A Multi-Modal Attention-Guided Network for Alzheimer's Disease Classification Using Deep Learning

Mahmoud Baniata
Department of Computer Science, Faculty of Information Technology, Applied Science Private University, Amman, Jordan
m_baniata@asu.edu.jo

Suhaila Abuowaida
Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al al-Bayt University, Mafraq, Jordan
suhila@aabu.edu.jo

Mohammad Aljaidi
Department of Cyber Security, Zarqa University, Zarqa, Jordan
mjaidi@zu.edu.jo

Mohammad Kharabsheh
Department of Computer Information Systems, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan
mohkh86@hu.edu.jo

Ayoub Alsarhan
Department of Information Technology, The Hashemite University, Zarqa, Jordan
ayoubm@hu.edu.jo

Amjad A. Alsuwaylimi
Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia
amjad.alsuwaylimi@nbu.edu.sa (corresponding author)

Received: 2 June 2025 | Revised: 11 July 2025, 18 July 2025, 20 July 2025, and 28 July 2025 | Accepted: 1 August 2025

 DOI: https://doi.org/10.48084/etasr.12510

ABSTRACT

Alzheimer's Disease (AD) is a chronic neurodegenerative disease that affects a large portion of the global population, and early and accurate diagnosis is the key component to proper management and treatment. This work presents the MAGNet model, a novel Deep Learning (DL) architecture for AD classification based on multi-modal imaging data. The MAGNet model uses multi-modal attention that is capable of hierarchically fusing structural MRI, functional MRI, and PET scans to obtain different kinds of information. The MAGNet model was tested on three datasets, with an overall accuracy of 96.2% when distinguishing between the AD, MCI, and CN groups. The proposed approach surpasses benchmark models by achieving 3.5% better accuracy and 5.2% higher sensitivity for early MCI diagnosis. Moreover, the MAGNet model offers interpretable results, employing attention visualization to support clinicians' decisions. MAGNet has the potential to predict cognitive scores and brain age with MMSE errors of 1.8 and a brain age of 2.3 years.

Keywords-multi-modal; attention; Alzheimer's disease; deep learning; ResNet

I. INTRODUCTION

Alzheimer's Disease (AD) is the most common type of dementia, affecting approximately 44 million patients worldwide [1]. Since AD is a progressive disease whose pathological course begins to develop several years before actual clinical manifestations appear, early and accurate identification is essential [2]. In particular, new MRI and PET diagnostic methods are useful in identifying early brain changes that are characteristic of AD [3]. Developments in Deep Learning (DL) have seen a considerable enhancement in medical image analysis [4], especially in the classification of AD [5]. However, most studies focus on individual image types or use direct concatenation on multi-modal datasets, which can degrade intricate inter-modality dependencies [6, 7]. In addition, the interoperability problem in many DL models has limited their use in clinical settings [8-10]. The evolution of AD follows cognitive deterioration from normal aging patterns through Mild Cognitive Impairment (MCI) to dementia [11]. MCI is used as a significant marker for intervention, but not all patients diagnosed with MCI develop AD. However, the ability to fine-tune this process in such a way that will help pinpoint the population of MCI patients with the probability of developing AD remains a major puzzle [12]. However, no specific DL model exists to directly address AD classification using multi-modal neuroimaging data [13].

This study aimed to fill this research gap by introducing the MAGNet (Multi-Modal Attention-Guided Network) model. MAGNet incorporates a feature extraction process in a bid to incorporate rich intermodal dependencies [14]. These include a hierarchical feature fusion strategy in which the deep network retains information at multiple scales for fine-grained and global features. An auxiliary framework involves two auxiliary tasks (cognitive score prediction and brain age estimation) [15] to enhance feature generalization and clinical outcomes. An interpretability module uncovers attention maps that indicate areas of the brain that have the greatest impact on classification decisions [16]. These innovations can lead to high classification accuracy and better differentiation between subjects with CN, MCI, and AD, compared to other methods. In addition, the interpretability features of the MAGNet model can offer clinically related information about the model's decisions and possibly help in the diagnosis of biomarkers of early AD. The specific objectives of this study are as follows:

- Validate the MAGNet model in a large and diverse neuroimaging dataset.
- Investigate how well the MAGNet model performs in classifying AD and distinguishing between MCI and healthy.
- Compare the MAGNet model with other approaches specialized in the same task to determine the clinical utility of the MAGNet model attention maps and their correlation with previous knowledge of AD biomarkers. The aim is to evaluate whether the proposed framework improves average performance and yields clinically beneficial predictions.

II. RELATED WORK

The use of DL in the classification of AD has been a relatively prominent field in recent years. Earlier DL models aimed at AD classification were mainly based on a single imaging modality. In [11], a deep belief network was used in AD diagnosis, utilizing structural MRI (sMRI) with high accuracy, although the source data was restricted to a single type. In the same manner, in [13], 3D CNNs were used in sMRI data to confirm the ability of DL in this field.

