MULTIMODAL APPROACH FOR ALZHEIMER'S DIAGNOSIS AND CLASSIFICATION

A PROJECT REPORT

Submitted by

AARTHI V (2022503537)

BHAVANA B (2022503305)

VEDALAKSHMI K (2022503301)

In partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

MADRAS INSTITUTE OF TECHNOLOGY CAMPUS

ANNA UNIVERSITY, CHENNAI 600 044

APRIL 2025

DEPARTMENT OF COMPUTER TECHNOLOGY

ANNA UNIVERSITY, MIT CAMPUS

CHROMEPET, CHENNAI 600044

BONAFIDE CERTIFICATE

Certified that this project report "**Multimodal Approach for Alzheimer's Diagnosis and Classification**" is the work of **Ms. Aarthi V (2022503537), Ms. Bhavana B (2022503305), Ms. Vedalakshmi K (2022503301)** in the Creative and Innovative Project Laboratory subject code CS6611 during the period January to April 2025.

SIGNATURE

Dr. Kottilingam K
SUPERVISOR
Assistant Professor
Department of Computer Technology
Anna University, MIT Campus
Chromepet 600 044.

SIGNATURE

Dr. Jayashree P
HEAD OF THE DEPARTMENT
Professor
Department of Computer Technology
Anna University, MIT Campus
Chromepet 600 044.

ABSTRACT

Alzheimer's Disease (AD) is the most common form of dementia and poses significant challenges for early detection due to its progressive and multifaceted nature. Traditional diagnostic approaches often rely on single-modality neuroimaging, which may not fully capture the complex structural and functional changes occurring in the brain. This project proposes a multimodal deep learning framework that integrates Magnetic Resonance Imaging (MRI), Diffusion Tensor Imaging (DTI), and Positron Emission Tomography (PET) data to enhance the accuracy of AD diagnosis.

The system is trained on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and aims to classify subjects into Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) categories. The proposed model includes several key components: data preprocessing tailored to each imaging type, image augmentation to increase training diversity, and feature extraction through Convolutional Neural Networks (CNN). By combining features from different imaging sources, the model captures complementary information that improves the differentiation between CN, MCI, and AD subjects.

Experimental results demonstrate that the proposed multimodal model achieved an accuracy of 98.09%, outperforming several unimodal models cited in the literature. The findings emphasize the potential of multimodal deep learning systems to provide more reliable and comprehensive support for Alzheimer's Disease diagnosis in clinical settings.

ACKNOWLEDGEMENT

We take this humble opportunity to thank the Dean, MIT Campus, Anna University, Dr. Ravichandran K, and Dr. Jayashree P, Professor & Head, Department of Computer Technology, MIT Campus, Anna University for providing all the lab facilities in pursuit of this project.

Undertaking this project has helped us learn a lot, and we would like to express our gratitude towards our supervisor Dr. Kottilingam K, Assistant Professor, Department of Computer Technology, MIT, Anna University, whose guidance, and directions helped shape this project perfectly.

We acknowledge the efforts and feedback of the panel members Dr.Muthurajkumar S, Associate Professor, Dr. Pabitha P, Associate Professor, Dr. Ponsy R.K. Sathia Bhama, Associate Professor, and Dr. Kathiroli R, Assistant Professor, Dr. Sudhakar T, Assistant Professor Department of Computer Technology, MIT, Anna University, in reviewing our work, providing constant valuable comments and encouraging us to view the different aspects of the project in successful implementation of the project.

We thank the teaching and non-teaching staff of the Department of Computer Technology, MIT, Anna University, for their invaluable support. We also extend our heartfelt gratitude to our friends and family for their love, patience, and encouragement, which kept us motivated and strong throughout this journey. This achievement is as much theirs as it is ours.

Thank you.

Aarthi V (2022503537)

Bhavana B (2022503305)

Vedalakshmi K (2022503301)

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADNI	-	Alzheimer's Disease Neuroimaging Initiative
AI	-	Artificial Intelligence
CNN	-	Convolutional Neural Network
DTI	-	Diffusion Tensor Imaging
DNN	-	Deep Neural Network
fMRI	-	Functional Magnetic Resonance Imaging
ML	-	Machine Learning
MRI	-	Magnetic Resonance Imaging
PET	-	Positron Emission Tomography
ReLU	-	Rectified Linear Unit
ROI	-	Region of Interest

CHAPTER 1

INTRODUCTION

1.1 DEEP LEARNING

Deep learning is a specialized subfield of machine learning that focuses on algorithms inspired by the structure and function of the human brain, known as artificial neural networks [7]. It enables models to automatically learn hierarchical feature representations from raw data, making it particularly powerful for complex tasks such as image recognition, natural language processing, and medical diagnosis. Unlike traditional machine learning methods, which often require manual feature extraction, deep learning algorithms are capable of learning features directly from data through multiple layers of abstraction [9].

At the heart of deep learning are deep neural networks (DNNs), which consist of input, hidden, and output layers. Each layer contains numerous interconnected nodes (neurons) that process and transform data through mathematical operations and activation functions. These networks can range from a few layers to hundreds, allowing them to model highly non-linear relationships and capture intricate patterns in large datasets. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers are common architectures used in various applications depending on the task and data type.

Deep learning thrives on large amounts of labeled data and computational power, often utilizing GPUs for efficient training. As more data becomes available, deep learning models continue to improve in performance, making them suitable for tasks like medical image analysis, autonomous

driving, voice assistants, and robotics. In healthcare, particularly, deep learning has shown great promise in diagnosing diseases from imaging modalities such as MRI, PET, and DTI, as it excels at recognizing subtle and complex patterns that may not be easily captured by traditional algorithms [7].

1.2 MULTIMODAL IMAGES

In the context of Alzheimer's diagnosis, multimodal imaging refers to the combination of different brain imaging modalities such as Magnetic Resonance Imaging , Diffusion Tensor Imaging and Positron Emission Tomography to obtain a more comprehensive understanding of the brain's structure, function, and pathology. Each modality captures unique and complementary information. MRI provides high-resolution structural details of brain tissue, DTI maps white matter integrity and connectivity, and PET reveals metabolic and molecular activity, such as amyloid plaque accumulation, which is a hallmark of Alzheimer's Disease.

Using these modalities together allows for a deeper and more holistic analysis, especially in early detection of Alzheimer's Disease where subtle changes may be missed by single-modality analysis. Multimodal images help improve classification accuracy by providing richer feature sets for machine and deep learning models, enhancing their ability to distinguish between Cognitively Normal individuals, those with Mild Cognitive Impairment and patients with Alzheimer's Disease .

In this project, multimodal data is preprocessed and aligned to ensure compatibility across modalities. A deep learning-based approach is used to extract relevant features from each image type and fuse them effectively, allowing the model to learn more discriminative patterns for accurate classification. This integration of multimodal imaging data plays a crucial role in developing a robust and intelligent diagnostic tool for Alzheimer's Disease.

