accuracy to 65%, demonstrating its efficacy in handling class imbalance. For Research Question 2, A Convolutional Neural Network (CNN) model was implemented, achieving 32% accuracy initially. However, the dataset was imbalanced, necessitating oversampling techniques like SMOTE and ADASYN, which improved model robustness and yielded accuracies of 50% and 57%, respectively. Lazy Predict, a Python library, was employed to benchmark multiple traditional models, with Linear Discriminant Analysis, Linear SVC, and Logistic Regression achieving accuracies of 68%, 62%, and 62%. Grad-CAM visualizations enhanced interpretability by highlighting key regions influencing model predictions. This study provides insights into the neural correlates of schizophrenia and auditory hallucinations, aiming to improve diagnostic accuracy, treatment personalization, and patient outcomes. The findings highlight the potential of combining MRI data with advanced computational techniques for enhancing clinical decision-making in schizophrenia diagnosis.

Keywords: Schizophrenia, Auditory Hallucinations, MRI, Deep Learning, CNN, SMOTE, ADASYN, Machine Learning, Lazy Predict.

1 Introduction

Schizophrenia comes under mental disorders that affects approximately 1 percent of the global population, people with schizophrenia often experience serious symptoms like hearing voices that aren't there (auditory hallucinations), believing things that aren't true (delusions), and having confused thoughts. Auditory hallucinations, where patients hear distressing voices, significantly impact their quality of life. Understanding the neural mechanisms underlying these symptoms is crucial for developing better diagnostic and therapeutic strategies. The primary research problem is to investigate the neural activation patterns in schizophrenia patients during speech perception tasks and develop machine learning models to classify these patterns. This study focuses on understanding how speech perception stimulated MRI scans in schizophrenia patients with hallucinations and without auditory hallucinations and predict susceptibility to schizophrenia and hallucinations using MRI data.

1.1 Research Questions:

1. Can deep learning & machine learning models predict if a person is schizophrenic or susceptible to schizophrenia based on speech perception stimulated MRI images?

2. Can deep learning & machine learning models predict hallucinations in schizophrenia patients based on speech perception stimulated MRI images?

1.2 Objective:

Objectives to answer the research questions:

1. Investigate machine learning and deep learning techniques for brain MRI scan image classification.

2. Design a comprehensive framework integrating schizophrenia classification with speech perception using MRI images.

3. Implement models that analyze the relationship between patterns developed by speech perception in brain.

4. Evaluate the accuracy, reliability, and generalizability of the proposed models using datasets of real patients (open source data).

The study is organized into five sections which are Introduction, Literature Review, Methodology, Results and Conclusion. Sequentially, these sections address the research objectives and various steps undertaken toward the eventual implementation and validation of the developed models in the healthcare environment. Therefore, this topic has many prospects for high contribution to the modeling of schizophrenia related tools because it can effectively combine technical and practical aspects.

2 Related Work

This section studies various papers by different authors to understand classification of Schizophrenic patients based on different Machine learning and Deep learning methods.

2.1 Auditory Hallucinations and Brain Activity in Schizophrenia

1. Auditory verbal hallucinations (AVH) in schizophrenia are linked to reduced activation in the left primary auditory cortex (Heschl's gyrus) during speech perception tasks. Studies suggest this reflects an early information processing deficit in the disorder. Schizophrenia patients show reduced activation in response to sentences and reversed speech,

regardless of AVH presence, with no significant differences between hallucinating and non hallucinating groups. AVH may alter superior temporal cortex activity, potentially causing competition for processing resources between hallucinations and real auditory stimuli. While some studies report reduced activation during AVH, others find no significant differences. Neuroimaging reveals structural deficits, including gray matter loss in superior temporal regions and bilateral thalami, correlating with auditory dysfunction. Functional disruptions, such as reduced mismatch negativity (MMN) amplitude and default mode network (DMN) dysregulation, are also implicated in AVH generation and broader cognitive deficits. Meta analyses highlight consistent left superior temporal gyrus activation reductions in AVH plus patients. Advanced machine learning approaches suggest neurobiological subtypes in schizophrenia, aiding diagnosis and treatment. Reduced auditory cortex activation appears independent of AVH frequency, pointing to widespread auditory and cognitive impairments in schizophrenia. These findings emphasize AVH as a modulator rather than the sole driver of auditory processing deficits in the disorder. Soler-Vidal et al. (2022)

2.2 Machine Learning in Schizophrenia Diagnosis

1. This study investigates the application of machine learning (ML) techniques for early and accurate diagnosis of schizophrenia and bipolar disorder (BD). By employing algorithms such as support vector machines (SVM), random forests (RF), and gradient boosting (GB), the research evaluates their performance in predicting these mental health conditions. Among the models, random forests demonstrated superior accuracy and sensitivity, making it particularly effective for identifying patients with schizophrenia and BD. Gradient boosting, on the other hand, exhibited the highest specificity, suggesting its strength in minimizing false positives. These findings highlight the potential of ML to enhance diagnostic precision, offering a valuable tool for early intervention and treatment planning. The study underscores the importance of algorithm selection based on diagnostic priorities, such as balancing sensitivity and specificity, in addressing the diagnostic challenges posed by complex and heterogeneous psychiatric conditions like schizophrenia and BD. Montazeri et al. (2022)

2. The study focused on using machine learning (ML) models to predict schizophrenia by analyzing polygenic risk scores (PRS) along with clinical and demographic factors. It compared the performance of ML models to traditional logistic regression. The findings revealed that while ML models had strong predictive ability, they did not perform significantly better than logistic regression. Additionally, the study found that polygenic risk scores were linked to traits commonly associated with schizophrenia. This suggests that while advanced models like ML can handle complex data, simpler methods like logistic regression may still be equally effective for certain medical predictions. Overall, the research highlights the potential of combining genetic and demographic data for better understanding schizophrenia risk, while also pointing out that the choice of method depends on the specific application and data used. Bracher-Smith et al. (2022)

3. This study reviewed 119 research papers that used advanced machine learning techniques to analyze brain imaging data. The goal was to find reliable biomarkers that could distinguish people with schizophrenia spectrum disorders (SSD) from healthy individuals. The analysis showed that the methods achieved a sensitivity of 79.1 percent (correctly

identifying patients) and a specificity of 80.0 percent (correctly identifying healthy individuals). These results highlight the potential of brain imaging to support the diagnosis of SSD. However, the study also found that factors related to the patients and the methods used in the studies, such as imaging techniques or analysis tools, influenced the results. This research provides a better understanding of how to use brain imaging and machine learning for diagnosing SSD, while also emphasizing the need to address patient and methodological differences for even more accurate outcomes. Di Camillo et al. (2024)

2.3 Deep Learning in Schizophrenia Diagnosis

1. The study introduces a 3D convolutional neural network designed to segment subcortical brain structures in MRI scans. This advanced model addresses challenges like high computational and memory demands, making it efficient and practical for use. By utilizing GPU based training, the model processes data faster and handles large datasets effectively. It has been tested on different datasets and showed consistent and reliable performance, proving its ability to work across various scenarios. This makes the approach highly suitable for large scale brain imaging studies, helping researchers analyze complex brain structures more accurately and efficiently. Dolz et al. (2017)