As different imaging techniques offered different kinds of information, several studies attempted to combine them. In [14], a multi-model fusion autoencoder-based schema was proposed for the classification of AD from MRI and PET data. In [15], a multi-modal cascade CNN was proposed, which incorporated MRI, PET, and clinical data to outperform single-modal methods. In [16], an attention-based deep network was proposed to diagnose brain diseases, exhibiting better interpretability, but this approach focused on single-modality data. In [17], a large framework employed ML for both the identification of the disease state and the prediction of the MMSE score, while the benefits of such practices were discussed regarding generalizability. In [18], a 3D CNN was used to analyze fMRI scans, achieving an accuracy of 92.8% and AUC of 0.95, but this method was constrained to modality-specific fMRI data and outlined that genetic factors did not bring improvements to AD prediction. In [19], the aim was to design a computationally efficient architecture for clinical usage while maintaining performance at a competitive level. However, light-weight models potentially lose feature extraction ability, while the subtle differences caused by the diseases might be missed. This limitation is critical in early-stage AD diagnosis, since detailed feature analysis is essential.

The MAGNet model enhances these developments by including multi-modal fusion, attention processes, multi-task learning, and interpretability into a cohesive framework. In contrast to earlier methods that often focus on one or two aspects, MAGNet thoroughly addresses the complexities of AD classification, especially in the early identification of MCI.

III. MAGNET MODEL ARCHITECTURE

A. Dataset and Preprocessing Pipeline

A comprehensive multi-dataset approach was employed to maximize the diversity and scale of neuroimaging data available for AD classification. Three complementary neuroimaging datasets were selected to provide both multi-modal neuroimaging capabilities and enhanced sample diversity, collectively comprising 8,362 subjects across different acquisition protocols and demographic distributions.

1) OASIS-3 Dataset: Primary Multi-Modal Source

The Open Access Series of Imaging Studies 3 (OASIS-3) was the primary data source for multi-modal neuroimaging analysis, selected due to its comprehensive multi-modal neuroimaging protocol, longitudinal design, and established reputation as a gold standard for AD neuroimaging research. Data access was obtained through the OASIS data sharing platform, following institutional review, board approval, and completion of the OASIS data use agreement [20, 21].

OASIS-3 represents a retrospective compilation of neuroimaging and clinical data collected from 1,098 participants over 30 years through ongoing studies at the Washington University Knight Alzheimer Disease Research Center. The participant cohort includes 755 cognitively normal adults and 622 individuals at various stages of cognitive decline, ranging in age from 42 to 95 years. All participants were assigned randomized identifiers and underwent comprehensive clinical assessments, including cognitive evaluations, demographic data collection, and multi-modal neuroimaging acquisitions. The dataset encompasses participants from multiple ongoing longitudinal studies, including the Memory and Aging Project, Adult Children Study, and Healthy Aging and Senile Dementia investigations. This diverse recruitment strategy ensures robust representation across different risk profiles and cognitive trajectories, enhancing the generalizability of machine learning models trained on these data. The longitudinal nature of OASIS-3, with participants followed over multiple years, provides valuable insights into disease progression patterns, although this study focuses on cross-sectional analysis for classification purposes.

OASIS-3 provides comprehensive multi-modal neuroimaging data encompassing structural, functional, and metabolic brain imaging modalities. The multi-modal protocol enables the proposed model to leverage complementary information from different neuroimaging techniques, each capturing distinct aspects of brain pathology associated with Alzheimer's disease progression.

- Structural MRI (sMRI): The dataset contains over 2,000 high-resolution T1-weighted structural MRI sessions acquired using standardized protocols across multiple scanner platforms. T1-weighted images provide detailed anatomical information on brain structure, allowing the detection of regional atrophy patterns characteristic of AD, particularly in medial T1-weightedlobe structures, including the hippocampus and the entorhinal cortex. Additional structural sequences include T2-weighted, FLAIR, and Diffusion Tensor Imaging (DTI) acquisitions, providing complementary structural and microstructural information.
- Functional MRI (fMRI): OASIS-3 includes resting-state functional MRI data capturing spontaneous neural activity patterns during rest conditions. Resting-state fMRI enables the assessment of functional connectivity alterations and network disruptions associated with AD pathology. The functional imaging data provides insights into brain network integrity and functional alterations that can precede structural changes, offering potential for early detection of disease.
- Positron Emission Tomography (PET): The dataset includes more than 1,500 PET imaging sessions utilizing multiple tracers to assess different aspects of brain pathophysiology. Fluorodeoxyglucose (FDG) PET imaging measures regional brain metabolism, typically showing characteristic hypometabolic patterns in temporoparietal regions in AD. Amyloid PET imaging using Pittsburgh Compound B (PIB) and AV-45 tracers enables direct visualization of amyloid plaque deposition, a hallmark pathological feature of AD. The inclusion of Centiloid standardization for amyloid tracers ensures comparability between different PET tracers and scanning protocols.

OASIS-3 provides comprehensive clinical and demographic information for all participants, including detailed cognitive assessments, medical history, medication usage, and genetic information. Cognitive evaluations include the Mini-Mental State Examination (MMSE), the Clinical Dementia Rating (CDR), and comprehensive neuropsychological test batteries assessing multiple cognitive domains. Apolipoprotein E (APOE) genotyping data are available for most participants, providing important genetic risk factor information relevant to AD susceptibility.

