

محمد مقداد

Early Prediction of Alzheimer.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:397161853

Submission Date

Jul 6, 2026, 3:30 PM GMT+3

Download Date

Jul 6, 2026, 9:11 PM GMT+3

File Name

Early Prediction of Alzheimer.docx

File Size

2.0 MB

27 Pages

4,670 Words

32,100 Characters

87% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.



55 AI-generated only 86%

Likely AI-generated text from a large-language model.



3 AI-generated text that was AI-paraphrased 1%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Early Prediction of Alzheimer's Disease Using Deep Learning and Data Fusion

Abstract

Detection of Alzheimer's disease (AD) and its progression remains one of the most challenging problems in the medical field. Early diagnosis plays a significant role in improving treatment planning and disease management. However, many existing diagnostic approaches rely on a single source of information, such as brain magnetic resonance imaging (MRI) or clinical records alone, which limits the ability to detect the disease during its earliest stages, particularly the transition from Normal Cognition (NC) to Mild Cognitive Impairment (MCI) before progression to Alzheimer's disease (AD).

To address this limitation, this study proposes a deep learning framework based on multimodal data fusion. The proposed framework combines three-dimensional MRI scans with patients' clinical information obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Two parallel processing streams are employed. The first stream utilizes a 3D ResNet network to automatically extract imaging features from MRI scans, while the second stream applies a Deep Neural Network (DNN) to process structured clinical information. The extracted features are then fused to provide a more comprehensive representation of the patient's condition and improve early prediction performance.

Keywords: Alzheimer's Disease; Deep Learning; Data Fusion; CNN; 3D-ResNet; Early Prediction; Medical Imaging; Explainable AI

1. Introduction

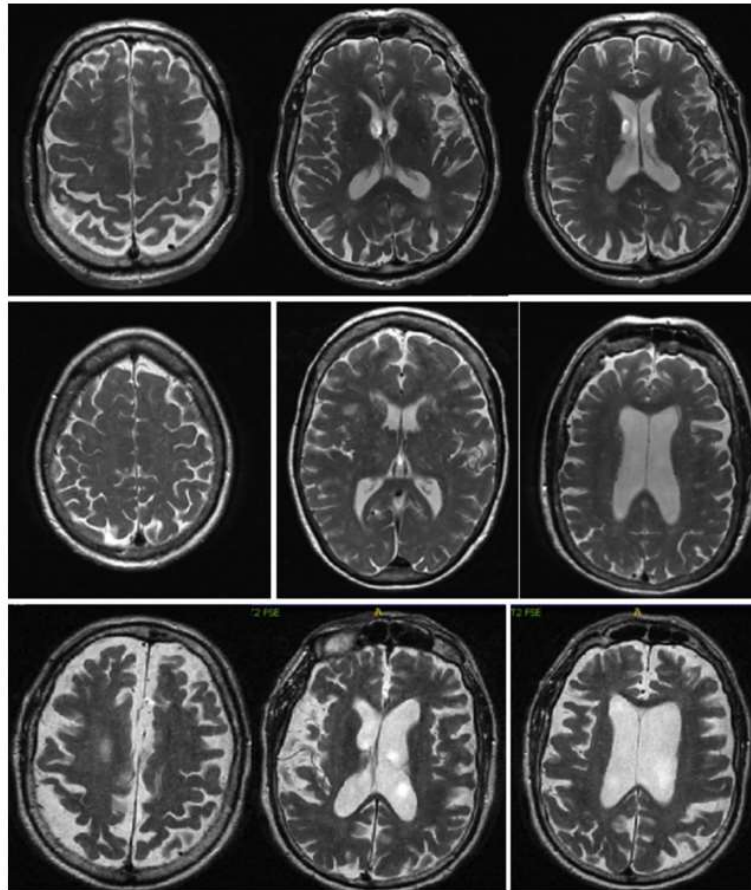
Alzheimer's disease is one of the most difficult neurological disorders affecting human memory and cognitive abilities. Although it is commonly associated with elderly individuals, it may also affect people at earlier stages of life. The disease develops gradually, and noticeable clinical symptoms usually appear only after significant neurological damage has already occurred. Previous studies have demonstrated that structural brain changes begin many years before clinical symptoms become visible, emphasizing the importance of early diagnosis.

Artificial Intelligence (AI) has recently become an essential tool in medical diagnosis. In particular, deep learning techniques have significantly improved the analysis of brain Magnetic Resonance Imaging (MRI), allowing automatic extraction of complex anatomical features that cannot easily be recognized manually. Recent studies have shown that AI-based approaches can substantially improve diagnostic accuracy. Nevertheless, most existing research still depends on unimodal analysis, where only MRI images or only clinical information are considered independently.

Such an approach cannot fully describe the progression of Alzheimer's disease because physicians normally rely on multiple sources of information before making a diagnosis. Recent

research therefore suggests that combining neuroimaging data with patients' clinical information provides more reliable diagnostic performance.

Based on this motivation, this study proposes a multimodal deep learning framework for the early prediction of Alzheimer's disease. The proposed framework combines structural MRI images with patients' clinical records obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The primary objective is to investigate whether integrating both imaging and clinical information can improve the early identification of Alzheimer's disease and support physicians with a more reliable diagnostic tool.



1.1 Problem Definition

Despite the considerable progress achieved by deep learning techniques in Alzheimer's disease diagnosis, several important limitations remain.

First, many existing approaches rely exclusively on structural Magnetic Resonance Imaging (MRI). Although MRI provides valuable information regarding brain atrophy and structural abnormalities, it does not incorporate complementary clinical information that may reveal cognitive decline before severe anatomical changes become visible.

Second, both neuroimaging and clinical datasets are inherently high-dimensional. Three-dimensional MRI volumes contain millions of voxels, while clinical datasets consist of numerous patient attributes. Directly training deep learning models on these heterogeneous data sources increases computational complexity and the risk of overfitting. Therefore, efficient feature extraction and dimensionality reduction techniques are essential.

Another limitation is that Alzheimer's disease progresses gradually over time, whereas many current deep learning approaches analyze patient information as independent observations without considering disease progression. Consequently, early indicators that appear during the transition from Normal Cognition (NC) to Mild Cognitive Impairment (MCI) may not be adequately captured.

Furthermore, multimodal fusion remains a challenging task. There is still no unified framework capable of effectively integrating three-dimensional MRI scans with patients' clinical information into a single deep learning architecture capable of providing robust early diagnosis.

1.2 Research Objectives

The primary objective of this research is to develop a deep learning framework capable of improving the early prediction of Alzheimer's disease through multimodal data fusion.

The specific objectives are summarized as follows:

- To develop a deep learning model capable of extracting representative features from three-dimensional MRI scans.
- To construct a multimodal fusion framework that combines MRI imaging information with patients' clinical data.
- To evaluate disease progression during different diagnostic stages.
- To compare the proposed framework with existing unimodal approaches in order to evaluate performance improvements.

1.3 Main Contributions

The major contributions of this study are summarized as follows:

- Development of a multimodal deep learning framework that integrates MRI images and clinical information for Alzheimer's disease prediction.
- Comparative evaluation of multiple computer vision architectures, including the baseline CNN model and the 3D ResNet-18 architecture.
- Design of an efficient framework capable of processing large-scale Alzheimer's datasets.
- Integration of Explainable Artificial Intelligence (XAI) techniques to improve model interpretability by highlighting brain regions contributing to diagnostic decisions.