2. This study introduces advanced deep learning methods to automatically diagnose schizophrenia using brain activity data from EEG signals. The researchers used a dataset collected at the Institute of Psychiatry and Neurology in Warsaw, Poland. Their model, combining Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) networks, achieved an impressive accuracy of 99.25 percent, outperforming earlier approaches. The study also compared the performance of this deep learning method with traditional machine learning techniques, highlighting its effectiveness and potential for improving schizophrenia diagnosis. This approach demonstrates how cutting edge AI can provide faster and more accurate medical diagnoses, offering new possibilities for mental health care. Shoeibi et al. (2021)

3. This study introduces a method for diagnosing schizophrenia using brain scan data from functional magnetic resonance imaging (fMRI) and a convolutional neural network (CNN) algorithm. The approach is designed to improve early detection of schizophrenia by identifying unique patterns in brain activity. The method not only enhances the accuracy of schizophrenia classification, achieving a high accuracy of 84.3 percent, but also improves the generalization ability of deep learning models, meaning it performs well on different datasets. This development highlights the potential of combining advanced imaging techniques with AI to support earlier and more reliable diagnoses, which can lead to better treatment outcomes. Zheng et al. (2021)

4. This study introduces a combined approach using 3D and 2D convolutional neural networks (CNNs) to analyze schizophrenia patients by processing fused features from different types of MRI data. By integrating information from multiple imaging techniques, the model offers a detailed and accurate understanding of brain differences in schizophrenia. The results are very promising, with the model achieving high performance metrics: 89.86 percent accuracy, 86.21 percent sensitivity (correctly identifying patients), 92.50 percent specificity (correctly identifying healthy individuals), an F1 score of 89.35 percent, and an average performance of 87.72 percent. These findings highlight

the potential of using advanced AI methods with multimodal imaging data to improve the diagnosis and understanding of schizophrenia. The approach could pave the way for more reliable and comprehensive diagnostic tools in mental health care. Guo et al. (2023)

5. This study investigates how magnetic resonance imaging (MRI) combined with deep learning, specifically graph convolutional networks (GCNs), can improve the diagnosis of schizophrenia. By analyzing brain connectivity patterns from MRI scans, the approach achieves higher accuracy, sensitivity, and specificity compared to traditional diagnostic methods. The use of GCNs allows the model to capture complex relationships in brain networks, which are often associated with schizophrenia. Additionally, the study identifies potential biomarkers in the brain, which could aid in understanding the underlying mechanisms of the disorder. These findings highlight the potential of advanced imaging and machine learning techniques in providing more reliable and early diagnosis of schizophrenia, offering a significant step forward compared to conventional methods. This approach not only improves diagnostic performance but also provides valuable insights into the brain regions and connections affected by the disorder, paving the way for future research and clinical applications.Gao et al. (2024)

6. This study introduces a novel method for analyzing functional MRI (fMRI) data to improve the detection of mental disorders. Instead of using traditional magnitude maps, the approach utilizes spatial source phase (SSP) maps, which are derived from complex-valued fMRI data. These SSP maps are less noisy and provide greater sensitivity to mental disorders. The study proposes a 3D convolutional neural network (3D-CNN) framework called SSPNet, designed to process SSP maps. SSPNet demonstrates superior performance in terms of accuracy and area under the curve (AUC) when compared to CNN models that use traditional magnitude maps as input. This approach highlights the advantages of leveraging phase information from fMRI data and advanced deep learning architectures to enhance the diagnosis of mental health conditions, offering a more precise and robust alternative to existing methods. Lin et al. (2022)

2.4 Neuroimaging Biomarkers and Functional Connectivity

1. Deep learning has proven effective in identifying brain changes related to schizophrenia by analyzing T1 weighted MRI scans. The model performs better than standard methods in distinguishing schizophrenia patients from healthy individuals. It highlights that specific brain regions, such as subcortical areas and ventricles, play a key role in making accurate predictions. These findings not only improve the ability to diagnose schizophrenia but also help identify structural brain features associated with the disease. This approach offers a more advanced way of using brain imaging to support early diagnosis and a better understanding of how the brain is affected in schizophrenia. Zhang et al. (2023)

2. Schizophrenia is linked to noticeable brain structure abnormalities, but identifying reliable biomarkers requires analyzing complex data and ensuring consistent results. Using advanced analysis tools, researchers discovered a specific brain pattern unique to schizophrenia patients. This neuroanatomical signature was found to be highly reliable, producing consistent results across different locations, patient groups, and scanning equipment.

Importantly, a user friendly online tool has been created, allowing for individual patient assessments based on this signature. This approach not only improves our understanding of how schizophrenia affects the brain but also offers a practical way to support diagnosis and tailor treatment to individuals. Rozycki et al. (2017)

3. This paper investigates how brain activity patterns, captured through EEG signals, can be used to detect schizophrenia with the help of deep learning methods. By using a functional brain network approach and a 3D convolutional neural network, the study was able to achieve high levels of accuracy, sensitivity, and specificity. This means the method was not only good at identifying individuals with schizophrenia but also at distinguishing them from those without the condition. The analysis also revealed noticeable differences in brain activity patterns between patients and healthy individuals, offering valuable insights for diagnosis and understanding the disorder. This advanced approach shows promise for improving schizophrenia detection and making the process more reliable. Shen et al. (2023)

4. The referenced study examines the application of deep learning techniques to functional magnetic resonance imaging (fMRI) data for the diagnosis of schizophrenia. The researchers propose a novel method that utilizes spatial source phase (SSP) maps derived from complex-valued fMRI data as input for a three-dimensional convolutional neural network (3D-CNN) framework, termed SSPNet. This approach aims to enhance diagnostic accuracy by leveraging the less noisy and more sensitive nature of SSP maps in detecting mental disorders. The study demonstrates that SSPNet significantly outperforms traditional convolutional neural networks (CNNs) that rely on magnitude maps, achieving higher accuracy and area under the curve (AUC) metrics. These findings suggest that incorporating phase information from fMRI data, through the use of SSP maps, can improve the performance of deep learning models in diagnosing schizophrenia. The research highlights the potential of advanced neuroimaging techniques combined with deep learning frameworks to provide more reliable and precise diagnostic tools for mental health conditions. Yin et al. (2023)

3 Methodology

This research adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework to ensure a structured, repeatable, and verifiable approach to the investigation. The six phases of CRISP-DM guide every step, ensuring alignment with the research objectives.

3.1 Business Understanding

The primary goal of this study is to classify schizophrenia patients based on the stimuli of speech perception on the auditory cortex using deep learning and machine learning models. This research aims to:

1. Advance the understanding of auditory processing in schizophrenia patients by leveraging advanced computational methods.

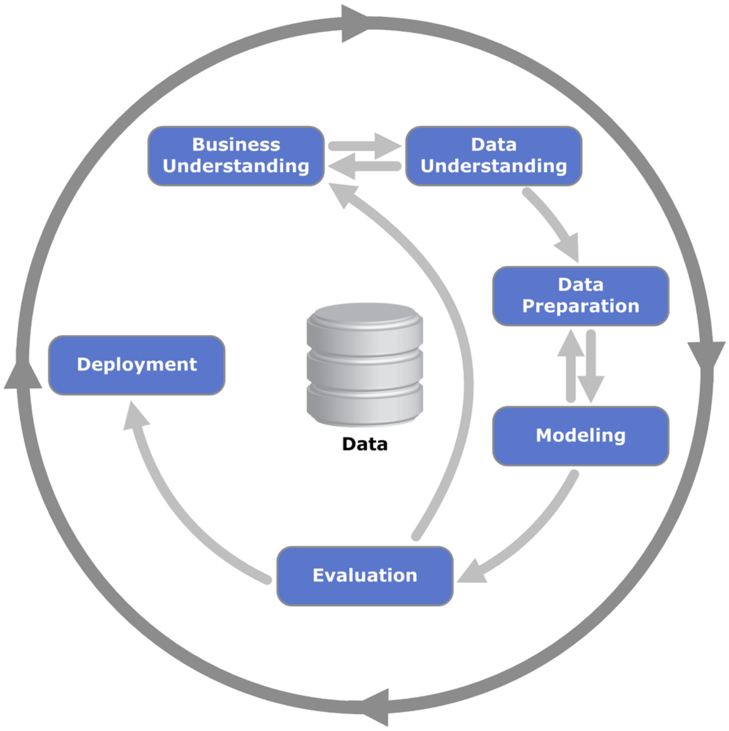

Figure 1: CRISP-DM Methodology contributors (2024)

2. Provide a robust classification framework to assist healthcare professionals in diagnosing schizophrenia based on neuroimaging data.