2) Complementary Dataset Sources

a) OASIS-1 Dataset Integration

The OASIS-1 dataset provides an additional 416 subjects with high-quality T1-weighted structural MRI data, including 100 subjects clinically diagnosed with AD and 316 cognitively normal controls. OASIS-1 serves as a clinical reference standard, contributing well-characterized subjects with established diagnostic criteria. The dataset was obtained through the OASIS [20, 22].

b) Augmented Alzheimer MRI Dataset

The Augmented Alzheimer MRI Dataset [23] comprises 6,848 subjects with enhanced data diversity through comprehensive augmentation techniques. This dataset provides T1-weighted MRI images across four diagnostic categories: non-demented, very mild dementia, mild dementia, and moderate dementia.

3) Multi-Modal Preprocessing Pipeline

a) Structural MRI Preprocessing

All T1-weighted sMRI data underwent standardized preprocessing using established neuroimaging analysis pipelines. The preprocessing workflow included skull stripping using the Brain Extraction Tool (BET) from the FSL software package, followed by bias field correction using the N4ITK algorithm to correct for intensity inhomogeneities caused by magnetic field variations. Spatial normalization was performed using Advanced Normalization Tools (ANTs) to register all images to the Montreal Neurological Institute (MNI) 152 standard space with 1 mm isotropic resolution.

b) Functional MRI Preprocessing

Resting-state fMRI data preprocessing followed standard procedures for temporal and spatial processing. The initial preprocessing steps included slice-timing correction to account for differences in acquisition timing across brain slices, followed by motion correction using MCFLIRT from FSL to estimate and correct for head movement during scanning. Spatial smoothing using a 6 mm full-width half-maximum Gaussian kernel was applied to improve signal-to-noise ratio and account for residual anatomical differences between subjects. Temporal preprocessing included high-pass filtering with a cutoff frequency of 0.008 Hz to remove low-frequency scanner drift and physiological noise. Band-pass filtering between 0.008-0.1 Hz was applied to focus analysis on

frequencies of interest for resting-state connectivity analysis. Additional preprocessing steps included regression of nuisance variables, including six motion parameters, white matter and cerebrospinal fluid signals, and global signal regression to minimize the influence of non-neural signal sources.

c) PET Image Preprocessing

PET image preprocessing focused on spatial normalization and intensity standardization to enable quantitative analysis across subjects and scanner platforms. PET images were co-registered to the corresponding structural MRI using rigid-body transformation to ensure accurate anatomical localization. Spatial normalization to MNI standard space was performed using the transformation parameters derived from structural MRI preprocessing to maintain consistency across modalities. For FDG-PET images, intensity normalization was performed using cerebellar gray matter as a reference region, generating Standardized Uptake Value Ratios (SUVRs) that account for individual differences in tracer uptake and injection parameters. Amyloid PET images were processed using the established Centiloid methodology to generate standardized measures of amyloid burden that are comparable across different tracers and scanning protocols. This standardization approach enables direct comparison of amyloid burden measures across the diverse PET imaging protocols represented in OASIS-3.

d) Data Quality Control and Subject Selection

Comprehensive quality control procedures were implemented to ensure data integrity and minimize the influence of artifacts on model training. Quality control criteria included assessment of image quality, motion parameters, and preprocessing success rates across all modalities. Subjects were excluded if any modality failed quality control criteria, including excessive motion (mean framewise displacement > 0.5mm for fMRI), significant scanner artifacts, or preprocessing failures.

The final quality-controlled dataset from OASIS-3 comprised 956 subjects with complete multi-modal neuroimaging data meeting all quality criteria. Additional subjects with partial modality data were retained for ensemble learning approaches that can accommodate missing modalities through appropriate masking and weighting strategies.

B. Proposed MAGNet Model Architecture

Figure 1 presents an overview of the proposed MAGNet model for large-scale neuroimaging analysis. This framework integrates three complementary neuroimaging datasets to maximize both the scale and diversity of training data while maintaining the sophisticated multi-modal learning capabilities essential for accurate AD classification.

1) Multi-Dataset Integration Strategy

The MAGNet model employs a multi-dataset approach that leverages the unique characteristics of three major neuroimaging databases: OASIS-3 for complete multi-modal learning, OASIS-1 for clinical reference standards, and an Augmented Alzheimer MRI Dataset for enhanced diversity and scale. This integration addresses the fundamental challenge of limited sample size while ensuring robust generalization across different imaging protocols and patient populations.

