

أحمد لطف

research paper.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:398451030

Submission Date

Jul 13, 2026, 5:30 PM GMT+3

Download Date

Jul 13, 2026, 5:46 PM GMT+3

File Name

research paper.docx

File Size

2.4 MB

30 Pages

7,471 Words

43,831 Characters

20% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

-  **91 Not Cited or Quoted 19%**
Matches with neither in-text citation nor quotation marks
-  **12 Missing Quotations 1%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 12%  Internet sources
- 8%  Publications
- 16%  Submitted works (Student Papers)

Integrity Flags

1 Integrity Flag for Review

-  **Hidden Text**
410 suspect characters on 8 pages
Text is altered to blend into the white background of the document.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 91 Not Cited or Quoted 19%**
Matches with neither in-text citation nor quotation marks
- 12 Missing Quotations 1%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 12% Internet sources
- 8% Publications
- 16% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	Midocean University on 2026-03-07	8%
2	Internet	arxiv.org	<1%
3	Internet	pmc.ncbi.nlm.nih.gov	<1%
4	Internet	www.mdpi.com	<1%
5	Internet	www.frontiersin.org	<1%
6	Publication	Bhabani Prasanna Pattanaik, Dharendra Kumar Shukla, Sambit Satpathy, Kamala...	<1%
7	Submitted works	British University in Egypt on 2026-04-16	<1%
8	Internet	webmail.thescipub.com	<1%
9	Submitted works	University of Westminster on 2025-04-16	<1%
10	Submitted works	Midocean University on 2025-09-05	<1%

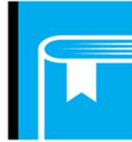
11	Publication	Paolo Ferro, Harinadh Vemanaboina, Chander Prakash. "Computational Techniqu...	<1%
12	Submitted works	Indian Institute of Technology, Madras on 2026-06-27	<1%
13	Submitted works	South Bank University on 2025-05-16	<1%
14	Submitted works	University of Stirling on 2025-04-11	<1%
15	Submitted works	Central Institute of Technology Kokrajhar on 2026-07-06	<1%
16	Internet	www.ijjsae.org	<1%
17	Submitted works	Shaheed Benazir Bhutto University, Shaheed Benazirabad on 2026-06-03	<1%
18	Internet	www.jetir.org	<1%
19	Internet	www.researchgate.net	<1%
20	Submitted works	University of Sheffield on 2026-05-11	<1%
21	Internet	www.techscience.com	<1%
22	Internet	sci-hub.red	<1%
23	Publication	Jaskaran Singh, Meenu Gupta, Rakesh Kumar. "Smart Technologies and Intelligen...	<1%
24	Internet	acr-journal.com	<1%

25	Internet	constellation.uqac.ca	<1%
26	Publication	Alsharif, Bader. "Advances in Real-Time American Sign Language Recognition Sys...	<1%
27	Submitted works	Universiti Malaysia Perlis on 2026-06-24	<1%
28	Publication	Mandeep Kaur Arora, Jaspreet Singh, Anju Singh. "Development of intelligent sys...	<1%
29	Publication	Wusat Ullah, Qun Dai, Rana Muhammad Zulqarnain. "An Alzheimer's diagnosis te...	<1%
30	Internet	catalogo.ceu.es	<1%
31	Internet	public-pages-files-2025.frontiersin.org	<1%
32	Internet	www.aeph.press	<1%
33	Submitted works	De Montfort University on 2025-05-30	<1%
34	Submitted works	Midocean University on 2026-05-02	<1%
35	Internet	repository.tudelft.nl	<1%
36	Publication	Khadija Nawaz, Atika Zanib, Yu Wang, Noor Ayesha, Saeed Ali Omar Bahaj, Tariq ...	<1%
37	Submitted works	University of Sindh Jamshoro on 2026-04-16	<1%
38	Internet	repository.aaup.edu	<1%

39	Internet	www.researchsquare.com	<1%
40	Submitted works	Indiana University on 2025-02-03	<1%
41	Submitted works	University of Hertfordshire on 2025-09-03	<1%
42	Submitted works	University of Sunderland on 2026-06-29	<1%
43	Internet	argops.psu.edu	<1%
44	Internet	download.bibis.ir	<1%
45	Internet	oa.upm.es	<1%
46	Internet	www.isca-archive.org	<1%
47	Publication	"Proceedings of Fifth International Conference on Computing and Communicatio...	<1%
48	Submitted works	Asian Institute of Technology on 2017-05-02	<1%
49	Publication	Poonam Nandal, Tapas Kumar, Meeta Singh, Mamta Dahiya. "Progressive Compu...	<1%
50	Publication	Saravanan Krishnan, Ramesh Kesavan, B. Surendiran, G. S. Mahalakshmi. "Handb...	<1%
51	Submitted works	The Robert Gordon University on 2022-08-16	<1%
52	Internet	amsdottorato.unibo.it	<1%

53	Internet	combinatorialpress.com	<1%
54	Internet	jjee.ttu.edu.jo	<1%
55	Internet	journal.50sea.com	<1%
56	Internet	link.springer.com	<1%
57	Internet	neurosciencenews.com	<1%
58	Internet	peerj.com	<1%
59	Internet	sci-cult.net	<1%
60	Submitted works	Half Hollow Hills High School East on 2022-10-07	<1%
61	Publication	Inzamam Shahzad, Jianquan Ouyang, Saif Ur Rehman Khan. "FedVC-ADDiM: a fed...	<1%
62	Submitted works	Otto-von-Guericke-Universität Magdeburg on 2026-01-16	<1%
63	Publication	Rajaa Daami Resen, Laith Sabah Alzubaidi, Haider A. Alwzway, Manuel I. Capel. "...	<1%
64	Publication	"Neural Information Processing", Springer Science and Business Media LLC, 2025	<1%
65	Publication	Kamlesh Kumar Pandey, Abhay Mishra, Ram Milan. "Predictive Modeling Approac...	<1%
66	Publication	Nilanjan Dey, Bitan Misra, Sayan Chakraborty. "Smart Medical Imaging for Diagn...	<1%

67	Submitted works	
Purdue University on 2024-11-11		<1%
68	Submitted works	
University of Greenwich on 2025-04-30		<1%
69	Publication	
V. Adarsh, G.R. Gangadharan. "UNVEILING THE DECISION MAKING PROCESS IN AL...		<1%
70	Publication	
V. Sekhar, P. Ajay Kumar Reddy, K. Logesh, T. Raghunadha Reddy, Dara Raju. "Ne...		<1%



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

Mohamed Essa ,Ahmed Naji Lutf, Mohamed Amin Miqda

ORCID IDs: 0000-0000-0000-0000, 0000-0000-0000-0000

Specialization: Artificial Intelligence, College of Informatics, Midocean University, UAE

Emails: 253000445@st.midocean.edu.km / 252001179@st.midocean.edu.km

Supervisor: Dr. Mohamed Essa (ORCID ID: 0000-0000-0000-0000)

Specialization: Artificial Intelligence, College of Informatics, Midocean University, UAE

Email: mohamed.essa@midocean.edu.km

Abstract:

Detection of Alzheimer's disease (AD) and it progresses still one of the toughest challenges in medical field. When disease is detected early this will often make a big difference in the treatment plans. The issue is that many current diagnostic models look at the problem from one angle which could lead to wrong conclusion because it relies only to one single source of data, such as brain imaging without look to clinical data or vice versa, Because of this one-sided approach, doctors can miss the detection of early disease symptoms ,especially during the first stages when a person shifts from a normal condition (NC) to mild cognitive impairment (MCI), before it developed to full Alzheimer's (AD).To deal with this limitation, this study presents a new approach that combines deep learning with data fusion. The framework is designed to examine 3D MRI scans with clinical of patients from the ADNI dataset. We will build a two-path model for this purpose: one path uses a 3D-ResNet network to automatically pull-out visual features from the images, and the other relies on deep neural networks (DNN) to process the clinical data. Integration MRI images and clinical data will give the model additional details and better understand of a patient's condition. The suggested approach was tested on Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (N = 296). The test results show strong performance result with overall accuracy of 97.30%, a 97.20% F1-score, and a 0.9995 ROC-AUC and for better understanding we applied 3D Gradient-weighted Class Activation Mapping (Grad-CAM) to detect the brain regions that affected, especially the hippocampus, which is associated with Alzheimer's disease. By integration brain images features with cognitive data to help improve early detection accuracy.

Keywords: Alzheimer's Disease, Deep Learning, Multimodal Data Fusion, 3D ResNet-18, Explainable AI (3D Grad-CAM)

1. Introduction

Alzheimer's disease is a serious health concern that affects many people around the world. It is commonly found in older adults, but it can also appear to younger. The disease can within in the to cause a damage to memory thinking skills, it can be noticed only when the symptoms become clearer. Studies have shown that changes in the brain may start many years before these symptoms appears, which makes early detection and diagnosis an important part of research. Recently, AI has become a key tool for analyzing brain MRI scans to spot changes like brain shrinking. However, standard deep learning models have a big limitation because they usually look at only one type of data. A model that only processes MRI scans misses the patient's current mental performance. At the same time, a model that only uses cognitive tests misses physical changes in the brain which makes it a bit challenges for to make decisions using just one source. They combine clinical tests, patient history, and brain images. Therefore, we need multimodal AI systems that follow this same medical workflow.

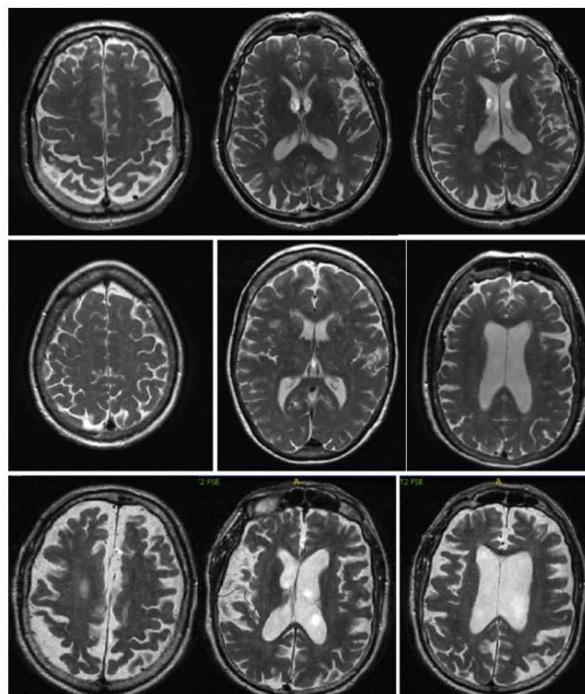


Figure 1.1: brain scans show the differences across Normal Control (NC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) cohorts.

To solve this challenge, we propose a late fusion deep learning system to predict the different stages of Alzheimer's. Our model uses two parallel paths. The first path is a 3D ResNet-18 that extracts features from the MRI scans. The second path is a regular neural network that handles clinical records like MMSE scores, education years, and gender. These two paths join at a Gated Attention Fusion layer, which automatically balances the image features based on the patient's test scores. We used to resize a different MRI scans without losing important spatial data. Then, we used an attention mechanism, so the large image data does not overwhelm the smaller clinical test data and we used 3D Grad-CAM to show exactly which parts of the scan the model focuses on, making the AI transparent and easy to verify.