3. Address gaps in previous studies by demonstrating the practical utility of combining neuroimaging and deep learning techniques in predictive models.

This research builds on findings from prior work Soler-Vidal et al. (2022) and seeks to develop interpretable models capable of classifying schizophrenia patients accurately.

3.2 Data Understanding

Data were derived from two primary sources OpenNeuro (2022) & OpenNeuro (2018). The combined dataset had 170 brain MRI scans, 104 of class 1 (schizophrenic) and 66 of class 0 (healthy).

1. Training and Testing Data: MRI scans sourced from publicly available neuroimaging repositories, focusing on auditory cortex activity during speech perception.

2. Validation Data: A secondary dataset collected via clinical trials was used to validate the model's performance.

The datasets included structural and functional MRI data along with associated labels indicating the patient groups (e.g., schizophrenia and healthy control). Initial exploratory data analysis (EDA) included visualizing neuroimaging data to understand voxel distributions, statistical correlation analysis to identify relevant features influencing classification and dimensionality reduction techniques to summarize brain regions of interest. These steps ensured the dataset's alignment with research objectives and suitability for deep learning-based classification models.

3.3 Data Preparation

Data preparation included the following steps:

1. Preprocessing: Decompression and validation of neuroimaging data in NIfTI format. Resizing 3D images to a uniform shape (96x96x96) for compatibility with convolutional neural networks (CNNs). Normalizing voxel intensities to ensure uniform data scaling.

2. Data Augmentation: Techniques like 3D rotation, flipping, and Gaussian noise addition were applied to enhance model robustness and generalizability.

3. Feature Engineering: Extracted regions of interest (ROI) focusing on the auditory cortex. Applied dimensionality reduction techniques like PCA to reduce computational overhead while retaining critical features.

4. Handling Imbalance: Applied SMOTE (Synthetic Minority Oversampling Technique) and ADASYN to address class imbalance in schizophrenia vs. healthy control samples.

The dataset was split into training, validation, and testing subsets, ensuring an 80-20 distribution.

3.4 Data Modeling

Two modeling approaches were employed:

1. Machine Learning Models (via Lazy Predict): Linear Discriminant Analysis, Random Forest, Logistic Regression, and more were evaluated. Lazy Predict provided quick benchmarking of multiple models.

2. Deep Learning Models: CNN (Extracted spatial patterns in 3D MRI data), Hybrid Models (Combined CNN with SMOTE or ADASYN for balanced model), referring to Dolz et al. (2017), Guo et al. (2023), Zheng et al. (2021)

Different model enhancement techniques were used to balance the data like SMOTE and ADASYN which boosted minority class representation, the Custom Loss Function which used focal loss to handle class imbalance effectively, cross validation which ensured reliability using 5-fold validation and hyperparameter Tuning where Grid search optimized hyperparameters for CNNs and traditional models were implemented.

3.5 Evaluation

To evaluate the models, different metrics were used, like accuracy, precision, recall, F1 score, and ROC AUC. Comparative analysis identified Random Forest and CNN-XGBoost as the top-performing models. Lazy Predict highlighted linear models for baseline comparisons, while CNN-based models excelled in learning complex spatial patterns.

3.6 Deployment

The deployment phase integrated the trained CNN model into a user friendly application where the best performing CNN model (trained on Google Colab) was saved as a '.pth' file using PyTorch. A Flask based web app was built which used the saved model which enables users to upload '.nii' MRI files for real time classification as either "Healthy" or "Schizophrenic" for Research Question 1 or "Hallucinating" or "Non Hallucinating" for Research Question 2. 1

Key features:

1. Grad-CAM Visualization: Highlights brain regions influencing predictions. Gao et al. (2024)

2. Insights Dashboard: Displays dataset statistics and a confusion matrix for performance analysis.

The MRI datasets were stored on Google Cloud Storage (GCS) and accessed directly during training and testing. Whereas the Flask app runs locally, leveraging the saved model for predictions with low latency and interpretability features. This deployment combines model accessibility, real-time insights, and visualization to bridge the model development with real world application.

4 Design Specification

This design specification describes the architecture, methodology, and deployment details for implementing the classification system for schizophrenia detection using MRI scans and deep learning techniques. The solution was developed using the CRISP DM (Cross Industry Standard Process for Data Mining) framework for a structured and repeatable approach.

4.1 System Architecture

1. The system architecture for this research incorporates multiple components for data handling, model training, evaluation, and deployment through a web application. The key components are:

2. Data Handling: MRI data in '.nii' format stored on Google Cloud Storage (GCS) for centralized access. Preprocessed and augmented datasets using techniques like ADASYN for class balancing.

3. Deep Learning Model: A Convolutional Neural Network (CNN) trained on augmented data to classify MRI scans into "Healthy" or "Schizophrenic".

4. Model Deployment: The trained CNN model was saved as a '.pth' file and integrated into a Flask based web application for local deployment. Real time predictions with Grad CAM visualizations for interpretability.

5. Visualization and Insights: The Flask app includes dataset statistics, prediction confidence, class distribution, and confusion matrix visualizations.

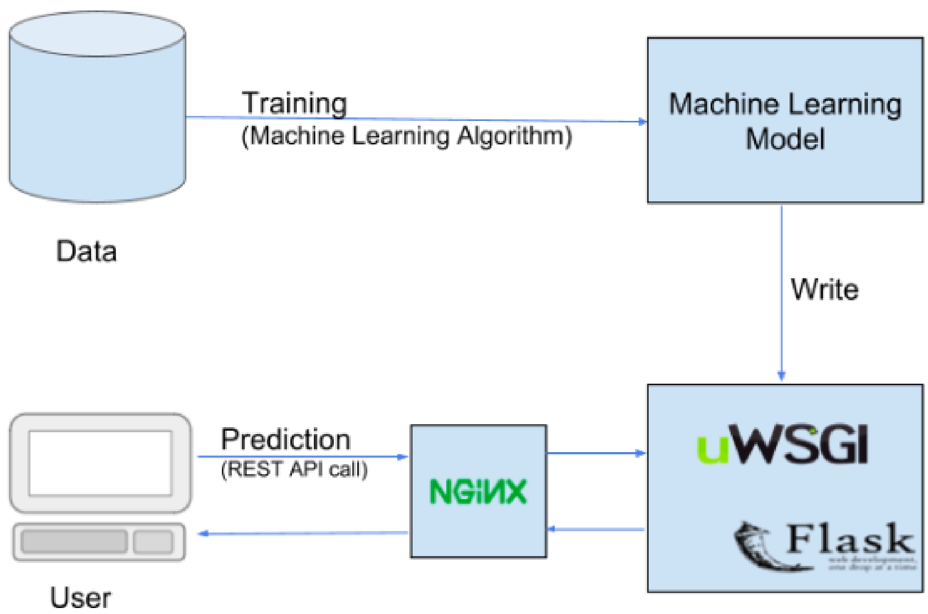

Figure 2: System Architechture KDnuggets (2019)