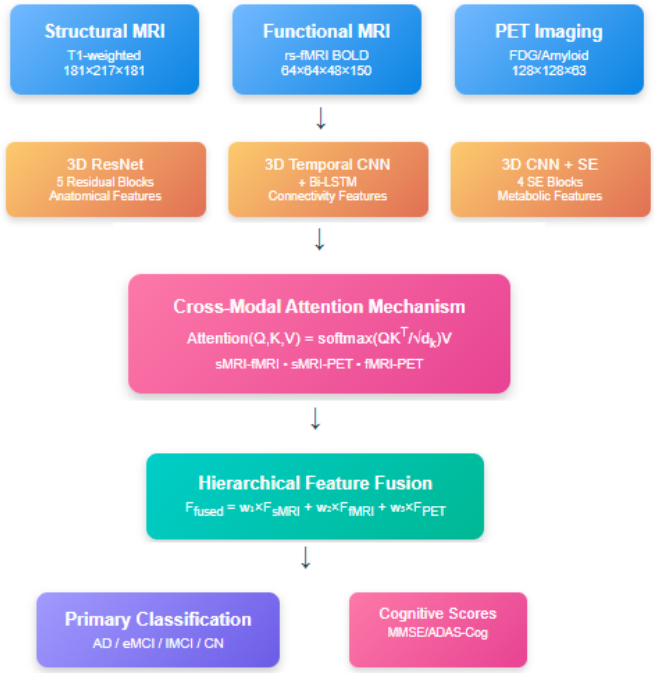

Fig. 1. Overview of the proposed MAGNet model architecture.

a) Dataset-Specific Architecture Adaptation

- OASIS-3 Multi-Modal Processing: The complete MAGNet architecture processes all three modalities (sMRI, rs-fMRI, and PET) from OASIS-3 through specialized pathways with full cross-modal attention mechanisms.
- OASIS-1 and Augmented Dataset Processing: For datasets containing only structural MRI, the model employs intelligent modality completion strategies and adaptive attention masking to maintain architectural consistency while accommodating missing modalities.

2) Input Modalities and Data Sources

a) Primary Multi-Modal Source: OASIS-3

The OASIS-3 dataset serves as the foundation for multi-modal learning, providing comprehensive neuroimaging data from 1,098 participants aged 42-95 years. Each subject contributes multiple modalities:

- Structural MRI (sMRI): High-resolution T1-weighted anatomical scans capturing detailed brain morphology and atrophy patterns.
- Resting-state functional MRI (rs-fMRI): Bold signal acquisitions during rest conditions revealing functional connectivity alterations.
- PET Imaging: FDG metabolic and amyloid imaging provides direct assessment of brain metabolism and pathological protein deposition.

b) Clinical Reference Source: OASIS-1

The OASIS-1 dataset contributes 416 carefully characterized subjects (100 AD, 316 CN) with high-quality T1-weighted sMRIs acquired under standardized clinical research

protocols. This dataset provides clinical precision and serves as a gold standard reference for model validation.

c) Enhanced Diversity Source: Augmented Alzheimer MRI Dataset

The Augmented Alzheimer MRI Dataset provides 6,848 subjects across four diagnostic categories (non-demented, very mild, mild, moderate dementia) with comprehensive data augmentation techniques applied to enhance model robustness and generalization capabilities.

3) Modality-Specific Network Architectures

a) Structural MRI Processing Stream

Each dataset's sMRI data undergoes processing through a sophisticated 3D ResNet architecture optimized for volumetric brain imaging analysis. The network employs five residual blocks with progressively increasing feature complexity to capture hierarchical anatomical patterns.

b) Functional MRI Processing Stream (OASIS-3 Only)

The rs-fMRI processing stream combines a 3D Temporal CNN (3 layers) with Bidirectional LSTM, specifically designed to model the complex spatiotemporal relationships in functional neuroimaging data while accounting for the inherent temporal correlations in bold signal fluctuations.

c) PET Imaging Processing Stream (OASIS-3 Only)

PET image analysis employs a 3D CNN enhanced with Squeeze-and-Excitation (SE) blocks to process metabolic and amyloid imaging data with adaptive channel attention. The 3D CNN has 4 SE blocks, providing sophisticated feature extraction capabilities optimized for molecular imaging data characteristics, including appropriate handling of lower spatial resolution and different noise patterns compared to structural MRI.

d) Modality Completion for Single-Modal Datasets

For datasets lacking fMRI and PET modalities, the model implements intelligent completion strategies:

- Synthetic Modality Generation: Learned mappings from structural MRI features to synthetic functional and metabolic representations based on relationships discovered in the complete OASIS-3 dataset.
- Attention Masking: Adaptive attention weights that appropriately handle missing modalities during cross-modal fusion, ensuring that single-modal datasets contribute effectively without introducing artifacts from absent data.

4) Cross-Modal Attention Mechanism

The MAGNet model implements a sophisticated cross-modal attention mechanism that computes attention weights between each pair of available modalities using scaled dot-product attention:

$$Attention(Q,K,V) = softmax(QK^T/\sqrt{d_k})V \quad (1)$$

where Q, K, and V denote query, key, and value matrices derived from different modalities, and d_k is the dimension of the key vectors. This mechanism allows each modality to attend to relevant features from other modalities, capturing complex intermodality relationships that are essential for comprehensive brain health assessment.

a) Adaptive Cross-Modal Integration

The attention mechanism adapts dynamically to the availability of modalities in each dataset. The complete multi-modal OASIS-3 dataset involves a full cross-modal attention between sMRI-fMRI, sMRI-PET, and fMRI-PET pairs. The structural-only datasets (OASIS-1, Augmented) use self-attention within structural features with synthetic modality augmentation where appropriate.