The typical workflow for multimodal neuroimaging in Alzheimer's diagnosis is shown in Figure 1.1.

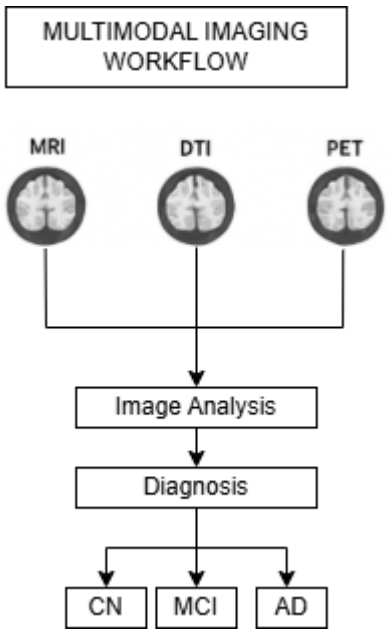

Figure 1.1 Multimodal Imaging Workflow

1.2.1 Advantages of Multimodal Images

i. **Improved diagnostic accuracy**: Multimodal imaging combines data from multiple imaging techniques (such as MRI, DTI, and PET), providing a more comprehensive view of the brain. This leads to better differentiation between Alzheimer's and other neurodegenerative disorders compared to single-modal imaging.

ii. **Enhanced tissue and brain structure visualization**: MRI offers detailed anatomical images, DTI highlights white matter integrity, and PET provides metabolic and functional information. Combining these modalities allows for a more accurate depiction of brain structures and function, essential for assessing Alzheimer's progression.

iii. **Comprehensive disease monitoring**: Multimodal imaging enables tracking both structural and functional changes in the brain over time.

This is crucial for evaluating disease progression, response to treatment, and identifying early signs of Alzheimer's, which may not be visible with a single imaging technique.

iv. **Improved biomarkers for early diagnosis**: By integrating data from different imaging techniques, multimodal images help identify early biomarkers of Alzheimer's, such as amyloid plaques, tau tangles, and neuronal atrophy, that may not be detectable with just one modality.

v. **Enhanced understanding of disease mechanisms**: The combination of modalities provides a multi-dimensional approach to understanding the complex pathological processes underlying Alzheimer's, from changes in brain metabolism to disruptions in white matter connectivity.

vi. **Versatility in clinical applications**: Multi-modal imaging supports a wide range of clinical applications, including early diagnosis, longitudinal studies, therapeutic monitoring, and precision medicine, by offering a more detailed and reliable understanding of Alzheimer's disease.

1.2.2 Properties of Multimodal Images

Multimodal imaging involves the integration of multiple imaging techniques to capture a comprehensive view of the subject, providing more detailed information than any single modality alone. In the context of brain imaging, commonly used modalities include MRI, DTI, and PET, each offering distinct information.

- Magnetic Resonance Imaging provides high-resolution anatomical images of the brain, helping to map brain structure, detect a trophy, and identify structural changes associated with Alzheimer's disease.

- Diffusion Tensor Imaging is a specialized MRI technique that visualizes the integrity of white matter fibers by tracking water molecule movement. It is particularly useful for detecting disruptions in brain connectivity, which are prominent in Alzheimer's.
- Positron Emission Tomography measures brain activity by detecting the distribution of radio labelled tracers that highlight regions of metabolic activity. PET can reveal the accumulation of amyloid plaques or tau tangles, which are hallmarks of Alzheimer's pathology.

Multimodal imaging systems typically integrate data from at least two or three different imaging modalities, enabling the simultaneous visualization of anatomical, structural, and functional changes within the brain. This fusion of data provides a more complete understanding of the disease process. For example, combining MRI and PET allows for the identification of both brain structure abnormalities and functional disruptions, leading to earlier diagnosis and better treatment monitoring.

The multimodal approach allows clinicians and researchers to visualize brain activity, connectivity, and tissue integrity simultaneously, which aids in tracking disease progression and understanding underlying mechanisms. This enhanced capability supports applications such as early detection, individualized treatment planning, and longitudinal studies of Alzheimer's disease that would not be possible with a single imaging modality.

1.3 IMAGE PROCESSING

1) **Preprocessing of Multimodal Data:** Preprocessing is a critical step in handling multimodal imaging data for Alzheimer’s diagnosis. Each modality (MRI, DTI, PET) provides distinct information about the brain, and preprocessing ensures that these data can be analyzed together effectively. The preprocessing steps involve:

a. **Spatial alignment** to ensure that all images from different modalities are correctly positioned in the same anatomical space.
b. **Resampling** to standardize the resolution and dimensions across modalities.
c. **Intensity normalization** to bring the pixel values from different imaging techniques to a comparable scale, ensuring that each modality contributes appropriately to the analysis.

These preprocessing steps allow for a more accurate and consistent comparison across MRI, DTI, and PET data, which is crucial for detecting changes related to Alzheimer's disease.

2) **Noise Reduction and Artifact Removal:** Raw multimodal images are often affected by noise and artifacts, which can interfere with accurate analysis. MRI scans may suffer from noise due to magnetic field variations, while PET scans may show artifacts due to photon attenuation or motion during the scan.
3) **Intensity Normalization:** Intensity normalization ensures that pixel values across different modalities are comparable, which is crucial when combining data from MRI, DTI, and PET. Since each modality has different intensity scales (e.g., MRI uses a scale of 0-1000, PET uses radiolabeled tracers), normalization adjusts the values to a consistent range, typically between 0 and 1.
4) **Region of Interest Extraction and Alignment:** To focus analysis on regions most affected by Alzheimer's, such as the hippocampus and cortical structures, ROI masks were applied to each modality. Additionally, images were spatially registered to a common anatomical space using affine transformations to ensure that features across modalities aligned correctly for fusion.

1.4 OBJECTIVES

- To develop a deep learning-based multimodal framework that integrates MRI, DTI, and PET imaging data for the diagnosis and classification of Alzheimer's Disease.
- To design and train convolutional neural network (CNN) models for each imaging modality to extract discriminative features.
- To implement an effective feature fusion strategy to combine modality-specific features and improve classification performance.
- To classify subjects into three categories — Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD).
- To evaluate the proposed system using performance metrics such as accuracy, precision, recall, F1-score and compare it with unimodal approaches.
- To demonstrate the effectiveness of multimodal integration in enhancing early diagnosis and supporting clinical decision-making.

CHAPTER 2

LITERATURE SURVEY

An, X. et al. [1] introduced a dynamic spatiotemporal graph neural network that utilizes rs-fMRI to identify MCI. Their method captures spatial and temporal information from brain connectivity patterns, achieving superior classification accuracy. However, it does not address PET or structural imaging integration. Our project expands on this by combining spatial, structural, and metabolic information from multiple modalities.

Faisal, Fazal Ur Rehman, and Goo-Rak Kwon [2] developed a whole-brain MRI-based model for automated detection of Alzheimer's and MCI. While they achieve strong results with MRI, they do not explore multimodal fusion or compare performance across modalities. Our approach fills this gap by integrating and evaluating complementary imaging types.