4.2 Techniques and Methodology

MRI scans were preprocessed by resizing the 3D volumes to ‘(96, 96, 96)‘ dimensions. Normalization was applied to standardize pixel intensity values. Since the data was imbalanced ADASYN (Adaptive Synthetic Sampling) was employed to address class imbalance by generating synthetic samples for minority classes. Later, a CNN architecture was designed with multiple 3D convolutional layers for spatial feature extraction. Batch normalization to stabilize learning. Max pooling for dimensionality reduction. Dropout layers to prevent overfitting. The model used a softmax activation function in the output layer for binary classification. The CNN model was then evaluated using metrics like accuracy, precision, recall, F1 score, and AUC ROC. A Grad CAM visualization was implemented to highlight brain regions contributing to model predictions.

4.3 Software and Hardware Specifications

1. Software: Python (version 10 and above) on Google Colab for training and local machine for deployment. Libraries used include Data Handling (‘nibabel‘, ‘numpy‘, ‘scipy‘), Modeling (‘PyTorch‘, ‘imblearn‘, ‘torchvision‘), Visualization (‘matplotlib‘, Grad CAM implementations), and Web Deployment (Flask).

2. Hardware: GPU enabled Google Colab, GCS Bucket to store datasets, local machine with moderate computational power, preferably a Windows OS, with at least 16 GB of RAM, 8 GB Graphics Memory, Intel i5/i7/i9 generation 10 and above (10 cores and above) or AMD Ryzen 5/7 (10 cores and above) processor.

4.4 Deployment Pipeline

1. Model Training: The CNN model was trained on Google Colab using the ADASYN augmented dataset. The model achieved high recall and accuracy for the minority

class (Schizophrenic patients). The trained CNN model was saved as a '.pth' file using PyTorch.

2. Flask Application: The Flask app enables real time predictions on user uploaded MRI scans. Users can upload '.nii' MRI files and the web app will give output "Healthy" or "Schizophrenic" with confidence scores. The app also give insights like Class distribution and percentage breakdown, Confusion matrix for model evaluation and Grad CAM heatmap for interpretability.

3. Hosting: Data is fetched from the GCS bucket during training, while the Flask app runs locally for predictions.

4.5 Design Advantages

1. Scalability: The use of Google Cloud Storage and GPU enabled training makes the design scalable for larger datasets.

2. Interpretability: Grad CAM heatmaps enhance model interpretability, supporting clinical validation of predictions.

3. User Centric Deployment: The Flask app simplifies MRI classification for non technical users, promoting real world usability.

4. Robustness: ADASYN augmentation improves the model's robustness to class imbalances.

This design demonstrates the effective use of deep learning and augmentation techniques to classify schizophrenia patients based on MRI scans, with a seamless integration into a user friendly web application for clinical utility.

5 Implementation

5.1 Data Preparation

MRI scans were stored in '.nii' and '.nii.gz' formats and hosted on Google Cloud Storage (GCS) for centralized access. The dataset included brain scans classified into two groups: "Non-Hallucinating" (Class 0) and "Hallucinating" (Class 1) for Research Question 1 and "Healthy" (Class 0) and "Schizophrenic" (Class 1) for Research Question 2. Files were downloaded and decompressed using a Python-based pipeline. MRI data was resized to a uniform shape of '(96, 96, 96)' using Scipy's 'zoom' function and intensity normalization was applied to standardize the pixel values. Imbalanced data was addressed using ADASYN (Adaptive Synthetic Sampling), generating synthetic samples for the minority class while preserving data variability. To handle imbalanced data augmentation was applied, and for it added more realistic augmentations like RandomRotation3D and GaussianNoise3D. Increased convolutional filter sizes and added more fully connected layers for better feature extraction and computed class weights dynamically based on the dataset to handle class imbalance better.

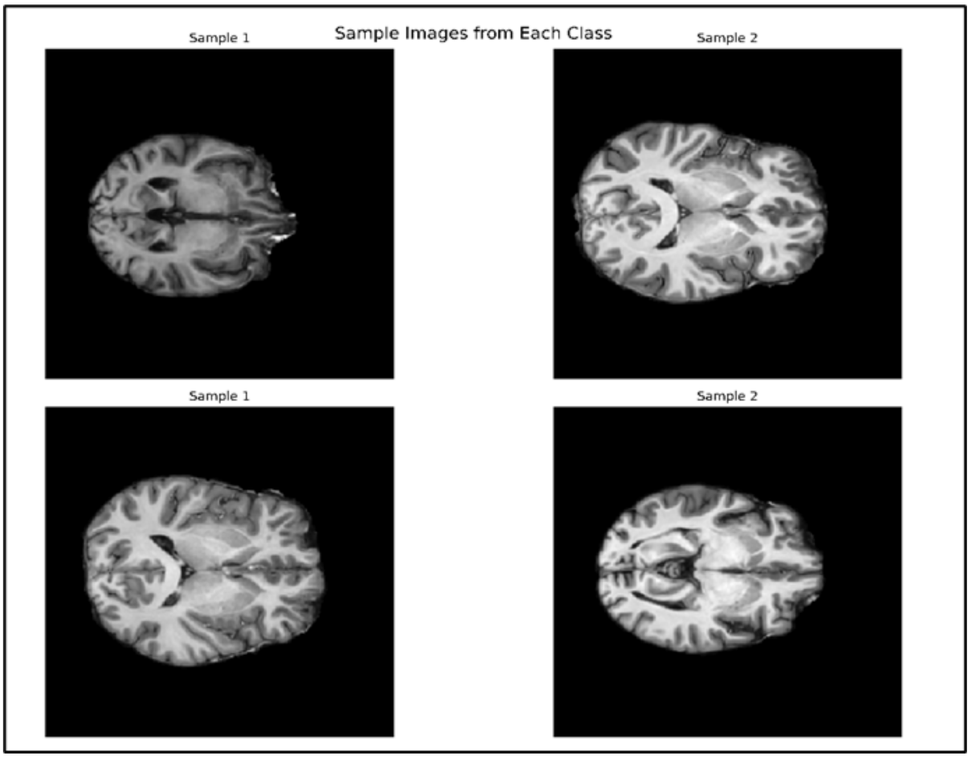

Figure 3: Sample Images of Class 0 (Non Schizophrenic) and class 1 (Schizophrenic)

5.2 Exploratory Data Analysis

Data inspection began by checking the shape and structure of the MRI scan data and labels to ensure they were loaded correctly. This provided an understanding of the dataset's dimensions. displayed sample images from each class (e.g., normal and affected brain scans) to visually inspect the data and observe any noticeable differences between the classes.

1. Missing Values Check: there were no missing values in the dataset, which could have impacted further analysis or model training.

2. Class Distribution Visualization: visualized the distribution of the classes (labels) using a bar chart (countplot). This helped determine if the dataset was balanced or imbalanced. There were 46 labels of class 1 (Hallucinating) and 25 labels of class 0 (Non-Hallucinating) for Dataset 1. And there were 104 labels of class 1 (Schizophrenic) and 66 labels of class 0 (Healthy) for Dataset 1 and 2 combined.

