NeuroFusion for Robust and Explainable Multimodal Deep Learning in Fine-Grained Staging of Alzheimer's Disease Across Imaging and Clinical Biomarkers

Garima Shukla
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Mumbai, Maharashtra, India
garimashukla0719@gmail.com

Vanshaj Awasthi
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Mumbai, Maharashtra, India
awasthivanshaj@gmail.com

Prashant Dubey
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Mumbai, Maharashtra, India
prashant.dubey@s.amity.edu

Sakshi Nipane
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Mumbai, Maharashtra, India
sakshi.nipane@s.amity.edu

Sampurna Roy
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Mumbai, Maharashtra, India
sampurna.roy1@s.amity.edu

Rajiv Iyer
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University Mumbai, Maharashtra, India
rajivkjs@gmail.com (corresponding author)

Sumendra Yogarayan
Faculty of Information Science and Technology (FIST), Multimedia University (MMU), Jalan Ayer Keroh Lama, Ayer Keroh, Melaka, Malaysia
sumendra@mmu.edu.my (corresponding author)

Received: 20 August 2025 | Revised: 4 October 2025 and 17 October 2025 | Accepted: 19 October 2025

 DOI: https://doi.org/10.48084/etasr.14186

ABSTRACT

The subtle transition from healthy cognition to very mild dementia often leaves even experienced clinicians uncertain, highlighting the need for computational frameworks that integrate heterogeneous data sources

while providing clinically meaningful explanations. Existing Artificial Intelligence (AI) approaches frequently reduce Alzheimer's Disease (AD) prediction to binary classification, rely on a single modality, or lack interpretability, thereby limiting their translational impact. This study introduces a multimodal Deep Learning (DL) framework designed for fine-grained, four-stage staging of AD across Non-Demented, Very Mild, Mild, and Moderate states. The framework integrates structural Magnetic Resonance Imaging (MRI) and structured clinical biomarkers within a unified architecture, employing pre-trained convolutional and transformer-based networks for imaging alongside gradient-boosted decision trees for tabular features such as demographics, cognitive scores, and laboratory measures. To promote transparency and clinical trust, the framework incorporates complementary interpretability strategies: Gradient-weighted Class Activation Mapping (Grad-CAM) to identify discriminative neuroanatomical regions and attention mechanisms to highlight influential clinical variables. Evaluation was conducted on the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, with diagnostic labels determined at the subject level using clinician consensus informed by Clinical Dementia Rating (CDR) and Mini-Mental State Examination (MMSE) scores. The dataset comprised 229 Non-Demented, 398 Very Mild, 192 Mild, and 176 Moderate cases. GhostNet achieved an F1-score of 0.888 for Non-Demented and 0.847 for Mild Dementia, whereas Very Mild remained the most challenging stage (best F1-score ≤ 0.628). On structured clinical features, Extreme Gradient Boosting (XGBoost) attained an accuracy of 95.35%. Despite reduced performance for intermediate stages due to class imbalance and overlapping phenotypes, model training remained stable with minimal overfitting. This work provides one of the first interpretable multimodal frameworks for four-stage AD staging, advancing beyond conventional binary models and demonstrating the value of integrating complementary imaging and clinical modalities. By combining robust diagnostic accuracy with transparent, clinician-facing explanations, the approach offers a scalable and trustworthy pathway toward AI-enabled dementia staging, with particular promise for deployment in resource-limited healthcare systems.

Keywords-Alzheimer's Disease (AD) staging; multimodal Deep Learning (DL); structural MRI; clinical biomarkers; interpretable artificial intelligence; Convolutional Neural Networks (CNNs); transformer architectures

I. INTRODUCTION

Alzheimer's Disease (AD) imposes profound medical, social, and economic challenges, with dementia ranked among the leading causes of disability and dependency in older adults worldwide [1]. More than 55 million people currently live with dementia, and projections estimate this number will reach 139 million by 2050 [2]. The annual global cost of dementia care already exceeds USD 1.3 trillion, a figure expected to nearly double by 2030, placing immense strain on healthcare systems [3]. In India, more than 9 million older adults are affected, yet many cases remain undiagnosed or detected only in advanced stages, especially in rural areas with limited neurological infrastructure and specialist availability [4].

Timely detection is crucial, particularly during Mild Cognitive Impairment (MCI), when interventions can slow cognitive decline and significantly improve quality of life [5]. Structural Magnetic Resonance Imaging (MRI) has emerged as a valuable non-invasive modality, capable of identifying neuroanatomical biomarkers such as hippocampal atrophy and cortical thinning, but diagnostic reproducibility is often hindered by variability in imaging protocols and reliance on expert interpretation [6].