Givian, Helia and Jean-Paul Calbimonte [3] provided a detailed review of machine learning methods applied to brain MRI for Alzheimer's diagnosis. Their survey categorizes various ML techniques and highlights the need for integrating multiple imaging biomarkers. Their work motivates our fusion-based approach by outlining the limitations of unimodal models.

Habuza, Tetiana et al. [4] investigated structural MRI and cognitive tests to model deviations from normal aging. Their CNN-based regression framework predicts cognitive scores from MRI, identifying early signs of dementia. Though promising, their method does not include PET or DTI inputs, which our project leverages for enhanced accuracy.

J Bai et al. [5] proposed LGG-NeXt, a hybrid CNN-Transformer model for Alzheimer's diagnosis using 2D structural MRI slices. Their model extracts local and global features through specialized network blocks, achieving high performance. However, it requires complex computation and lacks multimodal adaptability. Our work overcomes this by employing modality-specific networks and a fusion strategy to balance performance and efficiency.

Jabason, Emimal et al. [6] designed a lightweight deep CNN that captures both local and global contextual features from structural MRI for Alzheimer's classification. Their method reduces model complexity while preserving accuracy. However, their approach focuses only on MRI and binary classification tasks. In contrast, our model supports multiclass classification and fuses multiple modalities.

Murugan, Suriya et al. [7] presented DEMNET, a CNN model tailored to classify multiple stages of dementia using MRI. They addressed class imbalance and generated high-resolution disease probability maps. While effective, DEMNET is limited to MRI and lacks integration with other modalities. Our research extends this work by applying a multimodal approach that includes PET and DTI imaging.

Nan, Fengtao et al. [8] introduced a reproducible evaluation framework for multimodal learning in Alzheimer's diagnosis, emphasizing the comparative utility of various modality combinations. They demonstrate performance improvements when using SNP data alongside imaging, but do not include DTI or PET. Our work builds upon this by incorporating MRI, DTI, and PET data, offering a more comprehensive diagnostic approach.

Ning, Zhenyuan et al. [9] proposed a relation-induced multimodal shared representation model for Alzheimer's diagnosis. By learning shared representations across modalities, they enhance diagnostic performance. However, their focus is more on shared embedding rather than individual feature extraction and late fusion. Our architecture retains modality-specific learning while applying fusion techniques to maintain individual feature strengths.

Rehman, Abdul et al. [10] proposed a novel deep learning model called ResGLPyramid that utilizes 18F-FDG PET imaging to enhance early detection of Alzheimer's disease. Their architecture integrates convolutional operations with global-local attention modules to improve the extraction of metabolic features in PET images. Although effective, their model is limited to PET modality and lacks integration with structural imaging or cognitive data, which our multimodal framework addresses.

The related works were surveyed, and the limitations were identified to propose an organized work for our research, as described in the following elaborative headings.

CHAPTER 3

PROPOSED WORK

Dementia is a group of progressive symptoms leading to chronic cognitive deterioration, with Alzheimer's Disease (AD) being the most predominant type. Traditional diagnostic methods rely on single-modality imaging, which may fail to capture the full spectrum of pathological changes. This project aims to improve early AD detection by integrating multimodal neuroimaging data (MRI, PET, DTI) for more accurate classification. A deep learning-based approach will be explored to enhance automated diagnosis and support clinical decision-making.

3.1 INTRODUCTION

The early and accurate diagnosis of Alzheimer's Disease (AD) is crucial due to its progressive nature and significant impact on cognitive and functional abilities. Alzheimer's is the most common form of dementia, characterized by structural brain atrophy, disruption of white matter pathways, and metabolic dysfunction. Traditional diagnostic approaches often rely on single-modality imaging, which may not capture the complex and multifaceted nature of the disease.

To address this limitation, this project aims to integrate multimodal neuroimaging data, including Magnetic Resonance Imaging (MRI), Diffusion Tensor Imaging ,and Positron Emission Tomography (PET). These modalities together provide complementary information — structural, functional, and metabolic — for a more comprehensive understanding of Alzheimer's pathology.

By applying deep learning techniques, the project leverages the power of Convolutional Neural Networks to automate feature extraction and classification, allowing for enhanced diagnostic accuracy. The system aims to classify subjects into different stages such as cognitively normal, mild cognitive impairment, and Alzheimer's Disease, thereby aiding in clinical decision-making and early intervention. The architecture of the proposed multimodal classification system is depicted in Figure 3.1.

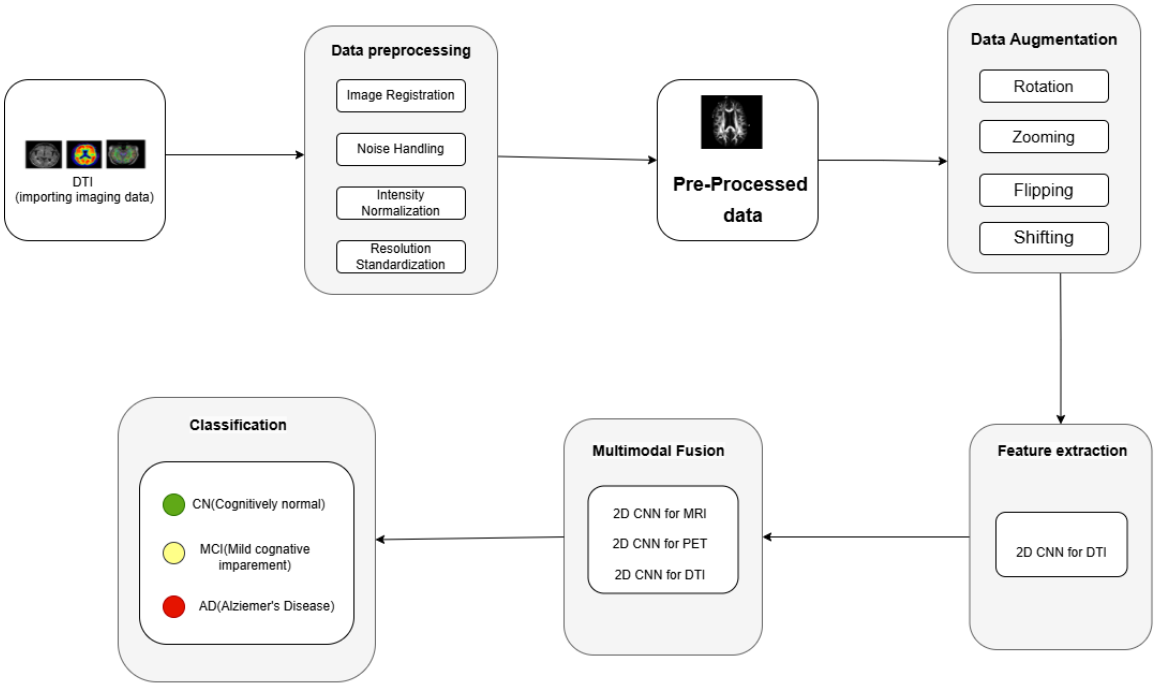

Figure 3.1 Proposed architecture diagram

I. Data Collection

Multimodal brain imaging data is collected from publicly available medical imaging datasets, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), which provides co-registered MRI, DTI, and PET scans for subjects at various stages of cognitive decline.

