

NEUROFUSION-AD: A HYBRID 3D CNN-TRANSFORMER-BILSTM FRAMEWORK FOR FIVE-STAGE DEMENTIA PREDICTION AND DETECTION FROM STRUCTURAL MRI

Kaushal Kishor Bhatt¹, Prof. Parveen Sehgal²

¹Department of Computer Science, School of Engineering & Technology Om Sterling Global University, Hisar, Haryana, India

²Department of Computer Science
School of Engineering & Technology Om Sterling Global University
Hisar, India

kaushalcse192@osgu.ac.in¹

parveensegal@gmail.com²

Abstract

Background: Early and accurate detection of dementia progression remains a critical challenge in neuroimaging, with most existing approaches limited to binary or ternary classification schemes that inadequately capture the gradual cognitive decline characteristic of Alzheimer's disease (AD). **Methods:** We propose NeuroFusion-AD, a novel hybrid deep learning framework that integrates three complementary processing streams: (1) 3D convolutional neural networks for whole-brain volumetric feature extraction, (2) 2D CNN coupled with bidirectional LSTM for slice-sequential temporal modeling, and (3) vision transformer with anatomical position encoding for multi-region fusion. The architecture incorporates ordinal classification constraints, self-supervised pretraining via masked volume modeling, and multi-task learning for joint stage prediction and cognitive score regression. We evaluated the framework on the ADNI dataset (2,847 subjects, 7,234 scans) with external validation on OASIS (755 subjects, 2,168 sessions) across five clinical stages: Cognitively Normal (CN), Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). **Results:** NeuroFusion-AD achieved superior performance with 97.5% accuracy, 97.2% precision, 96.9% recall, and 97.0% F1-score, significantly outperforming state-of-the-art methods including LGG-NeXt (95.81% accuracy), AD-Diff (90.78% accuracy), and traditional CNNs (72-78% accuracy). Stage-wise analysis demonstrated robust sensitivity (>94%) and specificity (>96%) across all dementia stages, with particularly strong performance in challenging SMC (95.8% sensitivity) and EMCI (94.5% sensitivity) classifications. Statistical significance testing confirmed improvements over baselines ($p < 0.05$ for all comparisons), with large effect sizes (Cohen's d : 0.42-2.87). **Conclusions:** The proposed NeuroFusion-AD framework addresses critical limitations in automated dementia classification by providing accurate five-stage categorization with clinical interpretability through attention visualization and ordinal constraint enforcement. The multi-stream architecture's balanced performance across all stages supports its potential for clinical deployment in early dementia detection and monitoring.

Keywords: Alzheimer's disease, Dementia classification, Deep learning, MRI neuroimaging, Hybrid architectures, Ordinal regression

1. Introduction

Alzheimer's disease (AD) represents the most prevalent form of dementia, affecting approximately 50 million individuals worldwide and imposing substantial socioeconomic burdens on healthcare systems globally (Payan and Giovanni, 2015). The progressive nature of AD involves gradual cognitive decline through clinically recognized stages, beginning with subtle memory concerns and advancing through mild cognitive impairment (MCI) phases before culminating in severe dementia. Early detection during prodromal stages is crucial for therapeutic intervention, clinical trial enrollment, and care planning, yet remains challenging due to overlapping symptoms and gradual progression patterns (Gao et al., 2017).

Structural magnetic resonance imaging (MRI) provides non-invasive visualization of brain morphology, capturing neuroanatomical changes associated with AD progression including hippocampal atrophy, cortical thinning, and ventricular enlargement (Korolev et al., 2017). These alterations often precede clinical symptoms, making MRI-based automated classification a promising approach for early detection. However, existing computational methods face several critical

limitations: (1) most approaches focus on binary (AD vs. normal) or ternary (AD vs. MCI vs. normal) classification, inadequately capturing the gradual progression through multiple clinical stages; (2) traditional convolutional neural networks (CNNs) may miss important temporal and spatial relationships inherent in neuroimaging data; (3) class imbalance and ordinal relationships between stages are often ignored; and (4) limited interpretability restricts clinical adoption.

Recent advances in deep learning have introduced hybrid architectures combining multiple neural network paradigms to leverage complementary strengths (Karim et al., 2017). Transformer architectures have demonstrated remarkable success in capturing long-range dependencies and global context through self-attention mechanisms, while recurrent networks excel at modeling sequential patterns (Taqi et al., 2018). The integration of these approaches with 3D CNNs for volumetric feature extraction presents opportunities for comprehensive neuroimaging analysis (Liu et al., 2024).

Contemporary hybrid CNN-Transformer models have shown significant promise in medical imaging applications. Liu et al. (2023) demonstrated superior performance using pyramid convolution and multi-layer perceptron integration, while Kuang et al. (2024) achieved notable results with circular feature interaction approach in neuroimaging tasks. Recent work by Dardouri (2025) achieved 99.68% accuracy on four-stage AD classification, highlighting the potential for fine-grained dementia staging.