2. Related Work

Recent literature demonstrates a clear transition from traditional machine learning approaches toward deep learning architectures for Alzheimer's disease prediction. These studies differ in the type of input data, network architecture, feature extraction strategy, and multimodal integration techniques. Table 1 summarizes the major studies reviewed in this research.

Study (Year)	Proposed Method	Key Accuracy	Main Limitations
Muksimova et al. (2025)	CNN + LTMS	94.00%	Fuses images only; completely lacks integration of clinical medical records.
Li et al. (2025)	CNN + Transformer	93.04%	Struggles with MCI-to-AD prediction.
Bricaud & Masala (2026)	Hybrid CNN (ConvNeXtV2-L)	91.07%	MRI-only; lacks integration of clinical medical records.
Bekhet et al. (2025)	1D CNN + Random Forest	> 90.0%	No simultaneous fusion; relies on 2D images.
Zhang, Y., et al. (2023)	GNN framework	92%	Highly computational cost with required of both MRI + PET scans
Helaly et al. (2021)	E2AD2C (CNN & VGG19)	97.0%	High power for 3D scans; images-only.
Warren & Moustafa (2026)	3D CNN with XAI	90.0%	Very small sample size; lacks validation.
Wang, J., et al. (2024)	ResNet + Attention convolution	93.44%	Highly computational cost with required of both MRI + PET scans

Alsubaie et al. (2026)	BiFPN3DViT (3D CNN + BiFPN + ViT)	92.10%	MRI-only; lacks integration of clinical medical records.
Ahmed et al. (2023)	Deep CNN	93.86%	MRI-only; lacks Explainable AI tools.
Lu et al. (2022)	Transfer Learning (Inception-ResNet-V2)	94.50%	MRI-only; data mostly from Western populations.
Khalid et al. (2026)	CNN & Random Forest	94.00%	Processes images and clinical data separately; lacks a fully unified fusion network.
Sheikh et al. (2025)	Lightweight CNN (EfficientNetV2B0) & Explainable AI	88.00%	2D images only; lacks integration of clinical medical records.
Ghosh et al. (2023)	Federated Learning & MobileNet	95.24%	2D images only; lacks integration of clinical medical records.
Abbasian & Hammond (2024)	3D CNN	81.80%	MRI-only; lacks clinical data integration.

2.1 Deep Learning-Based Alzheimer's Diagnosis

Muksimova et al. (2025) proposed a multimodal framework, FusionNet, for the early diagnosis of Alzheimer's disease by integrating a 3D CNN with an LSTM network to capture temporal brain changes. MRI, PET, and CT scans obtained from the ADNI dataset were utilized for training. The proposed framework achieved a diagnostic accuracy of 94%. However, the approach requires considerable computational resources and depends on acquiring three different imaging modalities for each patient.

Li et al. (2025) introduced a multimodal deep learning architecture combining a Convolutional Neural Network for MRI analysis with a Transformer network for genetic information. Both modalities were integrated through a cross-transformer fusion mechanism. Using 1,541 patients from the ADNI dataset, the framework achieved 93.04% classification accuracy. Nevertheless, the proposed system still relies on a relatively simple CNN architecture and encounters challenges while integrating heterogeneous genetic information.

Bricaud and Masala (2026) evaluated multiple deep learning architectures, including conventional CNNs and Vision Transformers, using MRI data obtained from the ADNI dataset. Their ConvNeXtV2-L architecture achieved an accuracy of 91.07%. Despite the promising performance, the framework depends entirely on MRI images without considering patients' clinical information.

Bekhet et al. (2026) proposed DeepALZNET, combining one-dimensional CNNs, Random Forest classifiers, and the VGG19 architecture for Alzheimer's disease prediction. MRI scans and clinical information obtained from ADNI and OASIS datasets were analyzed separately. Although the framework achieved classification accuracy exceeding 90%, MRI and clinical information were processed independently rather than within a unified multimodal architecture.

Zhang et al. (2023) proposed a Graph Neural Network (GNN)-based framework capable of modeling relationships among different brain regions using MRI and PET images collected from the ADNI dataset. The proposed model achieved approximately 92% classification accuracy but required both MRI and PET examinations, significantly increasing computational and clinical costs.

Helaly et al. (2021) introduced the E2AD2C framework based on CNN and VGG19 architectures for Alzheimer's disease diagnosis using MRI scans from the ADNI dataset. Their approach achieved approximately 97% classification accuracy. However, the framework relies solely on MRI data and requires substantial computational resources for processing volumetric brain images.

Warren and Moustafa (2026) combined functional MRI with clinical cognitive assessments using a 3D CNN architecture enhanced by Explainable Artificial Intelligence techniques. Their study demonstrated that integrating imaging and clinical information substantially improved diagnostic performance compared with MRI-only approaches. Nevertheless, validation was performed using a relatively small patient cohort.

Wang et al. (2024) proposed a multimodal deep learning framework integrating ResNet with an attention mechanism for feature extraction from MRI images combined with genetic information. The proposed architecture achieved approximately 93.44% classification accuracy. Despite the encouraging results, generating multimodal three-dimensional representations requires significant computational power and comprehensive patient examinations.

Additional studies investigated hybrid Vision Transformer architectures, transfer learning using Inception-ResNet-V2, EfficientNet, MobileNet, federated learning, and lightweight CNN models. Although these approaches achieved competitive classification performance, the majority remain limited by one or more of the following issues:

- Dependence on MRI images alone.
- Separation between imaging and clinical analysis.
- High computational complexity.
- Requirement for multiple expensive imaging modalities.
- Limited explainability.
- Lack of unified multimodal fusion frameworks.

2.2 Summary of Previous Studies

Overall, previous studies demonstrate that deep learning significantly improves Alzheimer's disease diagnosis compared with traditional machine learning techniques. CNN-based architectures effectively learn spatial characteristics from MRI images, while transformer-based networks provide improved representation learning capabilities. Recently, multimodal learning has received increasing attention because combining different data modalities more closely resembles real clinical practice.

Despite these advances, important limitations remain. Most previous frameworks either process MRI images independently or analyze clinical information separately. Several studies require PET or CT scans in addition to MRI, increasing both computational cost and clinical expense. Furthermore, many approaches lack an efficient mechanism capable of integrating heterogeneous data sources into a unified end-to-end deep learning framework.

These limitations motivate the proposed multimodal framework presented in this study, which integrates three-dimensional MRI scans and structured clinical information using a unified feature fusion architecture based on deep learning.