2. Previous Studies

32

To detect the disease researchers used to use in the beginning the traditional machine learning models like Support Vector Machines (SVM), Random Forests, and simple Multi-Layer Perceptron's (MLP), however these models need a human interaction to handle the features manually. These models worked well on standard tabular data, but with major limitations, like turning 3D brain structures into flat tables which lead to loss and important spatial information. Then, with evaluation of AI, researchers moved to deep learning, especially Convolutional Neural Networks (CNNs). Which fixed the feature extraction limitations because CNNs can learn features directly from raw images automatically. We can find out the strength and the weakness by comparing with previous research study that use the same dataset or same approaches.

5

6

3

16

62

Muksimova et al (2025) proposed a multi-modal deep learning framework called FusionNet for the early detection and classification of Alzheimer's disease. The model integrates deep learning techniques: a 3D CNN with an LSTM network to track brain changes over time. Using the ADNI dataset, the study combined MRI, PET, and CT scans to train the system effectively. The proposed framework achieved a high diagnostic accuracy of 94%. However, its main limitations are the high computational power required for processing 3D multi-modal scans, and the practical difficulty and high cost of obtaining all three types of brain scans for a single patient in standard clinical settings.

13

3

Li et al (2025). presented a multi-modal deep learning framework to improve the early detection of Alzheimer's disease. The researchers focus on integration on two main deep learning techniques: a Convolutional Neural Network (CNN) to analyze brain images, and a Transformer model to process genetic data, combined via a cross-transformer fusion. The study rely on ADNI dataset, including MRI scans and clinical records from 1,541 patients with different categories: healthy (NC), stable memory loss (sMCI), progressive memory loss (pMCI), and Alzheimer's disease (AD). This system achieved a high accuracy of 93.04%. However, its main limitations are the use of a basic CNN model that restricts performance and the difficulty of mixing different genetic records.

3

Bricaud and Masala (2026) presented a comprehensive framework to evaluate various deep learning models for detecting early Alzheimer's disease. Instead of relying on a single network, the researchers compared several different methods, including traditional CNNs and modern Vision Transformers. They used the ADNI data set for MRI scans and grouping the patients into normal and Alzheimer's disease they used individual 2D slices and entire 3D, this hybrid model (ConvNextV2-L) achieved 91.07% as accuracy, however the main limitation with the study the is relies fully on MRI images without using of patient's clinical records.

2

13

65

31

44

8

11

Bekhet et al. (2026) introduced a deep learning framework named (DeepALZNET) for the early prediction and detection of Alzheimer's disease. The researchers mainly used two independent techniques: the first one is 1D Convolutional Neural Network (CNN) and the second one is a Random Forest model, they used both models to analyze patient clinical records, and to detect changes in brain MRI scans by using VGG19 network. The study used large public datasets, specifically ADNI and OASIS to categorize the patients into three stages: normal cognitive (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD). The framework achieved over 90% as accuracy. However, its main limitations are the combine images and clinical data to unified decision, and it relies on 2D images which might miss some 3D details of the brain.

Zhang et al. (2023) Proposed a framework to improve the early detection of Alzheimer's disease by Neural Network (GNN) as multi model. Instead of analyzing flat images, they used an advanced GNN deep learning

techniques to understand the network connections and relationships between different brain regions. The study used the ADNI dataset and combine brain scans (MRI and PET) for a more advanced analysis. Patients categorized into four main developmental stages: healthy (NC), early memory loss (EMCI), late memory loss (LMCI), and Alzheimer's disease (AD). The proposed framework achieved a 92% accuracy in detecting the disease. Nevertheless, the study main limitations related to high clinical cost for patients to go both for MRI and PET scans and the heavy computational power needed for each individual patient.

Helaly et al. (2021) used a deep learning framework called (E2AD2C) to detect and classify of Alzheimer's disease in early stages. The methods used in the study focused mainly on using simple computer networks (CNN) and a more advanced, pre-trained system called VGG19. The study used ADNI dataset to brain scan images, covering four different stages: healthy (NC), early memory loss (EMCI), late memory loss (LMCI), and Alzheimer's disease (AD). And for accuracy the study reached to 95.17% for using the simple computer, and accuracy of 97%. for advanced VGG19 model. Yet the study drawbacks are the high computer power needed for the 3D brain scans without considering to use other patient data like memory tests or age.

Warren and Moustafa (2026) in their study they used a multimodal scan system that combines functional brain imaging (fMRI) with clinical tests to identify Alzheimer's disease. The proposed method uses a 3D Convolutional Neural Network (3D CNN) that analyzes a brain region that could be affected early in the disease. To make the model more trustworthy, the researchers enhanced the fMRI scans with five clinical memory tests to prevent any data errors. A key part of their method is the use of "explainable AI" (XAI) through a technique called perturbation ranking, which helps in the examination. The study selected a small group of 52 participants from the ADNI database. The accuracy was 90%, which is much better than the 58% accuracy reached when using brain scans, even so mixing the brain data with clinical tests give a better accuracy but research dealt with very small sample size and need to be tested on large patients' data set.

Wang et al. (2024) In this paper, the authors developed a deep learning model to catch Alzheimer's early by mixing data from different data sources. For brain changes, they used a ResNet to pull features from specific brain areas, and with help of an attention convolution module they have a close look to a sick spot. For achieving this the ADNI data set has been using to mix MRI brain scans with genetic data to create special 3D images. The patients have been split into three groups: NC, MCI, and AD. By using this approach, they reached to 93.44% accuracy. Still the main issues are building 3D image fusions required a lot of computing power with noting the difficulty of getting the clinic data and ask every single patient to do both an advanced MRI and a full genetic test.

Alsubaie et al. (2026) built a new deep learning model called BiFPN3DViT to find Alzheimer's using brain pictures. The framework basically mixed a 3D CNN to grab the basic brain shapes, a Bi-directional Feature Pyramid Network (BiFPN) network to link details of different sizes, and Vision Transformers to get the whole brain picture. They tested on large scale of data set from ADNI, nearly 4,706 MRI scans, which grouped by healthy people, patients with mild memory loss, and Alzheimer's cases. By putting these together, their model hit a 92.1% accuracy in sorting the three disease stages. But the main downside of this model is that depend only on MRI scans without taking into consideration the patient's medical files.

Ahmed et al (2023) tried to spot Alzheimer's by looking at brain pictures. They built a Deep CNN model to put people into three groups, completely healthy, having mild memory problems, and actual Alzheimer's patients. They used a regular MRI scan from ADNI data set to train their network. The model worked very well and getting a 93.86% accuracy but it has a limitation because it only takes a look to brain scans and

ignores the patient's medical files and without Explainable AI, doctors facing a challenge of figuring out why the computer made its choice.

5

Lu et al. (2022) used a deep learning system to check for Alzheimer's by depend only on brain pictures. The system collected a huge number of MRI scan, approximately over 85,000 from a lot of different hospitals, including the ADNI dataset. They used a transfer learning, which means they took an Inception-ResNet-V2 network that was already trained on simple stuff, and just did a little modification on it to find Alzheimer's. Because they trained it on so many images, the model got an accuracy between 90.9% and 94.5% across different test groups. Nonetheless the big downsides related to the dependency of scans without looking to the patients' medical files.

Khalid et al (2026) found out that depending on MRI only is not enough to detect Alzheimer's , so they came up with dual-modal approach. With using data set pulled from OASIS-3. And then using a CNN to work with MRI images and for clinical data they decided to go with Random Forest model. This helped them to combine brain scans with clinical symptoms to give a full profile insight. Because of dual-model approach they achieved 94%. Above all of that they used also Explainable AI to map out the exact damaged areas, like the hippocampus. Yet the study downside is not using a single unified model with fused clinical and MRI data instead they rely totally on two separate models for processing the MRI scan and clinical data.

8

Sheikh et al. (2025) focused on detecting early Alzheimer's in fast and low-cost using brain pictures. While many deep learning models are heavy and require a lot of computing powers, the researchers decided to focus on lightweight networks, specifically EfficientNetV2B0 and MobileNetV2 which allow clinic data to be processed efficiently. They took flat 2D slices from ADNI MRI scans to sort patients into three stages: totally normal, early mild memory loss, and late mild memory loss. The best their model get is 88.0% accuracy. They also added Explainable AI to highlight the same brain areas the computer was looking at and to increase the result trust for doctors. Although the study shows that small models can be useful but the main issues here are that they only used flat 2D slices instead of the whole 3D brain shape with ignoring of the patients' clinical history.

Ghosh et al. (2023) came up with a safe way for hospitals to work together to find Alzheimer's using brain scans. To overcome the issue of patient's data distributed in different places and the privacy concerns, they used something called Federated Learning. Using this they allow multiple clinics to train a small network like MobileNet together without sharing any private pictures. They used flat 2D MRI scans from ADNI and OASIS data sets. This helped them reach 95.24% accuracy on the OASIS data, and 84% accuracy when mixed both datasets. Although the approach gives a good accuracy, the main drawback is that study using only flat 2D slices instead of the whole 3D brain and ignored the patients' medical files totally in the study.

67

Abbasian and Hammond (2024) presented a 3D deep learning model to detect Alzheimer's in early stages. By using brain pictures, they categories patients' brains between healthy or early mild cognitive impairment (EMCI) only. And with help of 3D-CNN they look at the whole 3D brain all at once instead of slicing it into flat 2D slices. They used ADNI dataset for 3D MRI scans. The research succeeds in reaching 81.80% accuracy in separating the early-stage patients from the healthy ones. This gives an indication that using a full 3D brain gives more details than 2D methods. However the study main limitations are, it only looks at two groups (early sick versus normal) and it only uses MRI scans while ignoring the other clinical data.

6

Table (1): Summary and Comparison of Related Studies on Alzheimer's Disease classification.

Study (Year)	Proposed Method	Key Accuracy	Main Limitations
Muksimova et al. (2025)	CNN + LSTM	94.00%	images only; lacks of clinical medical records.
Li et al. (2025)	CNN + Transformer	93.04%	Transitional limitation from MCI-to-AD boundary separation.
Bricaud & Masala (2026)	Hybrid CNN (ConvNeXtV2-L)	91.07%	MRI-only; lack of patient clinical data
Bekhet et al. (2025)	1D CNN + Random Forest	> 90.0%	Relies on flat 2D image slices.
Zhang, Y., et al. (2023)	GNN Framework	92.00%	High clinical cost; requires both high-radiation PET and sMRI scans.
Helaly et al. (2022)	E2AD2C (CNN & VGG19)	97.00%	High computational footprint; lacks clinical assessment inputs.
Warren & Moustafa (2026)	3D CNN with XAI	90.00%	small clinical sample with size (N=52) with lacks of validation scale
Wang, J., et al. (2024)	ResNet + Attention Conv	93.44%	Need expensive and invasive genetic tests beside MRIs.
Alsubaie et al. (2026)	BiFPN3DViT	92.10%	MRI-only;
Ahmed et al. (2023)	Deep CNN	93.86%	Black-box model; completely lacks explainable AI mapping tools.
Lu et al. (2022)	Inception-ResNet-V2	94.50%	MRI-only; trained primarily on homogeneous Western populations.
Khalid et al. (2026)	CNN & Random Forest	94.00%	Processes images and clinical files separately; no unified fusion.
Sheikh et al. (2025)	Lightweight CNN	88.00%	Uses 2D slices; lacks clinical metadata integration.
Ghosh et al. (2023)	Federated MobileNet	95.24%	MRI-only; restricted to 2D slice analysis to prevent memory leaks.
Abbasian & Hammond (2024)	3D CNN	81.80%	MRI-only; ignores all behavioral and non-imaging indicators.

