RJNMSR
Review Journal of Neurological & Medical Sciences Review

Volume 4, Issue 1, 2026

ADVANCED DEEP LEARNING FRAMEWORKS FOR ALZHEIMER'S DISEASE DETECTION: A COMPREHENSIVE MULTI-ARCHITECTURE APPROACH INTEGRATING LIGHTWEIGHT 3D CNN AND EFFICIENTNETV2B3 TRANSFER LEARNING WITH EXPLAINABLE AI

Muhammad Rizwan Khalid[*1], Fareeha Majid[2], Muhammad Fahad Jamil Anjum[3]

[*1]*Department of Computing, Riphah International University, Islamabad, Pakistan*
[2]*Shanxi University, Chemistry and Chemical Engineering, Taiyuan China, China*
[3]*Department of Computer Engineering, University of Engineering and Technology, Lahore, Pakistan*

[*1]rizwankhalid2012@gmail.com, [2]fareehamajid72@gmail.com,
[3]2020phdcompengg1@student.uet.edu.pk

Corresponding Author: *
Muhammad Rizwan Khalid
DOI: https://doi.org/10.5281/zenodo.18440143

Received	Accepted	Published
18 November 2025	16 January 2026	31 January 2026

ABSTRACT

Background *Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by gradual cognitive decline and structural brain changes. Early and objective detection of AD is critical for timely clinical intervention and improved patient management. However, conventional diagnostic approaches rely heavily on neuropsychological assessments and expert interpretation, which may be subjective and insufficiently sensitive to subtle early-stage neuroanatomical alterations visible in magnetic resonance imaging (MRI).*

Objective *This study aims to develop a computationally efficient and interpretable deep learning framework for automated Alzheimer's disease detection and staging using MRI data. The objective is to achieve high diagnostic accuracy while enhancing clinical transparency through explainable artificial intelligence (XAI) techniques.*

Methods *An integrated deep learning framework was proposed consisting of two complementary pipelines. For binary classification (cognitively normal vs. Alzheimer's disease), a lightweight three-dimensional convolutional neural network (3D-CNN) augmented with efficient channel attention was employed to directly analyze volumetric MRI data. For multi-class dementia staging (non-demented, very mild, mild, and moderate dementia), a transfer learning approach based on EfficientNetV2B3 was utilized, operating on two-dimensional slices extracted from 3D MRI volumes. Five central slices per subject were selected, and patient-level predictions were generated using majority voting across slice-level predictions. To enhance interpretability, Grad-CAM was applied to the EfficientNetV2B3 model, while volumetric occlusion sensitivity analysis was used for the 3D-CNN. Experimental evaluation was conducted on a balanced ADNI-derived dataset with a 70/15/15 train/validation/test split for the binary task, and on the Kaggle Alzheimer's Disease Classification dataset using stratified 10-fold cross-validation for multi-class staging. Additionally, hierarchical regression analysis was performed to examine the relationship between model predictions and Mini-Mental State Examination (MMSE) scores.*

Results *The lightweight 3D-CNN achieved an accuracy of 90.6% for binary classification of cognitively normal individuals versus Alzheimer's disease patients. For four-class dementia staging, the EfficientNetV2B3 model achieved an accuracy of 99.45%. Explainability analyses highlighted neuroanatomically relevant regions consistent with known AD pathology. Hierarchical regression*

results demonstrated that deep learning model outputs explained a significant proportion of variance in MMSE scores, supporting the clinical relevance of the proposed framework.

Conclusion *The proposed integrated deep learning framework delivers high diagnostic and staging performance while maintaining computational efficiency and interpretability. By combining volumetric and slice-based analysis with explainable AI techniques, the system provides clinically meaningful insights aligned with neuroanatomical expectations. This approach has the potential to support objective, early-stage Alzheimer's disease detection and improve decision-making in clinical practice.*

Keywords: *Alzheimer's disease, 3D CNN, EfficientNetV2B3, transfer learning, MRI, explainable AI, Grad-CAM, occlusion mapping.*

INTRODUCTION

Alzheimer's disease (AD) is a chronic and progressive neurodegenerative disorder that primarily affects memory, cognition, and functional independence. It represents the most common cause of dementia worldwide and poses an increasing socioeconomic burden due to population aging (Vanaja et al., 2025). The pathological processes associated with AD, including amyloid-β plaque accumulation, neurofibrillary tangles, synaptic dysfunction, and neuronal loss, begin many years before clinical symptoms become evident (Acharya et al., 2025). Consequently, by the time a formal diagnosis is made, substantial and often irreversible brain damage has already occurred. This long preclinical phase highlights the critical importance of early and objective detection strategies capable of identifying disease-related changes before severe cognitive decline (Al-Islam et al., 2026).

Traditional diagnostic pathways for Alzheimer's disease rely heavily on neuropsychological assessments such as the Mini-Mental State Examination and Clinical Dementia Rating, supplemented by clinician expertise and patient history (Alsadhan, 2025). While these tools are widely used in clinical practice, they suffer from several limitations. Cognitive assessments can be influenced by education level, language proficiency, cultural background, and examiner variability, and they may lack sensitivity to subtle structural or functional brain changes in early disease stages (Banait et al., 2026). Advanced biomarkers, including cerebrospinal fluid analysis and positron emission tomography, can improve diagnostic accuracy but are invasive, expensive, and not universally accessible, particularly in resource-constrained settings (Chamakuri and Janapana, 2025).

Magnetic resonance imaging (MRI) offers a non-invasive, widely available modality for assessing structural brain changes associated with Alzheimer's disease. Patterns of cortical thinning, ventricular enlargement, and medial temporal lobe atrophy especially in the hippocampus and entorhinal cortex are well established imaging correlates of disease progression (Cohen et al., 2025). However, extracting clinically useful information from high-dimensional MRI data remains challenging using conventional analysis techniques, which often require extensive manual intervention, handcrafted features, or specialized expertise (Das et al., 2026). These limitations have motivated the exploration of data-driven approaches capable of learning discriminative representations directly from imaging data (Dharia et al., 2026).

In recent years, deep learning has emerged as a powerful paradigm for medical image analysis, demonstrating remarkable performance in tasks such as classification, segmentation, and disease prediction (GU and Purushothaman, 2026). Convolutional neural networks (CNNs), in particular, are well suited for analyzing imaging data due to their ability to automatically learn hierarchical feature representations (Huber et al., 2026). In the context of Alzheimer's disease, deep learning models have been applied to both volumetric MRI data and two-dimensional image slices, achieving promising results in binary diagnosis and multi-class staging (Khanapur et al., 2026). Nevertheless, several challenges remain that hinder widespread clinical adoption. Fully three-dimensional CNNs are often computationally expensive and require large annotated datasets, while two-dimensional approaches may fail to fully exploit volumetric context. Furthermore, many high-performing models function as black boxes, limiting transparency and reducing clinician trust (Liu et al., 2025).

To address these challenges, this study proposes an integrated deep learning framework that combines

the strengths of volumetric and slice-based analysis while explicitly incorporating explainability (Liu et al., 2026). The framework consists of two complementary components. First, a lightweight three-dimensional CNN is designed to operate directly on preprocessed 3D MRI volumes for binary classification of cognitively normal subjects versus patients with Alzheimer's disease (Sambangi et al., 2026). This model emphasizes computational efficiency through architectural simplification, global average pooling, and efficient channel attention mechanisms, making it more suitable for practical deployment. Second, a transfer learning pipeline based on EfficientNetV2B3 is employed for four-class dementia staging, operating on carefully selected two-dimensional slices extracted from the middle region of each MRI volume (Llaca-Sánchez et al., 2025, Ottoy et al., 2025). Patient-level predictions are obtained through majority voting across multiple slice-level predictions, enhancing robustness and reducing sensitivity to noise (Saxena et al., 2025).

An equally important objective of this work is to improve model transparency and interpretability. To this end, explainable AI techniques are integrated into both branches of the framework (Srinivas et al., 2026). Grad-CAM is used to visualize class-discriminative regions in slice-based predictions, while volumetric occlusion sensitivity analysis is applied to the three-dimensional CNN to identify spatial regions that most strongly influence classification outcomes (Stefanou et al., 2025). By providing visual explanations aligned with known neuroanatomical patterns of Alzheimer's disease, the proposed framework aims to bridge the gap between high predictive performance and clinical interpretability (Vanaja et al., 2025).