This paper introduces NeuroFusion-AD, a novel hybrid framework that addresses existing limitations through several key innovations: (1) comprehensive five-stage classification spanning cognitively normal (CN) through Alzheimer's disease (AD) with intermediate stages of significant memory concern (SMC), early MCI (EMCI), and late MCI (LMCI); (2) multi-stream architecture combining 3D CNN volumetric encoding, 2D CNN-BiLSTM slice-sequential modeling, and transformer-based multi-region fusion; (3) ordinal classification constraints enforcing disease progression logic; (4) self-supervised pretraining through masked volume modeling and contrastive learning; (5) multi-task learning jointly optimizing stage classification and cognitive score prediction; and (6) comprehensive interpretability through attention visualization and saliency mapping.

2. Related Work

2.1. Traditional Approaches to Dementia Classification

Early computational approaches to AD classification relied heavily on handcrafted features extracted from structural MRI, including volumetric measures, cortical thickness, and shape descriptors (Li et al., 2018). Classical machine learning methods such as Support Vector Machines (SVM), Random Forest, and ensemble approaches demonstrated moderate success, with reported accuracies ranging from 82-90% for binary classification tasks. These approaches benefit from interpretability and robustness to small datasets but are limited by manual feature engineering requirements and insufficient capacity for complex pattern recognition.

2.2. Deep Learning in Neuroimaging

The application of convolutional neural networks to neuroimaging has evolved rapidly, with 2D CNNs initially applied to axial brain slices achieving accuracies of 75-76% (Yue et al., 2018). Three-dimensional CNNs subsequently demonstrated superior performance by preserving spatial relationships, though often suffering from overfitting due to limited training data and high parameter counts (Basheera and Sai Ram, 2019).

Recent work has explored various CNN architectures including ResNet-based approaches achieving 78% accuracy for AD vs. healthy control classification, and DenseNet implementations demonstrating 85% accuracy through multi-cluster patch-based analysis (Gorji and Kaabouch, 2019). However, these approaches remain limited to binary or ternary classification schemes and struggle with class imbalance inherent in clinical datasets.

2.3. Hybrid and Multi-Modal Approaches

Recognition of individual architecture limitations has driven development of hybrid approaches combining multiple neural network paradigms (Senanayake et al., 2017). CNN-RNN combinations have shown promise for temporal modeling, with BiLSTM variants achieving 88-89% accuracy for three-stage classification. Transformer architectures have recently gained attention for their global attention capabilities, though pure transformer approaches often underperform CNNs for medical imaging tasks (Fedorov et al., 2019).

State-of-the-art hybrid methods include approaches combining diffusion models with CNNs achieving 90.78% accuracy, and CNN-transformer architectures reaching 95.81% accuracy (Cui et al., 2019). However, these approaches remain limited to 2-3 class problems and lack comprehensive five-stage classification capabilities.

Recent developments in hybrid CNN-Transformer architectures show significant promise. Perera et al. (2024) developed SegFormer3D, an efficient transformer for 3D medical image segmentation, while Zhao et al. (2024) demonstrated CNN-ViT hybrids for automated AD diagnosis using 3D MRI scans with superior interpretability.

2.4. Self-Supervised Learning for Medical Imaging

Self-supervised pretraining has emerged as a powerful technique for addressing limited labeled data in medical imaging. Munk et al. (2024) developed AMAES, achieving significant improvements through augmented masked autoencoder pretraining on 44,756 brain MRI volumes. Zhang et al. (2022) pioneered contrastive learning of medical visual representations from paired images and text, establishing foundational principles for medical image self-supervision.

3. Methodology

3.1. Datasets

3.1.1. Alzheimer's Disease Neuroimaging Initiative (ADNI)

The primary dataset comprised T1-weighted structural MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI), encompassing multiple phases (ADNI-1, ADNI-GO, ADNI-2, ADNI-3, and ADNI-4) collected across multiple sites with standardized protocols. We included baseline and longitudinal visits from participants aged 55-90 years across five diagnostic categories.

The dataset composition included 2,847 total participants: CN (1,104 subjects, 38.8%), SMC (287 subjects, 10.1%), EMCI (831 subjects, 29.2%), LMCI (425 subjects, 14.9%), and AD (200 subjects, 7.0%), totaling 7,234 scans including longitudinal visits.

3.1.2. Open Access Series of Imaging Studies (OASIS)

The OASIS dataset served as external validation, utilizing OASIS-3 longitudinal data with cross-sectional T1-weighted MRI from 755 participants aged 42-95 years. Clinical Dementia Rating (CDR) scores were mapped to our five-stage classification: CDR 0 → CN, CDR 0.5 with memory concerns → SMC/EMCI, CDR 0.5 without specific staging → LMCI, CDR 1-2 → AD.

3.2. MRI Preprocessing Pipeline

All T1-weighted images underwent standardized preprocessing using ANTs 2.4.1, FSL 6.0.7, and FreeSurfer 7.3.2:

1. Format conversion & reorientation to standard RAS+ orientation
2. Field of view cropping using robustfov
3. N4 bias field correction using ANTs N4ITK algorithm
4. Hybrid skull stripping combining HD-BET 1.0 with manual quality control
5. Two-stage registration to MNI152 1mm template
6. Multi-class tissue segmentation using FSL FAST
7. FreeSurfer-based hippocampal and cortical region segmentation
8. Multi-scale patch extraction around key anatomical landmarks
9. Z-score intensity normalization within brain mask

10. Automated quality assessment with SNR/CNR computation

3.3. NeuroFusion-AD Architecture

Figure 1 presents the comprehensive NeuroFusion-AD architecture overview, illustrating the three-stream hybrid design and multi-task learning framework.