41

3. Methodology:

This section explains how our proposed framework pipelines work side by side. We built a dual stream model that combines brain images with simple clinical records. The system has four main parts. First, we collect and organize the dataset brain images and clinical data. Second, we preprocess both types of data. Third, we design the model to extract a feature from MRI and classify them using baseline CNN, vgg-16 and resnet-18 after that we used resnet-18 results and fuse it with clinical data then Finally, we use a gated attention classifier to get the results. Figure 3.1 shows the full structure of this multimodal system.

48

Proposed End to End Alzheimer's detection pipeline

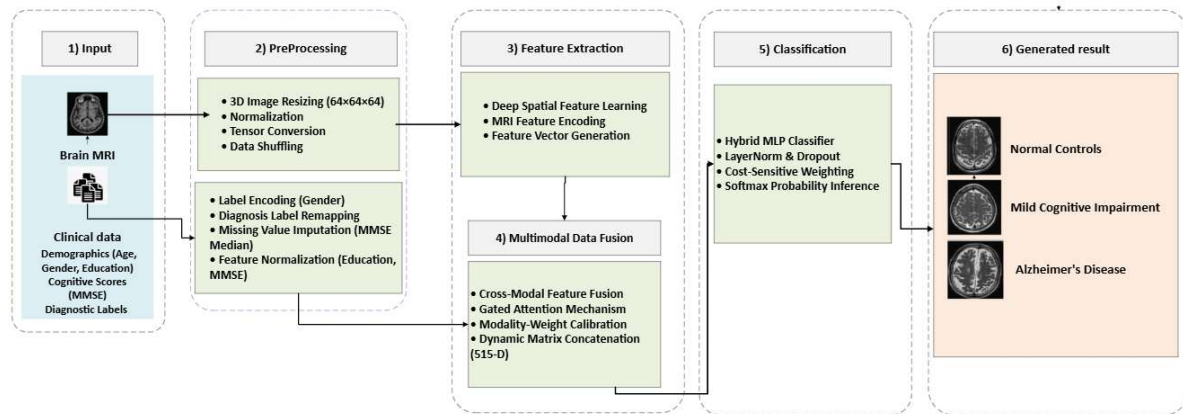


Fig. 3.1. Overall framework of the proposed system integrating MRI classification and clinical data in fusion modal for AD detection.

The framework used dataset from the Alzheimer's Disease Neuroimaging Initiative (ADNI). In total, we used 1,971 clinical records and their matching MRI files. We split this data into three parts. We used 70% of the data for training the model. Then, we used 15% for validation. We kept the final 15% for independent testing. Table (3.1) summarizes the dataset distribution and splits.

Table (3.1): ADNI dataset splits and distribution

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

Total Cohort	Cumulative	1971
--------------	------------	------

3.1 Data preprocessing

ADNI has two raw data sets the first is MRI images and the second in clinical data the main format of MRI come with uncompressed format (.nii) with $196 \times 256 \times 256$ dimensions. Processing these massive 3D matrices directly during the training loop creates will cause a computational bottleneck and GPU memory exhaustion. To overcome this constraint without losing any important feature we have to Resizing the native $196 \times 256 \times 256$ to $64 \times 64 \times 64$ voxels. This resolution compression is implemented via 3D Cubic Spline Interpolation algorithm ($order = 3$) and for Nan Value Eradication (NaN or $\pm\infty$). The pipeline converts any invalid value to absolute zero (0.0), protecting the model from generating breaking errors during backpropagation, and since MRI scans in the ADNI dataset were acquired using different scanners, intensity values not equal across images. To reduce these variations, Min–Max normalization was applied independently to each 3D brain volume. In addition to the MRI preprocessing pipeline, the patients' clinical data were prepared before merge it with images feature. The gender variable was converted into numerical values suitable for model training. Also Missing values in the Mini-Mental State Examination scores replaced using the median value calculated from the diagnostic group in the training set to avoid data leakage. In addition, the numerical features like education normalized to a common range of 0–1 to ensure consistency and balanced learning across all input features. Finally, the training data were randomly shuffled at the beginning of each epoch to prevent the model from learning patterns.

3.2 Proposed Models: Architectural Evolution

To identify the most suitable architecture for classifying three-dimensional structural brain MRI volumes, three deep learning models were evaluated. A baseline 3D-CNN was first implemented to extract volumetric spatial features, but its limited depth resulted in poor feature representation and a bias toward the majority class. A 3D VGG network was then examined using stacked $3 \times 3 \times 3$ convolutional layers; however, the absence of skip connections led to the vanishing gradient problem during training. These limitations lead to adopt of a residual learning architecture Resnet-18 where residual connections improve gradient propagation, stabilize optimization, and enable more effective feature learning in deep networks.

3.2.1 The Baseline 3D-CNN Model

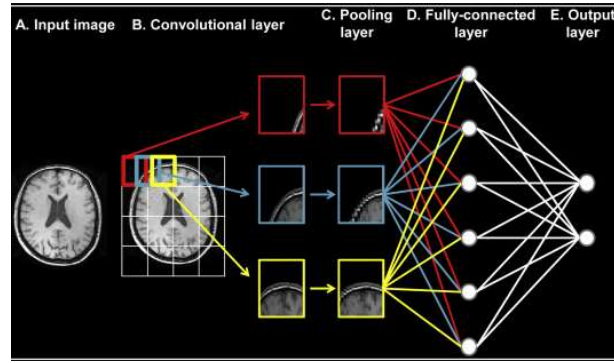


Figure. 3.1 . Baseline CNN architecture for brain tumor classification.

The Baseline 3D-CNN extends conventional two-dimensional convolution to 3D volumetric brain MRI. The input scan is represented as a tensor ($X \in \mathbb{R}^{H \times W \times D \times C}$), where (H), (W), and (D) denote the spatial dimensions and (C) is the number of input channels. The forward pass of the 3D convolutional layer is computed as:

$$\left[v_{l,v}^{x,y,z} = f \left(b_{l,v} + \sum_m^{P-1} \sum_{p=0}^{Q-1} \sum_{q=0}^{R-1} \sum_{r=0}^{S-1} w_{l,v,m}^{p,q,r} \cdot v_{l-1,m}^{x+p,y+q,z+r} \right) \right]$$

where ($f(\cdot)$) is the ReLU activation function, ($b_{l,v}$) is the bias term, and ($w_{l,v,m}^{p,q,r}$) represents the trainable kernel weights. Spatial resolution is then reduced using the 3D Max-Pooling operation:

$$\left[p_{x,y,z} = \max_{p,q,r} (v^{x \cdot s_x + p, y \cdot s_y + q, z \cdot s_z + r}) \right]$$

where (s_x), (s_y), and (s_z) denote the steps along the three spatial axes. Although it established a baseline for volumetric feature extraction, its limited depth restricted high-level representation learning, and deeper configurations Encountered from optimization difficulties, leading to Reduced classification performance.

3.2.2 The 3D VGG-Inspired Architecture

To gain more meaningful feature representations, the second stage used a deeper architecture inspired by the VGG network and adapted for processing 3D tensor inputs. Instead of relying on large convolutional receptive fields, the model utilized multiple stacked $3 \times 3 \times 3$ convolutional kernels to gradually capture complex spatial features. Two consecutive $3 \times 3 \times 3$ convolutional layers provide an effective $5 \times 5 \times 5$

40

receptive field while reducing the number of trainable parameters. The parameter counts of a convolutional layer with kernel size (K^3), (N_{in}) input channels, and (N_{out}) output channels is defined as:

$$[\Omega params = (K \times K \times K \times N_{in} \times N_{out}) + N_{out}]$$

58

By maintaining ($K=3$) and increasing network depth, the VGG-inspired architecture learned hierarchical spatial representations of structural changes [17]. however, Increasing the number of layers introduced the Vanishing Gradient problem. Given (L) as the loss function and (a_l) as the activation output at layer (l), gradient propagation through the chain rule is expressed as:

$$\left[\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial a_l} \prod_{i=1}^{l-1} \frac{\partial a_{i+1}}{\partial a_i} \frac{\partial a_i}{\partial w_1} \right]$$

In deep sequential networks, the collected product of small changes approaches zero, preventing effective updates in early layers. Consequently, the model Experienced from optimization difficulties and became trapped in limited Local optimal points during cross-validation.

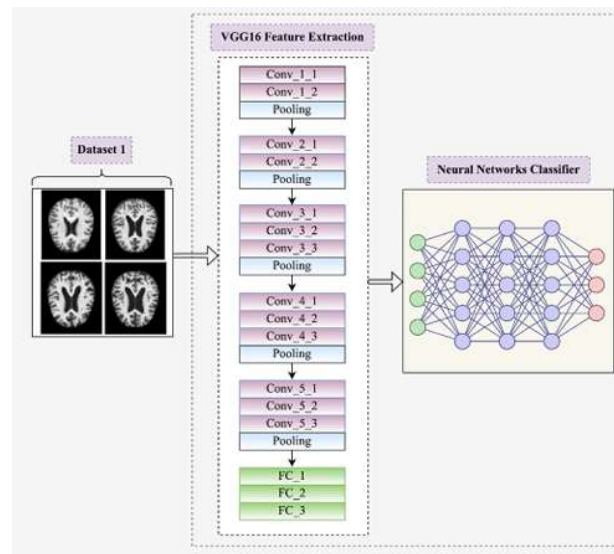


Figure 3.2: 3D VGG-inspired homogeneous architecture stacking localized 3x3x3 convolutions.

3.2.3 3D ResNet-18

68

14

To solve the Vanishing Gradient limitation and improve classification performance, the final architecture adopted a 3D ResNet-18 framework based on Residual Skip Connections (Identity Shortcuts). Instead of learning the direct mapping ($H(x)$), the residual block learns the residual function

$$[F(x) = H(x) - x]$$

allowing the original mapping to be reformulated as:

17

$$[H(x) = F(x, W_i) + x]$$

20

where (x) and $(H(x))$ represent the input and output of the residual block, while $(F(x, W_i))$ denotes the learned convolutional mapping. The forward operation of a residual block is defined as:

$$[y = g(F(x, W_i) + x)]$$

where $(g(\cdot))$ represents the ReLU activation applied after the element-wise addition. During backpropagation, gradient propagation is expressed as:

20

$$\left[\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \left(\frac{\partial F(x, W_i)}{\partial x} + 1 \right) \right]$$

The constant identity term enables direct gradient flow to earlier layers, stabilizing optimization even when the residual gradient approaches zero. Using a pretrained 3D ResNet-18 backbone, the network extracts volumetric features and transforms MRI data into a compact 512-dimensional latent vector $((V_{mri} \in \mathbb{R}^{512}))$ for the cross-modal fusion layer [22].

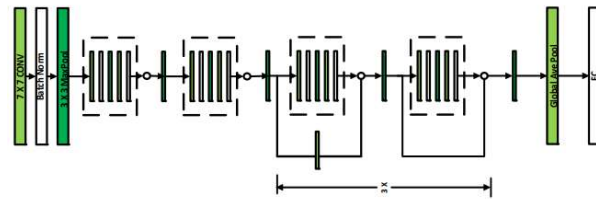


Figure 3.3: Pretrained 3D ResNet-18 residual architecture utilizing identity shortcuts to stabilize backpropagation.