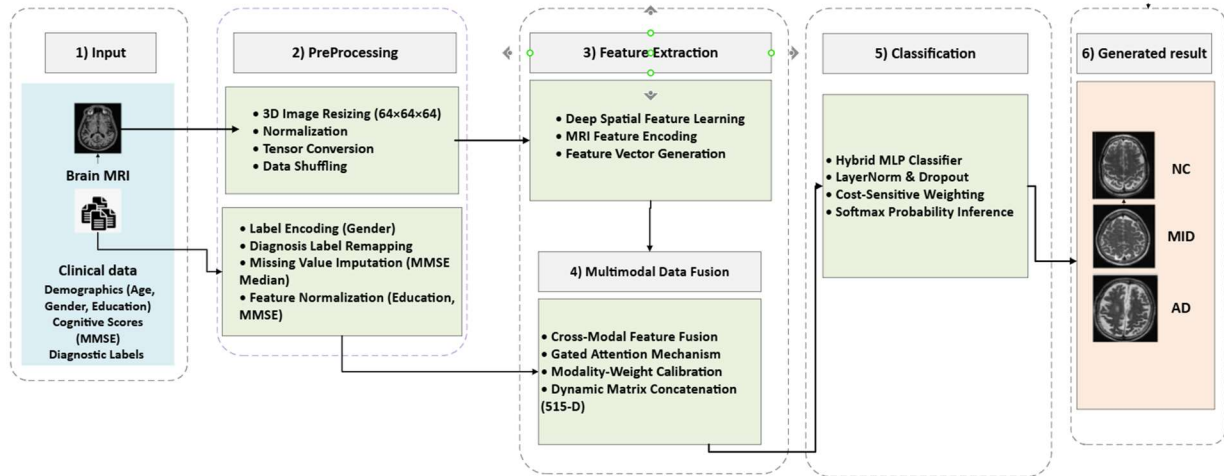
3. Proposed Methodology

3.1 Proposed Framework

The proposed framework aims to improve the early prediction of Alzheimer's disease through a multimodal deep learning architecture capable of integrating structural MRI scans with patients' clinical information. Rather than depending on a single source of information, the proposed system simultaneously analyzes volumetric brain images and structured clinical records to provide a more comprehensive representation of the patient's condition.

The overall methodology follows a sequential engineering pipeline beginning with data acquisition, followed by preprocessing of both MRI and clinical information. Feature extraction is then performed independently for each modality before both feature representations are fused into a unified latent feature space. Finally, the fused representation is passed to a hybrid multilayer perceptron classifier for disease prediction.

Proposed End to End Alzheimer's detection pipeline



3.2 Dataset Description

3.2.1 ADNI Dataset

The proposed framework utilizes data collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). This dataset is considered one of the most widely used public resources for Alzheimer's disease research because it provides standardized structural MRI scans together with longitudinal clinical information.

Using the ADNI dataset allows direct comparison with previously published work while ensuring the reliability and reproducibility of the experimental evaluation.

3.2.2 Dataset Split

To prevent data leakage and ensure unbiased evaluation, the dataset was divided into three independent subsets:

- Training Set (70%)
- Validation Set (15%)
- Testing Set (15%)

The complete dataset contains **1,971** patient samples distributed across the three diagnostic classes:

- Normal Control (NC)
- Mild Cognitive Impairment (MCI)
- Alzheimer's Disease (AD)

Medical Group (Diagnosis)	Data Split	Number of Patients (N)
Normal Control (NC)	Training (70%)	335
	Validation (15%)	72
	Testing (15%)	72
Early MCI (MID)	Training (70%)	536
	Validation (15%)	115
	Testing (15%)	115
Alzheimer's (AD)	Training (70%)	508
	Validation (15%)	109
	Testing (15%)	109
Total Cohort	Cumulative	1971

3.3 Data Preprocessing

Since both MRI scans and clinical information originate from different sources and possess different characteristics, two independent preprocessing streams were implemented before feature extraction.

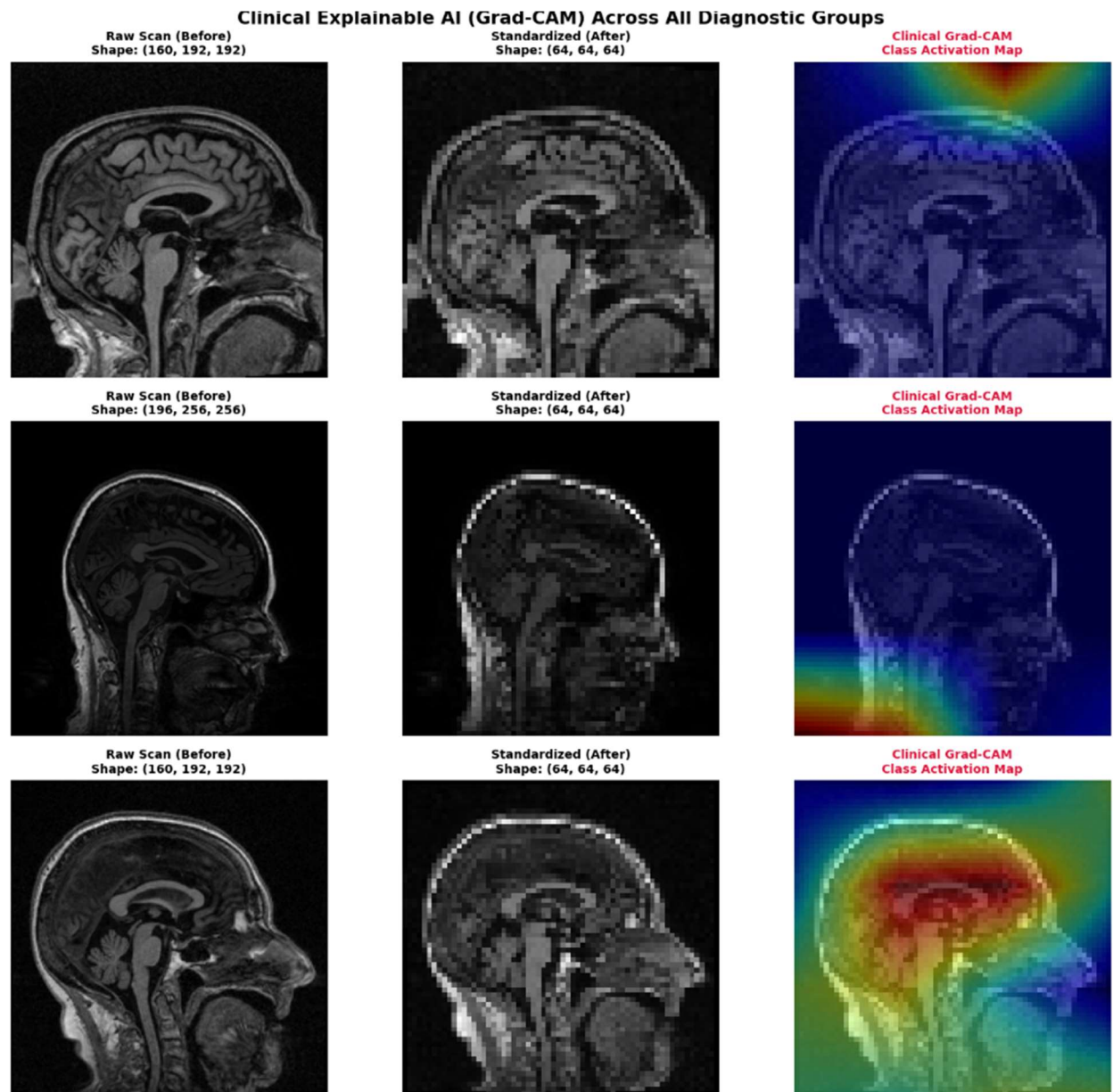
3.3.1 MRI Preprocessing

The structural MRI scans are stored using the NIfTI (.nii) format. Directly processing these large three-dimensional volumes significantly increases computational complexity and GPU memory requirements. Therefore, several preprocessing operations are applied before training.

The preprocessing pipeline consists of the following steps:

- Spline interpolation resizing.
- Removal of invalid voxel values.
- Intensity normalization using Min-Max normalization.

These operations preserve the anatomical structure of the brain while reducing computational cost and improving numerical stability during training.



3.3.2 Clinical Data Preprocessing

Clinical information is processed independently before fusion.

The preprocessing steps include:

- Gender label encoding.
- Missing value imputation using the median value.
- Feature normalization.
- Random data shuffling before every training epoch.