Artificial Intelligence (AI) and Deep Learning (DL) methods are increasingly applied to automate AD diagnosis. Convolutional neural networks (CNNs) can extract spatial patterns from MRI, whereas transformer architectures capture long-range dependencies in imaging and clinical features [7]. However, most existing approaches are limited to binary classification (Healthy vs. AD), overlooking intermediate stages such as Very Mild and Mild Dementia. In addition, unimodal imaging or clinical models face inherent limitations: CNN-based imaging systems frequently lack interpretability and generalizability, whereas clinical-only models fail to account for neuroanatomical patterns.

Several studies have sought to address these challenges by developing multimodal and explainable AD diagnostic frameworks. Authors in [8] proposed a multimodal Alzheimer's detection system combining 3D MRI and amyloid Positron Emission Tomography (PET), which improved diagnostic performance but required complex resources that limit clinical scalability. Authors in [9] introduced an explainable 3D residual self-attention deep neural network for MRI-based diagnosis, advancing interpretability but not achieving full multimodal integration. Authors in [10] systematically reviewed vision transformers for AD neuroimage analysis, highlighting both their potential and persistent challenges related to preprocessing, dataset size, and interpretability.

Further efforts combined modalities: authors in [11] developed DiaMond, a vision transformer using MRI and PET, achieving balanced accuracy but limited interpretability, whereas authors in [12] applied a multimodal surface-based transformer for early AD detection, with strong results but limited dataset diversity. Similarly, authors in [13] presented an interpretable Machine Learning (ML) framework fusing imaging and genetic data, though without robust multi-site validation. Data augmentation approaches, such as the use of generative AI with latent diffusion by authors in [14] to enhance MRI-based detection, showed performance gains but raised concerns about clinical relevance. In addition to these multimodal and transformer-based approaches, recent work by authors in [15] introduced an explainable ensemble framework for AD diagnosis that integrates Shapley values with ensemble learning techniques. Their study demonstrated that ensemble

methods, when combined with model-agnostic explainability, can deliver both strong predictive performance and transparent clinical insights. This work reinforces the growing consensus that accuracy alone is insufficient for real-world adoption of AI in healthcare; interpretability and clinician trust are equally critical. From this perspective, their findings provide a valuable complementary direction to multimodal frameworks such as the proposed model, which similarly emphasizes the integration of predictive robustness with transparent, clinically meaningful explanations.

To address these limitations, this study introduces a multimodal DL framework for fine-grained staging of AD across four clinically meaningful categories: Non-Demented, Very Mild, Mild, and Moderate Dementia. The framework integrates pretrained CNNs and transformer models for MRI with gradient-boosted decision trees for structured clinical features, incorporating interpretability mechanisms such as Gradient-weighted Class Activation Mapping (Grad-CAM) and attention visualizations. By combining accuracy, transparency, and scalability, this approach advances the development of clinically trustworthy AI systems for dementia staging.

II. METHODOLOGY

The methodology overcomes challenges in staging AD through a structured approach integrating clinical and imaging data. Three different architectures were chosen: GhostNet and Swin Transformer for MRI imaging data, and gradient-boosted decision trees for structured clinical features. These architectures were selected to take advantage of their previous successes and compatibility with interpretability methods such as Grad-CAM and attention maps. The hybrid approach leverages complementary data and the advantages of each model to increase classification accuracy while still being transparent, which is needed for clinical decisions.

A. Data Preprocessing

Two benchmark datasets were accessed to support multimodal analysis of AD. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset provided numerical data including cognitive scores, demographic details, clinical biomarkers, and genetic information, making it particularly suitable for modeling disease progression through clinical and biological factors [16]. To complement this, the Open Access Series of Imaging Studies (OASIS) dataset offered high-resolution structural MRI scans with clinical status labels covering healthy controls, MCI, and AD cases, enabling robust evaluation of image-based DL models [17]. The combined use of ADNI and OASIS ensured a balanced framework that integrates both clinical biomarkers and neuroimaging features for reliable and interpretable AD staging.

Preprocessing consisted of data cleaning, MRI normalization, skull stripping, spatial registration, wavelet-based feature extraction, and clinical data normalization with missing-value imputation. Clinical features were reduced by Principal Component Analysis (PCA) retaining 95% variance, and K-means clustering (k=3) was applied based on silhouette scores and expert input to cluster MRI features. Table I summarizes the computational methods for preprocessing and classification, including local contrast enhancement, tissue segmentation, feature extraction, class imbalance adjustment, and dimensionality reduction, with a brief description of each computational step in the workflow.