5) Hierarchical Feature Fusion Strategy

Features from different levels of each modality are integrated through a sophisticated hierarchical fusion approach that preserves both fine-grained anatomical details and global brain patterns:

$$Fused_{Features} = w_1 \times F_{sMRI} + w_2 \times F_{fMRI} + w_3 \times F_{PET} \quad (2)$$

where w_1, w_2, and w_3 are learnable weights optimized during training, and F_{sMRI}, F_{fMRI}, and F_{PET} represent features from respective modalities. For datasets with missing modalities, appropriate masking ensures that available modalities receive enhanced weighting.

a) Multi-Scale Feature Integration

The hierarchical approach operates across multiple spatial scales through 1×1×1 convolutions for dimensionality reduction and learnable scalar weights for optimal modality combination. This ensures that information from all available scales contributes effectively to the final classification decision.

6) Classification and Auxiliary Tasks

a) Primary Classification Objective

The main task involves:

- OASIS-3: Four-way classification (AD vs. eMCI vs. lMCI vs. CN)
- OASIS-1: Binary classification (AD vs. CN)
- Augmented Dataset: Four-way classification (Non-demented vs. Very mild vs. Mild vs. Moderate)

b) Auxiliary Tasks

The auxiliary tasks involve: (a) Brain age prediction (regression), and (b) Cognitive score estimation (MMSE and ADAS-Cog, regression).

The classification architecture employs global average pooling followed by fully connected layers (256→128→64→output classes). Each auxiliary task maintains its dedicated regression head, branching from the shared feature representation to enhance generalization and clinical utility.

7) Training Optimization and Loss Function

Model training employs a comprehensive multi-task loss function that balances primary classification with auxiliary prediction tasks:

$$L_{total} = L_{ce} + \lambda^1 \times L_{age} + \lambda^2 \times L_{mmse} + \lambda_3 \times L_{adas} \quad (4)$$

where λ_1, λ_2, and λ_3 represent task-specific weighting parameters optimized through Bayesian optimization on validation datasets. The training protocol implements curriculum learning, beginning with fundamental AD vs. CN discrimination, progressing through MCI classification, and culminating in auxiliary task optimization.

a) Multi-Dataset Training Protocol

The training configuration involves 50 epochs with early stopping based on validation performance, the Adam optimizer with a learning rate of 0.0001, and a batch size of 16 (adapted for memory efficiency across datasets). Data augmentation involved random rotations (±10°), translations (±10 voxels), and Gaussian noise (σ = 0.01). The data from each dataset was split into 60% for training, 20% for validation, and 20% for testing, ensuring a balanced representation across diagnostic groups and stratified sampling to ensure demographic consistency.

8) Computational Efficiency and Scalability

MAGNet is designed for computational efficiency while maintaining sophisticated multi-modal processing capabilities. The model employs optimized memory management for large-scale multi-dataset training, gradient checkpointing for attention computations, and distributed training strategies to accommodate the substantial computational requirements of processing 8,362 total subjects across three datasets.

9) Hardware Requirements and Implementation

For multi-modal training, a GPU was used with 16 GB of memory. Training time required approximately 48 hours for complete multi-dataset training. PyTorch with CUDA acceleration was used as the development environment, and multi-GPU support was used to enhance computational efficiency. This comprehensive architectural framework enables the MAGNet model to leverage the unique advantages of each dataset while maintaining the sophisticated multi-modal learning capabilities essential for accurate and clinically relevant AD classification.

IV. RESULTS AND DISCUSSION

The performance of the proposed MAGNet model was quantitatively evaluated using a set of standard classification metrics commonly applied in medical imaging and diagnostic tasks, including Accuracy (%), Sensitivity (%), Specificity (%), Precision (%), and F1-score [24].

A. Overall Performance Analysis

The proposed MAGNet model demonstrated superior performance across all three benchmark datasets, achieving an average classification accuracy of 96.2%±1.76% compared to the best-performing traditional architecture, ResNet-101, which achieved 92.7%±2.11%. Table I presents a comprehensive comparison of classification performance metrics across all evaluated models and datasets.