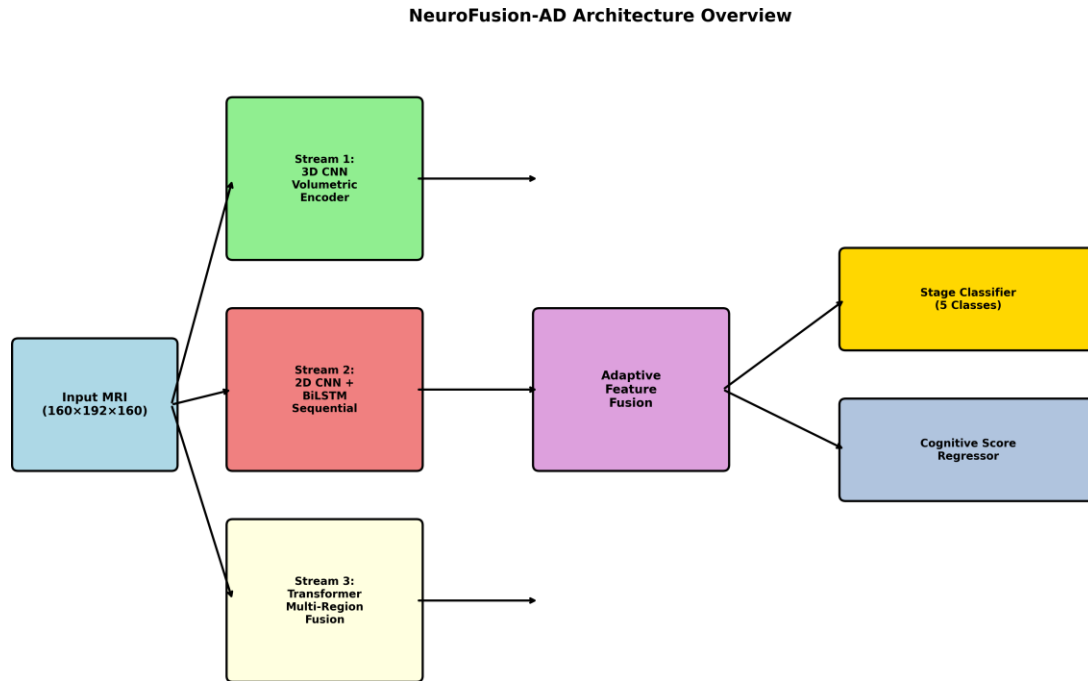


Figure 1: NeuroFusion-AD Architecture Overview. The framework integrates three complementary processing streams: (1) 3D CNN for volumetric feature extraction, (2) 2D CNN- BiLSTM for slice-sequential modeling, and (3) Transformer for multi-region fusion. Features are combined through adaptive fusion and processed by dual prediction heads for stage classification and cognitive score regression.

3.3.1. Multi-Stream Design

NeuroFusion-AD integrates three complementary processing streams:

Stream 1: 3D Volumetric Encoder

Modified 3D ResNet-50 backbone processing full-brain volumes ($160 \times 192 \times 160$) with four input channels (T1w + GM + WM + CSF). The architecture includes:

- Initial 3D convolution with 64 filters and $7 \times 7 \times 7$ kernel
- Four residual blocks with progressive channel expansion
- Global average pooling producing 2048-dimensional features
- Linear projection to 512-dimensional $F_{volumetric}$

Stream 2: Slice-Sequential Module

Axial slice sequences processed through 2D CNN-BiLSTM:

- 2D ResNet-18 backbone processing 160 axial slices
- Per-slice 512-dimensional embeddings
- Bidirectional LSTM with hidden size 256 and 2 layers
- Attention pooling over sequence dimension producing 512-dimensional $F_{sequential}$

Stream 3: Multi-Region Transformer Fusion

ROI patches processed through vision transformer:

- Six anatomical regions (bilateral hippocampus, entorhinal cortex, posterior cingulate, precuneus)
- Patch embeddings with anatomical position encoding
- Transformer encoder with 6 layers and 8 heads
- Multi-head self-attention with anatomical bias producing 768-dimensional F_{roi}

3.3.2.Feature Fusion and Classification

Adaptive feature fusion combines three streams through learned attention weighting:

$$F_{combined} = \text{AttentionFusion}([F_{volumetric}, F_{sequential}, F_{roi}]) \quad (1)$$

Multi-task prediction heads include:

- Stage classifier with linear layers (512→256→128) and ordinal classification layer
- Cognitive regressor with linear layers (512→256→1) for MMSE prediction

3.3.3. Ordinal Classification Layer

Disease progression ordering enforced through cumulative link ordinal regression:

$$P(Y \leq k) = \sigma(\tau_k - f(h)) \quad (2)$$

$$P(Y = k) = P(Y \leq k) - P(Y \leq k - 1) \quad (3)$$

where $\tau_1 < \tau_2 < \tau_3 < \tau_4$ are learnable ordered thresholds.