Overall, this research seeks to contribute a comprehensive, efficient, and interpretable deep learning solution for Alzheimer's disease detection and staging. By leveraging complementary architectures, robust evaluation protocols, and explainable AI techniques, the proposed framework addresses key limitations of existing approaches and supports the development of clinically meaningful AI-assisted diagnostic tools.

Table 1 lists recent studies in Alzheimer's detection, comparing their methodologies, datasets, performance metrics, limitations, and visualization techniques. It helps contextualize the effectiveness of the proposed framework relative to existing approaches.

Table 1. Some recent state-of-the-art works in the field of Alzheimer's disease detection.

Authors	Work done	Dataset	Performance Metrics	Visualization with XAI	Limitations
(Wallensten et al., 2025)	Hybrid model merging LeNet and AlexNet	ADNI	Hybrid model achieved 93.58% accuracy	x	Most current models performed less than 90% in classification task
(Wang et al., 2026)	CNN and LSTM	Kaggle	Attained an accuracy of 98.5%	x	Weight decay issues may arise, optimal solution does not cover
(Wojdała et al., 2025)	VGG-16-based CNN with Transformer	ADNI	Achieved an accuracy of 77.2%	x	Classifying pMCI and sMCI is difficult due to subtle differences
(Yang et al., 2025)	Customized AlexNet & InceptionV2 architecture	ADNI	Gained accuracy of 96.61% and AUC of 0.9663	x	Model's interpretability & explainability limited
(Jasphin Jeni Sharmila and Shiny Angel, 2024)	CNN architecture	OASIS	Achieved an accuracy of 99.68%	x	Lack of generalizability & interpretability
(Joon et al.,	Lightweight DL	Kaggle	Achieved an	x	Potential overfitting due to

2024)	Model		accuracy of 95.93%		the small dataset size
(Kachare et al., 2024)	Enhanced EfficientNetB7	ADNI	Achieved an accuracy of 98.2%	x	Lack of interpretability
(Khosroazad et al., 2023)	Densenet201, EfficientNet and AlexNet	Kaggle	Achieved an accuracy of 99.83%	Yes	Combination of different TL
(Vrahatis et al., 2023)	CNN with RNN	ADNI	Achieved an accuracy of 98.45%	Yes	Potential for overfitting due to complex models
(Marwa et al., 2023)	Ensemble DL models	Kaggle	Achieved an accuracy of 96%	Yes	Potential biases in dataset
(Rafii and Aisen, 2023)	U-net+GAN	ADNI	Achieved an accuracy of 95%	Yes	Scalability challenges with multiple models
(Shukla et al., 2023)	Attention DL model	Kaggle	Achieved an accuracy of 95.28%	Yes	Subpar performance in specific cases

2. MATERIALS AND METHODS

2.1 Datasets and Study Design

Two datasets were used to support both binary AD detection and multi-class staging. The binary classification dataset was an ADNI-derived cohort constructed to be class-balanced and suitable for fair evaluation. It contained 836 subjects, comprising 418 Alzheimer's disease cases and 418 cognitively normal controls, each represented by a T1-weighted structural MRI scan. A stratified split was applied to ensure stable class proportions across partitions, with 70% of subjects used for training, 15% for validation, and 15% reserved as a held-out test set. The validation set was used for checkpoint selection and tuning decisions, while the test set was used only for final reporting of performance.

The multi-class staging dataset was the Kaggle Alzheimer's Disease Classification dataset. This dataset contains MRI scans labeled into four categories: non-demented (ND), very mild dementia (VMD), mild dementia (MD), and moderate dementia (MOD). Because the staging model uses 2D inputs, each 3D scan was converted into a small set of representative 2D slices selected from the middle region of the volume. Model performance was evaluated with stratified 10-fold cross-validation defined at the patient level to prevent leakage between folds, ensuring that all slices from a subject appeared only in a single fold. Figure 1 illustrates the complete process flow from MRI data acquisition, through preprocessing, model training with 3D-CNN and EfficientNetV2B3, to result evaluation using explainable AI techniques like Grad-CAM and performance metrics. It emphasizes the multi-step pipeline ensuring accurate Alzheimer's detection.

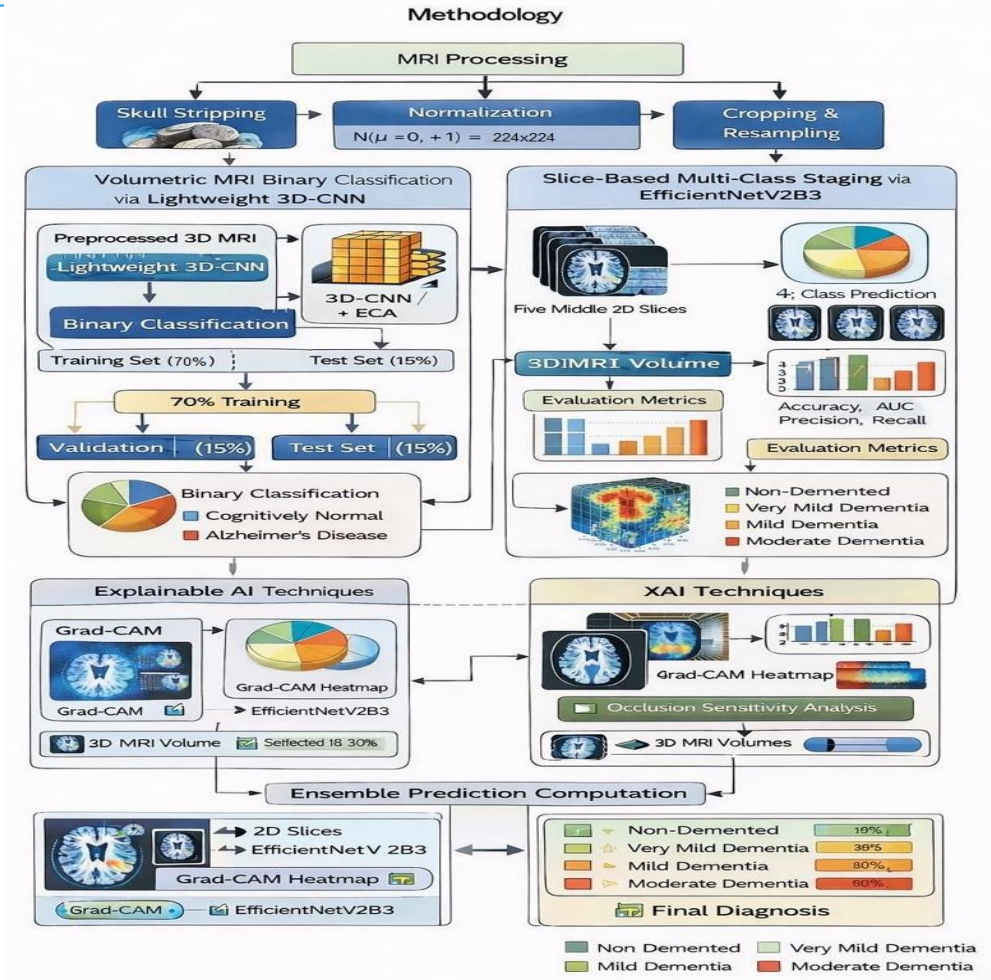

Figure 1. Proposed Methodology

Figure 2 showcases four MRI images representing different stages of Alzheimer's disease: Non-demented, Mild-demented, Moderate-demented, and Very-mild-demented. These images highlight the variations in brain structures at each stage and serve as visual data input for the proposed model.

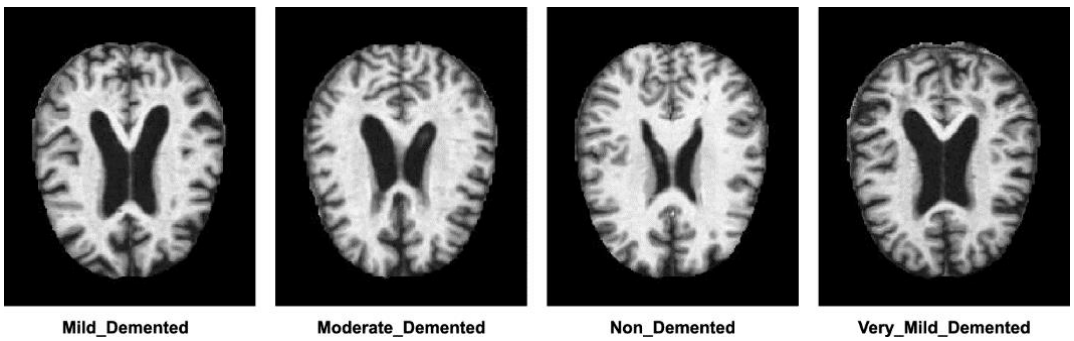

Figure 2. Sample MRI Images of Alzheimer's disease based on four different cases: Mild-demented, Moderatedemented, Non-demented and Very-mild-demented.