TABLE I. OVERVIEW OF APPLIED IMAGE PROCESSING AND DATA ANALYSIS TECHNIQUES

Method / Tool	Details
Contrast Limited Adaptive Histogram Equalization (CLAHE)	Enhances local contrast and highlights features like White Matter Lesions (WMLs)
K-means clustering	Segments tissues into 3 classes using k = 3
HAAR wavelet transform	Extracts low-frequency features using wavelet approximation
SMOTE / SMOTENC	Handles class imbalance using SMOTE for purely numerical data and SMOTENC for mixed-type features (numerical + categorical)
PCA	Retains components explaining 95% of the total variance

B. Model Development

Three architectures were selected for the three stages of AD, with each architecture matched to the corresponding data type. For MRI imaging data, GhostNet, a lightweight CNN, was applied to preprocessed 256 × 256 normalized images (fixed size) using the Adam optimizer and dropout layers to reduce overfitting. A Swin Transformer was also applied to MRI data, employing hierarchical patch-based self-attention with the Adam optimizer and dropout for overfitting reduction. For structured clinical data, Extreme Gradient Boosting (XGBoost) was tuned via grid search for optimal hyperparameters. Additional tabular models, including TabTransformer, FT-Transformer, and Hybrid CNN + Transformer, were incorporated for comparative evaluation of structured clinical features.

This setup brings together the CNNs that take advantage of the spatial imaging features of MRI data, the transformers that benefit from the complex dependencies that exist in the data (in particular the clinical data), and the gradient boosting for reliable tabular learning. Table II shows MRI and clinical models using CNNs, transformers, and XGBoost, all optimized for accurate, interpretable AD staging.

C. Model Interpretability

Interpretability analysis employed both visual and quantitative strategies to provide deeper insights into the decision-making process of the models and to enhance clinical trust. The quantitative methods included sanity checks, in which model weights or inputs were randomized to assess whether explanations remained consistent, occlusion sensitivity scoring (measuring prediction shifts when regions are hidden), and clinician review to ensure alignment with AD pathology. For visual interpretability, Grad-CAM heat maps generated from RegNet highlighted important MRI regions (Figure 1), whereas GhostNet attention maps, averaged and scaled over final layer channels, provided visual explanations of spatial focus (Figure 2). Table III color codes indicate significance levels: red/orange for highest, yellow/green for moderate, and blue for least importance.

TABLE II. ARCHITECTURE AND IMPLEMENTATION FOR EACH MODEL

Model	Type	Architecture & Implementation
RegNetY-400MF	MRI data	ImageNet-pretrained, ADNI-tuned 4-class dementia classifier with dropout and Grad-CAM
GhostNet-100	MRI data	AD MRI pipeline with Ghost modules; final attention maps show key regions
ConvNeXt-Tiny	MRI data	ImageNet-pretrained ConvNeXt-Tiny fine-tuned on ADNI MRI as a comparative CNN baseline for dementia staging
XGBoost	Clinical data	ADNI tabular baseline; auto-missing handling, 5-fold grid- tuned
TabTransformer	Clinical data	Fine-tuned Hugging Face model for ADNI with custom embeddings and attention tuning for dementia
FT-Transformer	Clinical data	ADNI features as tokens; transformer tuned for class imbalance and dementia-specific interactions
Hybrid CNN + Transformer	Clinical data	1D CNN for local patterns, Transformer for context; trained with dropout, batch normalization, and Adam
Swin Transformer	MRI data	Modified Swin Transformer for dementia stages; Conv2D, patch merging, dropout, and high ADNI separability

TABLE III. COLOR CODING

Color	Significance	Explanation
Red/Orange	Highest importance	Most influential prediction regions
Yellow/Green	Moderate importance	Relevant but less critical regions
Blue	Lowest importance	Minimal impact on predictions

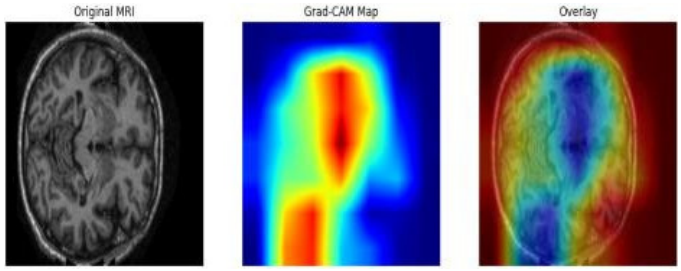

Fig. 1. Grad-CAM visualization for Very Mild Dementia showing the original MRI, Grad-CAM map, and overlay.