II. Data Preprocessing

Each imaging modality undergoes specific preprocessing steps. This includes skull stripping, spatial normalization, and intensity correction.

III. Feature Extraction and Fusion

From each modality, relevant features such as gray matter volume (MRI), white matter tract integrity (DTI), and glucose metabolism (PET) are extracted. These features are then fused either at the image level or feature level to generate a unified representation of brain condition.

IV. Classification

The fused data is fed into a deep learning classifier — typically a CNN or hybrid network — trained to categorize the subject into one of the diagnostic categories: cognitively normal, MCI, or AD. The model is trained using labeled data and validated on unseen subjects to evaluate performance.

3.2 DATA COLLECTION

The data collection for this project is centered around acquiring multimodal neuroimaging data from the kaggle and Alzheimer's Disease Neuroimaging Initiative , a comprehensive and widely used repository that provides standardized imaging and clinical datasets for Alzheimer's research. The selected imaging modalities are:

I. Magnetic Resonance Imaging – to capture structural changes such as cortical thinning, hippocampal atrophy, and ventricular enlargement.

II. Diffusion Tensor Imaging – to analyze white matter microstructural integrity by measuring the diffusion of water molecules in brain tissue.

III. Positron Emission Tomography– to evaluate functional and molecular characteristics, such as glucose metabolism or amyloid-beta deposition.

These modalities were collected for subjects spanning across the 2004–2024 period, covering multiple ADNI phases (ADNI-1, GO, 2, and 3), ensuring temporal diversity and a broad representation of Alzheimer's progression. The collected datasets are organized into three clinically validated diagnostic categories:

CN – Cognitively Normal

MCI – Mild Cognitive Impairment

AD – Alzheimer’s Disease

Table 3.1 DTI image distribution

Class	**Number of DTI Images**
CN	19.020
MCI	34,802
AD	8,553

This dataset enables the training of deep learning models that can leverage the complementary strengths of structural, functional, and microstructural information to improve the early diagnosis and classification of Alzheimer's disease. The image preprocessing pipeline is illustrated in Figure 3.2.

3.3 DATA PREPROCESSING

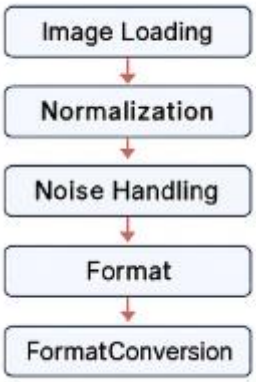

Figure 3.2 Preprocessing pipeline

3.3.1 Image Loading

Utilize a medical imaging library, such as SimpleITK or pydicom, to read the DICOM images. Ensure that the data is correctly organized into class-wise subdirectories (Cognitive Normal, Mild Cognitive Impairment and Alzheimer's disease) for efficient model training and analysis.

3.3.2 Normalization

Convert pixel values to float32 format to improve numerical precision and avoid overflow issues during further processing. Apply min-max normalization to scale the intensity values of the images between 0 and 1. This step ensures that the data is consistent and within a manageable range for subsequent deep learning models. Special care is taken to handle edge cases where the minimum and maximum pixel values are identical. In these instances, all intensity values are set to zero to avoid division by zero errors.

3.3.3 Noise Handling

In raw imaging data, certain pixel values may be corrupted or invalid, often represented as NaN (Not a Number). These NaN values are replaced with zeros to prevent distortion and ensure that the images remain usable in the next processing steps.

Table 3.2 Preprocessing step

Step	**Purpose**	**Tools/Methods Used**
Image Loading	Read DICOM images	SimpleITK, pydicom
Normalization	Scale values between 0 and 1	Min-Max Scaling
Noise Handling	Replace NaNs	NumPy
Resolution Standardization	Resize images to 224x224	OpenCV
Format Conversion	Convert to PNG	SimpleITK, OpenCV

3.3.4 Resolution Standardization

To ensure consistent input dimensions for the deep learning model, all images are resized to a fixed resolution of 224x224 pixels. This standardization is crucial to maintain uniformity across the dataset, optimizing the performance of neural networks, especially Convolutional Neural Networks (CNNs).

3.3.5 Format Conversion

The preprocessed images are saved in the PNG format, as it supports high-quality, lossless compression while maintaining a manageable file size.

PNG is also widely supported in most machine learning frameworks, making it an ideal choice for the next stages of analysis.

3.4 IMAGE AUGMENTATION

In this step, images are augmented using techniques like rotation, zooming, flipping, fill mode and brightness adjustment.

3.4.1 Image Loading

The first step in the image augmentation process is to load the preprocessed grayscale DTI images from the dataset. These images have already undergone normalization and noise handling, ensuring that they are ready for further modifications to improve model generalization.

3.4.2 Augmentation Techniques Applied

I. Rotation: Random rotation of images within a specified range is applied. This technique introduces variability in the dataset by simulating different orientations of the brain, which is essential for helping the model recognize patterns regardless of the image's orientation.

II. Zooming: Small zoom variations are introduced to the images, which alters the scale of the image content. This variation enhances the model's robustness to changes in the apparent size of brain structures, reflecting different levels of magnification during imaging.

III. Horizontal and Vertical Flipping: Random horizontal and vertical flipping of the images imics different perspectives of the brain scans. By doing so, the model learns to recognize features from different viewpoints, making it more adaptable to real-world scenarios where orientation may vary.

IV. Brightness Adjustment: Random adjustments are made to the image's brightness, normalizing the contrast. This helps simulate different lighting conditions that might occur in clinical settings or during different scanning sessions, making the model more robust to variations in intensity.

Table 3.3 Augmentation techniques summary

Technique	**Description**	**Benefit**
Rotation	Rotate image upto +-20	Orientation invariance
Zooming	Zooming upto 15%	Scale robustness
Flipping	Horizontal flip	Viewpoint flexibility
Fill mode	Fill empty pixels using nearest strategy	Maintain visual consistency
Brightness Adjustment	Vary brightness between 80% to 120%	Lighting robustness

3.4.3 Batch Generation

Augmentation is performed dynamically during the training process. By applying the augmentations in real-time, the model is exposed to a wider variety of images within each batch, ensuring it doesn’t memorize specific patterns and instead learns more generalized features.

It is crucial that the augmentation techniques applied do not distort the images to the point where they no longer retain their medical relevance. Therefore, careful attention is given to the range of transformations applied,

making sure that the augmentations preserve the integrity of key anatomical features and do not alter the spatial relationship of critical brain structures.