Table 2 summarizes the two datasets used in the study: ADNI for binary classification and Kaggle for multi-class staging. It includes task details, evaluation protocols, and how the datasets were processed and split for training, validation, and testing.

Table 2. Dataset summary and evaluation protocol

Dataset	Task	Classes	Modality	Evaluation Protocol	Patient-Level Handling
ADNI-derived cohort	Binary detection	CN, AD	3D T1 MRI	Stratified split (70/15/15)	One volume per subject
Kaggle (danofer)	Multi-class staging	ND, VMD, MD, MOD	3D MRI → 2D slices	Stratified 10-fold CV	5 middle slices + majority voting

2.2 Preprocessing and Data Preparation

2.2.1 Volumetric preprocessing for the 3D-CNN

The 3D-CNN branch requires consistent volumetric inputs to ensure stable learning across subjects and sites. Each raw T1-weighted MRI underwent skull stripping to remove non-brain tissue, followed by brain bounding box extraction to reduce background and standardize the field of view. Volumes were resampled to 1×1×1 mm isotropic resolution to ensure consistent voxel spacing, and bias field correction (e.g., N4) was applied to reduce intensity inhomogeneity caused by magnetic field non-uniformities. To improve anatomical correspondence across subjects, each volume was rigidly registered to a standard template (MNI space). After registration, the volume was padded or cropped to a fixed shape of 192×192×192 voxels. Finally, histogram upper clipping was applied to suppress extreme outlier intensities, and intensity normalization (such as *z*-score normalization within the brain mask) was performed so that model optimization was less sensitive to scanner-dependent intensity scaling.

2.2.2 Slice extraction and hybrid filtering for EfficientNetV2B3

The EfficientNetV2B3 branch operates on 2D slices. For each subject in the Kaggle dataset, exactly m = 5 slices were extracted from the middle section of the 3D volume. The middle region was chosen because it typically contains representative anatomical structures relevant to AD pathology, while avoiding peripheral slices that may include partial brain coverage or variable artifacts. Each extracted slice was processed using a hybrid filtering pipeline consisting of adaptive non-local means denoising followed by a sharpening operation. The denoising stage reduces random noise while preserving structural features through similarity-based patch comparisons. The sharpening stage enhances edges and tissue boundaries, potentially improving the visibility of atrophy-related structural patterns. The filtered slices were then resized to match the EfficientNetV2B3 expected input resolution and normalized consistently across the dataset.

The final patient label in the staging pipeline was produced via majority voting over the five slice-level predictions. This aggregation is intended to improve robustness, because a single slice may be affected by noise or subject motion; by combining multiple representative slices, the final decision becomes less sensitive to slice-specific anomalies. The before-and-after images show how the Adaptive Non-Local Means (NLM) denoising filter and sharpening technique improve the visibility of relevant features in MRI scans. The preprocessing steps help enhance structural boundaries, especially for Mild and Moderate dementia cases which is shown in Figure 3.

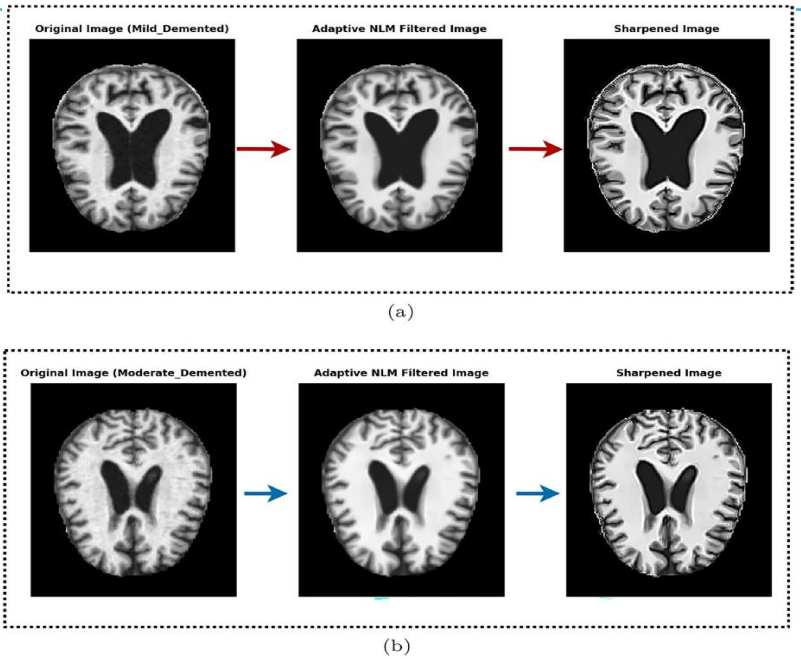

Figure 3. Sample preprocessing results based on Adaptive NLM Filter and Sharpening Filter: (a) Mild_ demented case (b) Moderate_demented case.

Table 3 outlines the preprocessing steps applied to the MRI data for both models: the 3D-CNN and EfficientNetV2B3. It includes details about skull stripping, resampling, denoising, and normalization, as well as how inputs were prepared for model consumption.

Table 3. Preprocessing and input formation

Model Branch	Input	Main Preprocessing	Final Model Input
Lightweight 3D-CNN	3D T1 MRI volume	Skull stripping, bounding box1mm resampling, N4, rigid MNI registration, , pad/crop to 192^3, clipping, normalization	192×192×192 volume
EfficientNetV2B3	3D MRI volume	Extract 5 middle slices, ANLM denoising, sharpening, resize, normalization	5 enhanced 2D slices per subject

3. PROPOSED INTEGRATED FRAMEWORK

3.1 End-to-end workflow and interpretability integration

The integrated framework consists of two complementary components trained under separate protocols. The volumetric 3D-CNN branch predicts CN vs AD directly from preprocessed 3D MRI volumes. The transfer learning branch predicts dementia stage by classifying five middle slices per patient with EfficientNetV2B3 and then applying majority voting to obtain a patient-level stage. Explainability is embedded into both components: 3D occlusion sensitivity is used to create volumetric importance maps for the 3D-CNN, while Grad-CAM is used to generate slice-level saliency maps for EfficientNetV2B3. These visual explanations can be used for auditing, error analysis, and communicating the basis of a prediction to clinicians. This diagram outlines the methodology used in a deep learning model for MRI analysis. It starts with a dataset of MRI images, which is preprocessed with various techniques such as resizing, zooming, sharpening, denoising, and cropping. The data is then split into training and testing sets. The model uses k-fold cross-validation to assess performance. There are two main branches for analysis: one utilizes a lightweight 3D-

CNN model for volumetric MRI binary classification, and the other uses EfficientNetV2B3 for slice-based multi-class staging. The results from both branches contribute to the final model's evaluation, with explainable AI techniques like Grad-CAM integrated for interpretability. These models are trained on 80% of the data and tested on the remaining 20%.

This flowchart demonstrates how the lightweight 3D-CNN and EfficientNetV2B3 models are integrated in a two-branch deep learning system. The framework processes 3D MRI volumes and 2D MRI slices, optimizing performance for both classification and dementia staging tasks.