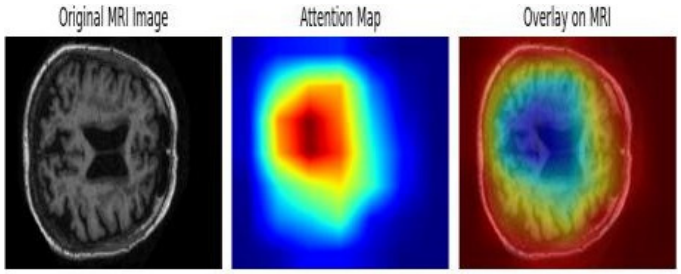

Fig. 2. Attention map for Mild Dementia showing the original MRI, attention map, and overlay.

III. RESULTS

A. *Performance Overview*

All performances were evaluated at each of the four stages of AD through accuracy, precision, recall, weighted F1-score, and class-specific Area Under the Curve (AUC), with McNemar's and DeLong's tests providing statistical significance. As presented at Table IV, Non-Demented and Mild Dementia achieved the highest overall performance (GhostNet F1-score = 0.888; RegNet marginally higher for Mild Dementia). Very Mild Dementia had significantly lower scores ($p < 0.05$) due to class imbalance, even with SMOTE. The bootstrap resampling consisting of 1,000 repetitions with 95% Confidence Interval (CI) yielded a GhostNet AUC of 0.953 (0.941–0.964) for Non-Demented and 0.781 (0.742–0.820) for Very Mild Dementia.

TABLE IV. F1-SCORES (IMAGE MODELS)

Class	RegNet	GhostNet
Non-Demented	0.810	0.888
Very Mild Dementia	0.628	0.598
Mild Dementia	0.804	0.847
Moderate Dementia	0.713	0.722

B. *Confusion Matrix Insights*

Confusion matrices were thoroughly analyzed to investigate class-specific prediction trends and to assess model sensitivity in detecting the Very Mild and Mild Dementia stages, which are particularly critical for early intervention and timely clinical decision-making.

Figures 3 and 4 include the confusion matrices for RegNet and GhostNet, respectively. GhostNet achieved better true positive rates than RegNet for Non-Demented (0.95) and Mild Dementia (0.916) but fell behind in Moderate Dementia (0.44 compared to 0.493).

Figures 5 and 6 include the confusion matrices for XGBoost and Hybrid CNN + Transformer, respectively. For tabular models, the CNN + Transformer revealed balanced classification with fewer misclassifications. XGBoost performed well, with FT-Transformer providing better recall and less overlap compared to TabTransformer.

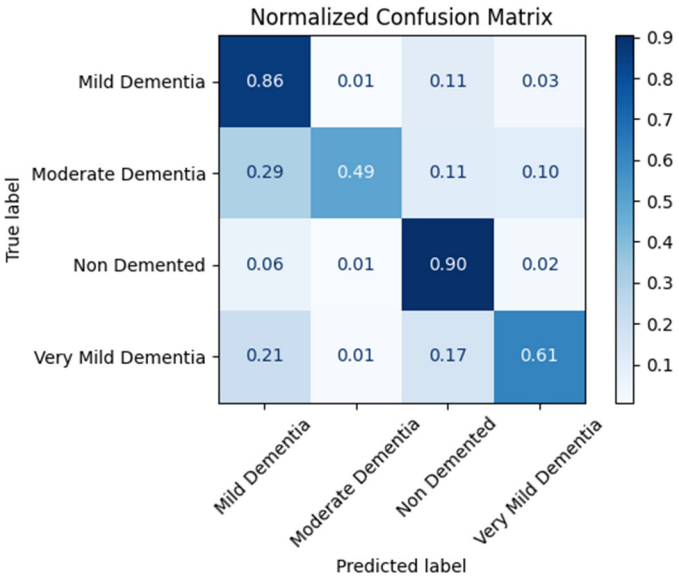

Fig. 3. Confusion matrix for RegNet.

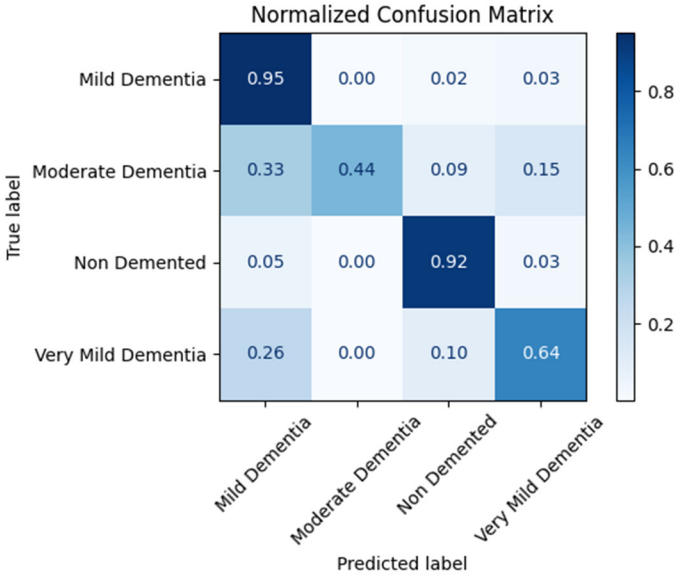

Fig. 4. Confusion matrix for GhostNet.