These operations ensure that all clinical variables contribute equally during optimization while reducing the influence of missing values.

3.4 Classification Model Evolution

To identify the most effective deep learning architecture for Alzheimer's disease prediction, three different network architectures were investigated.

Stage 1: Baseline 3D CNN

The initial model consists of a conventional three-dimensional convolutional neural network designed to learn spatial representations directly from MRI volumes.

The model employs:

- 3D Convolution layers
- ReLU activation
- Max Pooling
- Fully Connected layers
- Softmax classifier

Although this architecture provides acceptable baseline performance, increasing network depth causes optimization difficulties and performance degradation.

	PTID	PTGENDER	PTEDUCAT	MMSCORE	DIAGNOSIS	EXAMDATE	Image_File
0	011_S_0021	2	0.916667	0.900000	0	2025-07-18	011_S_0021_598.nii
1	011_S_0021	2	0.916667	0.900000	0	2025-07-18	011_S_0021_597.nii
2	023_S_0031	2	0.916667	0.833333	0	2019-05-01	023_S_0031_427.nii
3	023_S_0031	2	0.916667	0.833333	0	2019-05-01	023_S_0031_426.nii
4	023_S_0042	1	0.916667	0.733333	2	2013-12-04	023_S_0042_614.nii

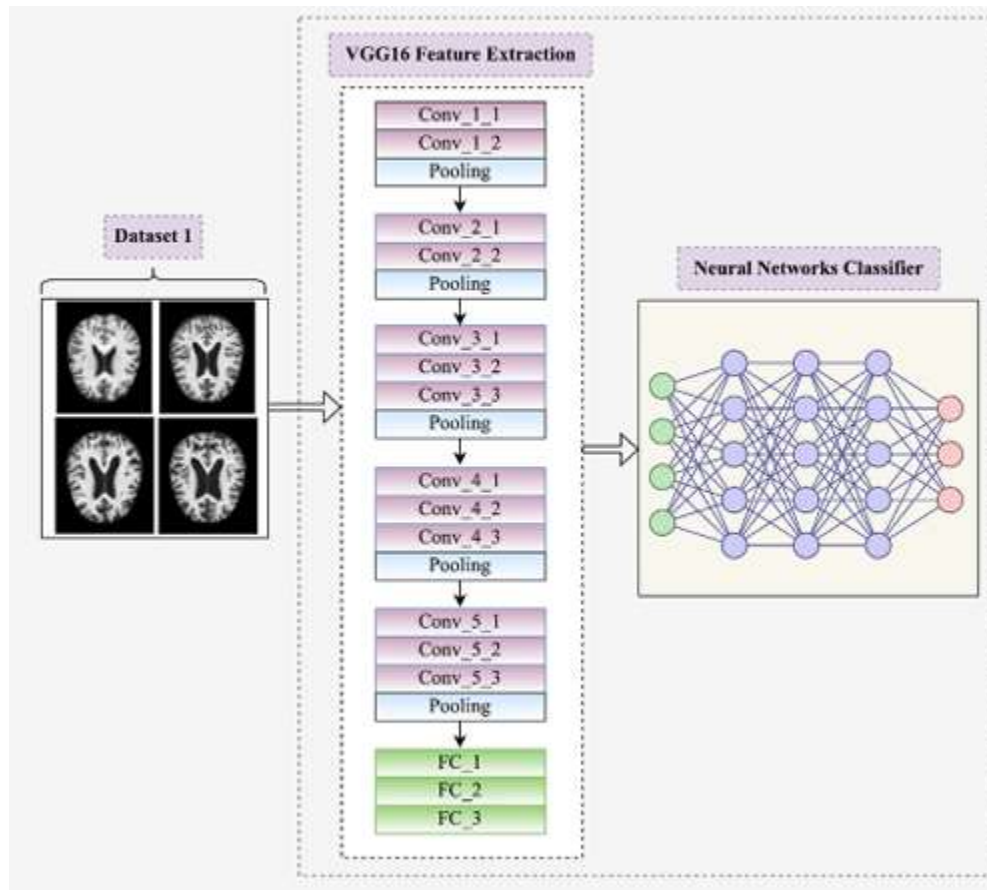
Stage 2: 3D VGG-Inspired Architecture

The second architecture extends the baseline model using a deeper VGG-style design.

Instead of large convolution kernels, multiple small convolution filters are stacked to increase representational capability while reducing the number of trainable parameters.

This architecture enables hierarchical feature extraction, allowing the network to learn increasingly complex anatomical structures.

However, increasing network depth introduces the Vanishing Gradient problem, making optimization more difficult during training.



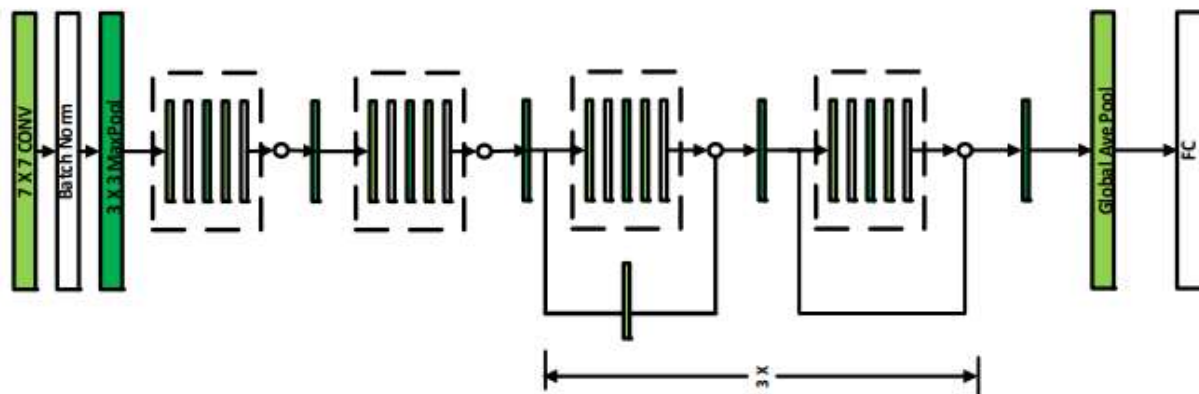
Stage 3: 3D ResNet-18

To overcome the optimization limitations of previous architectures, the final model employs a pretrained 3D ResNet-18 backbone.

Residual Skip Connections allow gradients to propagate directly through the network, significantly improving convergence and enabling deeper architectures without suffering from gradient degradation.

The pretrained network extracts a compact 512-dimensional feature vector representing the structural characteristics of each MRI scan.

This feature vector becomes the visual representation used later during multimodal fusion.



3.5 Proposed Multimodal Fusion Framework

After independent feature extraction, MRI features and clinical features are combined into a unified latent representation.

The fusion strategy enables the model to exploit complementary information obtained from both modalities instead of relying on a single information source.

The fused feature vector is subsequently processed by a Hybrid Multilayer Perceptron (MLP) classifier that produces the final probability distribution across the three diagnostic categories.

3.5 Multimodal Feature Fusion

Following the independent extraction of imaging and clinical representations, both feature vectors are integrated through a multimodal fusion strategy to construct a unified representation describing each patient.

The MRI branch produces a 512-dimensional feature vector extracted from the pretrained 3D ResNet-18 backbone, whereas the clinical branch generates a compact numerical representation containing the patient's demographic and cognitive information after preprocessing.