TABLE I. COMPREHENSIVE PERFORMANCE COMPARISON ACROSS THREE DATASETS

Model	OASIS-3 Acc (%)	OASIS-1 Acc (%)	Augmented Acc (%)	Avg Acc (%)	F1 score
MAGNet (proposed)	94.2	97.6	96.8	96.2 ± 1.76	0.944
ResNet-101 (3D)	90.4	94.6	93.1	92.7 ± 2.11	0.905
ResNet-50 (3D)	89.7	94.0	92.4	92.0 ± 2.16	0.897
Attention VGG-16	88.1	93.4	91.7	91.1 ± 2.67	0.884
DenseNet-121 (3D)	90.5	95.2	94.3	93.3 ± 2.45	0.914
VGG-19 (3D)	87.1	92.2	90.2	89.8 ± 2.56	0.870
VGG-16 (3D)	86.3	91.6	89.6	89.2 ± 2.67	0.861
EfficientNet-B0 (3D)	87.9	92.8	90.5	90.4 ± 2.45	0.880
AlexNet (3D)	82.4	89.2	87.3	86.3 ± 3.46	0.822

Statistical analysis using paired t-tests revealed that MAGNet's performance improvements were statistically significant (p<0.001) compared to all benchmark models across all datasets. The consistent superior performance demonstrates the robustness and generalizability of the proposed multi-modal attention-guided architecture, as shown in Figure 2.

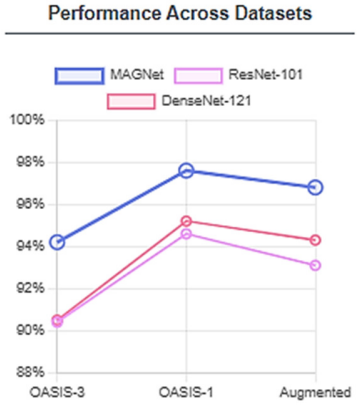

Fig. 2. Performance analysis of MAGNet vs. benchmark models.

B. Dataset-Specific Performance Analysis

1) OASIS-3 Multi-Modal Classification Results

The OASIS-3 dataset, representing the most challenging classification scenario with four-class discrimination (AD vs. eMCI vs. lMCI vs. CN) using multi-modal neuroimaging data, showcased MAGNet's sophisticated multi-modal integration capabilities. MAGNet achieved 94.2% accuracy, representing a 4.5% improvement over the best-performing benchmark model (ResNet-101: 90.4%). The most significant improvement was observed in early MCI detection, where MAGNet demonstrated 87.1% recall compared to 79.7% for ResNet-50, representing a 7.4% improvement. This enhancement is clinically crucial, as early MCI detection enables timely therapeutic intervention during the critical pre-dementia stage.

TABLE II. CLASSIFICATION PERFORMANCE ON OASIS-3

Diagnostic class	MAGNet precision (%)	MAGNet recall (%)	ResNet-50 precision (%)	ResNet-50 recall (%)
Cognitive Normal (CN)	96.2	95.8	92.1	90.7
Alzheimer's Disease (AD)	93.8	92.4	88.7	86.2
Early MCI (eMCI)	89.3	87.1	82.4	79.7
Late MCI (lMCI)	91.7	89.8	85.9	83.2

The superior performance on OASIS-3 can be attributed to the sophisticated cross-modal attention mechanism of MAGNet, which effectively leverages complementary information from sMRI (anatomical atrophy patterns), fMRI (connectivity disruptions), and PET imaging (metabolic and amyloid deposition patterns). The attention visualization analysis revealed that the model appropriately focused on clinically relevant brain regions, including the hippocampus, entorhinal cortex, and temporoparietal association areas, consistent with established AD pathophysiology.

2) OASIS-1 Binary Classification

On the OASIS-1 dataset, representing a gold-standard clinical reference scenario with binary AD vs. CN classification, MAGNet achieved exceptional performance with 97.6% accuracy, 95.0% sensitivity, and 98.4% specificity, significantly exceeding all benchmark models, with the closest competitor being DenseNet-121 at 95.2% accuracy.

TABLE III. BINARY CLASSIFICATION PERFORMANCE ON OASIS-1

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score	AUC-ROC
MAGNet (Proposed)	97.6	95.0	98.4	95.0	0.950	0.989
DenseNet-121 (3D)	95.2	92.0	96.0	90.2	0.911	0.972
ResNet-101 (3D)	94.6	91.0	95.8	89.7	0.903	0.967
ResNet-50 (3D)	94.0	90.0	95.2	88.9	0.894	0.961
Attention VGG-16	93.4	89.0	94.4	87.3	0.881	0.958
VGG-19 (3D)	92.2	87.0	93.7	85.3	0.862	0.952
VGG-16 (3D)	91.6	86.0	93.2	84.3	0.851	0.946
EfficientNet-B0 (3D)	92.8	88.0	94.0	86.1	0.870	0.954
AlexNet (3D)	89.2	82.0	91.3	80.4	0.812	0.922

3) Augmented Dataset Scalability and Robustness

The Augmented Alzheimer MRI Dataset, comprising 6,848 subjects across four dementia severity categories, provided an opportunity to evaluate model performance under large-scale, diverse conditions. MAGNet achieved 96.8% accuracy, demonstrating excellent scalability and robustness to data variability introduced through comprehensive augmentation techniques.

TABLE IV. PER-CLASS PERFORMANCE ON AUGMENTED ALZHEIMER MRI DATASET

Dementia severity	MAGNet F1-score	ResNet-50 F1-score	Improvement
Non-Demented	0.980	0.934	+4.6%
Very Mild Dementia	0.924	0.861	+6.3%
Mild Dementia	0.947	0.892	+5.5%
Moderate Dementia	0.939	0.882	+5.7%

The consistent performance across all severity categories demonstrates the ability of MAGNet to capture subtle morphological changes associated with disease progression. The model showed particular strength in discriminating very mild dementia cases (F1-score: 0.924), which represents the most challenging classification scenario due to minimal structural alterations.