3. Basic Statistics Calculation: calculated the mean and standard deviation of pixel intensities for each class, it was observed that most images have a standard deviation in the range of 250–300 and that there were no significant changes to the standard deviation distribution after resampling.

4. Histogram of Pixel Intensities: plotted histograms of pixel intensities for each class, which helped explore how the pixel values were distributed across different brain scan images.

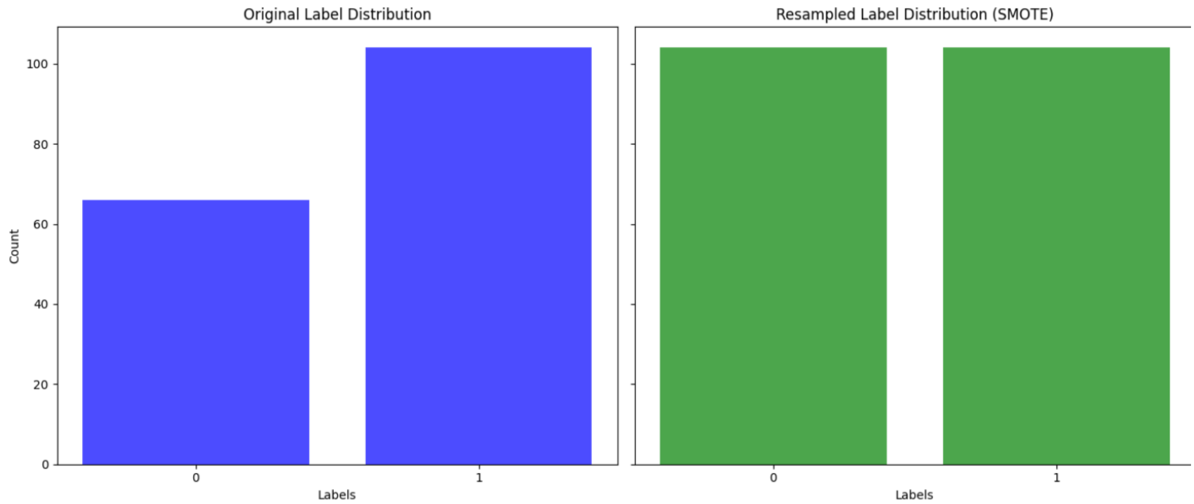

Figure 4: Class Distribution (Dataset 1 and Dataset 2)

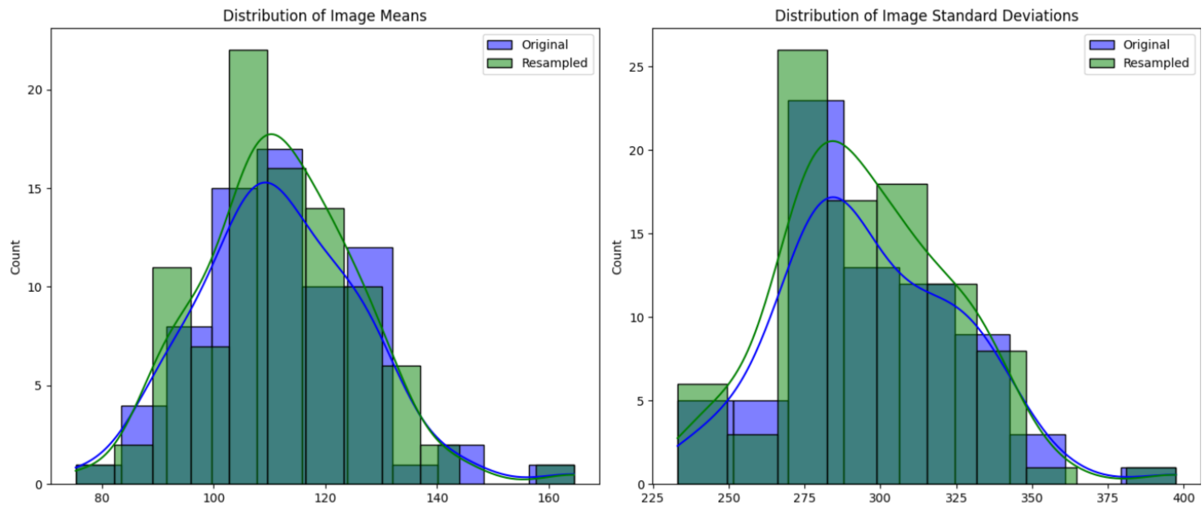

Figure 5: Distribution of Mean and Standard Deviation

5.3 Model Training

A 3D Convolutional Neural Network (CNN) was designed with Three convolutional layers with batch normalization and max-pooling, dropout layers to reduce overfitting and a fully connected layer for binary classification. Training was conducted on Google Colab, leveraging GPU acceleration for faster computation. Optimizer Adam was used with a learning rate of '0.001'. Focal Loss was used to handle class imbalance effectively and validation used an 80-20 train-test split. Apart from that metrics such as Accuracy, Precision, Recall, and F1- Score were used and Grad-CAM visualizations were generated to interpret the CNN model by highlighting brain regions influencing predictions. Hyperparameter tuning was conducted to optimize model performance in classifying individuals as schizophrenic or healthy. For deep learning models, the Convolutional Neural Network (CNN) architecture was tuned by experimenting with the number of convolutional layers, filter sizes, and dropout rates to balance learning capacity and overfitting. Batch size (16, 32) and learning rates (0.001, 0.0005) were systematically adjusted to improve convergence during training. The hyperparameter optimization focused on maximizing accuracy and recall to address class imbalance.

5.4 Deployment

The trained CNN model was saved as a '.pth' file using PyTorch. This model was then used in a Flask web application developed to enable real-time MRI classification. Key features include:

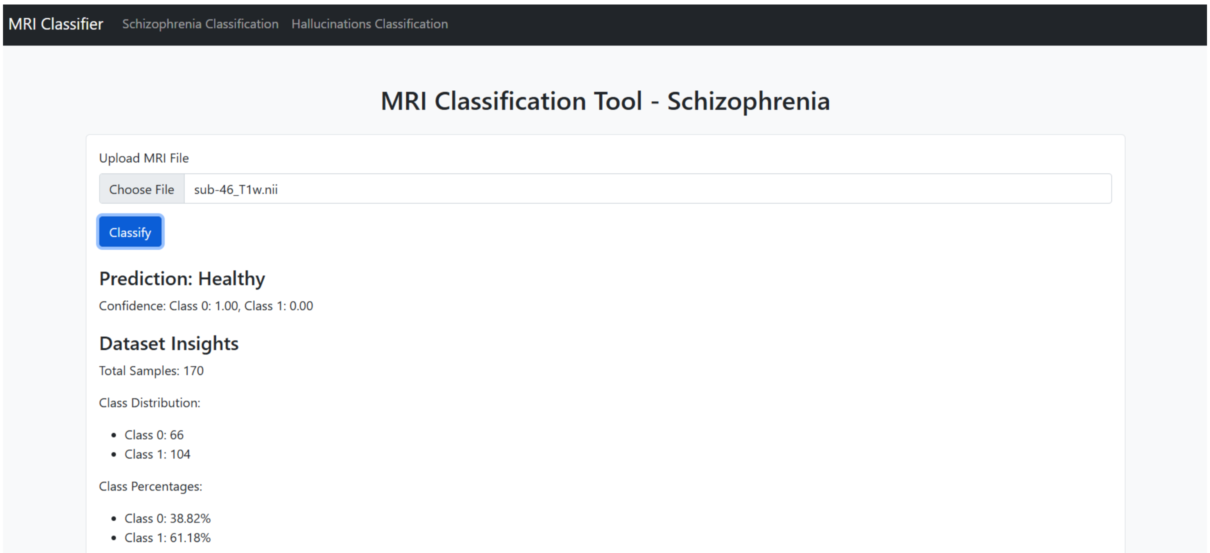

Figure 6: Locally Hosted Flask Based Web App for Classification 1

1. File Upload: Users can upload '.nii' files for predic- tion. Prediction Output: Displays classification ("Healthy" or "Schizophrenic") with confidence scores.
2. Visualization: Grad-CAM highlights the brain regions contributing to the decision.
3. Insights Dashboard: Shows dataset statistics, class distribution, and confusion matrix.