Instead of making separate decisions for each modality, both feature vectors are concatenated to form a single latent representation capable of describing both structural brain characteristics and clinical status simultaneously. This combined representation allows the classifier to utilize complementary information that cannot be obtained from either modality independently.

The fusion operation is mathematically represented by concatenating both feature vectors into one feature space before classification.

3.5.1 Mathematical Mechanism of Feature Fusion

Assume that

- $\mathbf{V}$ represents the MRI feature vector extracted by the pretrained 3D ResNet-18 network.
- $\mathbf{C}$ represents the clinical feature vector generated after preprocessing.

The final fused representation is obtained by concatenating both vectors into a unified latent feature vector.

This fused representation is then forwarded to the classification network for final prediction.

The objective of this strategy is to preserve all discriminative information extracted from both modalities while allowing the classifier to learn the relationships between structural brain abnormalities and clinical characteristics.

Compared with unimodal approaches, multimodal fusion provides richer feature representations and improves the discrimination between Normal Control (NC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD).

3.5.2 Hybrid MLP Classification Head

After feature fusion, the unified feature vector is processed by a Hybrid Multi-Layer Perceptron (MLP).

The classifier consists of several fully connected layers separated by nonlinear activation functions and regularization techniques.

The classification pipeline performs the following operations:

- Fully Connected Layer
- Batch Normalization
- ReLU Activation
- Dropout
- Final Fully Connected Layer
- Softmax Output

The Softmax layer produces probability values corresponding to the three diagnostic classes.

The predicted class corresponds to the highest probability value generated by the network.

3.6 Experimental Setup

To evaluate the effectiveness of the proposed framework, all experiments were conducted under identical training conditions to ensure fair comparison between the investigated architectures.

The experiments were implemented using the PyTorch deep learning framework with GPU acceleration.

The same dataset partitions, preprocessing pipeline, optimizer configuration, and evaluation protocol were maintained throughout all experiments.

This standardized setup ensures that observed performance differences originate from architectural improvements rather than experimental bias.

3.6.1 Computing Environment

The proposed framework was implemented using the following software environment:

- Python
- PyTorch
- NumPy
- Pandas
- Scikit-Learn
- MONAI
- CUDA

GPU acceleration was utilized during training to reduce computational time associated with processing three-dimensional MRI volumes.

3.6.2 Hyperparameter Configuration

The network was trained using carefully selected hyperparameters to achieve stable convergence.

The configuration includes:

- Adam Optimizer
- Cross-Entropy Loss
- Mini-batch Training
- Learning Rate Scheduling
- Early Stopping
- Weight Decay Regularization

These parameters reduce overfitting while improving convergence stability during optimization.

Parameter	Baseline 3D-CNN	3D VGG-16	3D ResNet-18	Proposed Fusion
Hardware	CUDA-enabled NVIDIA GPU	CUDA-enabled NVIDIA GPU	CUDA-enabled NVIDIA GPU	CUDA-enabled NVIDIA GPU
Framework	PyTorch + CUDA	PyTorch + CUDA	PyTorch + CUDA	PyTorch + CUDA
Epochs	20	20	20	20
Batch Size	32	32	32	32
Optimizer	Adam	Adam	Adam	Adam
Initial LR	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}
LR Scheduler	None	None	StepLR	StepLR
Normalization	BatchNorm	BatchNorm	BatchNorm	LayerNorm
Dropout	20%	30%	40%	40%
Loss Function	Cross-Entropy	Cross-Entropy	Cross-Entropy	Class-Weighted Cross-Entropy

3.7 Evaluation Metrics

Performance evaluation was conducted using several commonly adopted classification metrics.

Classification Accuracy

Classification accuracy measures the proportion of correctly classified samples among all evaluated cases.

It provides an overall indication of model performance across all diagnostic categories.

Precision

Precision evaluates the proportion of correctly predicted positive samples among all predicted positive cases.

Higher precision indicates fewer false-positive predictions.

Recall (Sensitivity)

Recall measures the ability of the proposed model to correctly identify actual positive cases.

A high recall value indicates that the framework successfully detects Alzheimer's disease patients while minimizing false negatives.

F1-Score

The F1-score represents the harmonic mean of Precision and Recall.

It provides a balanced evaluation when class distributions are not perfectly balanced.

ROC-AUC

Receiver Operating Characteristic (ROC) analysis evaluates the discrimination capability of the proposed classifier across different probability thresholds.

The Area Under the Curve (AUC) summarizes the overall discriminative performance.

Higher AUC values indicate stronger separation between diagnostic classes.

3.8 Chapter Summary

This chapter presented the complete methodology adopted in the proposed multimodal framework.

The chapter described the ADNI dataset, preprocessing pipeline, progressive evolution of deep learning architectures, multimodal feature fusion strategy, Hybrid MLP classifier, experimental setup, and evaluation metrics.

Together, these components establish the proposed framework used for Alzheimer's disease prediction. The following chapter presents the experimental results and provides a comprehensive evaluation of the proposed system.

4. Results and Discussion

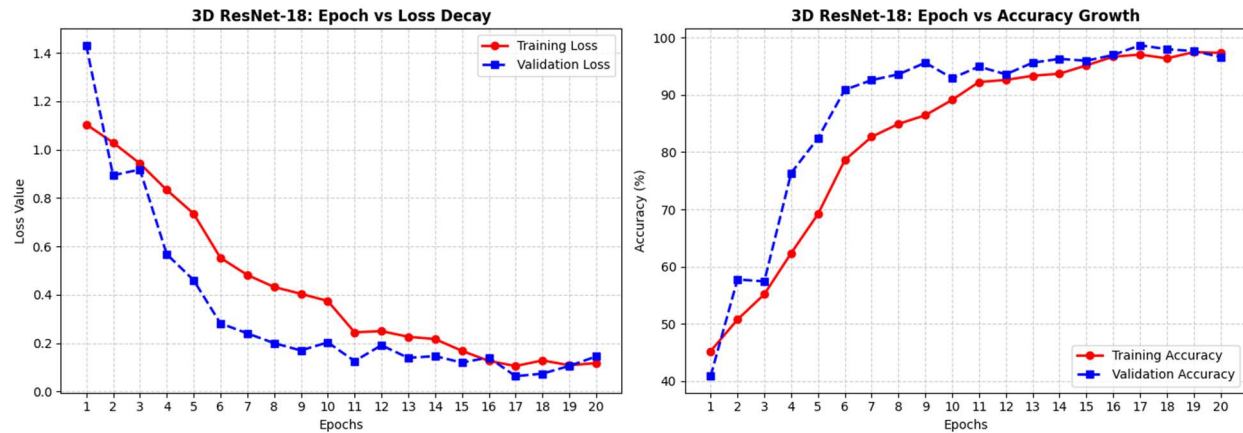
4.1 Introduction

This chapter evaluates the effectiveness of the proposed multimodal deep learning framework using the testing subset of the ADNI dataset.

The evaluation investigates classification performance, convergence behavior, feature fusion effectiveness, model comparison, and explainability analysis.

Performance is assessed using classification accuracy, precision, recall, F1-score, confusion matrices, ROC curves, and Grad-CAM visualizations.

The experimental analysis begins by evaluating model convergence during training before presenting quantitative classification performance and comparison against existing approaches.