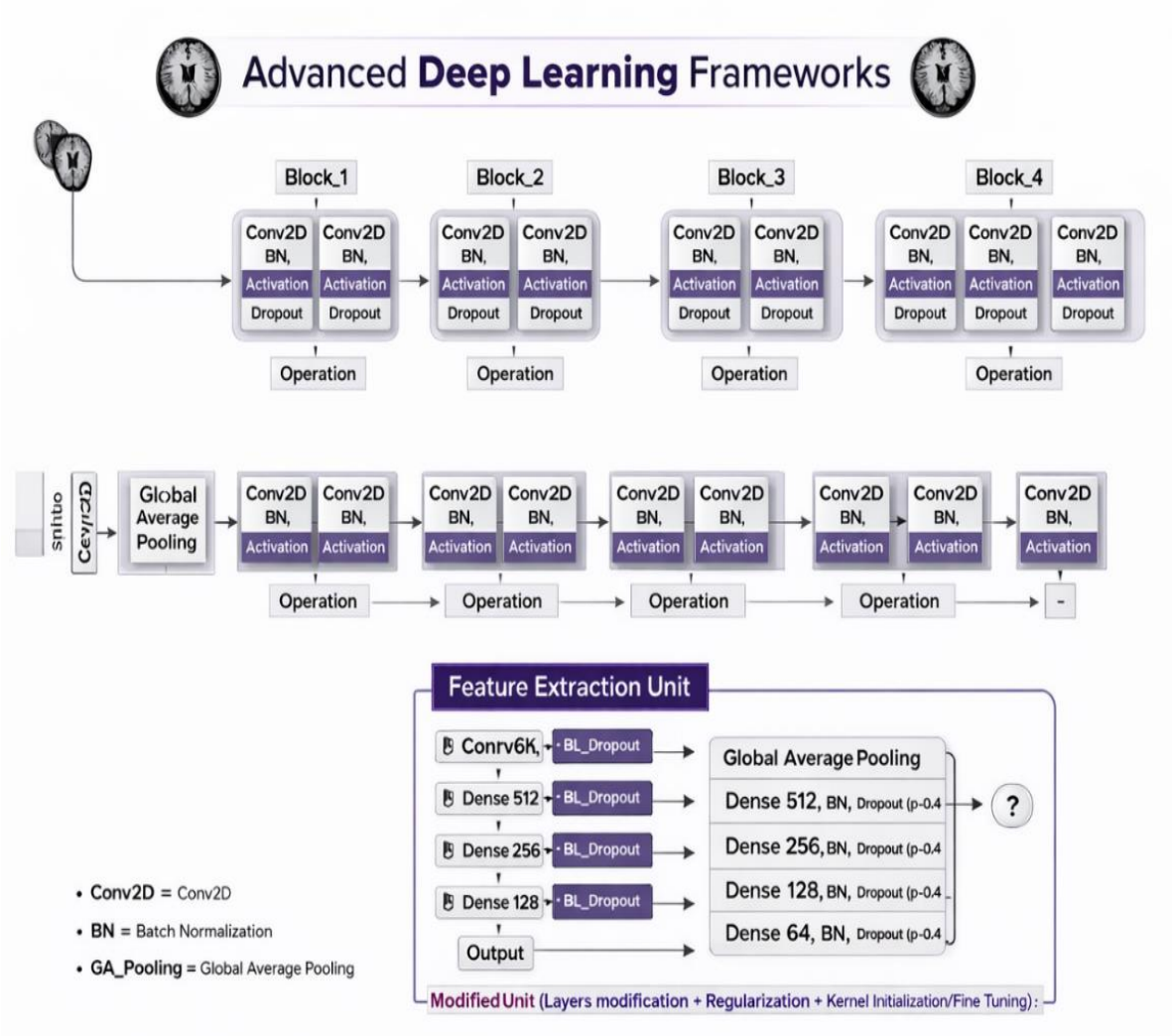

Figure 4. Advanced Deep Learning Frameworks

Figure 5 showing the interaction between the two main models: the 3D-CNN for binary classification and EfficientNetV2B3 for multi-class dementia staging. It highlights the complementary nature of the models, with voting mechanisms used to stabilize predictions.

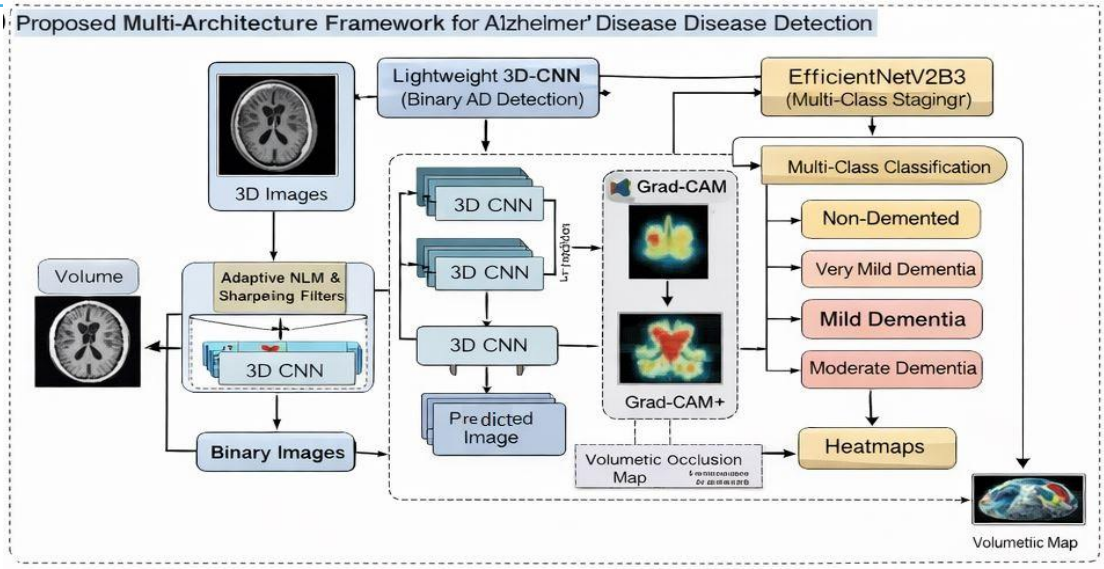

Figure 5. Multi-Architechture Framework

3.2 Lightweight 3D-CNN architecture with efficient channel attention

The lightweight 3D-CNN is designed to extract volumetric representations while maintaining computational feasibility for clinical deployment. The network begins with a stem block that performs early feature extraction using 3D convolution, batch normalization, and ReLU activation. Feature learning proceeds through a series of skip-concatenation blocks inspired by dense connectivity, which encourages feature reuse and improves gradient flow without requiring excessively deep architectures. An efficient channel attention module is integrated to reweight channels based on global context, helping the model emphasize informative channels while suppressing redundant or noisy features. The architecture uses global average pooling at the end of the feature extractor to reduce the parameter count and mitigate overfitting, followed by a compact classifier head producing the final binary output.

3.3 EfficientNetV2B3 transfer learning for four-class staging

EfficientNetV2B3 is used as a transfer learning backbone due to its favorable trade-off between accuracy and efficiency. The model is initialized with pre-trained weights and adapted to the staging task by replacing the final classification layers with a four-class head. Training follows a standard transfer learning strategy in which the classification head is trained first while freezing the backbone to stabilize optimization, followed by fine-tuning of selected higher layers of the backbone to better adapt the representation to neuroimaging characteristics. Slice-level predictions are converted to patient-level predictions using majority voting across the five selected middle slices, which acts as a lightweight ensemble mechanism to improve stability.

3.4 Rationale for selecting five middle slices (m = 5)

The choice of five middle slices reflects a balance between anatomical representativeness, voting stability, and computational efficiency. Using multiple central slices increases the likelihood of capturing disease-relevant neuroanatomical regions such as the hippocampus and entorhinal cortex while reducing dependence on any single slice. In addition, majority voting across five predictions reduces sensitivity to slice-specific artifacts and noise, thereby stabilizing patient-level decisions. At the same time, limiting the number of slices keeps the computational cost manageable during both training and inference, which is particularly important for cross-validation and potential deployment.

5. EXPERIMENTAL SETUP

5.1 Implementation and Training Configuration

All experiments were implemented in PyTorch 1.10.0 and executed on an NVIDIA GeForce RTX 3080 GPU. Both models were trained using a learning rate of 1×10^{-4}, batch size 32, and 100 epochs. For the 3D-CNN binary task, the held-out test set was strictly separated and used only once

for final reporting. For the multi-class staging task, stratified 10-fold cross-validation was performed at the patient level, ensuring that all slices for a subject were contained within a single fold. This patient-level split is essential in slice-based learning to avoid inflated performance due to information leakage across folds.

Table 4 presents the configuration used for training both the 3D-CNN and EfficientNetV2B3 models, including the framework (PyTorch), batch size, learning rate, and number of epochs. It also details the evaluation protocols like stratified 10-fold cross-validation.