4. Integration: The saved model was loaded in the Flask app for inference. The app was hosted locally, making it accessible via a web browser.

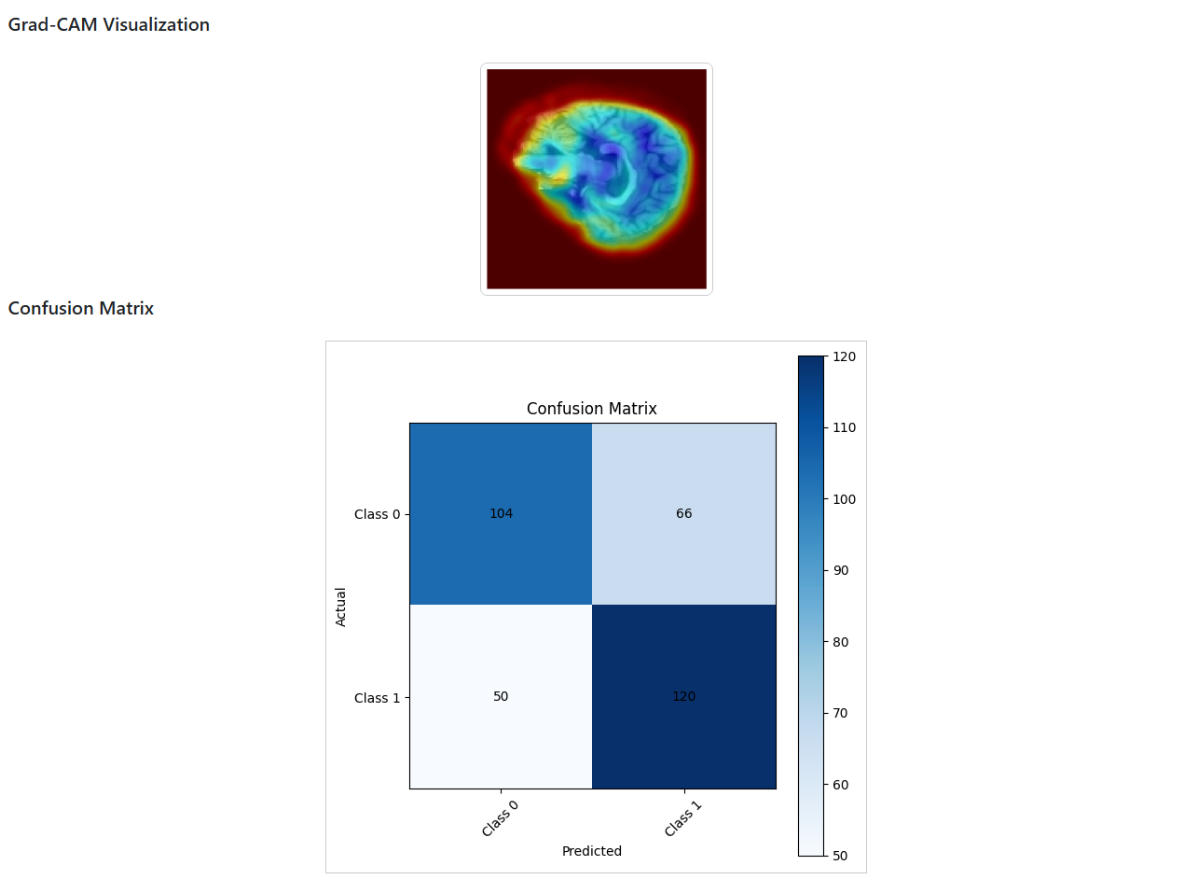

Figure 7: Locally Hosted Flask Based Web App for Classification 2

5.5 Tools & Technologies

1. Frameworks and Libraries: PyTorch (Model development and training), Flask (Backend framework for the web application), Scipy and Nibabel (Preprocessing MRI data), Matplotlib (Visualization of confusion matrices and Grad-CAMs).

2. Cloud Resources: Google Cloud Storage for hosting MRI datasets for seamless access during training and Google Collab for Computing Resources.

5.6 Challenges & Solutions

The biggest challenge was the imbalanced class distribution which led to biased predictions. the solution to it was to use ADASYN to oversample the minority class dynamically. Second challenge was Large file sizes of MRI images and memory constraints. The solution to it was MRI files were resized and normalized to optimize memory usage. Third challenge was interpretability of CNN predictions. The solution to it was integrating Grad-CAM for visual explanations of the classification. Thus, this implementation effectively combined robust machine learning techniques with an accessible deployment platform, enabling practical applications of schizophrenia classification using MRI data.

6 Evaluation

Different models were applied on the datasets in consideration for both Research Question 1 and 2. Different metrices were used to evaluate the model's performance. The different models and their performance is discussed as follows:

6.1 Experiment1: Baseline Classifier Results

A baseline evaluation was performed using LazyPredict, which compared various traditional machine learning classifiers such as Logistic Regression, Light GBM Classifier, Linear SVC, etc.

1. Research Question 1: Predict if a schizophrenic patient hallucinates (Dataset 1)

   For research question 1, MRI scans labeled for hallucination classification, LightGBMClassifier achieved the highest accuracy 73%, Quadratic Discriminant Analysis, XGBoostClassifier, and DecisionTreeClassifier followed with 67% accuracy. Ensemble and boosting methods like AdaBoostClassifier, BaggingClassifier, and LabelPropagation demonstrated competitive results.

2. Research Question 2: Predict if a person is schizophrenic (Dataset 1 + Dataset 2).

   For research question 2, LinearDiscriminantAnalysis achieved the highest accuracy 68%, LinearSVC, PassiveAggressiveClassifier, and LogisticRegression followed with 62% accuracy. Models like RidgeClassifier, Perceptron, and ExtraTreesClassifier also demonstrated consistent performance.

6.2 Experiment 2: Model Training and Testing with Hyperparameter Tuning

The CNN model outperformed traditional classifiers in terms of accuracy and interpretability due to its ability to capture spatial and structural features in 3D MRI data. Evaluation for Research Question 1 1, The research question evaluates the ability to predict whether a schizophrenic patient hallucinates or not based on MRI data using two models (Dataset 1) OpenNeuro (2022)

1. The CNN Baseline achieved an accuracy of 32%. Precision, recall, and F1-score for Class 0 (Non-Hallucinating) are 0.00, indicating no correct predictions. For Class 1 (Hallucinating), the model has a recall of 1.00 but suffers from low precision (0.62), which results in overfitting toward the majority class. The model shows clear evidence of imbalance, struggling to generalize for both classes effectively.