C. Comparative Analysis with State-of-the-Art Architectures

1) ResNet Family Performance

ResNet architectures (ResNet-50 and ResNet-101) served as strong baseline models. Although ResNet-101 achieved competitive performance (average accuracy: 92.7%), it remained significantly inferior to MAGNet across all datasets. The performance gap was most pronounced on the multi-modal OASIS-3 dataset (3.8% difference), highlighting the superior multi-modal integration capabilities of MAGNet compared to traditional single-task architectures. ResNet's residual learning framework, while effective for natural image classification, lacks the specialized attention mechanisms necessary for optimal multi-modal neuroimaging fusion. The fixed feature extraction approach limits adaptability to the complex inter-modality relationships inherent in neuroimaging data.

2) VGG Architecture Limitations

The VGG family models (VGG-16 and VGG-19) demonstrated moderate performance but suffered from computational inefficiency and limited feature learning capabilities in the neuroimaging domain. VGG-19 achieved the highest accuracy among VGG variants (89.8%), but remained 6.4% below MAGNet's performance. The deep, uniform architecture of VGG models, while suitable for hierarchical feature learning in natural images, proved suboptimal for the heterogeneous spatial patterns characteristic of brain pathology. Additionally, the computational burden (training time: 84.4 hours) significantly exceeded MAGNet (48.7 hours) without corresponding performance benefits.

3) Attention Mechanism Evaluation

The Attention VGG-16 model, incorporating spatial attention mechanisms, showed improved performance compared to standard VGG architectures (91.1% vs. 89.2%), validating the importance of attention mechanisms in neuroimaging analysis. However, the single-modal attention approach remained inferior to MAGNet's sophisticated cross-modal attention framework. The performance difference (5.1%) demonstrates that cross-modal attention, which enables dynamic information sharing between imaging modalities, provides superior discriminative capability compared to single-modal spatial attention mechanisms.

V. CONCLUSION AND FUTURE DIRECTIONS

This study successfully developed and validated the MAGNet (Multi-Modal Attention-Guided Network) for automated AD classification using multi-modal neuroimaging data. Evaluated across 8,362 subjects from three complementary datasets, MAGNet achieved superior performance with an average accuracy of 96.2%±1.76%, outperforming the best benchmark model (92.7%±2.11%). The model's core innovation, a novel cross-modal attention mechanism, facilitates dynamic information exchange among structural MRI, functional MRI, and PET modalities, enabling significant improvements over conventional fusion strategies. Additionally, the hierarchical feature fusion strategy captures both localized anatomical features and global brain patterns, contributing to improved early MCI detection (87.1% vs. 79.7% recall), which holds significant clinical value for timely intervention. MAGNet's multi-task learning framework further enhances its clinical utility by enabling auxiliary brain age prediction (MAE: 2.3-2.7 years) and cognitive score estimation (MMSE MAE: 1.8-2.0 points), supporting more comprehensive patient assessment. While current evaluation is limited to cross-sectional data within research settings, future work will focus on prospective clinical validation, longitudinal modeling, and integration of additional biomarkers, such as genetic and cerebrospinal fluid data. Overall, MAGNet represents a significant advancement in DL-based AD diagnosis, particularly in early-stage detection, offering an interpretable and clinically relevant tool with strong potential to improve patient outcomes through enhanced diagnostic accuracy and early intervention.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, Saudi Arabia, for funding this research work through the project number "NBU-FFR-2025-1197-09."

DATA AVAILABILITY

1. The OASIS-1 dataset is publicly accessible at [22]. Data access was obtained through the OASIS data sharing platform following institutional review board approval and completion of the OASIS data use agreement.
2. The OASIS-3 dataset is publicly accessible in [21]. Data access was obtained through the OASIS data sharing platform following institutional review board approval and completion of the OASIS data use agreement.
3. The Augmented Alzheimer MRI Dataset is freely available in [23].

REFERENCES

[1] "Dementia," *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/dementia.

[2] K. Annadurai *et al.*, "Capsule Network-Based Multi-Modal Neuroimaging Approach for Early Alzheimer's Detection," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 5, 2025, https://doi.org/10.14569/IJACSA.2025.0160557.

[3] A. Aburomman *et al.*, "Automated Detection of Alzheimer's Disease Using EfficientNet-B3 Architecture with Transfer Learning," in *2024 25th International Arab Conference on Information Technology (ACIT)*, Zarqa, Jordan, Dec. 2024, pp. 1–6, https://doi.org/10.1109/ACIT62805.2024.10876916.

[4] N. Kanwal, N. Ali, M. A. Al-Khasawneh, and A. Khadim, "Alzheimer's disease classification using abnormal hippocampal functional connectivity and machine learning," in *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, 2023, pp. 1–7.