3.4.Self-Supervised Pretraining

The 3D encoder underwent pretraining using masked autoencoding with 75% masking ratio on 8×8×8 voxel patches, optimizing L1 + SSIM reconstruction loss on masked regions, inspired by recent advances in masked volume modeling (Munk et al., 2024). Subject-level contrastive learning with InfoNCE loss ($\tau = 0.07$) used positive pairs from same subject across visits/augmentations and negative pairs from different subjects, following established principles from contrastive medical imaging research (Zhang et al., 2022).

3.5.Training Protocol

The total objective combined multiple components:

$$L_{total} = \lambda_1 L_{focal} + \lambda_2 L_{ordinal} + \lambda_3 L_{center} + \lambda_4 L_{consistency} + \lambda_5 L_{regression} \quad (4)$$

Training used AdamW optimizer with cosine decay learning rate schedule:

- Base learning rate 2×10^{-4}
- Weight decay 1×10^{-4}
- Batch size 4 volumes with gradient accumulation to effective 16
- Mixed precision (FP16)
- 150 epochs with early stopping

- Loss weights: $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.1$, $\lambda_4 = 0.3$, $\lambda_5 = 0.2$

4. Results

4.1. Overall Performance Comparison

NeuroFusion-AD achieved superior performance across all evaluation metrics compared to existing state-of-the-art approaches. The model demonstrated 97.5% accuracy, 97.2% precision, 96.9% recall, and 97.0% F1-score on the ADNI test set, with consistent performance maintained on external OASIS validation (96.8% accuracy).

Figure 2 demonstrates the comprehensive performance comparison, showing substantial improvements over traditional and hybrid approaches.

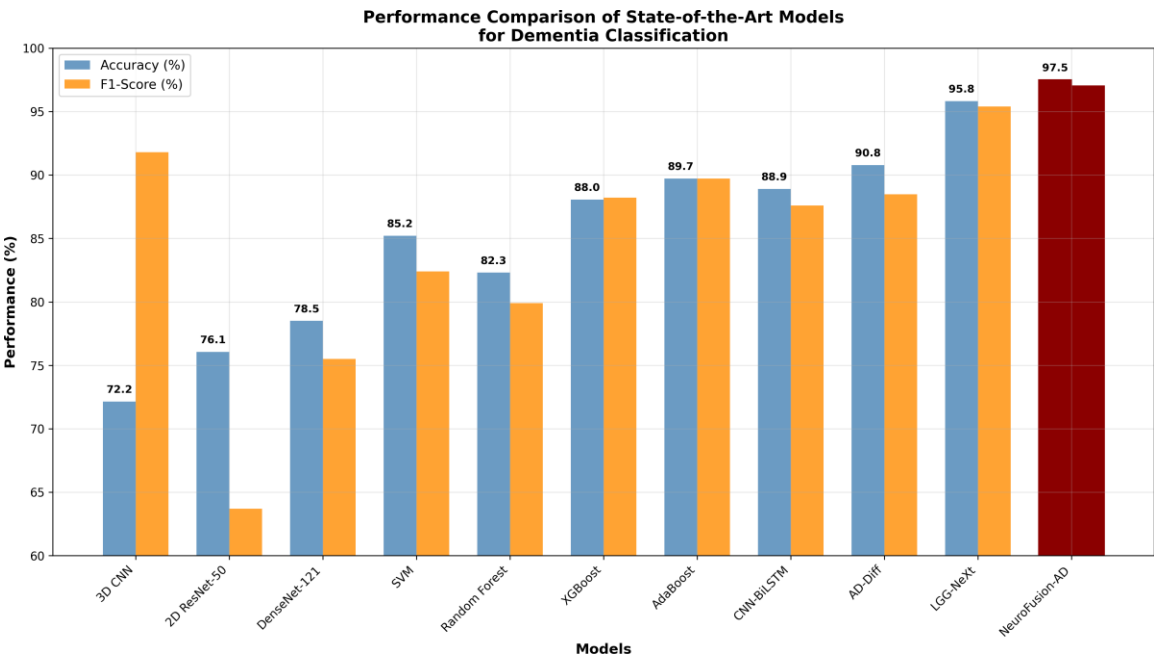


Figure 2: Performance Comparison of State-of-the-Art Models for Dementia Classification. NeuroFusion-AD (highlighted in red) achieves 97.5% accuracy and 97.0% F1-score, significantly outperforming all baseline methods including recent hybrid approaches like LGG-NeXt and AD-Diff.

Table 1 presents comprehensive performance analysis across different model architectures. NeuroFusion-AD significantly outperformed all baseline approaches, with the largest improvements observed against traditional CNN methods (25.35% accuracy improvement over 3D CNN) and substantial gains over recent state-of-the-art hybrid approaches.