2. In the CNN with ADASYN model accuracy improved to 65% after addressing the class imbalance using the ADASYN technique. For Class 0 (Non-Hallucinating) Precision increased to 0.69, recall improved to 0.75, and the F1-score is 0.72. For Class 1 (Hallucinating) Precision is 0.57, recall is 0.5, and the F1-score is 0.53. The model performs significantly better in balancing predictions for both classes, addressing the imbalance issue effectively.

3. The CNN baseline model suffers from severe class imbalance, with the model predicting the majority class (hallucinating patients) almost exclusively. ADASYN successfully mitigated the imbalance, leading to substantial improvements in accuracy and balanced classification performance.

Model	**Accuracy**	**Recall (H)**	**F1-Score (H)**	**Recall (S)**	**F1-Score(S)**	**Precision (H)**	**Precision (S)**
CNN Baseline	32%	0	0	1	0.77	0	0.62
CNN + ADASYN	65%	0.75	0.72	0.5	0.53	0.69	0.57

Table 1: Comparison of Model Performance Metrics (H - Healthy, S - Schizophrenic)

Evaluation for Research Question 2 1, Predicting if a Person is Schizophrenic or Not (Dataset 1 + Dataset 2) OpenNeuro (2022) OpenNeuro (2018)

1. For this research question, CNN models were implemented to classify individuals into schizophrenic and healthy categories based on MRI scans. Below are the key evaluation results for the applied models:
2. For the Baseline CNN (Without Balancing) the accuracy attained was 32%. The model performed relatively well in identifying the majority class (healthy individuals). Minority class (schizophrenic individuals) recall and precision were significantly low due to class imbalance.
3. For the CNN + SMOTE (Synthetic Minority Oversampling Technique) model the accuracy attained was 57%. SMOTE improved the recall for the minority class (schizophrenic individuals), ensuring better sensitivity. Overall accuracy improved due to oversampling artifacts, such as synthetic samples representing real data distributions well. The model showed improved balance in performance across classes but suffered from reduced precision.
4. For the CNN + ADASYN (Adaptive Synthetic Sampling) model the accuracy attained was 65%. ADASYN provided better balance between precision and recall for the minority class compared to SMOTE. Improved minority class predictions without as significant a drop in overall accuracy as seen with SMOTE. Demonstrated a slight edge in class balance, making it more suitable for imbalanced data in this context.
5. The experiments revealed that while baseline CNN achieved lowest accuracy, it failed to address the imbalance between schizophrenic and healthy classifications. Applying SMOTE and ADASYN helped improve minority class recall but reduced overall accuracy. Among the balancing techniques, ADASYN offered a better trade-off, indicating its potential for handling imbalanced MRI data in schizophrenia diagnosis.

Model	**Accuracy**	**F1 Score (S)**	**Recall (S)**	**F1 Score (H)**	**Recall (H)**	**Precision (S)**	**Precision (H)**
Baseline CNN	32%	0.77	1	0	0	0.62	0
CNN + SMOTE	57%	0.57	0.48	0.57	0.71	0.71	0.48
CNN + ADASYN	65%	0.53	0.5	0.72	0.75	0.57	0.69

Table 2: Performance Metrics of Various CNN Models (H - Healthy, S - Schizophrenic)

6.3 Explainable AI (XAI)

Grad-CAM Overview

Explainable AI (XAI) techniques aim to make complex models interpretable by highlighting the most influential regions of the input data contributing to predictions. Grad-CAM (Gradient-weighted Class Activation Mapping) is utilized in this research to identify significant areas in MRI scans that influence model decisions. Grad-CAM generates heatmaps to visualize which parts of Broca's area and other related regions are activated during classification.

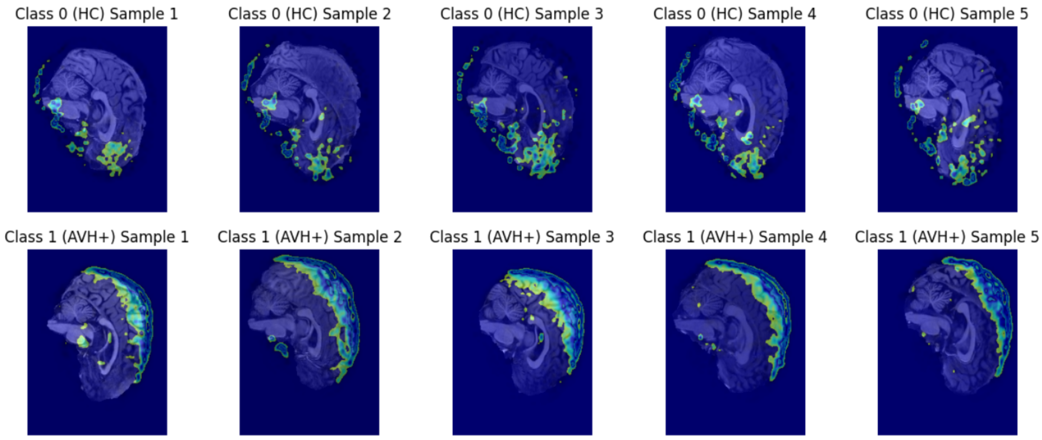

Figure 8: CNN Gradcam Images of Class 0 (Non Schizophrenic) and class 1 (Schizophrenic)

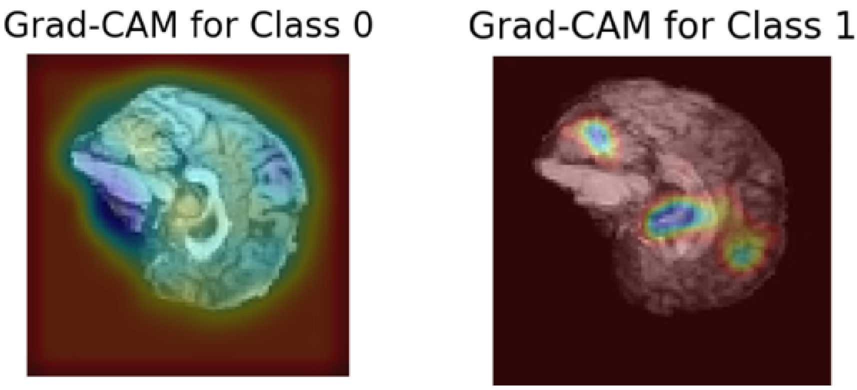

Figure 9: CNN + ADASYN Gradcam Images of Class 0 (Non Schizophrenic) and class 1 (Schizophrenic)

1. Baseline CNN: The Grad-CAM visualization for the baseline CNN model highlights key regions of the Broca's area but lacks focus due to the class imbalance. Predic-

tions for schizophrenic patients with and without hallucinations reveal overlapping regions, reducing the model's reliability for precise classification.

2. CNN + SMOTE: The Grad-CAM heatmaps for the CNN model with SMOTE show a noticeable improvement in distinguishing activated areas. However, regions outside the Broca's area were often activated, possibly due to noise introduced during oversampling.

3. CNN + ADASYN: The CNN + ADASYN model provides the most focused and interpretable Grad-CAM heatmaps. These visualizations show distinct activations within Broca's area, correlating well with schizophrenia-related auditory hallucination patterns. The technique successfully highlights unique patterns for hallucinating and non-hallucinating schizophrenic patients.

6.4 Discussion

The findings from the experiments for both research questions reveal critical insights and limitations. For Research Question 1 (predicting whether a schizophrenic patient hallucinates), the CNN baseline model suffered from severe class imbalance, achieving a low accuracy of 32% and poor performance for non-hallucinating patients. Applying ADASYN to address the imbalance significantly improved results, with an accuracy of 65% and balanced F1-scores for both classes. While ADASYN improved classification performance, the model's reliance on oversampling could still introduce noise into the dataset, potentially affecting generalizability. A better approach might include integrating hybrid resampling techniques like SMOTE-ENN or employing cost-sensitive learning to further improve classification balance.