3.5 FEATURE EXTRACTION

In this step, important features are extracted from the preprocessed DTI images using a convolutional neural network to prepare for classification.

3.5.1 Load the Preprocessed Image

The first step is to load the preprocessed grayscale DTI images, ensuring that each image is reshaped to the required input dimensions for the model. Specifically, the image is reshaped to the shape (1, 224, 224, 1), where 1 represents the batch size (single image), 224x224 is the resolution, and 1 indicates that the image is grayscale. This step prepares the image for feeding into the convolutional neural network (CNN).

3.5.2 Feature Extraction Using CNN Layers

i. **Conv Layer 1:** In this initial convolutional layer, 32 filters (3x3 kernel size) are applied with ReLU (Rectified Linear Unit) activation. This layer helps extract low-level features from the image, such as edges, textures, and simple shapes. These fundamental patterns serve as building blocks for more complex features in later layers.

ii. **Max Pooling 1 (2x2):** Max pooling is applied after the first convolutional layer with a pool size of 2x2. This operation reduces the dimensionality of the feature maps, effectively down-sampling the image while retaining the most important spatial features, such as the edges and textures learned in the previous step.

iii. **Conv Layer 2:** The second convolutional layer applies 64 filters (3x3 kernel size) with ReLU activation. This layer captures more complex patterns and structural features of the brain, such as shapes and

anatomical relationships, building on the low-level features extracted in the first layer.

iv. **Max Pooling 2 (2x2):** Similar to Max Pooling 1, Max Pooling 2 further reduces the image size, but this time the feature maps contain more abstract and complex patterns. The pooling operation ensures that the most significant features are preserved while reducing computational load and memory requirements.

v. **Flatten and Fully Connected Layers:** After pooling, the multi-dimensional feature maps are flattened into a 1D feature vector. This step prepares the data for fully connected layers, where high-level, abstract representations of the image are learned. A dense layer with 128 neurons is applied to the flattened vector, allowing the model to learn and represent complex relationships within the data.

vi. **Classification Layer:** The final classification layer consists of a softmax activation with 3 output neurons, corresponding to the three classes: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). The softmax function outputs a probability distribution, representing the likelihood that the image belongs to each of the three classes.

3.6 MODEL EVALUATION

Once the model is trained, its performance is evaluated on a separate test set using metrics like accuracy, precision, recall, and F1-score. These metrics help determine how well the model classifies different categories in the datasets.

The four fundamental parameters used to calculate these metrics are:

True Positive (TP): Correctly predicted positive cases

True Negative (TN): Correctly predicted negative cases

False Positive (FP): Incorrectly predicted positive cases

False Negative (FN): Incorrectly predicted negative cases

Accuracy

Accuracy measures the ratio of correctly predicted observations (both positives and negatives) to the total observations.

$$Accuracy = \frac{TP+TN}{TN+TP+FP+FN} \quad (3.1)$$

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP+FP} \quad (3.2)$$

Recall

Recall (also called sensitivity) is the ratio of correctly predicted positive observations to all actual positive observations.

$$Recall = \frac{TP}{TP+FN} \quad (3.3)$$

F1 Score

The F1 Score is the harmonic mean of precision and recall. It is a useful metric when you want to seek a balance between precision and recall.

$$F1\ Score = \frac{2*Recall*Precision}{Recall + Precision} \quad (3.4)$$

CHAPTER 4

IMPLEMENTATION

The implementation of this project involves a series of structured steps to effectively analyze brain changes and diagnose Alzheimer's Disease (AD) using multimodal imaging data, specifically MRI, Diffusion Tensor Imaging (DTI), and Positron Emission Tomography (PET). The project leverages advanced image processing techniques and machine learning algorithms to classify brain images into three categories: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD).

4.1 PLATFORM USED

Jupyter Notebook

It is an open-source, interactive web-based environment that allows users to write and execute Python code in a modular and user-friendly format. In this project, Jupyter Notebook was used extensively for developing and executing Python scripts related to data preprocessing, augmentation, model building, training, and evaluation. It enabled interactive visualizations and step-by-step analysis, which was particularly helpful for training convolutional neural networks (CNNs) for image classification tasks. The notebook format also made it easier to document the workflow and iterate on experiments efficiently. Figure 4.1 shows the Jupyter Notebook interface used for implementation and experimentation.

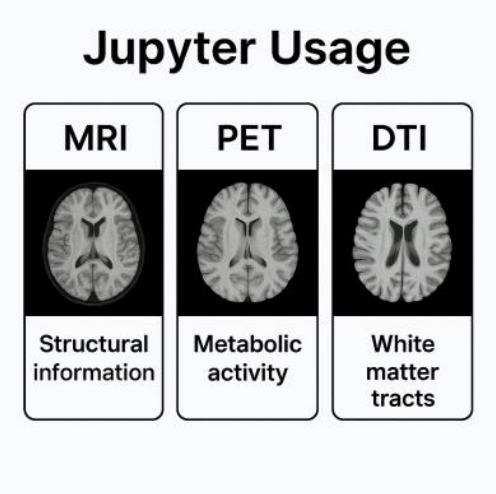

Figure 4.1 Jupyter usage

1.2 LIBRARIES USED

Table 4.1 Libraries used

Library	Description
numpy	Numerical computing library for Python, providing support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.
pandas	Data manipulation and analysis library, particularly useful for handling structured data in the form of dataframes. It provides easy-to-use data structures and data analysis tools for handling time-series, statistics, and other types of data.
matplotlib	Comprehensive library for creating static, animated, and interactive visualizations in Python. It provides an object-oriented API for embedding plots into applications.
seaborn	Python visualization library based on matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics.

Library	Description
tensorflow	Open-source machine learning framework developed by Google for building and training deep learning models. It is particularly known for its applications in neural networks.
keras	High-level neural networks API for prototyping deep learning models. It runs on top of TensorFlow and allows for easy and fast experimentation with deep learning models.
torch	Deep learning library developed by Facebook's AI Research lab (FAIR) that provides a flexible and efficient platform for building and training models, particularly for tasks involving deep learning and artificial intelligence.
sklearn	Scikit-learn, a machine learning library for Python, provides simple and efficient tools for data mining and data analysis. It is built on top of NumPy, SciPy, and matplotlib and provides access to various supervised and unsupervised learning algorithms.
cv2	OpenCV (Open Source Computer Vision Library) is an open-source computer vision and machine learning software library that contains various functions for image and video processing.

4.3 MODELS USED

Three different Convolutional Neural Network models were trained, each for a imaging modality , that includes MRI, PET and DTI.

4.3.1 CNN for MRI Classification

A Convolutional Neural Network (CNN) was trained specifically on grayscale MRI images, preprocessed to a size of 224×224. The CNN model excels at identifying spatial features such as structural atrophy and ventricular enlargement, which are key indicators in Alzheimer's progression. The MRI model outputs a probabilistic prediction over three classes: CN (Cognitively Normal), MCI (Mild Cognitive Impairment), and AD (Alzheimer’s Disease).