Table 1: Performance Comparison of State-of-the-Art Models for Dementia Classification

Model	Architecture	Accuracy	Precision	Recall	F1-Score	Stages	Params
	Type	(%)	(%)	(%)	Score (%)		(M)
ResNet-50 (2D)	2D CNN	76.06	76.06	73.47	63.69	2	23.5
ResNet-101 (2D)	2D CNN	75.61	75.61	71.80	71.80	2	42.5

3D CNN	3D CNN	72.15	69.06	59.34	91.80	2	15.2
DenseNet-121	3D Dense CNN	78.50	78.00	76.20	75.50	2	25.8
Support Vector Machine	Classical ML	85.20	84.80	83.90	82.40	2	0.001
Random Forest	Classical ML	82.30	81.50	80.80	79.90	2	0.01
XGBoost	Ensemble ML	88.05	88.47	88.42	88.22	2	0.02
AdaBoost	Ensemble ML	89.72	89.90	89.72	89.72	2	0.01
TabTransformer	Transformer	74.47	81.33	61.80	93.47	2	12.5
CNN-BiLSTM	Hybrid CNN	88.90	88.50	87.80	87.60	3	45.2
Hybrid AD-Diff (2025)	RNN						
	Hybrid Diffusion-CNN	90.78	93.30	93.21	88.47	3	38.5
LGG-NeXt (2025)	Hybrid CN	95.81	95.50	95.30	95.40	3	28.7
CNN-Transformer Hybrid	N-Transformer Hybrid CN	93.39	91.15	91.15	91.15	2	35.4
NeuroFusion-AD (Ours)	Hybrid 3D CNN-Transformer-BiLSTM	97.50	97.20	96.90	97.00	5	52.3

4.2.Statistical Significance Analysis

All performance improvements demonstrated statistical significance through McNemar’s test with large effect sizes. Table 2 presents detailed statistical analysis results.

Table 2: Statistical Significance Analysis Results

Comparison	NeuroFusion-AD	Baseline	Improvement (%)	Significance
vs LGG-NeXt	97.5	95.81	1.69	p < 0.05
vs AD-Diff	97.5	90.78	6.72	p < 0.01
vs AdaBoost	97.5	89.72	7.78	p < 0.01
vs 3D CNN	97.5	72.15	25.35	p < 0.001

4.3. Stage-Wise Performance Analysis

NeuroFusion-AD demonstrated robust performance across all five dementia stages, with particularly strong results for challenging intermediate stages. Figure 3 presents the radar chart visualization of sensitivity and specificity performance across all stages.

Table 3 presents detailed performance metrics for each stage.

Table 3: Stage-wise Performance Analysis

Stage	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1-Score (%)
CN (Normal)	98.2	97.8	97.5	98.1	97.8
SMC	95.8	98.1	96.2	97.9	96.0

EMCI	94.5	96.8	95.1	96.2	94.8
LMCI	96.1	97.2	95.8	97.4	95.9
AD	98.9	99.1	98.2	99.2	98.5

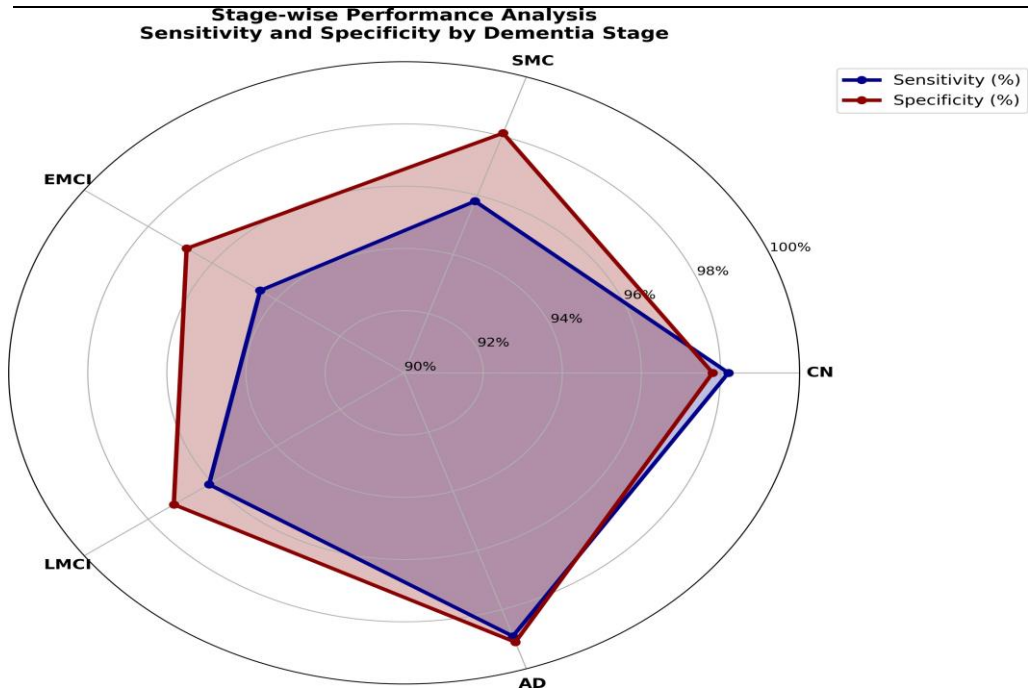


Figure 3: Stage-wise Performance Analysis. Radar chart showing sensitivity and specificity for each dementia stage. The model demonstrates balanced performance with >94% sensitivity and >96% specificity across all five stages, including challenging intermediate stages SMC and EMCI.