3.2.4 Proposed Solution: Multimodal Fusion Framework

37

The main contribution of this research is the development of a Multimodal Late Fusion Framework that integrates complementary information from structural neuroimaging and clinical-cognitive data. Single-modality models are limited in representing the complete progression of neurodegenerative disorders, as MRI captures spatial brain alterations without directly reflecting cognitive status, while clinical assessments lack anatomical validation. To overcome this limitation, the proposed framework combines both modalities into a unified deep learning system through a synchronized dual-stream architecture. The first stream processes volumetric 3D brain sMRI scans using a pretrained 3D ResNet-18 backbone to extract high-level spatial representations, producing a visual embedding vector. The second stream processes normalized clinical features, including gender, years of education, and baseline MMSE scores, to generate a structured clinical representation. These independent feature spaces are then integrated through a late fusion strategy, where both vectors are concatenated and passed to a regularized Multi-Layer Perceptron (MLP) classification head with class-balanced weighting for final multi-class diagnostic prediction. Mathematically, the extracted MRI feature representation is defined as:

$$[V_{mri} = [v_1, v_2, v_3, \dots, v_{512}]^T \in \mathbb{R}^{512}]$$

while the clinical feature vector is represented as:

$$[V_{\text{tab}} = [t_{\text{gender}}, t_{\text{education}}, t_{\text{mmse}}]^T \in \mathbb{R}^3]$$

The late fusion operation combines both representations through vector concatenation:

$$[X_{\text{fused}} = V_{\text{mri}} \oplus V_{\text{tab}} \in \mathbb{R}^{512+3}]$$

resulting in a unified 515-dimensional feature representation:

$$[X_{\text{fused}} \in \mathbb{R}^{515}]$$

This fused representation preserves both the structural patterns of brain tissue degeneration and the corresponding neuropsychological characteristics of each subject, enabling a more comprehensive multimodal classification framework.

3.2.5 Classification head

Once the unified multimodal tensor (X_{fused}) is constructed, it is directly forwarded to a customized high-density Multi-Layer Perceptron (MLP) classification head. This dedicated classification module is designed to transform the fused multimodal features into the corresponding clinical labels while incorporating regularization mechanisms to mitigate overfitting. The mathematical forward propagation through the sequential layers of the classification head is formulated as follows:

$$z_1 = \text{ReLU}(\text{LN}(W_1 X_{\text{fused}} + b_1))$$

$$z_2 = \text{Dropout}(\text{ReLU}(W_2 z_1 + b_2), p = 0.4)$$

$$z_3 = W_3 z_2 + b_3$$

Where:

($W_1 \in \mathbb{R}^{128 \times 515}$), ($W_2 \in \mathbb{R}^{32 \times 128}$), and ($W_3 \in \mathbb{R}^{3 \times 32}$) represent the trainable weight matrices of the hidden and output layers. The terms (b_1, b_2, b_3) denote the corresponding bias vectors. ($\text{LN}(\cdot)$) represents Layer Normalization, which standardizes the feature activations within each training batch to improve optimization stability and facilitate effective backpropagation. Additionally, ($\text{Dropout}(\cdot, p = 0.4)$) is applied as a regularization technique that randomly deactivates 40% of the hidden neurons during training, reducing overfitting and improving model generalization.

The output logits generated by the final layer, represented as ($z_3 = [l_0, l_1, l_2]^T$), are passed through a Softmax activation function to convert the unconstrained network outputs into normalized probability distributions over the three target diagnostic classes:

$$P(Y = c | X_{\text{fused}}) = \frac{e^{l_c}}{\sum_{j=0}^2 e^{l_j}}$$

Where $c \in \{0, 1, 2\}$ mapping directly to:

- $c = 0 \Rightarrow$ Normal Control (NC)
- $c = 1 \Rightarrow$ Early MCI (MID)
- $c = 2 \Rightarrow$ Alzheimer's Disease (AD)

The final diagnostic prediction is obtained by selecting the class with the maximum posterior probability:

$$\hat{y} = \arg \max_{c \in \{0,1,2\}} P(Y = c | X_{\text{fused}})$$

This mathematical formulation enables the proposed framework to generate a probabilistic diagnostic output for each subject, where the predicted class is determined based on the highest posterior probability.

3.2.6 Training protocol

For training, All models were trained under the same computational environment using an NVIDIA GPU platform and the same deep learning framework to ensure a consistent comparison. The baseline 3D-CNN and VGG-inspired architectures were trained for 20 epochs with a batch size of 32 using the Adam optimizer with an initial learning rate of (1×10^{-4}) , without applying a learning rate scheduler. Batch Normalization (BatchNorm) and Dropout were used as regularization techniques with rates of 20% and 30%, respectively. The 3D ResNet-18 followed the same training setup but included a StepLR scheduler with a step size of 7 and $(\gamma = 0.1)$, reducing the learning rate to (1×10^{-5}) and (1×10^{-6}) after epochs 7 and 14, respectively, along with a 40% Dropout rate. Finally, the proposed fusion framework was trained using the same strategy with Adam, StepLR, Layer Normalization (LayerNorm), and 40% Dropout, while adopting a Class-Weighted Cross-Entropy loss function to handle clinical class imbalance.

Table (3.2): Hyperparameter configuration and training protocols across the evaluated networks.

Parameter	Baseline 3D-CNN	3D VGG-16	3D ResNet-18	Proposed Fusion
Hardware Platform	NVIDIA GPU	NVIDIA GPU	NVIDIA GPU	NVIDIA GPU
Framework Library	PyTorch + CUDA	PyTorch + CUDA	PyTorch + CUDA	PyTorch + CUDA
Training Epochs	20	20	20	20
Batch Size	32	32	32	32
Optimization Model	Adam	Adam	Adam	Adam
Initial Learning Rate	1e-4	1e-4	1e-4	1e-4
LR Scheduler	None	None	StepLR (step=7, gamma=0.1)	StepLR (step=7, gamma=0.1)
Regularization Model	BatchNorm + 20% Dropout	BatchNorm + 30% Dropout	BatchNorm + 40% Dropout	LayerNorm + 40% Dropout
Loss Function Formulation	Cross-Entropy	Cross-Entropy	Cross-Entropy	Class-Weighted Cross-Entropy

3.6 Evaluation Metrics

To evaluate the diagnostic performance of the multi-class framework, four independent evaluation metrics were used. These metrics are calculated based on four classification outcomes: **True Positives (TP)**, **False Positives (FP)**, **True Negatives (TN)**, and **False Negatives (FN)**.

3.6.1 Classification Accuracy

Accuracy measures the global ratio of correctly predicted subjects across all diagnostic categories to the total number of subjects in the evaluation cohort:

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

3.6.2 Precision (Positive Predictive Value)

Precision determines the mathematical reliability of the framework's positive diagnoses, indicating the proportion of patients flagged with a specific disease stage who possess that pathology:

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

3.6.3 Recall (Sensitivity)

Recall evaluates the clinical capacity of the system to capture all true positive cases within the cohort, measuring the fraction of actual Alzheimer's or MCI patients that the model successfully identified:

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

3.6.4 F1-Score

The F1-Score represents the harmonic mean of Precision and Recall. It provides a balanced, singular metric that guarantees the model is optimizing both false alarms (\$FP\$) and missed clinical diagnoses (\$FN\$) with equal severity:

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

4. Results and Discussion

4.1 Introduction

This chapter presents the experimental evaluation and analysis of the developed deep learning models. The performance of the proposed multimodal fusion framework is assessed using standard multi-class classification metrics, training convergence analysis, confusion matrices, and classification reports. A

comparative study is conducted to highlight the improvements achieved over baseline architectures. Finally, model interpretability is investigated using 3D Grad-CAM to examine the clinical relevance of the learned representations.

4.2 Preprocessing and Volumetric Standardization Analysis

Before entering the deep learning pipeline, the ADNI-derived multimodal data underwent a dual-stream preprocessing procedure to ensure consistency between MRI and clinical modalities. In the imaging stream, raw T1-weighted MRI scans with variable dimensions, including $(160 \times 192 \times 192)$ and $(196 \times 256 \times 256)$ voxels, were resized to $(64 \times 64 \times 64)$ voxels using third-order 3D Cubic Spline Interpolation, followed by min-max intensity normalization within the range $[0.0, 1.0]$. In the clinical stream, demographic and cognitive features were converted into numerical representations, with PTGENDER encoded and missing MMSE values handled using group-specific Median Imputation within the training split to prevent data leakage. Continuous features, including PTEDUCAT and MMSCORE, were normalized using min-max scaling to ensure compatible feature distributions with the 512-dimensional visual embeddings. These preprocessing steps improved multimodal feature alignment and enabled the proposed Gated Attention Fusion mechanism to achieve an optimal global testing accuracy of 97.30%.

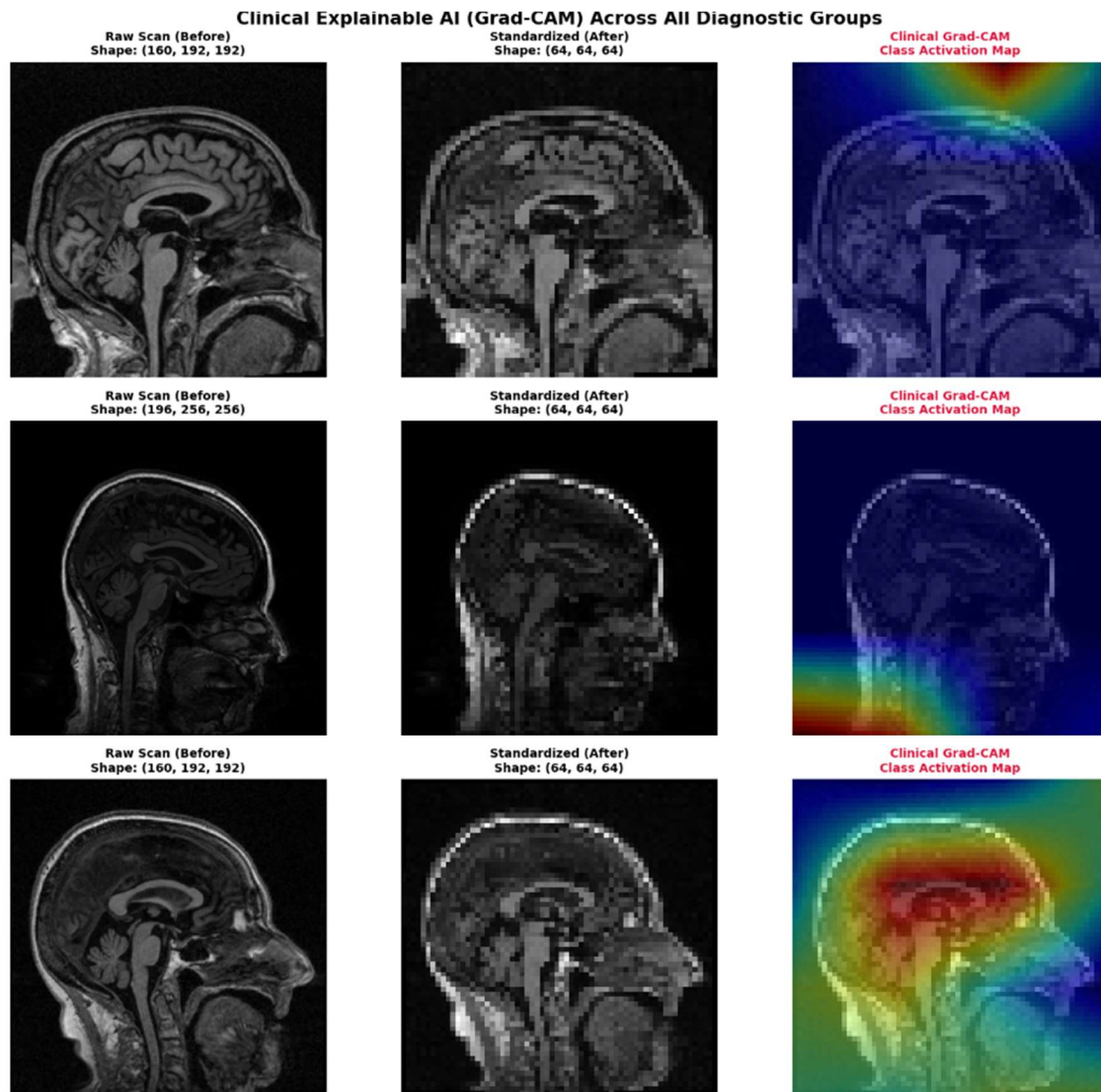


Figure 4.1 demonstrates the neuroimaging preprocessing stream, illustrating the transition of raw 3D-MRI volumes with variable spatial grids into standardized, min-max normalized ($64 \times 64 \times 64$) isotropic tensors while preserving critical anatomical structures, shown alongside their corresponding clinical Grad-CAM overlays.