4.3.2 CNN for PET Classification

The PET-based CNN model was designed to interpret metabolic activity and patterns in brain function, crucial for early-stage Alzheimer's detection. Images were normalized and resized for consistency across modalities. The model was trained to distinguish between the same three classes using the patterns of glucose uptake in brain regions.

4.3.3 CNN for DTI Classification

The DTI model focused on identifying changes in white matter integrity, often affected in early neurodegeneration. The CNN was trained on fractional anisotropy maps from DTI images, capturing microstructural changes in neural pathways. Like the other models, it outputs predictions across CN, MCI, and AD categories.

4.3.4 Multimodal Fusion Strategy

After individual predictions were obtained from the MRI, PET and CNN

models, a fusion strategy was applied to derive a unified diagnosis. The final class is decided by the most frequently predicted class across all modalities. This fusion approach improves overall classification reliability by leveraging information from structural, functional, and diffusion-based imaging.

4.4 PSEUDO CODE FOR IMAGE PREPROCESSING AND LOADING

Input: Raw DTI, MRI, and PET images with associated labels

Pseudo code

Step 1: begin

Step 2: modalities = {MRI, PET, DTI}

Step 3: for each modality in modalities

Step 4: image_paths[modality] = get_image_paths(modality_directory)

Step 5: for each path in image_paths[modality]

Step 6: image = load_image(path)

Step 7: resized_image = resize_image(image, target_size=(224, 224))

Step 8: normalized_image = normalize_image(resized_image)

Step 9: append normalized_image to preprocessed_data[modality]

Step 10: return preprocessed_data

Step 11: End.

Output: Preprocessed, resized image tensors ready for model training

4.5 PSEUDO CODE FOR DATA AUGMENTATION

Input: Preprocessed training images for each modality (MRI, PET, DTI)

Pseudo code

Step 1: Begin

Step 2: define augmentation_parameters = {
rotation_range=20,
zoom_range=0.15,
horizontal_flip=True,
brightness_range=[0.8, 1.2],
fill_mode="nearest"
)
Step 3: for each modality in {MRI, PET, DTI}
Step 4: data_generator[modality] =
ImageDataGenerator(augmentation_parameters)
Step 5: x_train, y_train = load_training_data(modality)
Step 6: augmented_data[modality] =
data_generator[modality].flow(x_train, y_train, batch_size=32)
Step 7: return augmented_data
Step 8: End.

Output: Augmented image dataset

4.6 PSEUDO CODE FOR CNN MODEL TRAINING PER MODALITY

Input: Preprocessed image tensors and labels

Pseudo code

Step 1: begin
Step 2: for each modality in {MRI, PET, DTI}
Step 3: x_train, x_test, y_train,
y_test = split_data(preprocessed_data[modality])
Step 4: y_train = to_categorical(y_train, num_classes=3)
Step 5: y_test = to_categorical(y_test, num_classes=3)
Step 6: model = Sequential()

Step 7: model.add(Conv2D(32, kernel_size=3, activation='relu', input_shape=(224, 224, 1)))

Step 8: model.add(MaxPooling2D(pool_size=2))

Step 9: model.add(Conv2D(64, kernel_size=3, activation='relu'))

Step 10: model.add(MaxPooling2D(pool_size=2))

Step 11: model.add(Flatten())

Step 12: model.add(Dense(128, activation='relu'))

Step 13: model.add(Dense(3, activation='softmax'))

Step 14: model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

Step 15: model.fit(x_train, y_train, validation_data=(x_test, y_test), epochs=10, batch_size=32)

Step 16: save_model(model, modality + "_model.h5")

Step 17: End.

Output: Trained CNN model for each modality

4.7 PSEUDO CODE FOR MULTIMODAL PREDICTION AND FUSION

Input: A single patient's MRI, PET, and DTI image

Pseudo code

Step 1: begin

Step 2: for each modality in {MRI, PET, DTI}

Step 3: image = preprocess(input_image[modality])

Step 4: model = load_model(modality + "_model.h5")

Step 5: prediction[modality] = model.predict(image)

Step 6: final_prediction = mode([argmax(pred) for pred in prediction.values()])

Step 7: return final_prediction

Step 8: End.

Output: Final predicted class (CN, MCI, AD)

4.8 PSEUDO CODE FOR PERFORMANCE METRICS CALCULATION

Input: True labels and predicted labels from the test set

Pseudo code

Step 1: begin

Step 2: y_true, y_pred = collect_predictions(test_data, model)

Step 3: accuracy = compute_accuracy(y_true, y_pred)

Step 4: precision = compute_precision(y_true, y_pred,
average='macro')

Step 5: recall = compute_recall(y_true, y_pred, average='macro')

Step 6: f1_score = compute_f1(y_true, y_pred, average='macro')

Step 7: print("Accuracy:", accuracy)

Step 8: print("Precision:", precision)

Step 9: print("F1-score:", f1_score)

Step 10: End.

Output: Accuracy, Precision, Recall, F1-score

CHAPTER 5

RESULT AND ANALYSIS

5.1 MRI MODEL

The MRI-based CNN model was trained on 10,000 images per class and showed strong performance in detecting structural abnormalities associated with AD. These include hippocampal shrinkage and ventricular enlargement—commonly observed in progressive stages of Alzheimer's.

As shown in Figure 5.1, the confusion matrix reveals high classification accuracy for CN and AD, with a slight overlap between MCI and AD, indicating subtle feature similarities between these states. Figure 5.2 presents the classification report, which includes precision, recall, and F1-score. The test accuracy of the model, as seen in Figure 5.3, reached 94.62%, demonstrating robust learning from MRI-based structural features.

The model's success can be attributed to the clarity of structural biomarkers in MRI, making it the most informative single modality in this study.

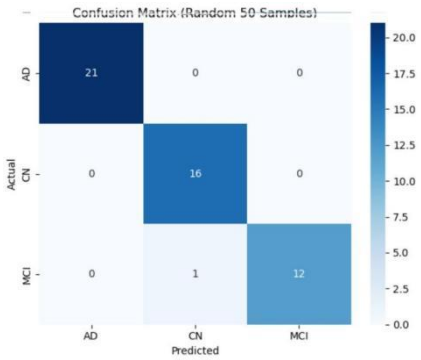

Figure 5.1 Confusion Matrix for MRI

Classification Report (Random 50 Samples):

	precision	recall	f1-score	support
AD	1.00	1.00	1.00	21
CN	0.94	1.00	0.97	16
MCI	1.00	0.92	0.96	13
accuracy			0.98	50
macro avg	0.98	0.97	0.98	50
weighted avg	0.98	0.98	0.98	50

Figure 5.2 Classification report for MRI

```
90/90 ━━━━━━━━━━━━━━━━━━━━ 96s 1s/step - accuracy: 0.9740 - loss: 0.0877

Test Accuracy: 94.62%
```

Figure 5.3 Test accuracy for MRI

5.2 PET MODEL

The PET model, trained on 10,000 images per class, analyzed glucose metabolism patterns in brain regions. PET scans provide functional insights, allowing for early detection of neurodegeneration even before structural damage appears.