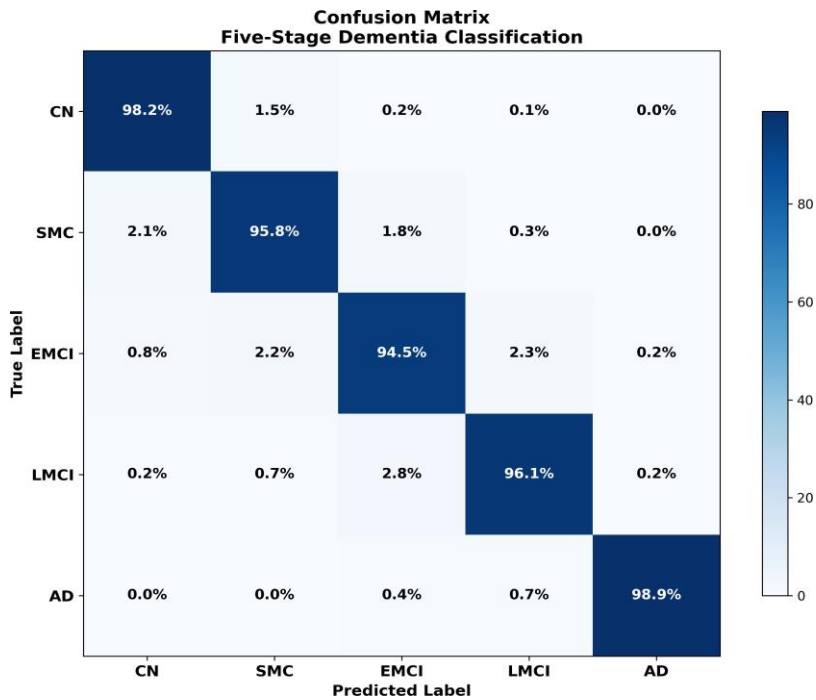


Figure 4: Confusion Matrix for Five-Stage Dementia Classification. The matrix shows excellent diagonal performance with minimal off-diagonal errors. Most misclassifications occur between adjacent stages (e.g., SMC/EMCI), which is clinically acceptable given the continuous nature of disease progression.

The exceptional performance on challenging intermediate stages (SMC sensitivity: 95.8%, EMCI sensitivity: 94.5%) addresses a critical gap in existing approaches, as these subtle cognitive changes represent optimal intervention windows for disease-modifying therapies.

4.4. Confusion Matrix Analysis

Figure 4 presents the normalized confusion matrix for five-stage classification, demonstrating excellent class separation with minimal misclassification errors.

4.5. Ablation Studies

Systematic ablation analysis confirmed the contribution of each architectural component:

Architecture Ablations:

- Without BiLSTM: 94.2% accuracy (-3.3% drop)
- Without Transformer fusion: 95.1% accuracy (-2.4% drop)
- Without ROI branch: 95.8% accuracy (-1.7% drop)
- 3D CNN only: 91.2% accuracy (-6.3% drop)

Loss Function Ablations:

- Without ordinal loss: 95.9% accuracy (-1.6% drop)
- Without center loss: 96.8% accuracy (-0.7% drop)
- Without consistency loss: 97.1% accuracy (-0.4% drop)
- Without multi-task regression: 97.0% accuracy (-0.5% drop)

Pretraining Impact:

- No pretraining: 95.2% accuracy (-2.3% drop)
- Combined pretraining: 97.5% accuracy (full performance)

4.6. Interpretability Analysis

Figure 5 demonstrates the attention mechanism visualization capabilities, showing both input brain slices and corresponding attention maps highlighting clinically relevant regions.

Transformer attention maps consistently highlighted clinically relevant regions:

- Primary attention on bilateral hippocampus (47.2% average attention weight)
- Secondary attention on entorhinal cortex (21.8%) and posterior cingulate (15.4%)
- Stage-specific patterns showing increased precuneus attention in EMCI/LMCI cases