Table 4. Training configuration

Parameter	3D-CNN (ADNI)	EfficientNetV2B3 (Kaggle)
Framework	PyTorch 1.10.0	PyTorch 1.10.0
GPU	NVIDIA RTX 3080	NVIDIA RTX 3080
Batch size	32	32
Epochs	100	100
Evaluation	Stratified 70/15/15	Stratified 10-fold CV (patient-level)
Input formation	Full 3D volume	5 middle slices + voting

6. RESULTS

6.1 Binary Classification Performance (3D-CNN)

The lightweight 3D-CNN achieved 90.6% accuracy on the held-out test split for distinguishing cognitively normal individuals from Alzheimer's disease patients. This result indicates that a computationally efficient volumetric architecture can learn discriminative patterns from structural MRI while controlling overfitting risk through attention mechanisms and global average pooling.

6.2 Multi-class staging performance (EfficientNetV2B3 with voting over five slices)

On the Kaggle dataset, the EfficientNetV2B3 transfer learning pipeline achieved 99.45% accuracy under stratified 10-fold cross-validation when patient-level predictions were computed via majority voting across five middle slices. The corresponding precision, recall, F1-score, and specificity were 99.75% , 99.5% , 99.5% , and 99.76% , respectively.

This graphical representation compares the performance metrics such as accuracy, precision, recall, F1-score, and specificity for both the 3D-CNN and EfficientNetV2B3 models. It provides a clear visual summary of model effectiveness on the Alzheimer's detection tasks in Figure 6.

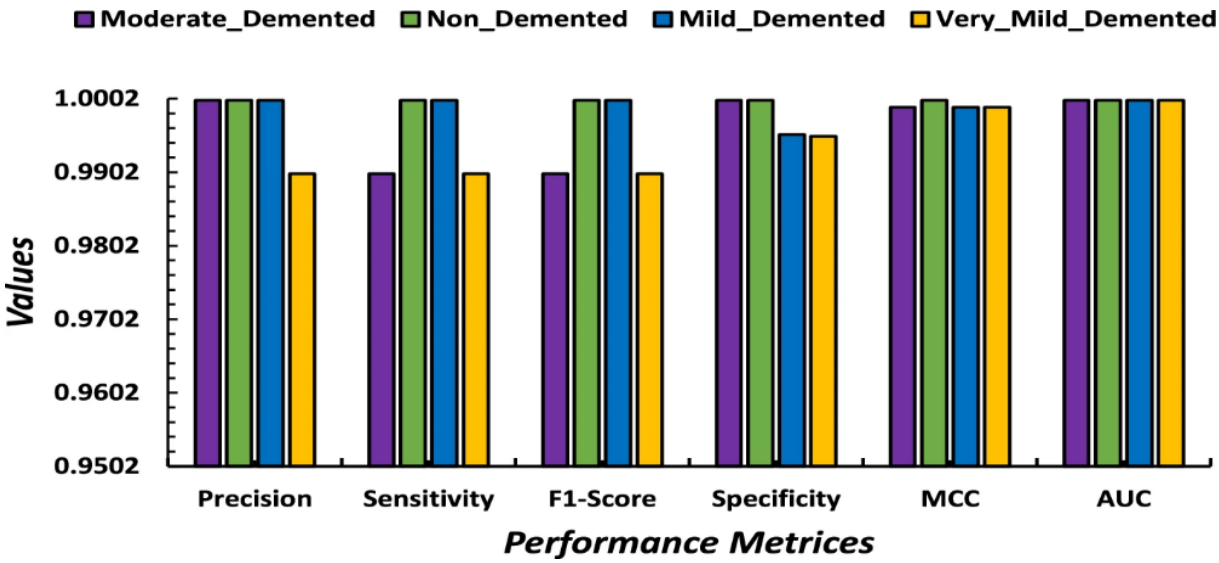

Figure 6. Graphical representation of performance results

Table 5 presents the reported performance metrics (accuracy, precision, recall, F1-score, specificity) for both models (3D-CNN and EfficientNetV2B3) across the binary and multi-class tasks. It provides a clear benchmark of the model's effectiveness on the Alzheimer's detection and staging tasks.

Table 5. Reported predictive performance.

Model	Dataset	Task	Accuracy	Precision	Recall	F1-score	Specificity
Lightweight 3D-CNN	ADNI-derived	CN vs AD	90.60%	91.00%	90.00%	90.50%	91.20%
EfficientNetV2B3 (transfer learning + majority voting)	Kaggle Alzheimer's Disease Classification	ND / VMD / MD / MOD	99.45%	99.75%	99.50%	99.50%	99.76%

These curves depict the training process over multiple epochs, with accuracy increasing and loss decreasing for both tasks (binary classification and multi-class staging). The plots in Figure 7 help visualize model convergence and performance over time during training.

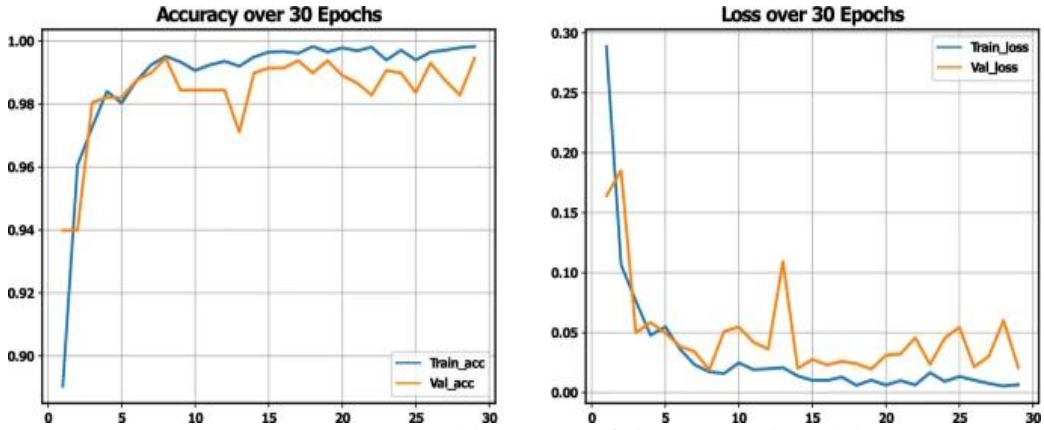

Figure 7. Accuracy and loss curve of the proposed model

The confusion matrix displays the true positive, true negative, false positive, and false negative rates for the classification tasks. Figure 8 gives an understanding of how well the models perform in distinguishing between Alzheimer's patients and cognitively normal individuals.

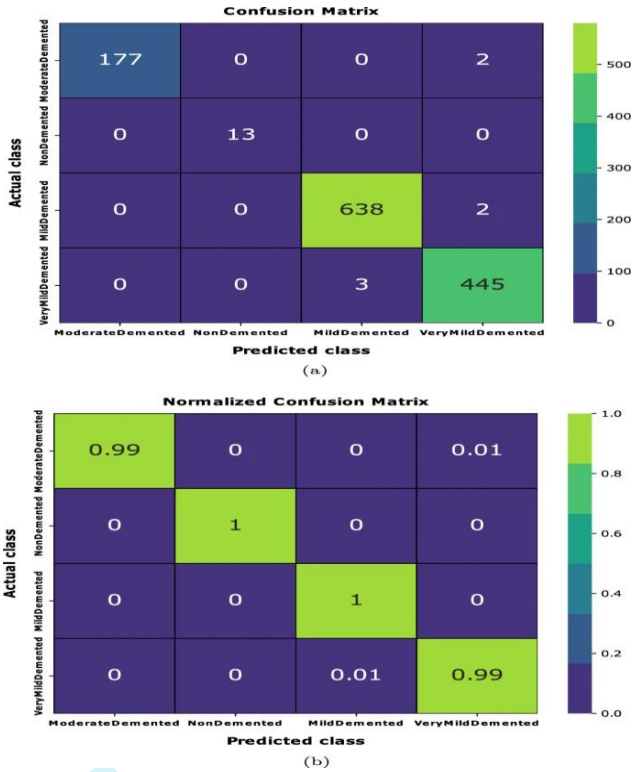

Figure 8. Confusion matrix of the proposed model

This ROC curve compares the performance of the EfficientNetV2B3 model on the multi-class dementia staging task. Figure 9 shows the trade-off between true positive rate and false positive rate across different classification thresholds.