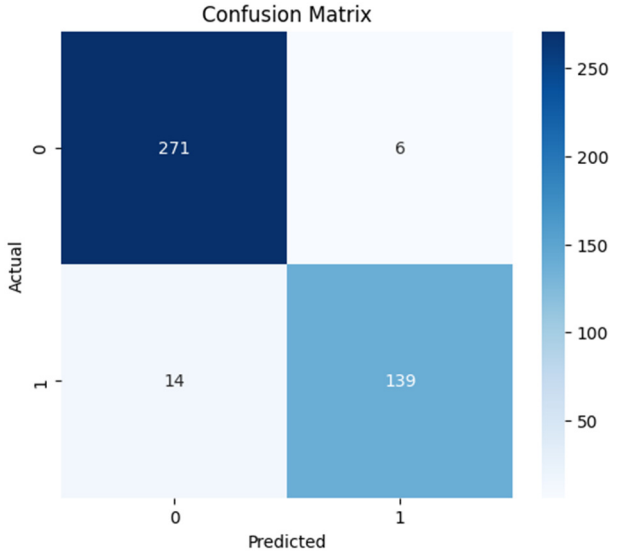

Fig. 5. Confusion matrix for XGBoost.

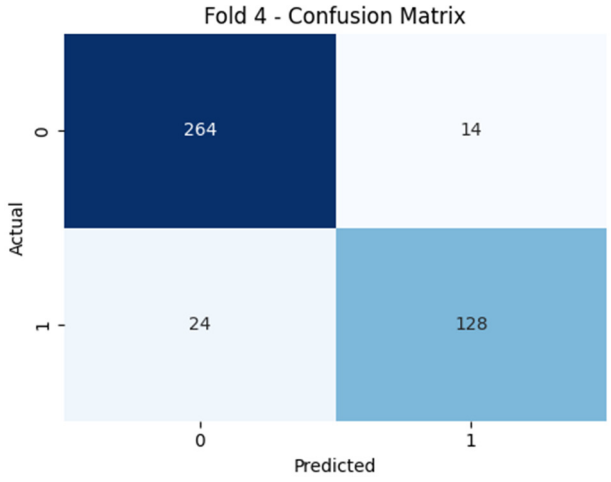

Fig. 6. Confusion matrix for hybrid CNN + Transformer.

Misclassifications in Very Mild Dementia highlight the clinical challenge of detecting early cognitive decline, where subtle symptoms often overlap with normal aging. The models still achieved high true positive rates for Non-Demented and Mild Dementia, supporting reliable distinction between diagnostic groups and timely intervention. These observations further highlight the need for additional subjects and more representative samples in the early disease stages to improve model sensitivity and robustness.

The imaging models applied a four-stage staging (Non-Demented, Very Mild, Mild, and Moderate Dementia); however, the confusion matrices of the tabular and hybrid models (Figures 5 and 6) depict a binary classification analysis. Specifically, the four clinical stages were grouped into two classes (Non-Demented and Demented) to evaluate the sensitivity of correctly identifying the healthy group and the symptomatic group and to reduce class imbalance in the ADNI cohort. This strategy supports the strength of clinical features to provide a high-confidence primary diagnosis, which in turn complements the detailed neuroanatomical staging achieved by the imaging architectures.

C. Precision–Recall and Receiver Operating Characteristic–Area Under the Curve

Precision–Recall and Receiver Operating Characteristic (ROC)–AUC analyses were performed to evaluate the discriminative capability of the imaging-based models across the four AD stages, with particular emphasis on the trade-off between sensitivity and specificity, which is a crucial consideration in clinical applications where both false negatives and false positives carry significant consequences. Both GhostNet and RegNet demonstrated strong performance for the Non-Demented and Mild Dementia categories, whereas their discrimination ability was reduced for the Very Mild and Moderate Dementia classes. Overall, GhostNet performed slightly better than RegNet across most categories. Table V presents the ROC-AUC values for all evaluated imaging models, including Swin Transformer and ConvNeXt.