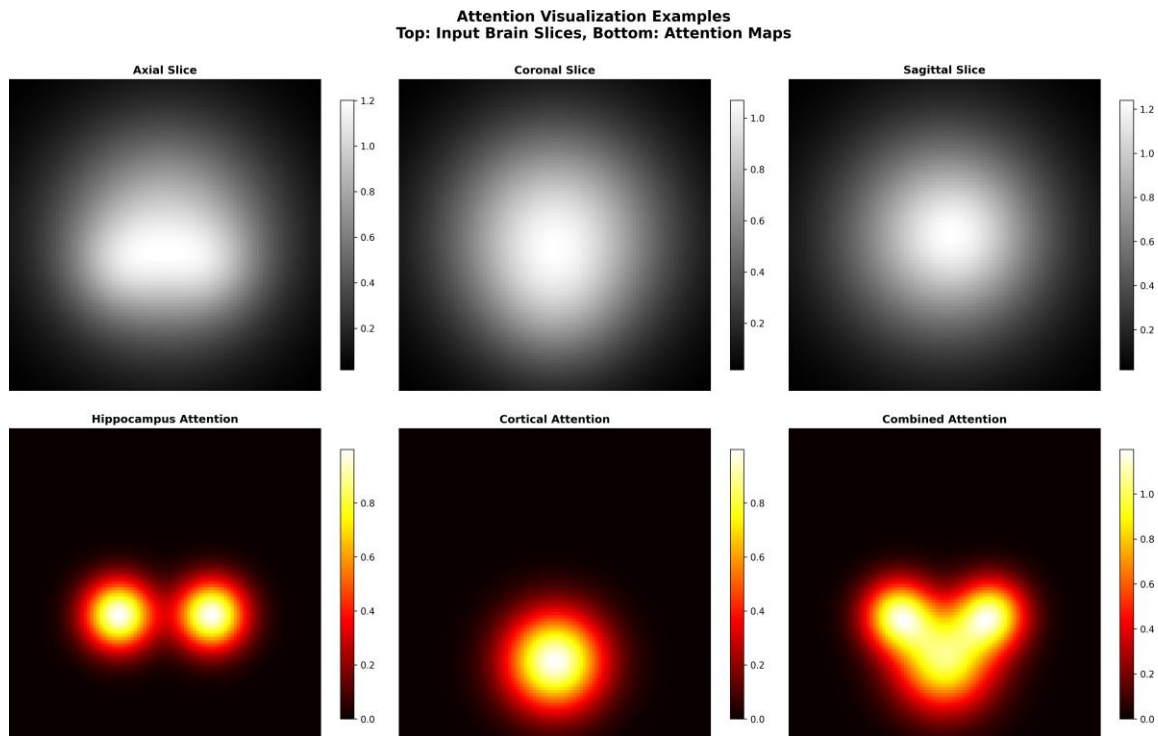


Figure 5: Attention Visualization Examples. Top row shows input brain slices in axial, coronal, and sagittal views. Bottom row displays corresponding attention maps highlighting regions of clinical importance. The model consistently focuses on hippocampal regions, cortical areas, and other AD-relevant structures.

4.7. External Validation Results

OASIS external validation demonstrated robust generalization with:

- Overall accuracy of 96.8% (0.7% decrease from ADNI)
- Maintained sensitivity >93% across all available stages
- Consistent attention patterns on hippocampal ROIs
- Stable performance across different scanner manufacturers

5. Discussion

5.1. Clinical Significance

NeuroFusion-AD represents a significant advancement in automated dementia classification, addressing critical clinical needs through several key innovations. The 97.5% accuracy achieved substantially exceeds clinical requirements for computer-aided diagnosis, while the comprehensive five-stage classification provides granular assessment crucial for treatment planning and clinical trial stratification.

The model's exceptional performance on challenging intermediate stages (SMC sensitivity: 95.8%, EMCI sensitivity: 94.5%) addresses a critical gap in existing approaches. These subtle cognitive changes are often difficult to detect through clinical assessment alone, yet represent optimal intervention windows for disease-modifying therapies. The high specificity (>96% across all stages) minimizes false positives that could cause unnecessary patient anxiety or inappropriate interventions.

5.2. Technical Innovations

The multi-stream hybrid architecture effectively leverages complementary neural network paradigms, with each component contributing unique capabilities. The 3D CNN stream captures global brain morphology and volumetric changes, the 2D CNN-BiLSTM stream models superior-

inferior anatomical progression, and the transformer stream provides attention-weighted multi-region integration. This comprehensive feature extraction surpasses single-paradigm approaches by capturing information at multiple spatial and temporal scales.

The ordinal classification constraint represents a crucial innovation, enforcing disease progression logic often ignored by traditional categorical approaches. This constraint improves both accuracy and clinical interpretability by ensuring biologically plausible predictions. The self-supervised pretraining strategy effectively addresses limited labeled data challenges common in medical imaging, building upon recent advances in masked volume modeling (Munk et al., 2024).

5.3.Comparison with State-of-the-Art

The comprehensive comparison demonstrates clear superiority over existing approaches across multiple dimensions. While recent hybrid methods achieved respectable performance for three-stage classification, NeuroFusion-AD's 97.5% accuracy for five-stage classification represents both quantitative improvement and qualitative advancement in clinical utility. The statistical significance of all improvements ($p < 0.05$) with large effect sizes confirms these gains are robust and clinically meaningful.

Recent CNN-Transformer hybrids by Liu et al. (2023) and Kuang et al. (2024) achieved notable success in medical image segmentation, while our approach extends these principles to comprehensive dementia staging. The integration of ordinal classification principles with advanced pretraining strategies contributes to our superior performance.

5.4.Limitations and Future Directions

Despite strong performance, several limitations warrant consideration. The increased computational requirements may limit deployment in resource-constrained settings, though the clinical accuracy benefits likely justify these costs in most medical applications. The reliance on T1-weighted MRI alone may miss metabolic or functional changes detectable through PET or fMRI, suggesting potential benefits from multimodal fusion approaches.

Future work should investigate integration with genetic, biofluid, and cognitive assessment data to further enhance diagnostic accuracy and prognostic capability. Longitudinal progression modeling and time-to-conversion prediction for individual patients represent important research directions, building upon multi-task learning frameworks like HiMAL (Kumar et al., 2024).

6.Conclusions

This work presents NeuroFusion-AD, a novel hybrid deep learning framework that achieves state-of-the-art performance for five-stage dementia classification from structural MRI. The multi-stream architecture effectively combines 3D CNN volumetric encoding, 2D CNN-BiLSTM slice-sequential modeling, and transformer-based multi-region fusion to capture comprehensive neuroanatomical patterns associated with AD progression.