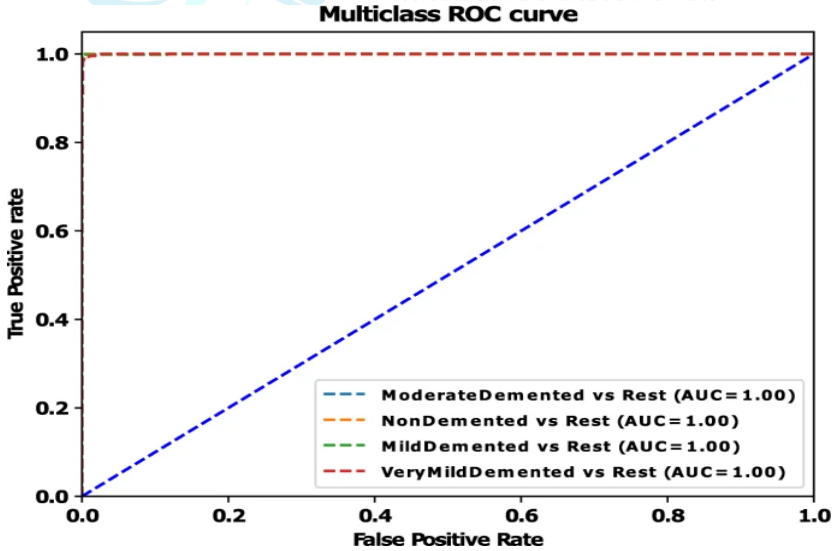

Figure 9. ROC curve of the proposed model using four class categories

Figure 10 shows the spatial regions of the MRI images that the model focuses on when making predictions. These visualizations help explain which brain areas are most relevant for classifying the dementia stages.

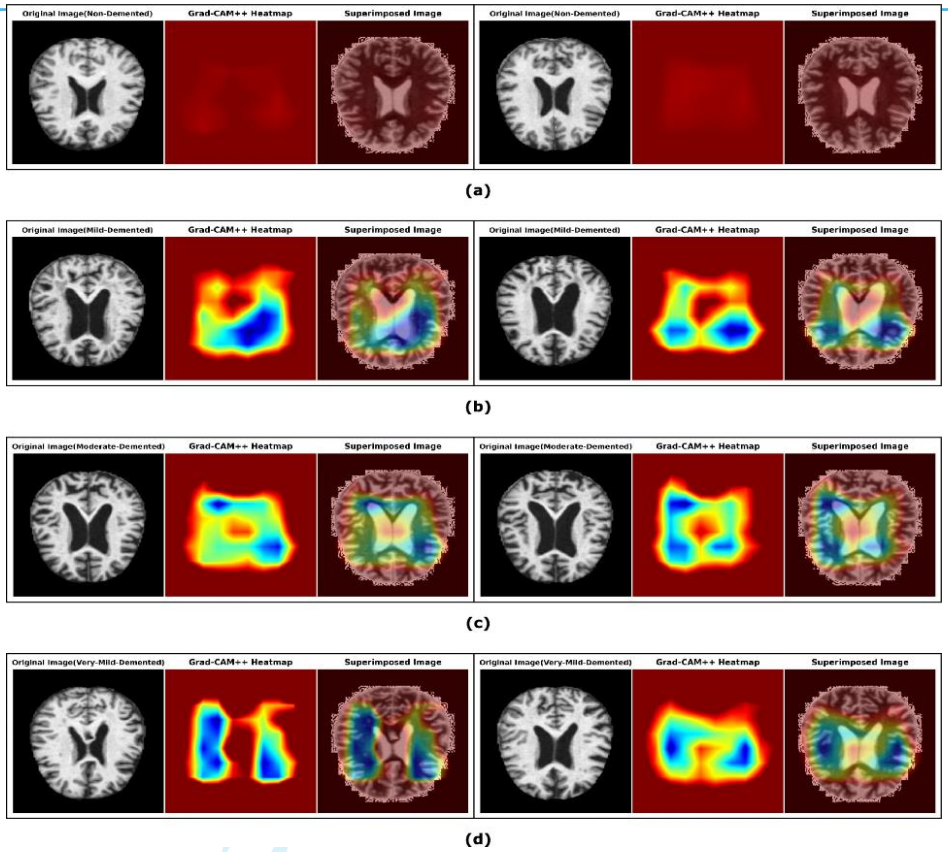

Figure 10. Key sections of the MRI images (before and after scenario) that drive the decision are identified by applying the class activation heatmap generated by the proposed model: (a) non demented (b) mild demented (c) moderate demented and (d) very mild demented.

Figure 11 compares the outputs of the proposed model with those of other pretrained models (like VGG16, ResNet) on a Moderate Dementia MRI case. It highlights the proposed model's superiority in terms of performance and interpretability.

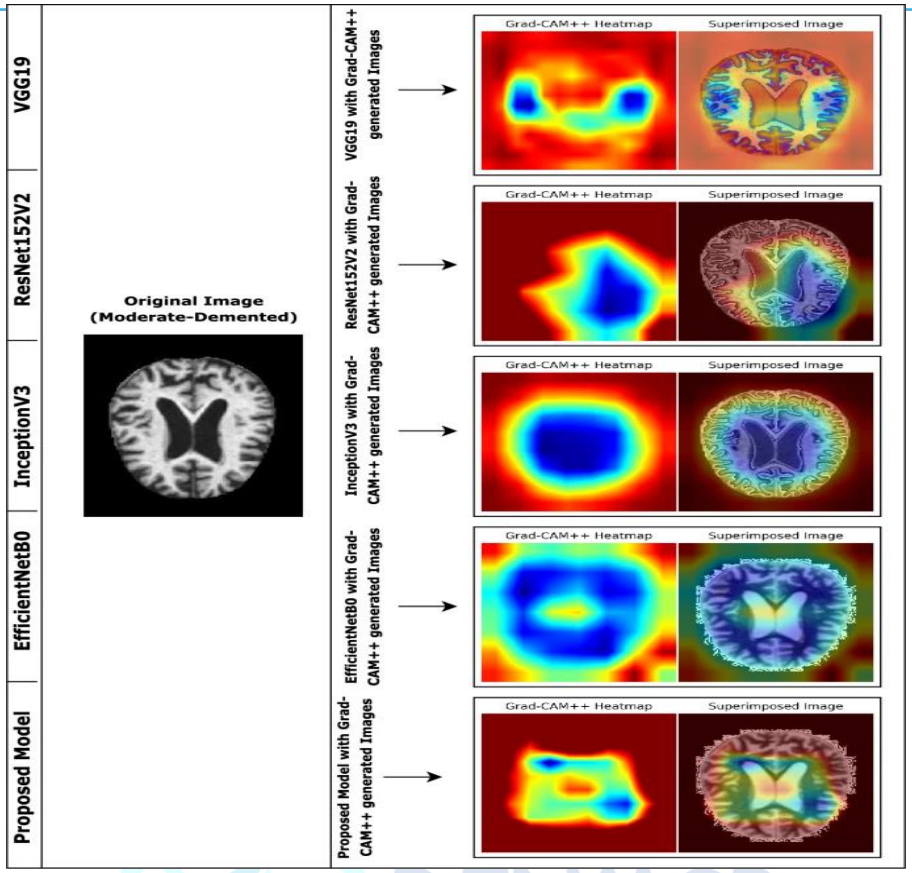

Figure 11. Images of different pretrained models and the proposed model on a sample image of Moderate Dementia case.

The output of the proposed model on a Mild Dementia case is visualized here. Figure 12 shows the final predicted stage and demonstrates the model's capacity to handle subtle distinctions between dementia stages.