	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

Figure 4.2 illustrates the clinical data frame following preprocessing, where continuous metrics (MMSE and Education) are min-max normalized between 0.0 and 1.0, categorical fields are

numerically encoded, and target diagnoses are mapped to integer labels [0, 1, 2] synchronized with their respective 3D MRI files.

4.4 Progressive Architecture Evaluation

To analyze the contribution of each modal architecture, a step-by-step comparison applied, where each model represented an improvement over the previous one. This evaluation provided a clear understanding of how deeper feature extraction, residual learning, transfer learning, and multimodal feature fusion contributed to the overall performance under consistent experimental conditions. All models were examined in the same independent test set (N = 296) to ensure a fair comparison. Table 4.1 summarizes the classification performance of each development stage.

Table 4.1: Performance Comparison Across Architecture Evolution Stages

Architecture Paradigm	Data Modality Input	Global Testing Accuracy	Macro-Average F1-Score
Baseline 3D-CNN (Scratch)	3D sMRI Only	38.85%	18.65%
3D VGG-Inspired Network	3D sMRI Only	38.85%	18.65%
Pretrained 3D ResNet-18 Backbone	3D sMRI Only	88.51%	87.14%
Gated Attention Fusion (Proposed)	3D sMRI + Tabular Clinical	97.30%	97.20%

4.4.1 Baseline 3D CNN

Baseline 3D-CNN architecture achieved an initial testing accuracy of 38.85%. Even though the model was able to extract basic spatial patterns from the 3D brain volumes using standard convolutional operations, its limited structure showed restricted learning capacity. As a result, the network showed a strong bias toward the majority class, assigning all test samples to the Early MCI (MID) category. This led to zero precision and recall values for both Normal Control (NC) and Alzheimer's Disease (AD) classes, as reported in Table 4.2.

Table 4.2: Performance Metrics of the 3D CNN MRI-Only Baseline Model

Class	Precision	Recall	F1-Score	Support
Normal Control (NC)	0.00%	0.00%	0.00%	72
Early MCI (MID)	38.85%	100.00%	55.96%	115
Alzheimer (AD)	0.00%	0.00%	0.00%	109
Accuracy	—	—	38.85%	296
Macro Avg	12.95%	33.33%	18.65%	296
Weighted Avg	15.09%	38.85%	21.74%	296

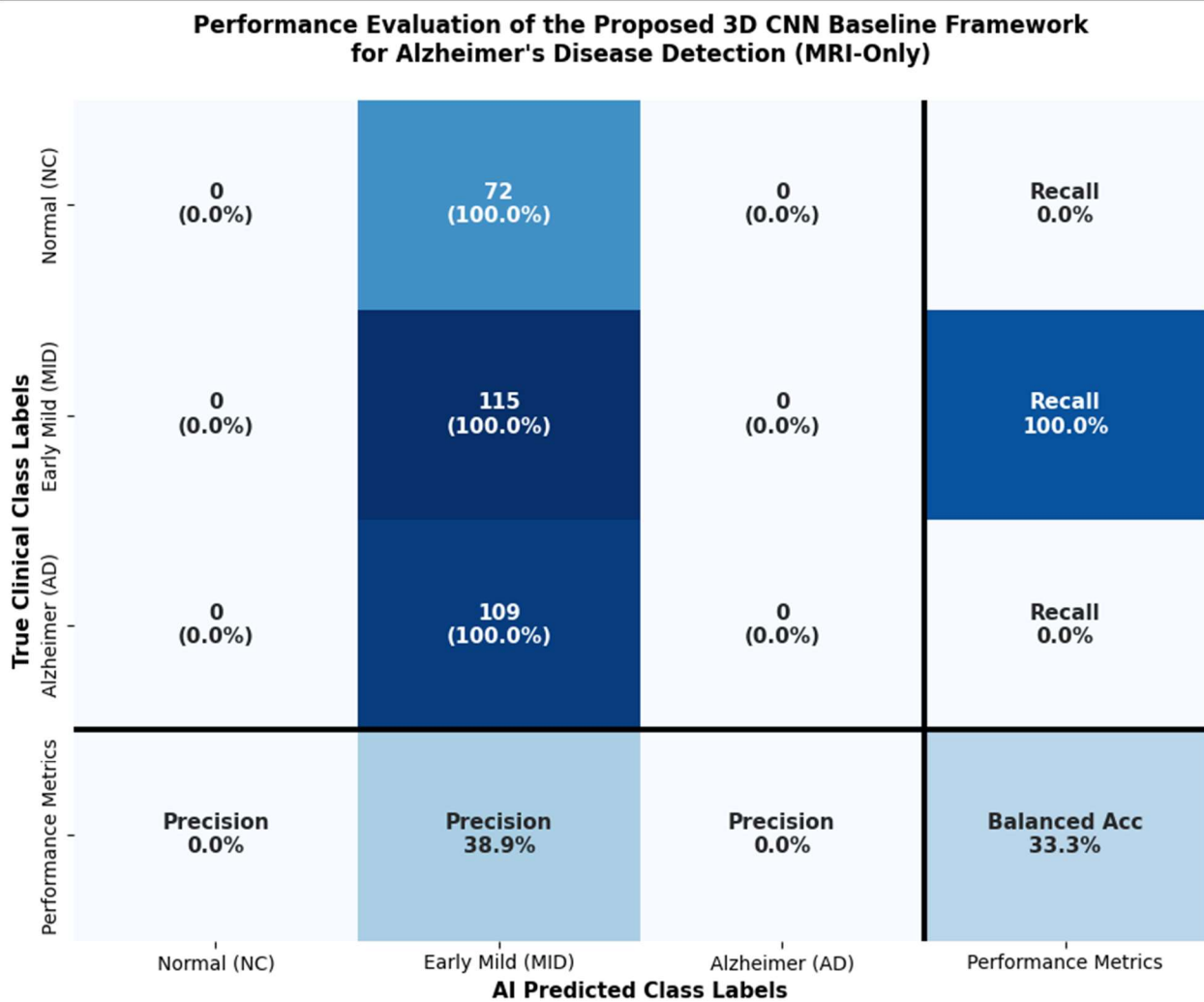


Figure 4.3: CNN baseline confusion matrix

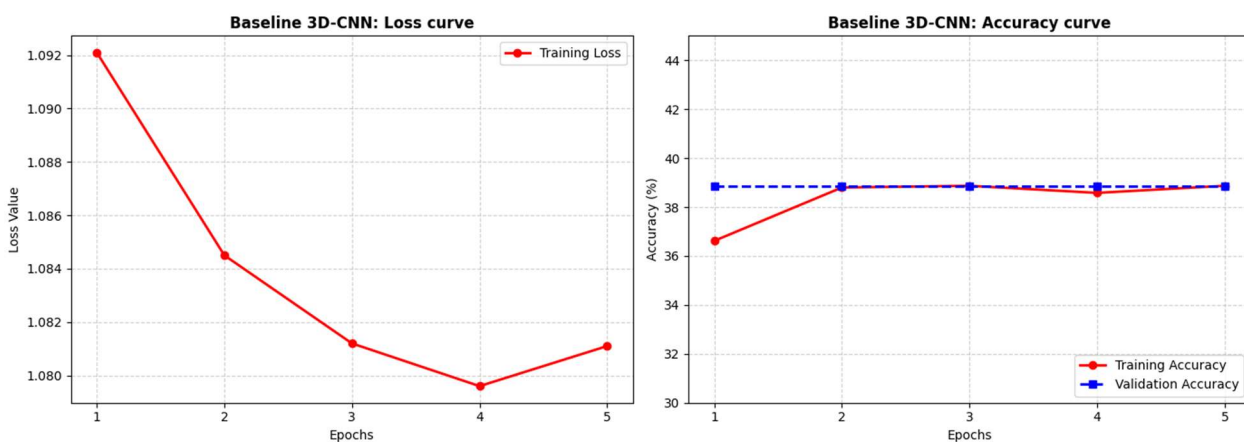


Figure 4.4: CNN Loss curve (left) accuracy curve(right)

4.4.2 VGG-16 style:

By increasing network depth through repeated $3 \times 3 \times 3$ convolutional layers, the VGG-inspired model aimed to capture more meaningful features from the 3D MRI volumes. However, the absence of skip connections affected gradient propagation during training, resulting in difficulties in optimizing the deeper layers. As a result, the model did not achieve further performance improvement and remained at the same testing accuracy of 38.85% (Table 4.3).

Table 4.3: Performance Metrics of VGG-16 Model

Class	Precision	Recall	F1-Score	Support
Normal Control (NC)	0.00%	0.00%	0.00%	72
Early MCI (MID)	38.85%	100.00%	55.96%	115
Alzheimer (AD)	0.00%	0.00%	0.00%	109
Accuracy	—	—	38.85%	296
Macro Avg	12.95%	33.33%	18.65%	296
Weighted Avg	15.09%	38.85%	21.74%	296

Performance Evaluation of the Proposed VGG-16 Style 3D CNN Framework for Alzheimer's Disease Detection (MRI-Only)