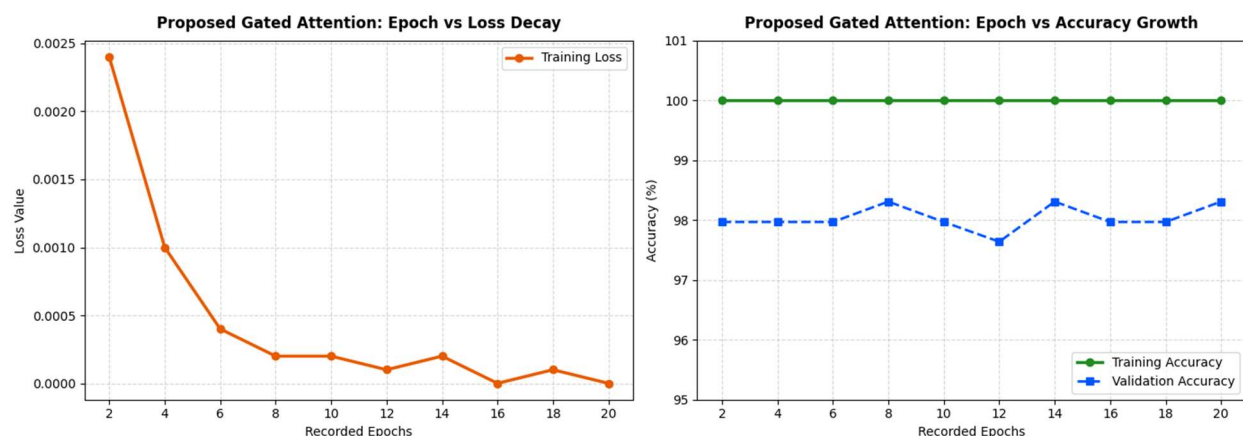


4.2 Training Convergence and Loss Curve Analysis

The training process was monitored continuously to evaluate the stability of the proposed multimodal framework throughout optimization. Training and validation losses were recorded after every epoch to determine whether the network successfully converged without exhibiting signs of underfitting or overfitting.

The obtained learning curves demonstrate that the proposed framework converged progressively as training advanced. During the early epochs, both training and validation losses decreased rapidly, indicating that the network successfully learned the dominant structural and clinical patterns associated with Alzheimer's disease. As training continued, the reduction in loss gradually stabilized until convergence was achieved.

The close agreement between the training and validation curves indicates that the proposed preprocessing strategy, feature extraction pipeline, and regularization techniques effectively improved the model's generalization capability while preventing excessive memorization of the training data.



4.3 Classification Results

Following convergence, the proposed multimodal framework was evaluated using the independent testing dataset.

Performance evaluation included several complementary metrics to provide a comprehensive assessment of classification capability. These metrics include classification accuracy, precision, recall, F1-score, confusion matrices, ROC analysis, and explainability visualization.

Overall, the proposed framework successfully differentiated between the three diagnostic categories:

- Normal Control (NC)
- Mild Cognitive Impairment (MCI)
- Alzheimer's Disease (AD)

The combination of MRI features and clinical information enabled the classifier to distinguish subtle differences between disease stages more effectively than approaches relying on a single data modality.

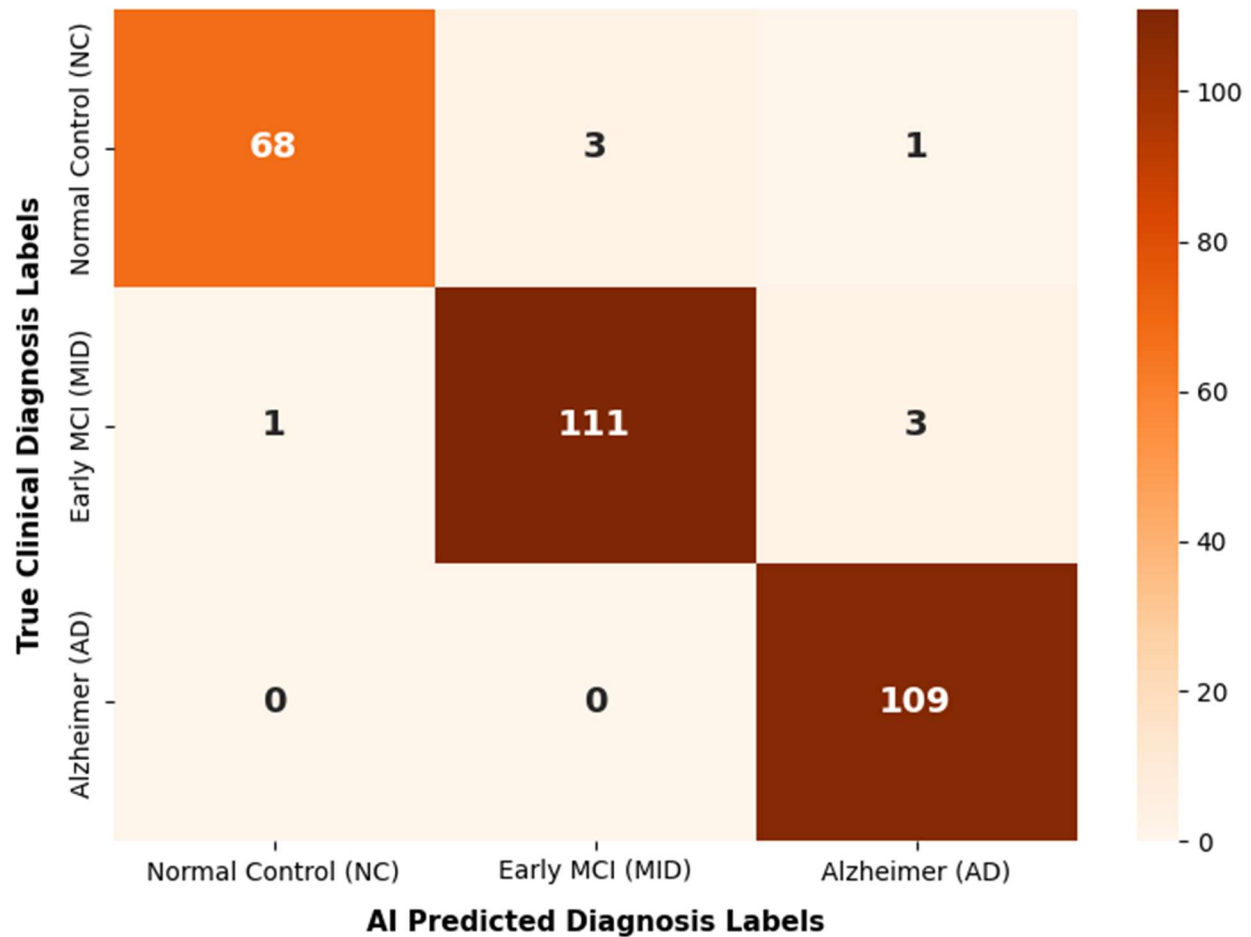
4.3.1 Confusion Matrix Analysis

The confusion matrix provides a detailed representation of prediction performance by comparing the predicted diagnostic labels with their corresponding ground-truth classes.

Correct classifications appear along the principal diagonal of the matrix, whereas misclassified samples appear outside the diagonal.

The confusion matrix indicates that the proposed multimodal framework correctly classified the majority of Normal Control, Mild Cognitive Impairment, and Alzheimer's Disease cases. Most prediction errors occurred between neighboring disease stages, particularly between Mild Cognitive Impairment and Alzheimer's Disease, reflecting the gradual progression of the disease.

The relatively small number of misclassifications demonstrates the effectiveness of combining MRI images with clinical information within a unified feature representation.



4.3.2 Classification Report

The classification report summarizes Precision, Recall, F1-score, and overall Accuracy for each diagnostic category.