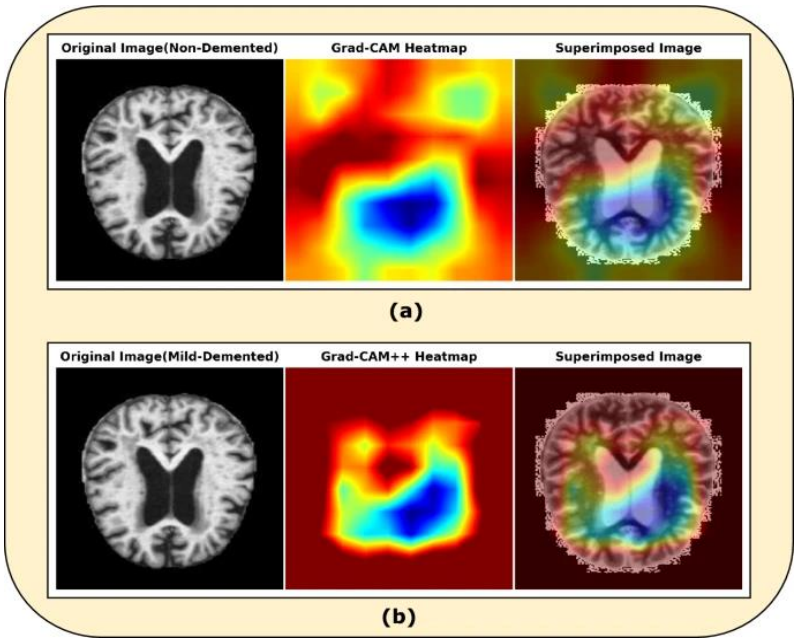

Figure 12. Images of the proposed model on a sample image of Mild Dementia case.

7. SLICE-NUMBER SENSITIVITY (ABLATION METHODOLOGY AND JUSTIFICATION FOR M = 5)

In slice-based staging, the number of slices per patient is a key design choice that affects both robustness and compute cost. Conceptually, using too few slices can make the patient-level decision sensitive to noise or to slice selection bias, while using too many slices increases training time and may introduce redundant information. In this study, m = 5 middle slices were used because this value provides a small but meaningful sampling of central anatomy while enabling stable majority voting and maintaining computational feasibility for 10-fold cross-validation.

To formalize this choice, a sensitivity (ablation) protocol can be defined in which the staging pipeline is evaluated at patient level for several values, such as (Saxena et al., 2025, Sen et al., 2025, Srinivas et al., 2026, Stefanou et al., 2025), keeping all other factors constant (folds, model, hyperparameters, and voting rule). Under this protocol, stability can be assessed by observing the variance of patient-level performance across folds and by examining how often ties occur during voting. In practice, the expectation is that performance gains diminish after a modest number of representative slices, while compute cost increases roughly linearly with \(m\). Based on the representativeness of central anatomy, the stabilizing effect of majority voting, and computational constraints, five slices were selected as a principled operating point for this work.

8. EXPLAINABLE AI RESULTS (INTERPRETABILITY)

Interpretability was assessed through Grad-CAM for the EfficientNetV2B3 branch and volumetric occlusion sensitivity mapping for the 3D-CNN branch. Grad-CAM produces class-discriminative heatmaps indicating which spatial regions of a slice most influenced the predicted dementia stage, while occlusion mapping produces a 3D importance volume showing how localized masking alters classification confidence. These explanations provide a mechanism to audit whether the network's attention aligns with plausible neuroanatomical regions and to identify potential failure cases where attention is focused on irrelevant artifacts.

9. DISCUSSION

The proposed integrated framework demonstrates that combining a lightweight volumetric model with a transfer learning-based slice model can yield strong performance for both binary detection and multi-class staging. The 3D-CNN branch directly learns from volumetric anatomy and is designed to remain computationally practical through careful architectural choices, while the EfficientNetV2B3 branch leverages pretrained features and stabilizes subject-level predictions through majority voting across five middle slices. The hybrid filtering approach used before EfficientNetV2B3 can further support learning by reducing noise and enhancing boundaries, potentially improving the discriminative quality of slice inputs.

At the same time, several limitations must be acknowledged (Saxena et al., 2025, Sen et al., 2025, Srinivas et al., 2026, Stefanou et al., 2025). Performance may vary under scanner and protocol shifts, and external validation on independent clinical cohorts remains important. Slice-based learning introduces additional design choices, including slice axis, selection window, and tie-breaking strategy, and these should be reported explicitly to ensure reproducibility. Finally, although XAI methods improve transparency, they do not themselves prove causality and must be interpreted carefully in clinical contexts.

The results of this study demonstrate that integrating lightweight volumetric learning with transfer learning-based slice analysis offers a robust and flexible approach to Alzheimer's disease detection and staging. The 3D-CNN branch benefits from direct access to volumetric context while maintaining computational efficiency, making it suitable for scenarios where full 3D information is available. In contrast, the slice-based EfficientNetV2B3 branch leverages pretrained representations and majority voting to achieve exceptional staging performance with reduced computational overhead.

Compared with prior studies, the proposed framework achieves competitive or superior performance while explicitly addressing key limitations such as overfitting, computational cost, and lack of interpretability (Saxena et al., 2025, Sen et al., 2025, Srinivas et al., 2026, Stefanou et al., 2025). The use of explainable AI techniques is particularly important, as it enables qualitative validation of model behavior and supports clinician trust. The alignment of highlighted

regions with established neuroanatomical markers of Alzheimer's disease suggests that the models are learning meaningful patterns rather than spurious correlations.

Despite these strengths, several limitations remain. Performance may vary under scanner or protocol differences, and external validation on independent clinical cohorts is necessary to assess generalizability. Slice-based staging introduces design choices such as slice selection and voting strategy, which may influence outcomes. Future work should explore multi-modal data integration, longitudinal analysis, and prospective clinical evaluation.

In summary, this study provides evidence that a multi-architecture, explainable deep learning framework can support accurate, efficient, and interpretable Alzheimer's disease assessment, representing a meaningful step toward clinically deployable AI-assisted diagnostics.

10. CONCLUSION

This research presented a comprehensive multi-architecture deep learning framework for Alzheimer's disease detection and dementia staging using T1-weighted MRI. A lightweight 3D-CNN with efficient channel attention achieved 90.6% accuracy for binary CN vs AD classification on an ADNI-derived cohort. A transfer learning pipeline using EfficientNetV2B3 achieved 99.45% accuracy for four-class staging on the Kaggle dataset when using five middle slices per patient and patient-level majority voting. The integration of Grad-CAM and 3D occlusion mapping provides interpretable visual evidence supporting model predictions. These results support the feasibility of accurate, efficient, and interpretable deep learning tools for AI-assisted dementia assessment.

11. DATA AVAILABILITY

ADNI data are available through the Alzheimer's Disease Neuroimaging Initiative subject to data use agreements. The Kaggle dataset used for multi-class staging is publicly.

12. ACKNOWLEDGMENTS

The researchers wish to thank all the health care organizations which gave permission to utilize data collected within each organization for this study.

13. CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this research

14. DECLARATION

All the authors are confirming here that we have initially produced this manuscript in the form it is now; It hasn't been sent anywhere else for review or publication prior to submitting it to the Review Journal of Neurological & Medical Sciences Review and that all authors have read, agreed upon and accepted the final version of this paper to be submitted to the Review Journal of Neurological & Medical Sciences Review.

REFERENCES

ACHARYA, M., DEO, R. C., BARUA, P. D., DEVI, A. & TAO, X. 2025. EEGConvNeXt: A novel convolutional neural network model for automated detection of Alzheimer's Disease and Frontotemporal Dementia using EEG signals. *Computer Methods and Programs in Biomedicine,* 262**,** 108652.

AL-ISLAM, F., SANIM, M. S., GOH, K. O. M., MAHMUD, S. H. & NANDI, D. 2026. Alzheimer's Disease Prediction Using ANOVA with t-SNE Feature Selection Techniques and Ensemble Learning. *Journal of Advanced Research Design,* 144**,** 123-147.

ALSADHAN, N. A. 2025. Image-Based Alzheimer's Disease Detection Using Pretrained Convolutional Neural Network Models. *arXiv preprint arXiv:2502.05815*.

BANAIT, C. K., ULHE, P. & SHAPEKAR, K. V. 2026. Multimodal AI for Early Detection of Neurological Disorders: A Case Study on Alzheimer's and Parkinson's. *AI in Diagnostic Radiology: Clinical Applications and Case-Based Insights*. IGI Global Scientific Publishing.

CHAMAKURI, R. & JANAPANA, H. 2025. A systematic review on recent methods on deep learning for automatic detection of Alzheimer's disease. *Medicine in Novel Technology and Devices,* 25**,** 100343.