For Research Question 2 (predicting schizophrenia vs. healthy individuals), Lazy Predict identified models such as LightGBM (73% accuracy) and Quadratic Discriminant Analysis (68% accuracy) as the most effective. However, these models rely heavily on hyperparameter tuning and require additional validation to ensure consistency across unseen datasets. The CNN models, even with resampling, did not surpass 62% accuracy, indicating that more sophisticated architectures or transfer learning techniques might be necessary to capture complex MRI features.

In both cases, the designs effectively utilized modern approaches but could be improved by incorporating multimodal data, advanced deep learning techniques (e.g., 3D CNNs or attention mechanisms), and robust validation protocols. Comparisons with prior research underscore the need for more nuanced models to enhance diagnostic reliability.

7 Conclusion and Future Work

This research addressed two key questions: identifying schizophrenia through MRI scans and distinguishing hallucinating from non-hallucinating patients 1. The objectives included applying deep and machine learning models, improving class balance, and achieving reliable classifications. Experiments utilized CNNs, Lazy Predict, SMOTE, and ADASYN. While the CNN achieved 48% accuracy initially, ADASYN improved it to 65% for hallucination prediction that is (Research Question 1), demonstrating the efficacy of resampling. For schizophrenia detection (Research Question 2), CNN with ADASYN achieved 57% accuracy, whereas LightGBM with (73% accuracy) outperformed CNN models, emphasizing the strength of ensemble techniques. However, limitations include reliance on oversampling, limited dataset size, and absence of multimodal features. The findings provide insights into schizophrenia diagnostics, but the efficacy is constrained by model complexity and dataset quality. Future work could incorporate multimodal data (EEG, fMRI) to enhance predictive power, explore transfer learning for MRI-specific feature extraction, and investigate longitudinal studies for robust generalization. These advancements could pave the way for clinical integration or commercial applications in mental health diagnostics.

References

Bracher-Smith, M., Rees, E., Menzies, G., Walters, J. T., O'Donovan, M. C., Owen, M. J., Kirov, G. and Escott-Price, V. (2022). Machine learning for prediction of schizophrenia using genetic and demographic factors in the uk biobank, *Schizophrenia Research* **246**: 156–164.

contributors, W. (2024). Cross-industry standard process for data mining, `https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining`. Accessed: 2024-12-08.

Di Camillo, F., Grimaldi, D. A., Cattarinussi, G., Di Giorgio, A., Locatelli, C., Khuntia, A., Enrico, P., Brambilla, P., Koutsouleris, N. and Sambataro, F. (2024). Magnetic resonance imaging–based machine learning classification of schizophrenia spectrum disorders: A meta-analysis, *Psychiatry and Clinical Neurosciences* .

Dolz, J., Desrosiers, C. and Ayed, I. B. (2017). 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study, *NeuroImage* **170**: 456–470.

Gao, J., Qian, M., Wang, Z., Li, Y., Luo, N., Xie, S., Shi, W., Li, P., Chen, J., Chen, Y., Wang, H., Liu, W., Li, Z., Yang, Y., Guo, H., Wan, P., Lv, L., Lu, L., Yan, J. and Jiang, T. (2024). Exploring schizophrenia classification through multimodal mri and deep graph neural networks: Unveiling brain region-specific weight discrepancies and their association with cell-type specific transcriptomic features, *Schizophrenia Bulletin* .

Guo, H., Jian, S., Zhou, Y., Chen, X., Chen, J., Zhou, J., Huang, Y., Ma, G., Li, X., Ning, Y., Wu, F. and Wu, K. (2023). Discriminative analysis of schizophrenia patients using an integrated model combining 3d cnn with 2d cnn: A multimodal mr image and connectomics analysis, *Brain Research Bulletin* **206**: 110846.

KDnuggets (2019). Easily deploy machine learning models using flask. Available at `https://www.kdnuggets.com/2019/10/easily-deploy-machine-learning-models-using-flask.html`.

Lin, Q., Niu, Y., Sui, J., Zhao, W., Zhuo, C. and Calhoun, V. D. (2022). Sspnet: An interpretable 3d-cnn for classification of schizophrenia using phase maps of resting-state complex-valued fmri data, *Medical Image Analysis* **79**: 102430.

Montazeri, M., Montazeri, M., Bahaadinbeigy, K., Montazeri, M. and Afraz, A. (2022). Application of machine learning methods in predicting schizophrenia and bipolar disorders: A systematic review, *Health Science Reports* **6**(1).

OpenNeuro (2018). Working memory in healthy and schizophrenic individuals. Dataset available at `https://openneuro.org/datasets/ds000115/versions/00001`.

OpenNeuro (2022). Brain correlates of speech perception in schizophrenia patients with and without auditory hallucinations. Dataset available at `https://openneuro.org/datasets/ds004302/versions/1.0.1`.

Rozycki, M., Satterthwaite, T. D., Koutsouleris, N., Erus, G., Doshi, J., Wolf, D. H., Fan, Y., Gur, R. E., Gur, R. C., Meisenzahl, E. M., Zhuo, C., Yin, H., Yan, H., Yue, W., Zhang, D. and Davatzikos, C. (2017). Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals, *Schizophrenia Bulletin* **44**(5): 1035–1044.

Shen, M., Wen, P., Song, B. and Li, Y. (2023). 3d convolutional neural network for schizophrenia detection using an eeg-based functional brain network, *Biomedical Signal Processing and Control* **89**: 105815.

Shoeibi, A., Sadeghi, D., Moridian, P., Ghassemi, N., Heras, J., Alizadehsani, R., Khadem, A., Kong, Y., Nahavandi, S., Zhang, Y. and Gorriz, J. M. (2021). Automatic diagnosis of schizophrenia in eeg signals using cnn-lstm models, *Frontiers in Neuroinformatics* **15**.

Soler-Vidal, J., Fuentes-Claramonte, P., Salgado-Pineda, P., Ramiro, N., García-León, M. , Torres, M. L., Arévalo, A., Guerrero-Pedraza, A., Munuera, J., Sarró, S., Salvador, R., Hinzen, W., McKenna, P. and Pomarol-Clotet, E. (2022). Brain correlates of speech perception in schizophrenia patients with and without auditory hallucinations, *PLoS ONE* **17**(12): e0276975.

Yin, G., Chang, Y., Zhao, Y., Liu, C., Yin, M., Fu, Y., Shi, D., Wang, L., Jin, L., Huang, J., Li, D., Niu, Y., Wang, B. and Tan, S. (2023). Automatic recognition of schizophrenia from brain-network features using graph convolutional neural network, *Asian Journal of Psychiatry* **87**: 103687.

Zhang, J., Rao, V. M., Tian, Y., Yang, Y., Acosta, N., Wan, Z., Lee, P., Zhang, C., Kegeles, L. S., Small, S. A. and Guo, J. (2023). Detecting schizophrenia with 3d structural brain mri using deep learning, *Scientific Reports* **13**(1).

Zheng, J., Wei, X., Wang, J., Lin, H., Pan, H. and Shi, Y. (2021). Diagnosis of schizophrenia based on deep learning using fmri, *Computational and Mathematical Methods in Medicine* **2021**: 1–7.