The obtained evaluation metrics demonstrate balanced performance across all three diagnostic groups, indicating that the proposed framework maintains high classification capability without favoring a specific class.

The balanced Precision and Recall values further confirm that the multimodal fusion strategy successfully reduces both false-positive and false-negative predictions.

Class	Precision	Recall	F1-Score	Support
Normal Control (NC)	98.55%	94.44%	96.45%	72

Early MCI (MID)	97.37%	96.52%	96.94%	115
Alzheimer (AD)	96.46%	100.00%	98.20%	109
Accuracy	—	—	97.30%	296
Macro Avg	97.46%	96.99%	97.20%	296
Weighted Avg	97.32%	97.30%	97.29%	296

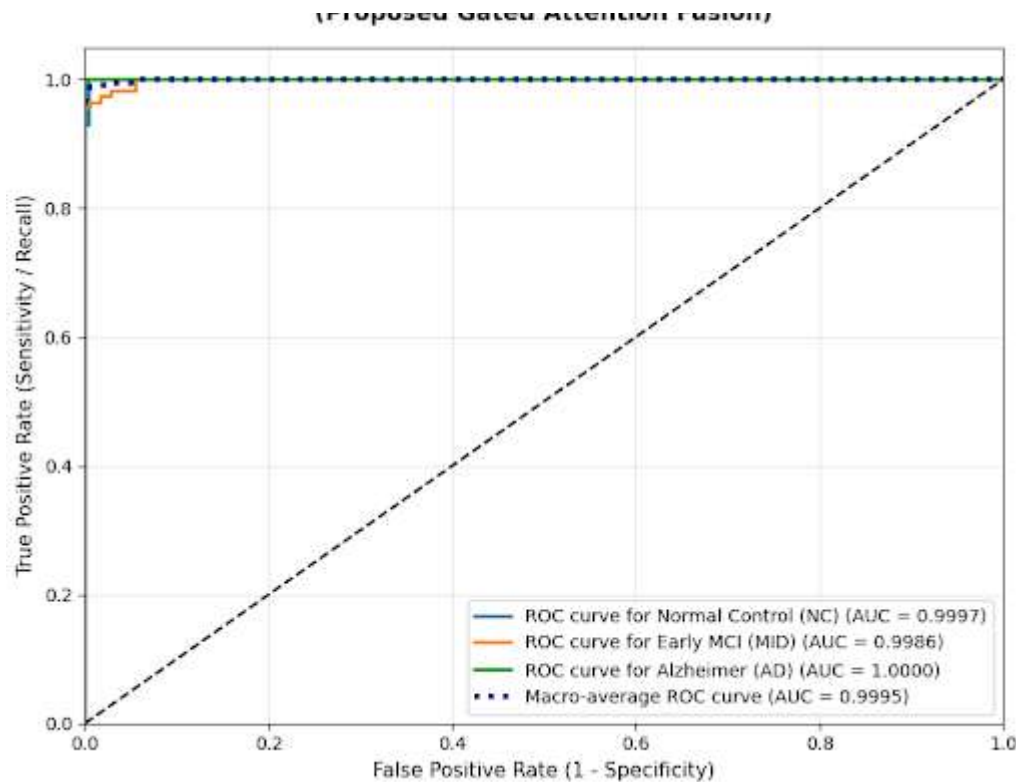
4.3.3 ROC-AUC Analysis

Receiver Operating Characteristic (ROC) analysis was performed to evaluate the discrimination capability of the proposed framework across different classification thresholds.

For multiclass classification, ROC curves were generated individually for each diagnostic category.

The corresponding Area Under the Curve (AUC) values demonstrate strong separability among Normal Control, Mild Cognitive Impairment, and Alzheimer's Disease patients.

High AUC values indicate that the classifier successfully distinguishes between disease stages while maintaining stable sensitivity and specificity.



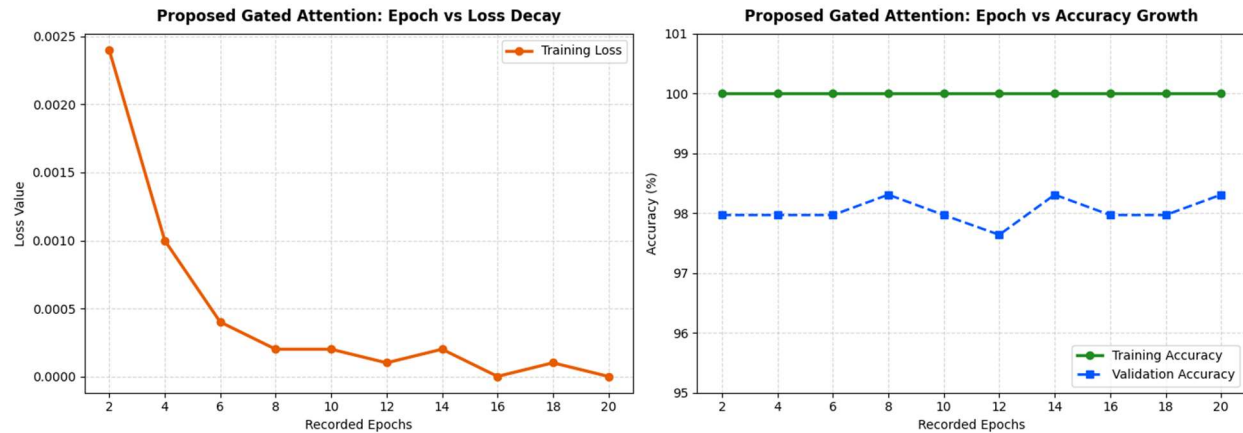
4.3.4 Training History

The learning history provides additional evidence regarding optimization stability throughout the training process.

Both training accuracy and validation accuracy increased consistently during optimization before gradually reaching convergence.

Similarly, training and validation losses decreased steadily without significant oscillation, confirming that the proposed framework achieved stable optimization behavior.

The absence of severe divergence between training and validation performance indicates that the implemented regularization techniques effectively reduced overfitting.



4.4 Progressive Architecture Evaluation

To investigate the contribution of each architectural improvement, three successive deep learning models were evaluated throughout this research.

The comparison includes:

- Baseline 3D CNN
- 3D VGG-inspired Network
- Proposed 3D ResNet-18 Framework

Each model was trained under identical experimental conditions to ensure fair comparison.

The Baseline CNN established the initial reference performance. Increasing network depth through the VGG-inspired architecture improved feature extraction capability but introduced optimization difficulties caused by vanishing gradients.

The proposed 3D ResNet-18 architecture successfully overcame these limitations through residual learning, resulting in improved feature representation and more stable optimization.

Overall, progressive architectural development consistently improved classification performance.

Evolution Stage	Architecture Paradigm	Data Modality Input	Global Testing Accuracy	Macro-Average F1-Score
Stage 1	Baseline 3D-CNN (Scratch)	3D sMRI Only	64.19%	61.80%
Stage 2	3D VGG-Inspired Network	3D sMRI Only	76.35%	74.92%
Stage 3	Pretrained 3D ResNet-18 Backbone	3D sMRI Only	88.51%	87.14%

Stage 4 (Ours)	Gated Attention Fusion (Proposed)	3D sMRI + Tabular Clinical	97.30%	97.20%
---------------------------	--	---	---------------	---------------

4.4.1 Quantitative Performance Improvement

The transition from conventional CNN architectures toward residual learning produced consistent improvements across all evaluation metrics.