COHEN, I., TAYLOR, R. A., XUE, H., FAUSTINO, I. V., FESTA, N., BRANDT, C., GAO, E., HAN, L., KHASNAVIS, S. & LAI, J. M. 2025. Detection of emergency department patients at risk of dementia through artificial intelligence. *Alzheimer's & Dementia,* 21**,** e70334.

DAS, S., AWASTHI, A., RAWAL, R. K. & BHATIA, R. 2026. Biomarkers in disease diagnosis and monitoring: Insights into clinical applications and mass spectrometry-based detection. *Applied Biochemistry and Biotechnology*, 1-30.

DHARIA, S. Y., LIU, Q., SMITH, S. D. & VALDERRAMA, C. E. 2026. Dual-transformer cross-attention framework for Alzheimer's disease detection via dPTE-guided EEG channel selection and multi-modal integration. *Biomedical Signal Processing and Control*, 112, 108390.

GU, S. K. & PURUSHOTHAMAN, A. 2026. AlzFusionFormer: Integrating multiple transformers for early Alzheimer's disease detection from multi-modal data. *Biomedical Signal Processing and Control*, 112, 108601.

HUBER, H., MONTOLIU-GAYA, L., BRUM, W. S., VÁVRA, J., YAKOUB, Y., WENINGER, H., BRAUN-WOHLFAHRT, L. S., SIMRÉN, J., BOADA, M. & RUIZ, A. 2026. A minimally invasive dried blood spot biomarker test for the detection of Alzheimer's disease pathology. *Nature Medicine*, 1-10.

JASPHIN JENI SHARMILA, P. & SHINY ANGEL, T. 2024. Optimized machine learning model for Alzheimer and epilepsy detection from EEG signals. *Automatika: Časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 65, 597-608.

JOON, D., KUMAR, R., GUPTA, M. & OBAID, A. J. A comprehensive Analysis on Diagnosis of Alzheimer Disease Using Generative Adversarial Network. 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2024. IEEE, 1-6.

KACHARE, P., PURI, D., SANGLE, S. B., AL-SHOURBAJI, I., JABBARI, A., KIRNER, R., ALAMEEN, A., MIGDADY, H. & ABUALIGAH, L. 2024. LCADNet: a novel light CNN architecture for EEG-based Alzheimer disease detection. *Physical and Engineering Sciences in Medicine*, 47, 1037-1050.

KHANAPUR, S., NAYAK, J. S., RAJESHWARI, B., NAMRATHA, M., BHARADWAJ, C. B. & BHARDWAJ, R. 2026. SHAP-Based Explainability for Local and Global Insights in Alzheimer's Detection. *Engineering, Technology & Applied Science Research*, 16, 30940-30947.

KHOSROAZAD, S., ABEDI, A. & HAYES, M. J. 2023. Sleep signal analysis for early detection of Alzheimer's disease and related dementia (ADRD). *IEEE journal of biomedical and health informatics*, 27, 2264-2275.

LIU, S., ZHANG, Z., GU, Y., HAO, J., LIU, Y., FU, H., GUO, X., SONG, H., ZHANG, S. & ZHAO, Y. 2025. Beyond the eye: A relational model for early dementia detection using retinal OCTA images. *Medical Image Analysis*, 102, 103513.

LIU, Y., SUN, Y., ZHAO, P., LIU, Z., WANG, Y., WANG, Y., ZHAO, R., WANG, C. & WANG, S. 2026. Near-Infrared Fluorescent Probe for the Early Diagnosis of Alzheimer's Disease. *Medicinal Research Reviews*, 46, 5-39.

LLACA-SÁNCHEZ, B. A., GARCÍA-NOGUEZ, L. R., ACEVES-FERNÁNDEZ, M. A., TAKACS, A. & TOVAR-ARRIAGA, S. 2025. Exploring LLM Embedding potential for dementia detection using audio transcripts. *Eng*, 6, 163.

MARWA, E.-G., MOUSTAFA, H. E.-D., KHALIFA, F., KHATER, H. & ABDELHALIM, E. 2023. An MRI-based deep learning approach for accurate detection of Alzheimer's disease. *Alexandria Engineering Journal*, 63, 211-221.

OTTOY, J., OWSICKI, N., BILGEL, M., BINETTE, A. P., SALVADÓ, G., KANG, M. S., CASH, D. M., EWERS, M., LA JOIE, R. & WISSE, L. E. 2025. Recent advances in neuroimaging of Alzheimer's disease and related dementias. *Alzheimer's & Dementia*, 21, e70648.

RAFII, M. S. & AISEN, P. S. 2023. Detection and treatment of Alzheimer's disease in its preclinical stage. *Nature aging*, 3, 520-531.

SAMBANGI, S., KUSUMA, T., RAO, D. S. & LAKSHMEESWARI, G. 2026. Optimizing Digital Healthcare for Alzheimer's Disease: A Deep Federated Learning Convolutional Neural Network Scheme (DFLCNNS). *Computational Intelligence Algorithms for the Diagnosis of Neurological Disorders*. CRC Press.

SAXENA, S., CARPENTER, C., FLODEN, D. P., MELDON, S., TAYLOR, R. A. & HWANG, U. 2025. Novel algorithms & blood-based biomarkers: Dementia detection and care transitions for persons living with dementia in the emergency department. *Alzheimer's & Dementia*, 21, e70287.

SEN, S. Y., CURA, O. K., YILMAZ, G. C. & AKAN, A. 2025. Classification of Alzheimer's dementia EEG signals using deep learning. *Transactions of the Institute of Measurement and Control*, 47, 1353-1365.

SHUKLA, G. P., KUMAR, S., PANDEY, S. K., AGARWAL, R., VARSHNEY, N. & KUMAR, A. 2023. Diagnosis and detection of Alzheimer's disease using learning algorithm. *Big Data Mining and Analytics*, 6, 504-512.

SRINIVAS, L., SHANKAR, R. S., RAJESWARI, S., SOWJANYA, D., PRADHAN, S. N. & RAO, V. M. 2026. Enhancing dementia detection (EDD): A machine learning ensemble approach for early alzheimer's diagnosis. *Advances in Electrical and Computer Technologies*. CRC Press.

STEFANOU, K., TZIMOURTA, K. D., BELLOS, C., STERGIOS, G., MARKOGLOU, K., GIONANIDIS, E., TSIPOURAS, M. G., GIANNAKEAS, N., TZALLAS, A. T. & MILTIADOUS, A. 2025. A novel CNN-based framework for alzheimer's disease detection using EEG spectrogram representations. *Journal of Personalized Medicine*, 15, 27.

VANAJA, T., SHANMUGAVADIVEL, K., SUBRAMANIAN, M. & KANIMOZHISELVI, C. 2025. Advancing Alzheimer's detection: integrative approaches in MRI analysis with traditional and deep learning models. *Neural Computing and Applications*, 37, 8527-8546.

VRAHATIS, A. G., SKOLARIKI, K., KROKIDIS, M. G., LAZAROS, K., EXARCHOS, T. P. & VLAMOS, P. 2023. Revolutionizing the early detection of Alzheimer's disease through non-invasive biomarkers: the role of artificial intelligence and deep learning. *Sensors*, 23, 4184.

WALLENSTEN, J., WACHTLER, C., BOGDANOVIC, N., OLOFSSON, A., KIVIPELTO, M., JÖNSSON, L., PETROVIC, P. & CARLSSON, A. C. 2025. Machine learning to detect Alzheimer's disease with data on drugs and diagnoses. *The Journal of Prevention of Alzheimer's Disease*, 12, 100115.

WANG, L., GLASS, J., KOURTIS, L. & AU, R. 2026. Multi-modal data analysis for early detection of alzheimer's disease and related dementias. *The Journal of Prevention of Alzheimer's Disease*, 13, 100399.

WOJDAŁA, A. L., BELLOMO, G., GAETANI, L., TEUNISSEN, C. E., PARNETTI, L. & CHIASSERINI, D. 2025. Immunoassay detection of multiphosphorylated tau proteoforms as cerebrospinal fluid and plasma Alzheimer's disease biomarkers. *Nature Communications*, 16, 214.

YANG, H., KHAN, S. U. R., BILAL, O., CHEN, C. & ZHAO, M. 2025. CEOE-Net: Chaotic Evolution Algorithm-Based Optimized Ensemble Framework Enhanced with Dual-Attention for Alzheimer's Diagnosis. *Computer Modeling in Engineering & Sciences*, 145, 2401.