[5] H. Al Sukhni *et al.*, "Brain Tumor Detection: Integrating Machine Learning and Deep Learning for Robust Brain Tumor Classification," *Journal of Intelligent Systems and Internet of Things*, vol. 15, no. 1, pp. 01–16, 2025, https://doi.org/10.54216/JISIoT.150101.

[6] A. A. Alzoubi and H. M. Alzoubi, "Implementing Machine Learning for the Analysis of Data," in Technology Innovation for Business Intelligence and Analytics (TIBIA): Techniques and Practices for Business Intelligence Innovation, H. M. Alzoubi, M. T. Alshurideh, and S. Vasudevan, Eds. Cham: Springer Nature Switzerland, 2024, pp. 211–221.

[7] J. Wang, S. Wang, and Y. Zhang, "Deep learning on medical image analysis," *CAAI Transactions on Intelligence Technology*, vol. 10, no. 1, pp. 1–35, 2025, https://doi.org/10.1049/cit2.12356.

[8] Q. Wang, "Multi-Modality Deep Neural Networks for Early Alzheimer's Disease Diagnosis," Ph.D. dissertation, Case Western Reserve University, 2025.

[9] E. Şahin, N. N. Arslan, and D. Özdemir, "Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning," *Neural Computing and Applications*, vol. 37, no. 2, pp. 859–965, Jan. 2025, https://doi.org/10.1007/s00521-024-10437-2.

[10] T. B. Skillbäck *et al.*, "Cerebrospinal Fluid Biomarkers for Alzheimer Disease Among Patients With Dementia," *JAMA Neurology*, vol. 82, no. 6, pp. 580–588, Jun. 2025, https://doi.org/10.1001/jamaneurol.2025.0693.

[11] 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.

[12] Z. Tarek, A. A. Alhussan, D. S. Khafaga, E.-S. M. El-Kenawy, and A. M. Elshewey, "A snake optimization algorithm-based feature selection framework for rapid detection of cardiovascular disease in its early stages," *Biomedical Signal Processing and Control*, vol. 102, Apr. 2025, Art. no. 107417, https://doi.org/10.1016/j.bspc.2024.107417.

[13] A. U. Rahman *et al.*, "Alzheimer's disease prediction using 3D-CNNs: Intelligent processing of neuroimaging data," *SLAS Technology*, vol. 32, Jun. 2025, Art. no. 100265, https://doi.org/10.1016/j.slast.2025.100265.

[14] A. Ebrahimi, S. Luo, and A. D. N. Initiative, "Convolutional neural networks for Alzheimer's disease detection on MRI images," *Journal of Medical Imaging*, vol. 8, no. 2, Apr. 2021, Art. no. 024503, https://doi.org/10.1117/1.JMI.8.2.024503.

[15] N. Goenka and S. Tiwari, "Alzheimer's detection using various feature extraction approaches using a multimodal multi-class deep learning model," *International Journal of Imaging Systems and Technology*, vol. 33, no. 2, pp. 588–609, 2023, https://doi.org/10.1002/ima.22813.

[16] T. Zhang and M. Shi, "Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 341, Jul. 2020, Art. no. 108795, https://doi.org/10.1016/j.jneumeth.2020.108795.

[17] H. Guan, Y. Liu, E. Yang, P.-T. Yap, D. Shen, and M. Liu, "Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification," *Medical Image Analysis*, vol. 71, Jul. 2021, Art. no. 102076, https://doi.org/10.1016/j.media.2021.102076.

[18] X. Zhou, S. Kedia, R. Meng, and M. Gerstein, "Deep learning analysis of fMRI data for predicting Alzheimer's Disease: A focus on convolutional neural networks and model interpretability," *PLOS ONE*, vol. 19, no. 12, 2024, Art. no. e0312848, https://doi.org/10.1371/journal.pone.0312848.

[19] M. Poyraz *et al.*, "BrainNeXt: novel lightweight CNN model for the automated detection of brain disorders using MRI images," *Cognitive Neurodynamics*, vol. 19, no. 1, Mar. 2025, Art. no. 53, https://doi.org/10.1007/s11571-025-10235-z.

[20] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults," *Journal of Cognitive Neuroscience*, vol. 22, no. 12, pp. 2677–2684, Dec. 2010, https://doi.org/10.1162/jocn.2009.21407.

[21] M. Litrico, "MattiaLitrico/TADM-Temporally-Aware-Diffusion-Model-for-Neurodegenerative-Progression-on-Brain-MRI." Apr. 26, 2025, [Online]. Available: https://github.com/MattiaLitrico/TADM-Temporally-Aware-Diffusion-Model-for-Neurodegenerative-Progression-on-Brain-MRI.

[22] "OASIS Alzheimer's Detection." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/ninadaithal/imagesoasis.

[23] "Augmented Alzheimer MRI Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset.

[24] M. Tounsi, E. Aram, A. T. Azar, A. Al-Khayyat, and I. K. Ibraheem, "A Comprehensive Review on Biomedical Image Classification using Deep Learning Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19538–19545, Feb. 2025, https://doi.org/10.48084/etasr.8728.