TABLE V. ROC-AUC (IMAGE MODELS)

Class	RegNet	GhostNet	Swin Transformer	ConvNeXt
Non-Demented	0.975	0.978	0.975	0.971
Very Mild Dementia	0.950	0.953	0.950	0.944
Mild Dementia	0.975	0.979	0.975	0.962
Moderate Dementia	0.954	0.957	0.954	0.930

Figure 7 illustrates the ROC curves for RegNet, showing high class separability for Non-Demented and Mild Dementia, with AUC values exceeding 0.95, whereas weaker discrimination is observed for Very Mild and Moderate Dementia.

Figure 8 illustrates the Precision–Recall curves, with strong precision–recall balance for Non-Demented and Mild Dementia, moderate performance for Moderate Dementia, and reduced recall for Very Mild Dementia, underscoring the difficulty of early-stage detection.

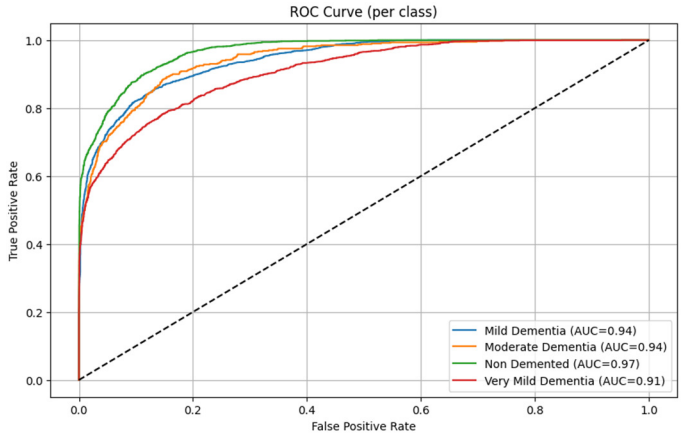

Fig. 7. ROC curve for RegNet.

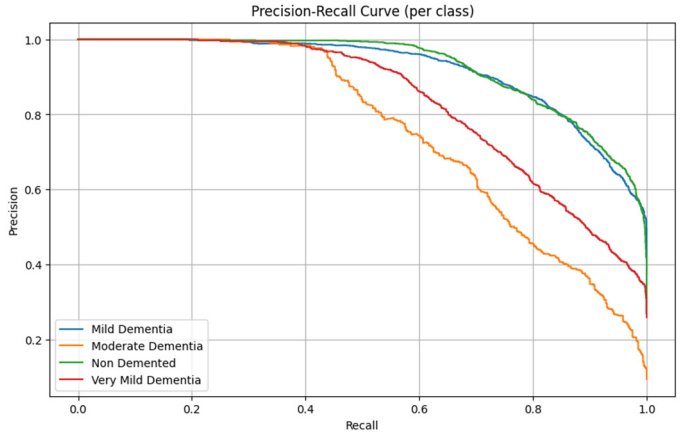

Fig. 8. Precision-Recall curve for RegNet.

For tabular models, XGBoost led with an AUC of 0.95, followed by CNN + Transformer, FT-Transformer, and TabTransformer, as seen in Table VI.

TABLE VI. TABULAR MODEL METRICS

Model	Accuracy (%)	F1-score	ROC-AUC
XGBoost	95.35	0.95	0.95
CNN + Transformer	90.00	0.85	0.94
FT-Transformer	85.58	0.80	0.91
TabTransformer	82.33	0.74	0.89

Figure 9 depicts the XGBoost loss curve, where log-loss quickly declines before stabilizing near 0.15, indicating fast convergence and minimal overfitting. Figure 10 shows the Hybrid CNN + Transformer training process, with a steep early drop in training loss and validation loss reaching its lowest point around epoch 30, demonstrating efficient learning and good generalization.

D. Training Dynamics

Training curves provided steady convergence across the board for all models. XGBoost registered stable log-loss (~15%) in 30 boosting iterations, indicating negligible overfitting. Hybrid CNN + Transformer reduced training loss steeply in the first 15 epochs and achieved minimum validation loss (~25%) at around epoch 30 when early stopping was initiated. TabTransformer registered steady loss decrease across 45 epochs with occasional minor dips in validation loss, demonstrating slow but reliable learning. In all models, no significant divergence was seen between training and validation curves, implying good generalization and not extreme overfitting.

IV. LIMITATIONS AND FUTURE SCOPE

While the proposed multimodal framework shows strong predictive performance on the ADNI dataset, generalizability to broader clinical populations with different imaging protocols and demographics may be compromised by data dependence on a single collection. Very Mild Dementia sensitivity remains reduced, further underscoring the difficulty of finding subtle early-stage changes. Although this study primarily focused on the ADNI dataset in combination with OASIS for multimodal analysis, further validation is necessary to establish the generalizability of the proposed NeuroFusion framework.