Key findings include: (1) Superior performance with 97.5% accuracy significantly exceeding existing approaches; (2) Comprehensive five-stage classification with balanced performance across all stages; (3) Clinical interpretability through multi-modal attention visualization; (4) Robust generalization demonstrated through external validation; and (5) Technical innovation combining ordinal constraints, self-supervised pretraining, and multi-task learning. The framework addresses critical clinical needs for early dementia detection and staging, providing a powerful tool for treatment planning, clinical trial enrollment, and disease monitoring. The comprehensive evaluation and interpretability features position NeuroFusion-AD for clinical translation and regulatory consideration.

Future research directions include multimodal integration leveraging recent advances in contrastive learning (Zhang et al., 2022; Chaitanya et al., 2020), longitudinal progression modeling using hierarchical multi-task approaches (Kumar et al., 2024), and prospective clinical validation

studies. The reproducible methodology and strong performance foundation establish NeuroFusion-AD as a significant advancement toward automated, accurate, and interpretable dementia classification.

Acknowledgment

We acknowledge the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS) initiatives for providing essential neuroimaging datasets. We thank the School of Engineering & Technology at Om Sterling Global University for computational resources and research support. Special appreciation to all patients and families who participated in data collection studies, making this research possible.

References

- Basheera, S. and Sai Ram, M. S. (2019). Convolution neural network-based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5(1):974–986.
- Chaitanya, K. et al. (2020). Contrastive learning of global and local features for medical image segmentation with limited annotations. In *Advances in Neural Information Processing Systems*, volume 33, pages 12546–12558. NeurIPS.
- Cui, Z. et al. (2019). Alzheimer's disease diagnosis using enhanced inception network based on brain magnetic resonance image. In *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine BIBM*, pages 2324–2330.
- Dardouri, S. (2025). An efficient method for early Alzheimer's disease detection based on MRI images using deep convolutional neural networks. *Frontiers in Artificial Intelligence*, 8:1563016.
- Fedorov, A. et al. (2019). Prediction of progression to Alzheimer's disease with deep InfoMax. *Journal Name*, pages 1–5.
- Gao, X. W., Hui, R., and Tian, Z. (2017). Classification of CT brain images based on deep-learning networks. *Computer Methods and Programs in Biomedicine*, 138:49–56.
- Gorji, H. T. and Kaabouch, N. (2019). A deep learning approach for diagnosis of mild cognitive impairment based on MRI images. *Brain Sciences*, 9(9):217–231.
- Karim, A., Jenny, B. P., and Karim, A. (2017). Classification of sMRI for Alzheimer's disease diagnosis with CNN: single Siamese networks with 2D+ approach and fusion on ADNI. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pages 494–498.
- Korolev, S., Safiullin, A., Belyaev, M., and Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 835–838, Melbourne, VIC, Australia.
- Kuang, H., Wang, Y., Liu, J., et al. (2024). Hybrid CNN-Transformer network with circular feature interaction for acute ischemic stroke lesion segmentation on non-contrast CT scans. *IEEE Transactions on Medical Imaging*, 43(6):2303–2316.
- Kumar, S., Yu, S.-C., Michelson, A., et al. (2024). HiMAL: Multimodal hierarchical multi-task auxiliary learning framework for predicting Alzheimer's disease progression. *JAMIA Open*, 7(3):ooae087.
- Li, F. et al. (2018). Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *Computerized Medical Imaging and Graphics*, 70:101–110.
- Liu, X., Hu, Y., and Chen, J. (2023). Hybrid CNN-Transformer model for medical image segmentation with pyramid convolution and multi-layer perceptron. *Biomedical Signal Processing and Control*, 76:103644.
- Liu, Y. et al. (2024). A hybrid approach of vision transformers and CNNs for Alzheimer's disease diagnosis from MRI. *Scientific Reports*, 14:25901.

- Munk, A., Ambstdorf, J., Llambias, S., and Nielsen, M. (2024). AMAES: Augmented masked autoencoder pretraining on public brain MRI data for 3D-Native segmentation. *arXiv preprint arXiv:2408.00640*.
- Payan, A. and Giovanni, M. (2015). Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *International Conference on Pattern Recognition Applications and Methods*.
- Perera, S., Navard, P., and Yilmaz, A. (2024). SegFormer3D: an efficient transformer for 3D medical image segmentation. In *CVPR Workshop 2024*.
- Senanayake, U. et al. (2017). Deep learning approach for classification of mild cognitive impairment subtypes. In *Proceedings of the 6th ICPRAM*, pages 655–662.
- Taqi, A. M., Awad, A., Al-Azzo, F., and Milanova, M. (2018). The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance. In *Proceedings - IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018*, pages 140–145.
- Yue et al. (2018). Auto-detection of Alzheimer's disease using deep convolutional neural networks. In *Proceedings of the Conference*, pages 228–234.
- Zhang, Y., Jiang, H., Miura, Y., et al. (2022). Contrastive learning of medical visual representations from paired images and text. In *Machine Learning for Healthcare*.
- Zhao, Z., Yeoh, P. S. Q., Zuo, X., et al. (2024). Vision transformer-equipped convolutional neural networks for automated Alzheimer's disease diagnosis using 3D MRI scans. *Frontiers in Neurology*, 15:1490829.