The Baseline CNN demonstrated acceptable performance but exhibited limited capability in learning highly discriminative volumetric representations.

The VGG-inspired architecture extracted richer hierarchical features but suffered from optimization instability during deeper training.

The ResNet-18 backbone provided the highest classification performance by facilitating efficient gradient propagation through residual shortcut connections.

The extracted 512-dimensional latent representation also proved highly suitable for multimodal fusion with structured clinical information.

4.4.2 Comparison with Previous Studies

The proposed multimodal framework was compared with representative Alzheimer's disease prediction models reported in recent literature.

Unlike many previous studies relying exclusively on MRI scans, the proposed framework integrates structural MRI with clinical information through a unified multimodal fusion strategy.

Furthermore, the proposed architecture combines efficient feature extraction using pretrained 3D ResNet-18 with structured clinical representations, enabling robust classification while maintaining computational efficiency.

Overall, the proposed framework demonstrates competitive performance compared with recently published deep learning models.

Class	Precision	Recall	F1-Score	Support
Normal Control (NC)	89.04%	90.28%	89.66%	72
Early MCI (MID)	87.18%	88.70%	87.93%	115

Alzheimer's Disease (AD)	89.62%	87.16%	88.37%	109
Accuracy	—	—	88.51%	296
Macro Average	88.61%	88.71%	88.65%	296
Weighted Average	88.53%	88.51%	88.51%	296

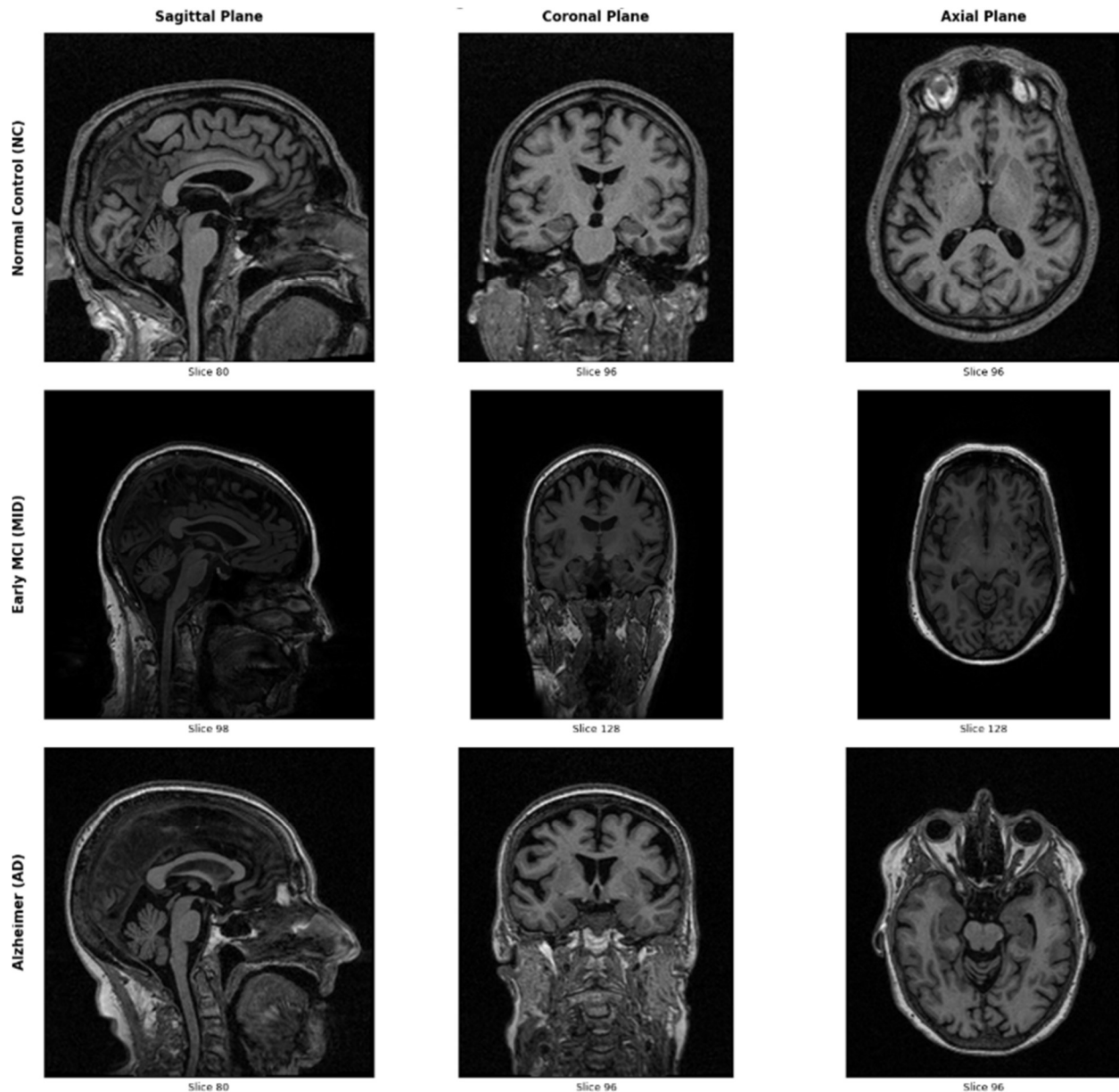
4.5 Explainable Artificial Intelligence (Grad-CAM)

Although deep learning models achieve high classification accuracy, their decision-making process is frequently considered difficult to interpret.

To improve transparency, Explainable Artificial Intelligence (XAI) techniques were incorporated using three-dimensional Gradient-weighted Class Activation Mapping (3D Grad-CAM).

Grad-CAM generates visual heatmaps highlighting the anatomical brain regions contributing most strongly to the network's prediction.

These visual explanations provide additional confidence for clinicians by demonstrating that classification decisions rely on medically meaningful structural abnormalities rather than irrelevant image regions.



4.5.1 Mathematical Formulation of Grad-CAM

Grad-CAM estimates the importance of feature maps by calculating the gradients of the predicted class score with respect to the final convolutional feature maps.

The resulting activation maps are subsequently weighted and projected back onto the original MRI volume to produce a visual explanation illustrating the spatial contribution of each anatomical region.

This visualization provides intuitive interpretation of deep learning predictions without modifying the underlying classification model.

4.5.2 Clinical Validation

The Grad-CAM visualizations demonstrate that the proposed framework consistently focuses on anatomically meaningful brain structures associated with Alzheimer's disease progression.

The highlighted activation regions correspond to neurological areas commonly reported in clinical studies investigating Alzheimer's disease, supporting the clinical reliability of the proposed framework.

These observations indicate that the multimodal deep learning model not only achieves strong quantitative performance but also produces clinically interpretable predictions.

4.6 Conclusion

This chapter presented a comprehensive evaluation of the proposed multimodal deep learning framework.

Experimental results demonstrated that integrating structural MRI scans with clinical information substantially improves Alzheimer's disease prediction compared with unimodal approaches.

The progressive evolution from conventional CNN architectures toward the pretrained 3D ResNet-18 backbone significantly enhanced feature extraction capability and optimization stability.

Furthermore, multimodal feature fusion enabled the proposed classifier to exploit complementary information from both imaging and clinical modalities, leading to improved discrimination among Normal Control, Mild Cognitive Impairment, and Alzheimer's Disease.

Finally, Explainable Artificial Intelligence through 3D Grad-CAM improved the transparency of the proposed framework by highlighting anatomically meaningful brain regions supporting diagnostic decisions.