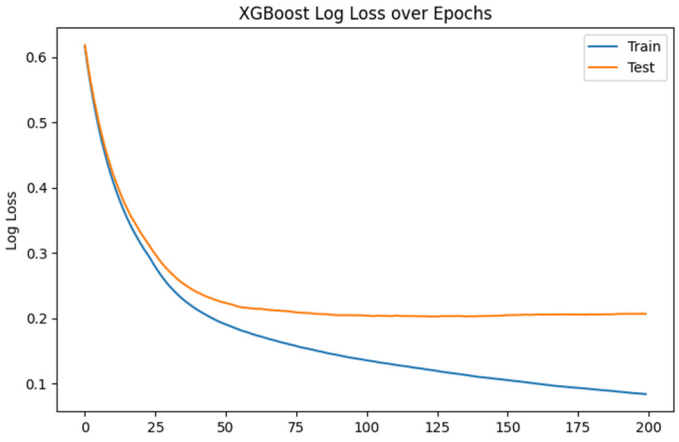

Fig. 9. XGBoost log-loss over epochs.

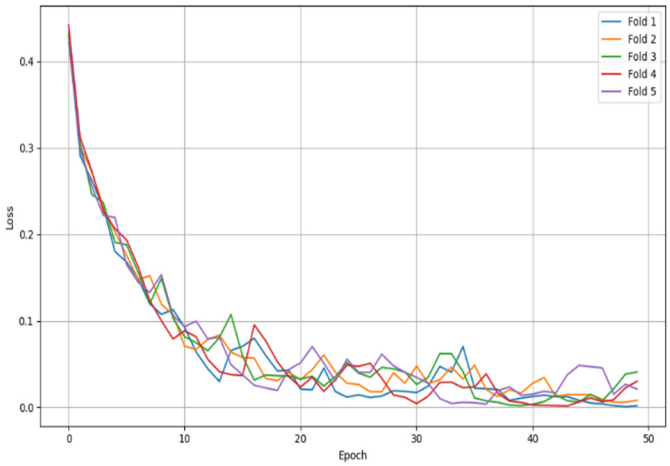

Fig. 10. Hybrid CNN + Transformer model training loss.

Future efforts will extend evaluation to external, multi-site, and demographically diverse cohorts such as the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing

(AIBL), the National Alzheimer's Coordinating Center (NACC), and other hospital-based or biobank repositories. Validation across heterogeneous imaging protocols, scanner types, and population subgroups will help address domain shift and site-specific variability that frequently limit clinical translation. In addition, longitudinal data from these cohorts will be incorporated to capture disease progression trajectories and enable predictive modeling of transition from MCI to AD. Exploring federated learning approaches may also allow leveraging multi-institutional data while preserving patient privacy. These directions are expected to enhance the robustness, scalability, and clinical applicability of NeuroFusion.

V. CONCLUSION

This study introduced NeuroFusion, an interpretable multimodal Deep Learning (DL) framework for fine-grained staging of Alzheimer's Disease (AD) across four clinically meaningful categories. The framework integrates structural Magnetic Resonance Imaging (MRI) with structured clinical biomarkers, leveraging convolutional, transformer-based, and gradient-boosted decision tree models within a unified architecture. To enhance clinical trust, interpretability is incorporated through Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations of neuroanatomical regions and attention mechanisms for tabular features.

The novelty of NeuroFusion lies in three core contributions: (i) moving beyond conventional binary or ternary staging by addressing four distinct stages of AD, (ii) achieving complementary fusion of imaging and clinical modalities to strengthen diagnostic robustness, and (iii) embedding interpretability at multiple levels to provide transparent, clinician-facing explanations. Compared with prior works that focus primarily on unimodal imaging [9], synthetic augmentation [14], or Explainable Artificial Intelligence (XAI)-based ensemble strategies [16], NeuroFusion delivers an integrated solution that combines multimodal fusion, stage-specific prediction, and explainability within a single framework.

Evaluation on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrates that NeuroFusion achieves strong classification performance, particularly for Non-Demented and Mild Dementia categories, while highlighting the persistent challenge of distinguishing Very Mild Dementia due to overlapping clinical phenotypes. These results emphasize both the promise and the limitations of current multimodal approaches in detecting early cognitive decline.

In summary, NeuroFusion contributes a novel, robust, and explainable pathway toward Artificial Intelligence (AI)-enabled dementia staging. By uniting predictive performance with transparency, the framework offers potential for deployment in real-world clinical workflows. Future research will extend validation across multi-site datasets, incorporate longitudinal modeling to capture disease progression, and explore counterfactual explanations to further strengthen clinical interpretability and adoption.