As shown in Figure 5.4, the confusion matrix reveals a moderate drop in classification accuracy for MCI, with some confusion between MCI and CN. The classification report in Figure 5.5 indicates that while precision remains high, the recall for MCI is relatively lower, suggesting difficulties in separating intermediate cases.

The final test accuracy, reported in Figure 5.6, was 93.02%. While PET is effective at highlighting hypometabolic regions linked to AD, variations in

metabolic activity between individuals may account for its lower specificity compared to MRI.

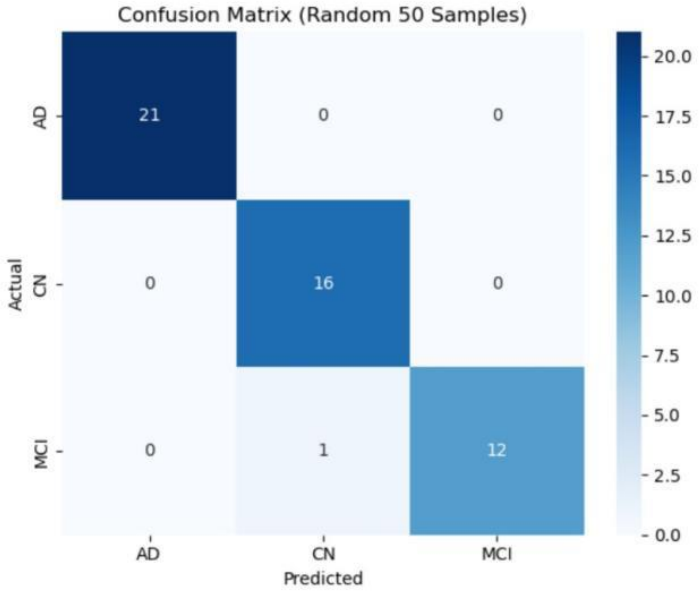

Figure 5.4 Confusion Matrix for PET

Classification Report (50 Random Samples):

	precision	recall	f1-score	support
AD	0.92	1.00	0.96	12
CN	0.96	1.00	0.98	23
MCI	1.00	0.87	0.93	15
accuracy			0.96	50
macro avg	0.96	0.96	0.96	50
weighted avg	0.96	0.96	0.96	50

Figure 5.5 Classification Report for PET

```
2683/2683 ━━━━━━━━━━━━━━━━━━━━ 1884s 698ms/step - accuracy: 0.9025 - loss: 0.3047

Test Accuracy: 93.02%
```

Figure 5.6 Test Accuracy for PET

5.3 DTI MODEL

The DTI model utilized 30,000 images per class, giving it an advantage in capturing microstructural white matter abnormalities, particularly useful in identifying early MCI stages.

The confusion matrix in Figure 5.7 shows a high success rate in detecting MCI and AD cases, while CN samples were sometimes misclassified as MCI due to overlapping features in diffusion patterns. As per Figure 5.8, the classification report suggests a strong F1-score for MCI detection. The test accuracy, seen in Figure 5.9, was 90.71%, showcasing the model's sensitivity to subtle diffusion-related changes.

DTI's high sensitivity in capturing early changes in white matter integrity makes it especially useful in the early detection of cognitive decline.

Figure 5.7 Confusion Matrix for DTI

Classification Report (50 Random Samples):

	precision	recall	f1-score	support
AD	1.00	1.00	1.00	19
CN	0.85	0.92	0.88	12
MCI	0.94	0.89	0.92	19
accuracy			0.94	50
macro avg	0.93	0.94	0.93	50
weighted avg	0.94	0.94	0.94	50

Figure 5.8 Classification Report for DTI

```
327/327 ━━━━━━━━━━━━━━━━━━━━ 68s 206ms/step - accuracy: 0.9099 - loss: 0.3274

Test Accuracy: 90.71%
```

Figure 5.9 Test Accuracy for DTI

5.4 MULTIMODAL FUSION RESULTS

To overcome the limitations of single-modality classification, predictions from MRI, PET, and DTI models were combined using a late-fusion ensemble strategy. Majority voting among the three modality-specific models determined the final predicted class. The final fused accuracy achieved was 98.09%, correctly classifying 2825 out of 2880 images, as shown in Figure 5.12.

Fusion results significantly surpassed the performance of individual models, indicating the complementary nature of structural, functional, and microstructural information in Alzheimer's diagnosis.Table 5.10 provides a detailed breakdown of individual and fused model performance metrics.

Fusion led to improved sensitivity and specificity, especially in detecting early MCI cases where single modalities struggled. The ensemble approach

better captured the complex multi-domain nature of Alzheimer's progression. The substantial improvement in accuracy demonstrates the critical importance of multimodal data integration in medical imaging-based AI diagnosis systems.

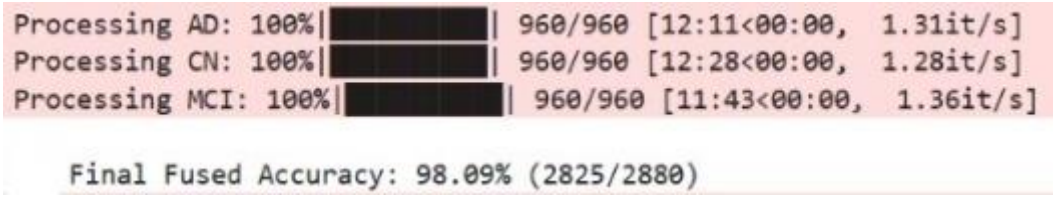

Figure 5.10 Final Fused Accuracy Display

Table 5.1 Comparison of individual and fused model performance

Approach	**Test Accuracy**
MRI	94.62%
PET	93.02%
DTI	90.71%
Fusion Approach	98.09%

5.5 COMPARITIVE ANALYSIS

The graph shown in Figure 5.5 compares the accuracy of the proposed multimodal model with notable systems from the literature survey. The proposed model achieved an accuracy of 98.09%, demonstrating superior performance by effectively leveraging MRI, DTI, and PET modalities for enhanced diagnosis. Although Abdul Rehman et al. [1] achieved an accuracy of 92.75% using PET images alone, Emimal Jabason et al. [2] reported 91.4% with a lightweight CNN extracting local and global features from MRI, and Jing Bai et al. [6] achieved 95.81% using a CNN-Transformer hybrid on MRI, these works primarily focused on single modalities.

Our approach integrates structural, diffusion, and metabolic imaging, resulting in improved classification accuracy and robustness across multiple stages of Alzheimer's disease.