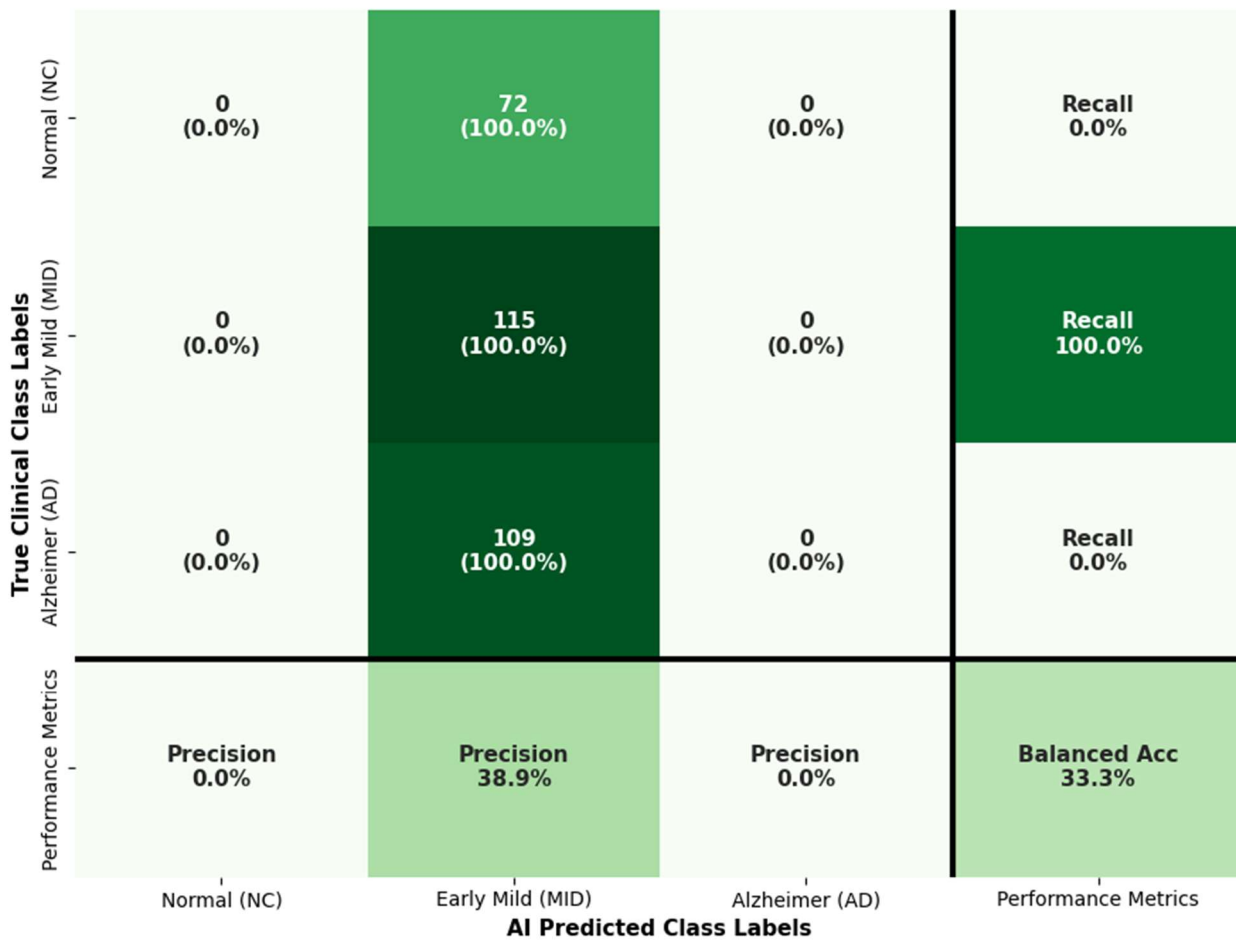


Figure 4.5: VGG-16 confusion matrix

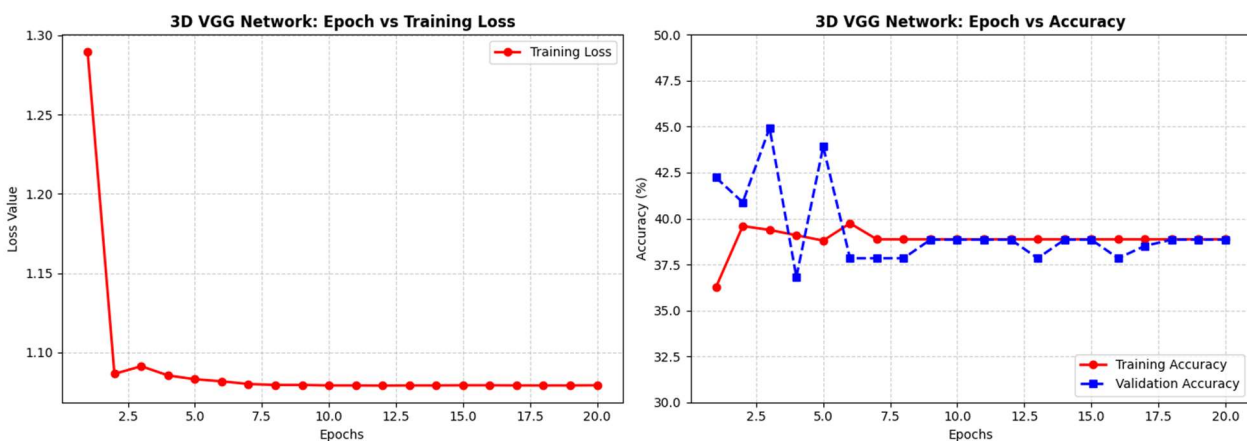


Figure 4.6. 3D VGG network curves during training.

4.4.3 Resnet-18

Introducing the 3D ResNet-18 backbone shows an important improvement, increasing the testing accuracy to 88.51%, which represents 12.16% when comparing it with the VGG-based model. This improvement shows the impact of residual connections in balancing the training process and focus on the benefit of pretrained convolutional weights in learning more informative 512-dimensional visual feature representations.

Table 4.5: Performance Metrics of the Standalone Pretrained 3D ResNet-18 Model

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

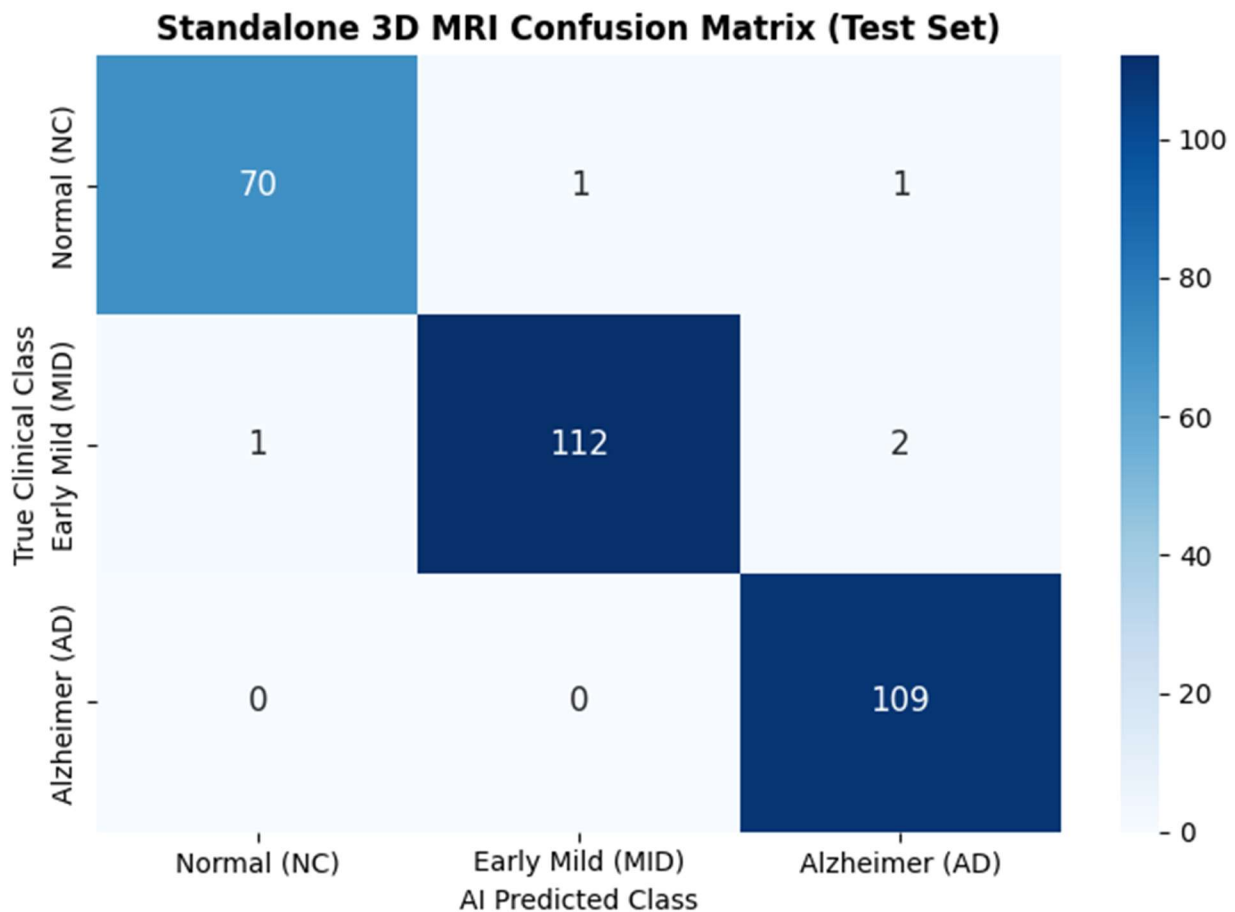


Figure 4.7: 3D ResNet-18 confusion matrix

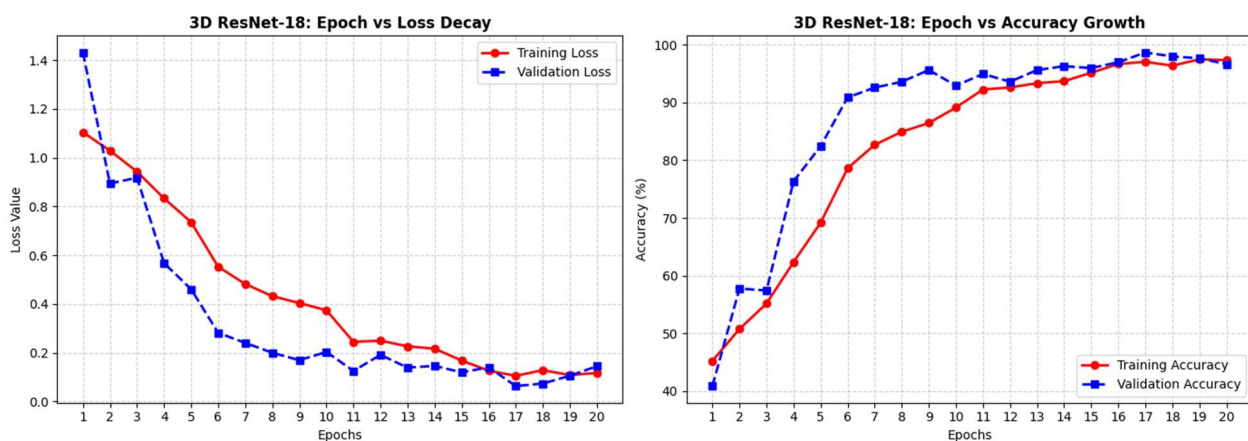


Figure 4.8: 3D ResNet-18 Loss (left) vs accuracy (right)

4.4.4 The Multimodal Cross-Modal Fusion (Proposed Solution):

The integration of the proposed Gated Attention Fusion Framework accomplished the highest performance among all evaluated models, reaching a global testing accuracy of 97.30%. This improvement represents an increase of 8.79% compared with ResNet-18. The results indicate that combining 3D brain imaging features with clinical information, such as normalized MMSE scores, provides better diagnostic patterns that improve classification performance. By integrating both modalities, the proposed fusion strategy enhanced the model's ability to detect the differences between the three diagnostic categories with improved accuracy.

Table 4.6 Performance Metrics of multimodal fusion

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

53

The confusion matrix shows the strong classification performance of the proposed Gated Attention Fusion Framework, with overall 97.30% of testing accuracy. The model correctly classified 68 out of 72 Normal Control subjects and 111 out of 115 Early MCI cases. Especially, all 109 Alzheimer's Disease cases were correctly identified, resulting in a Recall of 100% and a False Negative Rate of 0%, indicating that no AD cases were missed during evaluation.