REFERENCES

[1] J.-H. Shin, "Dementia Epidemiology Fact Sheet 2022," *Annals of Rehabilitation Medicine*, vol. 46, no. 2, pp. 53–59, Apr. 2022, https://doi.org/10.5535/arm.22027.

[2] E. Nichols and T. Vos, "The estimation of the global prevalence of dementia from 1990-2019 and forecasted prevalence through 2050: An analysis for the Global Burden of Disease (GBD) study 2019," *Alzheimer's & Dementia*, vol. 17, no. S10, Dec. 2021, Art. no. e051496, https://doi.org/10.1002/alz.051496.

[3] "World Alzheimer Report 2022: Life after diagnosis: Navigating treatment, care and support." Alzheimer's Disease International. https://www.alzint.org/resource/world-alzheimer-report-2022/.

[4] "ADI - Dementia statistics." Alzheimer's Disease International. https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/.

[5] Alzheimer's Association, "2024 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 20, no. 5, pp. 3708–3821, May 2024, https://doi.org/10.1002/alz.13809.

[6] G. Shukla, V. Awasthi, D. Theng, R. Gupta, S. Nipane, and S. Singh, "Hybrid 3D CNN and ResNet Deep Transfer Learning for High-Resolution Hippocampal Atrophy Mapping and Automated Alzheimer's MRI Diagnosis: Deep Hybrid 3D CNN and ResNet Transfer Learning for High-Resolution Hippocampal Atrophy Mapping and Automated Alzheimer's MRI Diagnosis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 26047–26053, Aug. 2025, https://doi.org/10.48084/etasr.11372.

[7] J. Zhou *et al.*, "A deep learning model for early diagnosis of alzheimer's disease combined with 3D CNN and video Swin transformer," *Scientific Reports*, vol. 15, no. 1, July 2025, Art. no. 23311, https://doi.org/10.1038/s41598-025-05568-y.

[8] G. Castellano, A. Esposito, E. Lella, G. Montanaro, and G. Vessio, "Automated detection of Alzheimer's disease: a multi-modal approach with 3D MRI and amyloid PET," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, Art. no. 5210, https://doi.org/10.1038/s41598-024-56001-9.

[9] X. Zhang, L. Han, W. Zhu, L. Sun, and D. Zhang, "An Explainable 3D Residual Self-Attention Deep Neural Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis Using Structural MRI," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5289–5297, Nov. 2022, https://doi.org/10.1109/JBHI.2021.3066832.

[10] V. Mubonanyikuzo, H. Yan, T. E. Komolafe, L. Zhou, T. Wu, and N. Wang, "Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis," *Journal of Medical Internet Research*, vol. 27, no. 1, Feb. 2025, Art. no. e62647, https://doi.org/10.2196/62647.

[11] Y. Li, M. Ghahremani, Y. Wally, and C. Wachinger, "DiaMond: Dementia Diagnosis with Multi-Modal Vision Transformers Using MRI and PET," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision*, Tucson, AZ, USA, 2025, pp. 107–116, https://doi.org/10.1109/WACV61041.2025.00021.

[12] Q. A. Duong, S. D. Tran, and J. K. Gahm, "Multimodal surface-based transformer model for early diagnosis of Alzheimer's disease," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, Art. no. 5787, https://doi.org/10.1038/s41598-025-90115-y.

[13] M. E. Vlontzou, M. Athanasiou, K. V. Dalakleidi, I. Skampardoni, C. Davatzikos, and K. Nikita, "A comprehensive interpretable machine learning framework for mild cognitive impairment and Alzheimer's disease diagnosis," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, Art. no. 8410, https://doi.org/10.1038/s41598-025-92577-6.

[14] N. J. Dhinagar, S. I. Thomopoulos, and P. M. Thompson, "Generative AI improves MRI-based Detection of Alzheimer's Disease by using Latent Diffusion Models and Convolutional Neural Networks," *Alzheimer's & Dementia*, vol. 20, no. S2, Dec. 2024, Art. no. e089958, https://doi.org/10.1002/alz.089958.

[15] B. K. Raghupathy, M. R. Reddy, P. Theeda, E. Balasubramanian, R. K. Namachivayam, and M. Ganesan, "Harnessing Explainable Artificial Intelligence (XAI) based SHAPLEY Values and Ensemble Techniques for Accurate Alzheimer's Disease Diagnosis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20743–20747, Apr. 2025, https://doi.org/10.48084/etasr.9619.

[16] "Alzheimer's Disease Neuroimaging Initiative." ADNI. https://adni-lde.loni.usc.edu/.

[17] "Open Access Series of Imaging Studies (OASIS)." Oasisbrains. https://sites.wustl.edu/oasisbrains/.