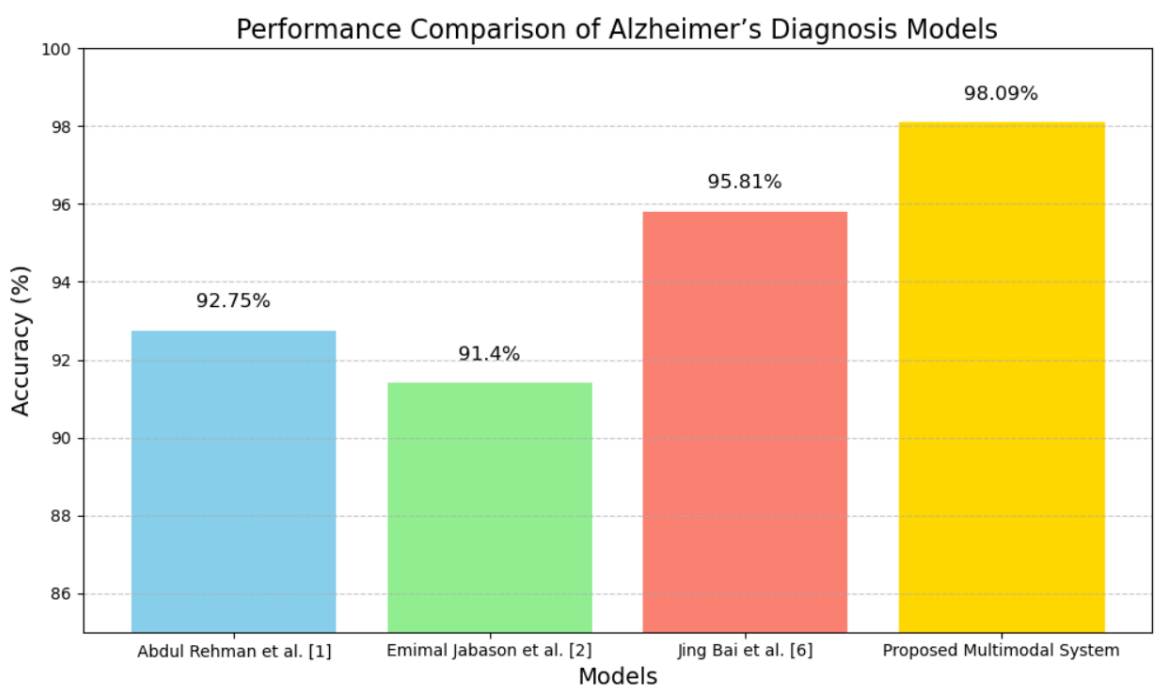

Figure 5.11 Performance Comparison Graph

CHAPTER 6

CONCLUSION AND FUTURE WORKS

This project work utilized deep learning techniques to analyze and classify Alzheimer’s disease using multimodal medical imaging data, including MRI, PET, and DTI scans. Convolutional Neural Networks (CNNs) were developed for each modality to accurately diagnose and categorize Alzheimer's disease across different stages of progression.

To further enhance the classification performance, a late fusion strategy was applied by combining the outputs of the individual MRI, PET, and DTI models, resulting in a final fused accuracy of 98.09%, significantly improving upon single-modality results. This fusion approach leveraged complementary structural, functional, and microstructural information from the different imaging techniques, thereby increasing the robustness and precision of the diagnosis.

The proposed CNN models demonstrated high classification accuracy, indicating their strong potential in assisting early diagnosis and clinical decision-making for Alzheimer’s disease. These findings have profound clinical relevance, offering a more reliable, automated, and interpretable approach to the early identification and classification of Alzheimer's disease stages. Furthermore, this study establishes a foundation for future advancements aimed at further improving diagnostic systems.

Potential future directions include integrating other data modalities such as genetic biomarkers or cognitive assessment scores to create a more comprehensive diagnostic model. Exploring advanced deep learning architectures, including 3D CNNs and transformer-based models, could offer additional improvements in classification performance. Real-time integration

into clinical settings through lightweight deployment models and the incorporation of explainable AI techniques would enhance trust and usability. Additionally, expanding the dataset with multi-center, multi-scanner imaging sources and enabling longitudinal analysis could pave the way for predicting disease progression and aiding personalized treatment planning in Alzheimer's disease care.

REFERENCES

[1] An, Xingwei, Yutao Zhou, Yang Di, Ying Han, and Dong Ming. "A Novel Method to Identify Mild Cognitive Impairment Using Dynamic Spatio-Temporal Graph Neural Network." IEEE Transactions on Neural Systems and Rehabilitation Engineering (2024).

[2] Faisal, Fazal Ur Rehman, and Goo-Rak Kwon. "Automated detection of Alzheimer's disease and mild cognitive impairment using whole brain MRI." IEEE Access 10 (2022): 65055-65066.

[3] Givian, Helia, and Jean-Paul Calbimonte. "A Review on Machine Learning Approaches for Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment Based on Brain MRI." IEEE Access (2024).

[4] Habuza, Tetiana, Nazar Zaki, Elfadil A. Mohamed, and Yauhen Statsenko. "Deviation from model of normal aging in alzheimer's disease: Application of deep learning to structural MRI data and cognitive tests." IEEE Access 10 (2022): 53234-53249.

[5] J Bai, Jing, Zhengyang Zhang, Yue Yin, Weikang Jin, Talal Ahmed Ali Ali, Yong Xiong, and Zhu Xiao. "LGG-NeXt: A Next Generation CNN and Transformer Hybrid Model for the Diagnosis of Alzheimer's Disease Using 2D Structural MRI." IEEE Journal of Biomedical and Health Informatics (2024).

[6] Jabason, Emimal, M. Omair Ahmad, and M. N. S. Swamy. "A Lightweight Deep Convolutional Neural Network Extracting Local and Global Contextual Features for the Classification of Alzheimer's Disease Using Structural MRI." IEEE Journal of Biomedical and Health Informatics (2024).

[7] Murugan, Suriya, Chandran Venkatesan, M. G. Sumithra, Xiao-Zhi Gao, B. Elakkiya, Muthuramalingam Akila, and S. Manoharan. "DEMNET: A deep

learning model for early diagnosis of Alzheimer diseases and dementia from MR images." Ieee Access 9 (2021): 90319-90329.

[8] Nan, Fengtao, Shunbao Li, Jiayu Wang, Yahui Tang, Jun Qi, Menghui Zhou, Zhong Zhao, Yun Yang, and Po Yang. "A multi-classification accessment framework for reproducible evaluation of multimodal learning in Alzheimer's disease." IEEE/ACM Transactions on Computational Biology and Bioinformatics (2022).

[9] Ning, Zhenyuan, Qing Xiao, Qianjin Feng, Wufan Chen, and Yu Zhang. "Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis." IEEE Transactions on Medical Imaging 40, no. 6 (2021): 1632-1645.

[10] Rehman, Abdul, Myung-Kyu Yi, Abdul Majeed, and Seong Oun Hwang. "Early diagnosis of alzheimer's disease using 18f-fdg pet with soften latent representation." IEEE Access (2024).