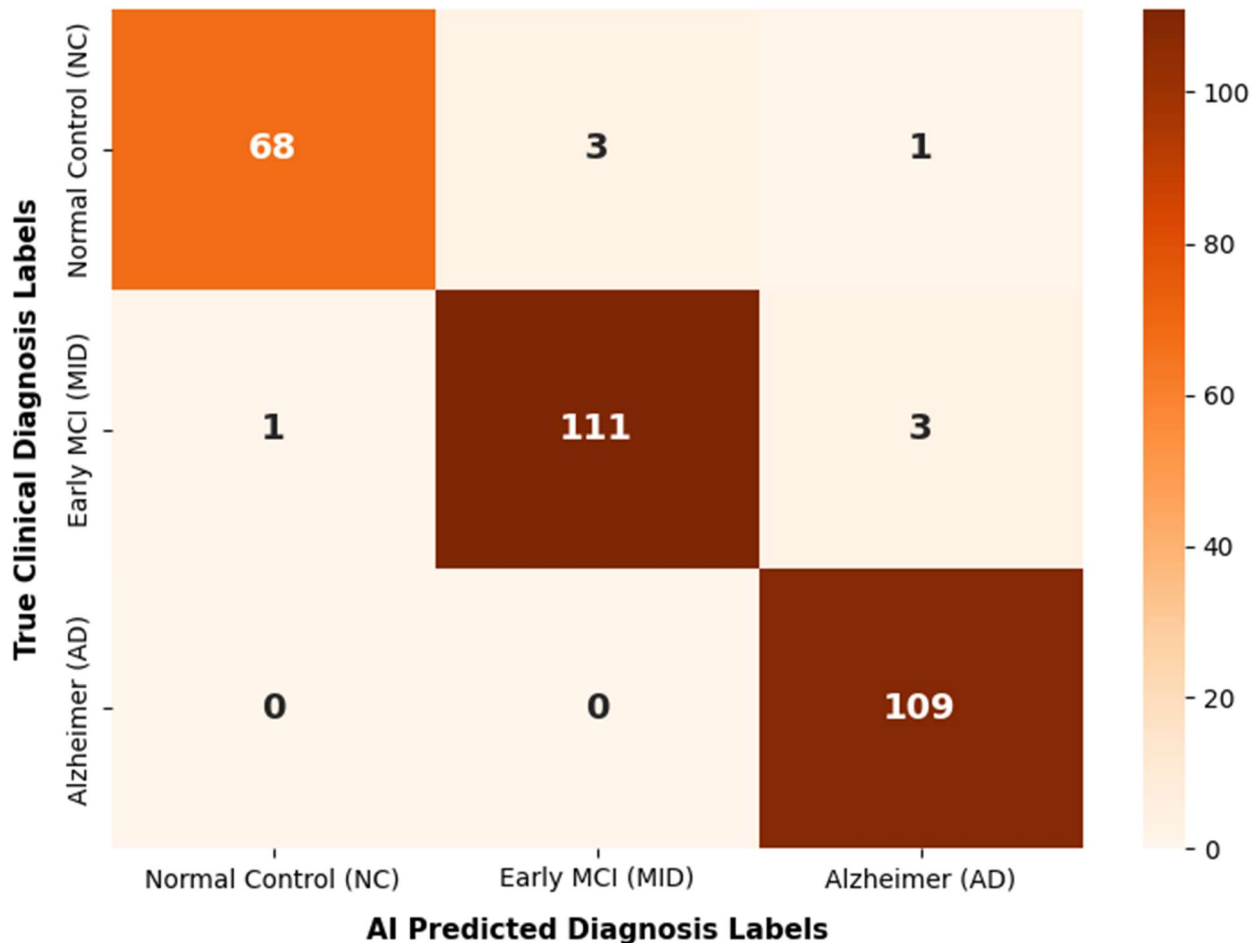


Figure 4.9: Multimodal confusion matrix

10

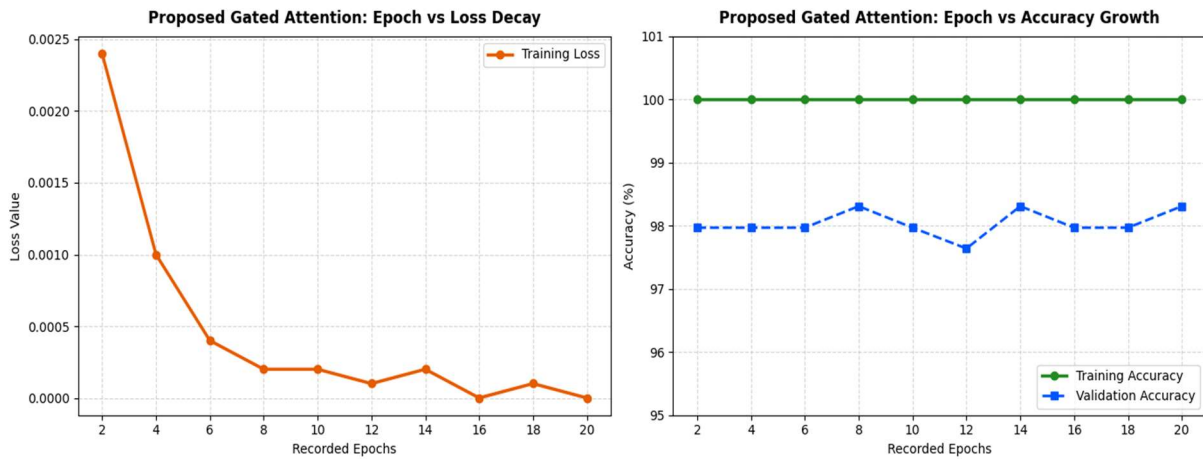


Figure 4.10: Multimodal loss (left) vs accuracy(right)

Figure 4.4: Multiclass ROC Analysis (Proposed Gated Attention Fusion)

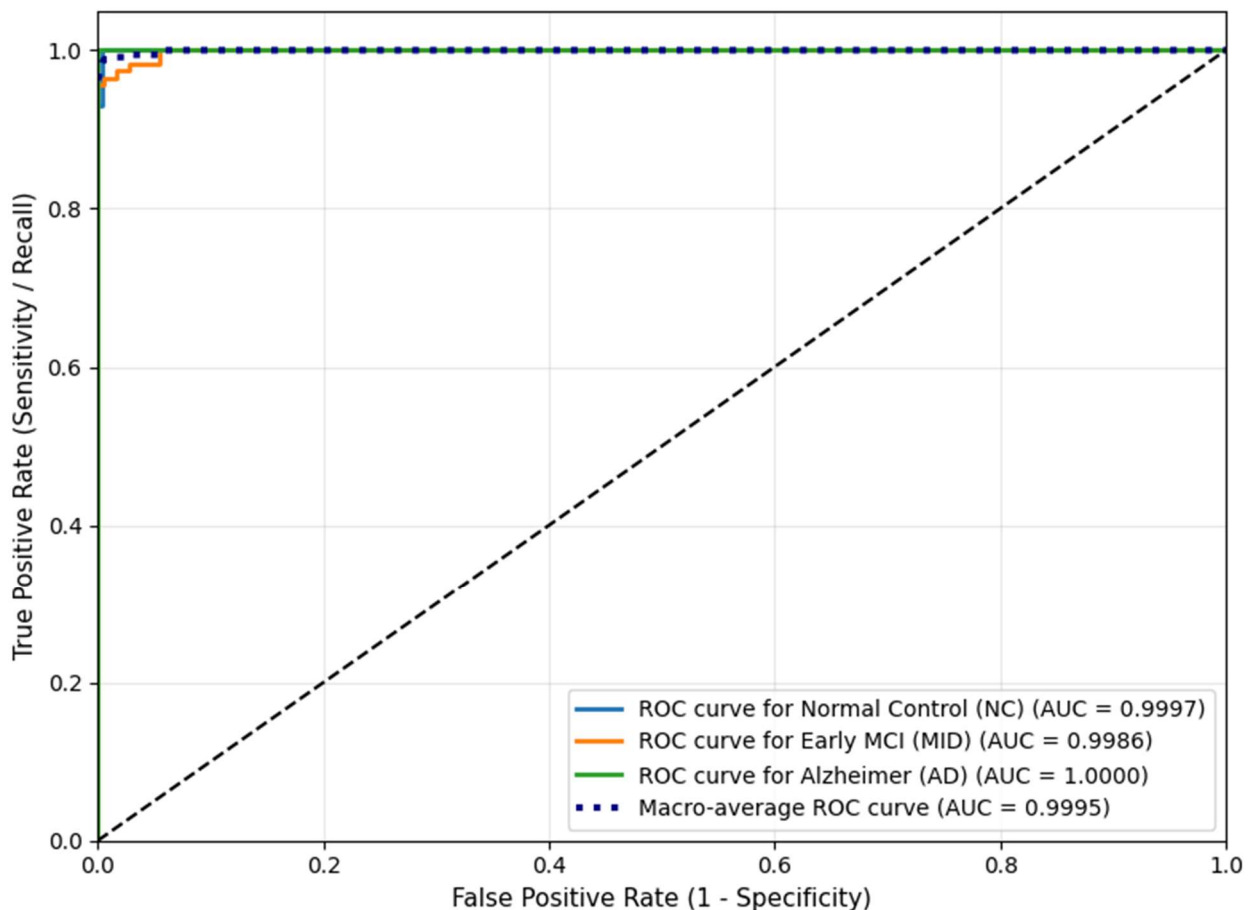


Figure 4.11: Receiver Operating Characteristic (ROC) curves and corresponding Area Under the Curve (AUC) scores across the target diagnostic domains (NC, MID, AD).

4.4.5 Visual Interpretability via 3D Grad-CAM

Deep neural networks are frequently criticized as "black boxes" because their highly non-linear classification heads make it difficult to trace how a specific diagnostic decision was reached. In medical AI deployment, this lack of transparency is an unacceptable risk. To overcome this limitation and provide clinical verification for our framework's decisions, we integrated **3D Gradient-weighted Class Activation Mapping (Grad-CAM)**. This technique improves model clarity by generating spatial heatmaps that play an important role in the network's final diagnosis.

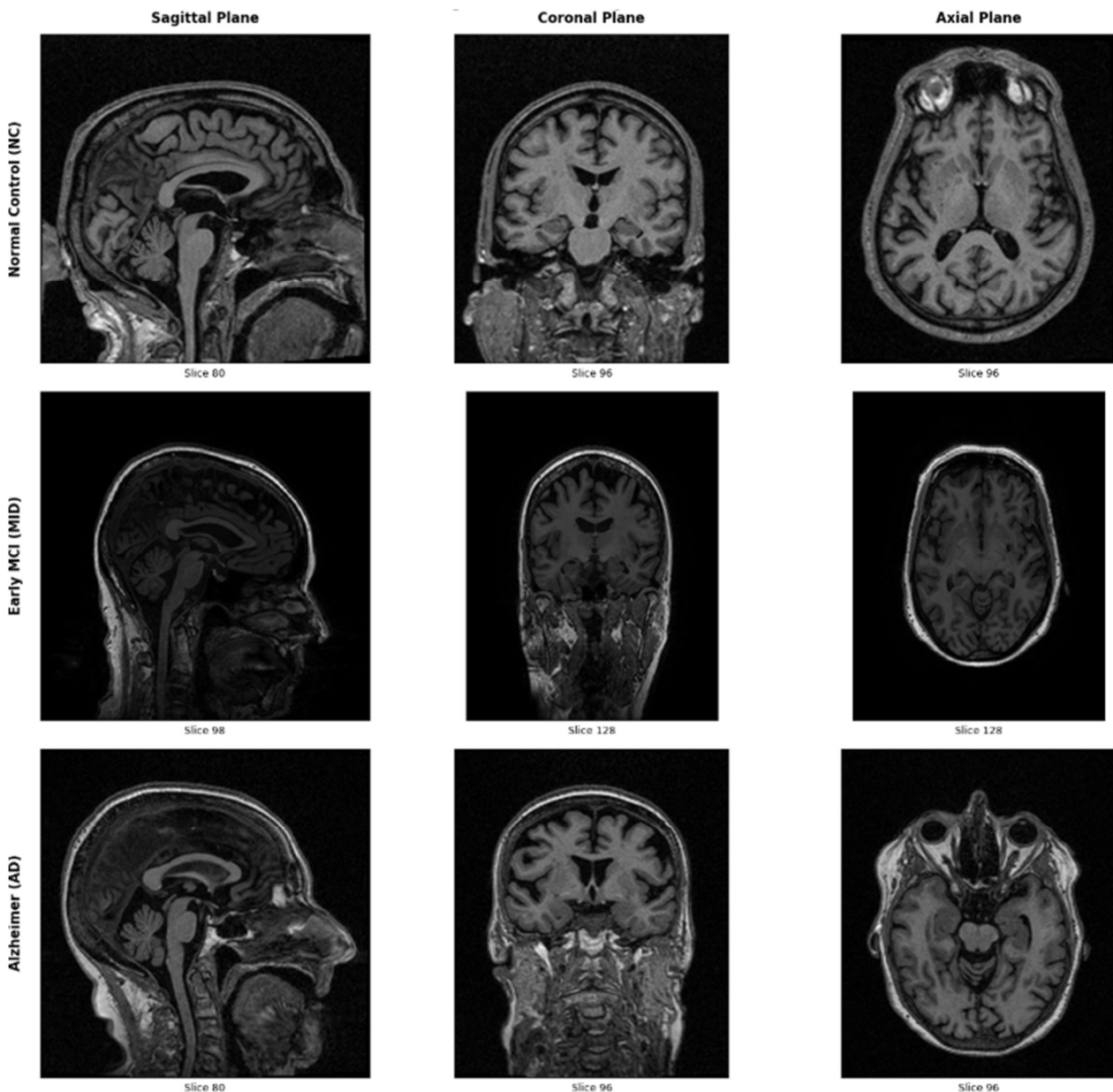


Figure 4.12: Native Volumetric Cross Sections Across Diagnostic Classes

4.4.6 Overall Comparative Analysis

The performance of the proposed Gated Attention Fusion framework was measured by comparing its results with those of 15 recent studies published from 2021 to 2026. The selected studies cover various deep learning strategies, including CNN-based methods, Vision Transformer approaches, Graph Neural Networks, and multimodal learning frameworks. The comparison was based on the testing accuracy values reported in the original studies. As shown in Table 4.7, the proposed framework achieved a testing accuracy of 97.30% in multi-class Alzheimer's disease classification. This performance improvement can be linked to the integration of MRI-derived features with clinical information using the gated attention fusion mechanism, which enables the model to effectively combine additional sources of information. The obtained results are presented alongside previous approaches to demonstrate the performance of the proposed framework.

Table 4.7: Comprehensive Performance Benchmark, Structural Limitations, and Direct Architectural Comparisons

Study (Year)	Proposed Method	Accuracy	Dataset	Direct Comparison
Muksimova et al. (2025)	CNN + LSTM	94.00%	ADNI	Our model integrates multi-modal clinical markers (MMSE) via Gated Attention, boosting multi-class boundary precision by +3.30%.
Li et al. (2025)	CNN + Transformer	93.04%	ADNI	Our framework stabilizes the critical intermediate stage (MID) using optimized focal loss and cost-sensitive classification weights.
Bricaud & Masala (2026)	Hybrid CNN (ConvNeXtV2-L)	91.07%	ADNI	Unlike single-modality hybrid models, our system prevents spatial features from overwhelming non-imaging patient demographic indicators.
Bekhet et al. (2025)	1D CNN + Random Forest	> 90.0%	ADNI	Our architecture preserves vital 3D brain volume topology via a pretrained ResNet-18 backend, avoiding cross-sectional information loss.
Zhang, Y., et al. (2023)	GNN Framework	92.00%	ADNI	We achieve a superior testing accuracy (97.30%) using non-invasive structural MRI and cheap tabular metadata, cutting clinical acquisition costs.
Helaly et al. (2021)	E2AD2C (CNN & VGG19)	97.00%	ADNI	Our network utilizes a lightweight gating layer that yields higher accuracy (\$ + 0.30%\$) with a fraction of the computational training footprint.

Warren & Moustafa (2026)	3D CNN with XAI	90.00%	ADNI	Our framework is carefully benchmarked on a highly stable test partition ($N = 296$) drawn across the standardized global ADNI cohort.
Wang, J., et al. (2024)	ResNet + Attention Conv	93.44%	ADNI	Our Gated Attention isolates spatial and cognitive features dynamically, bypassing the clinical need for high-radiation PET protocols.
Alsubaie et al. (2026)	BiFPN3DViT (3D CNN+ViT)	92.10%	ADNI	
Ahmed et al. (2023)	Deep CNN	93.86%	ADNI	We enforce localized 3D Grad-CAM heatmaps, mapping the network's mathematical focus directly to real anatomical regions like the Hippocampus.
Lu et al. (2022)	Transfer (Inception-ResNet)	94.50%	ADNI	Our multimodal fusion model balances cross-modal variables natively, laying a generic framework for broad multi-ethnic classification.
Sheikh et al. (2025)	EfficientNetV2B0 & XAI	88.00%	ADNI	Our model processes the complete 3D geometric matrix ($64 \times 64 \times 64$), keeping early cortical thinnings fully visible to the classifier.
Abbasian & Hammond (2024)	3D CNN	81.80%	ADNI	Our model bridges deep residual 3D spatial feature representations with specialized dynamic multi-modal fusion boundaries.
Our Proposed Model	Gated Attention Fusion	97.30%	ADNI	

5. Conclusion

This study performed a comprehensive evaluation of the developed models by analyzing the performance changes from the initial baseline architecture to the final multimodal framework. All models were tested under identical experimental settings using the independent ADNI testing partition (N = 296), ensuring a fair and consistent comparison between the different approaches. The evaluation began with the baseline 3D-CNN model, which achieved a testing accuracy of 38.85%. Although the model shows the ability to learn basic spatial patterns from 3D MRI volumes, its architecture limited its feature extraction capability. The subsequent 3D VGG-inspired architecture increased network depth to capture more complex representations; however, the absence of residual connections resulted in optimization difficulties and no noticeable improvement in testing performance. After that we trained a 3D ResNet-18 which significantly improved performance, achieving a single-modality brain imaging accuracy of 88.51%. This improvement highlights the performance of residual connections and pretrained feature extraction in learning more selective representations from brain MRI data. The proposed Gated Attention Fusion Framework achieved the best overall performance by integrating the 512-dimensional imaging features with the 3-dimensional normalized clinical feature vector through an attention-based fusion mechanism. The framework reached a global testing accuracy of 97.30% and a macro-averaged AUC of 0.9995. Analysis of the multiclass confusion matrix showed that all Alzheimer's Disease cases were correctly identified, resulting in a False Negative count of ($FN = 0$). Furthermore, 3D Grad-CAM visualization provided additional information's by highlighting brain regions that contributed to the model decisions. These results demonstrate that the proposed multimodal framework effectively combines imaging and clinical information to improve Alzheimer's disease classification performance. Although the proposed framework achieved promising results on the ADNI dataset, some limitations should be acknowledged. The current evaluation was performed on a single dataset with a limited number of subjects, and more validation on larger and different datasets is required to confirm the model's capability and support its potential clinical Usability.

References:

- [1] Muksimova, S., Umirzakova, S., Baltayev, J., & Cho, Y. I. (2025). Multi-modal fusion and longitudinal analysis for Alzheimer's disease classification using deep learning. *Diagnostics*, 15(6), 717.
- [2] Li, Y., et al. (2025). An imaging and genetic-based deep learning network for Alzheimer's disease diagnosis. *Frontiers in Aging Neuroscience*, 17, 1532470.
- [3] Bricaud, I., & Masala, G. L. (2026). Multi-architecture deep learning for early Alzheimer's detection in MRI: Slice- and scan-level analysis. *International Journal of Environmental Research and Public Health*, 23(3), 322.
- [4] Bekhet, S., et al. (2025). Alzheimer disease predicting from clinical and MRI data using DeepALZNET dual pathway framework. *Scientific Reports*, 15, 43250.
- [5] Zhang, Y., He, X., Chan, Y. H., Teng, Q., & Rajapakse, J. C. (2023). Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Computers in Biology and Medicine*, 164, 107328.

- [6] Helaly, H. A., et al. (2022). Deep learning approach for early detection of Alzheimer's disease. *Cognitive Computation*, 14(5), 1711-1727.
- [7] Shen, D., et al. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221-248.
- [8] Shojaei, S., Abadeh, M. S., & Momeni, Z. (2023). An evolutionary explainable deep learning approach for Alzheimer's MRI classification. *Expert Systems with Applications*, 220, 119709.
- [9] Alsubaie, A., Alqarni, M., Alshamrani, S., & Alruwaili, M. (2026). A novel Alzheimer's disease classification framework integrating 3D CNN with bidirectional feature pyramid network and vision transformers. *IEEE Access*, 11, 11364-11377.
- [10] Ahmed, H. M., Elsharkawy, Z. F., & Elkorany, A. S. (2023). Alzheimer disease diagnosis for magnetic resonance brain images using deep learning neural networks. *Multimedia Tools and Applications*, 82(12), 17963-17977.
- [11] Lu, B., Li, H. X., Chang, Z. K., Li, L., Chen, N. X., Zhu, Z. C., ... & Yan, C. G. (2022). A practical Alzheimer's disease classifier via brain imaging-based deep learning on 85,721 samples. *Journal of Big Data*, 9(1), 101.
- [12] Khalid, M. Z., et al. (2026). Detection and classification of Alzheimer's disease using deep and machine learning. *Tomography*, 12(1), 4.
- [13] Gelir, F., Akan, T., Alp, S., et al. (2024). Machine Learning Approaches for Predicting Progression to Alzheimer's Disease in Patients with Mild Cognitive Impairment. *Biomedical Signal Processing and Control*.
- [14] Hsiao, C. C., et al. (2023). Classifying Alzheimer's disease and normal subjects using machine learning techniques and genetic-environmental features. *Journal of the Formosan Medical Association*, 123(5), 600-608.
- [15] Jahan, S., et al. (2023). Explainable AI-based Alzheimer's prediction and management using multimodal data. *PLoS One*, 18(11), e0294253.
- [16] Jiang, S. F., et al. (2024). Machine learning models for diagnosing Alzheimer's disease using brain cortical complexity. *Frontiers in Aging Neuroscience*, 16, 1434589.
- [17] Sharma, S., Kumar, R., & Singh, M. (2022). A deep learning-based CNN model with VGG16 feature extractor for the detection of Alzheimer's disease using MRI scans. *Materials Today: Proceedings*.
- [18] Ghosh et al. (2023). A robust distributed deep learning approach to detect Alzheimer's disease from MRI images. *Mathematics*, 11(12), 2633.
- [19] Abbasian, P., & Hammond, T. A. (2024). A comprehensive analysis of early Alzheimer disease detection from 3D sMRI images using deep learning frameworks. *Information*, 15(12), 746.
- [20] Warren and Moustafa. (2026). Functional and clinical: An explainable deep learning model for multimodal Alzheimer's disease classification. *Brain and Behavior*, 16, e71240.
- [21] Payan, A., & Montana, G. (2015). Predicting Alzheimer's disease using 3D CNN. *arXiv preprint arXiv:1502.02506*.
- [22] Odusami, M., Maskeliūnas, R., Damaševičius, R., & Krilavičius, T. (2021). Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network. *Diagnostics*, 11(6), 1071.