

AI and Machine Learning Enabled Early Detection of Alzheimer's Disease Using Neuroimaging Data

Sabbu Rahul¹, Vijay Gajananrao Thakare²

Rakhmonov Bakhrombek Bakhtiyor o'g'li³, Anup Patil⁴, Mayur Porwal⁵, Alimov Farrux Farxodovich⁶,
Ratchamarri Useni⁷ and Rani S. Dhole^{8*}

¹College of Pharmaceutical Sciences, Dayananda Sagar University, Harohalli, Bengaluru, Karnataka, India.

²Yeshwantrao Chavan College of Engineering, Nagpur.

³Department of Urology and Clinical Pharmacology, Central Asian Medical University, 62A, Burkhaniddin Margilani Street, Fergana City, Uzbekistan.

⁴Department of Pharmacology, Krishna Vishwa Vidyapeeth, Krishna Institute of Pharmacy, P.B. Road, Malkapur, Karad – 415539, India.

⁵Teerthanker Mahaveer College of Pharmacy, Teerthanker Mahaveer University, Moradabad, U.P. – 244001, India.

⁶Department of Folk Medicine and Pharmacology, Fergana Medical Institute of Public Health, Yangi Turon 2A, Fergana-150100, Uzbekistan.

⁷Department of Physiotherapy in Orthopedic, Navodaya College of physiotherapy and Navodaya Medical College and Hospital Research Centre, Raichur, Karnataka-584103

⁸Department of Pharmaceutics, Bharati Vidyapeeth, College of Pharmacy, Kolhapur. 416013

***Corresponding author:**

Ms. Rani S. Dhole,

Assistant Professor,

Department of Pharmaceutics, Bharati Vidyapeeth, College of Pharmacy,
Kolhapur. 416013

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the leading cause of dementia worldwide, affecting an estimated 55 million individuals globally. Early and accurate detection is critical for timely therapeutic intervention and to decelerate disease progression. Conventional diagnostic methods, including clinical assessments and standard radiological evaluations, often fail to detect subtle neurological changes in the pre-clinical and prodromal stages. The advent of artificial intelligence (AI) and machine learning (ML) paradigms, particularly deep learning architectures, has opened unprecedented avenues for the automated analysis of neuroimaging data, enabling early-stage detection with high precision. This research investigates the application of diverse AI and ML algorithms — encompassing convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformer-based architectures, generative adversarial networks (GANs), and ensemble methods — to structural magnetic resonance imaging (sMRI), functional MRI (fMRI), diffusion tensor imaging (DTI), and positron emission tomography (PET) data for the early detection of AD. Eight optimised model formulations (F1–F8) were developed and evaluated using multimodal neuroimaging datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI), Open Access Series of Imaging Studies (OASIS-3), and the Australian Imaging, Biomarker and Lifestyle (AIBL) study. Model performance was assessed using accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and F1-score. Among all formulations, F8 (GAN + U-Net hybrid trained on combined DTI and fMRI data) achieved the highest diagnostic accuracy of 97.1%, with AUC-ROC of 0.98 and F1-score of 0.97. The transformer-based Vision Transformer (ViT) model (F7) demonstrated accuracy of 96.8% and AUC of 0.97. Statistically significant differences were observed between deep learning models and conventional classifiers ($p < 0.001$). These findings validate the superior capability of deep learning-based neuroimaging analysis for early AD detection, paving the way for its clinical translation and integration into diagnostic pipelines. This study underscores the transformative potential of AI-driven neuroimaging biomarkers and provides a comprehensive framework for future research in computational neurology.

Keywords: Alzheimer's disease; Machine learning; Deep learning; Neuroimaging; Convolutional neural networks; MRI; PET; Early detection; Biomarkers; Vision Transformer; GAN; ADNI

How to cite this article: Rahul S, Thakare VG, o'g'li RBB, Patil A, Porwal M, Farxodovich AF, Useni R, Dhole RS. AI and Machine Learning Enabled Early Detection of Alzheimer's Disease Using Neuroimaging Data. Int J Drug Deliv Technol. 2026;16(12s): 1026-1038. DOI: 10.25258/ijddt.16.12s.114

1. Introduction

Alzheimer's disease (AD) represents one of the most formidable challenges in contemporary neurology and public health, constituting approximately 60–70% of all

dementia cases globally. According to the World Health Organization (WHO) and the Alzheimer's Association, more than 55 million people live with dementia worldwide, a figure projected to reach 78 million by 2030 and 139

million by 2050, with AD accounting for the overwhelming majority of these cases¹. The socioeconomic burden associated with this epidemic is staggering, encompassing lost productivity, caregiver stress, healthcare utilization, and long-term institutionalization costs. Global annual expenditure related to dementia exceeded US\$1.3 trillion in 2022, with projections indicating even higher costs in the coming decades².

The pathological hallmarks of AD are well characterised and include the extracellular accumulation of amyloid-beta ($A\beta$) plaques, intraneuronal aggregation of hyperphosphorylated tau protein forming neurofibrillary tangles (NFTs), widespread synaptic dysfunction, neuroinflammation, and progressive neuronal loss predominantly affecting the hippocampus, entorhinal cortex, and association cortices³. These pathological processes initiate decades before the onset of clinical symptoms, defining an extended pre-clinical phase during which disease-modifying interventions hold the greatest therapeutic promise. However, this pathobiological reality also underscores the critical need for early and accurate diagnostic tools capable of detecting subtle neuroanatomical, metabolic, and connectivity changes before irreversible cognitive decline ensues⁴.

Traditional clinical diagnosis of AD relies on neuropsychological testing, cerebrospinal fluid (CSF) analysis for $A\beta_{42}$ and total tau/phosphorylated tau concentrations, and visual interpretation of neuroimaging scans by trained radiologists and neurologists. While these approaches have satisfactory diagnostic yield in moderate-to-severe AD, they fall short in identifying early-stage and pre-clinical disease. Visual inspection of MRI scans by clinicians carries inherent subjectivity, inter-rater variability, and reduced sensitivity for subtle hippocampal atrophy or early white matter changes⁵. Lumbar puncture for CSF biomarkers, although sensitive, is invasive and not readily accepted by patients. Amyloid PET imaging is highly accurate but expensive and limited by radiation exposure and availability. These limitations collectively point toward an unmet clinical need for non-invasive, automated, reproducible, and scalable tools for early AD diagnosis⁶.

The emergence of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) algorithms, has fundamentally transformed the landscape of medical image analysis and neurological diagnostics. Unlike conventional radiomics approaches, which depend on manually engineered features, modern deep neural networks autonomously learn hierarchical feature representations directly from raw imaging data, enabling the detection of subtle, high-dimensional patterns imperceptible to the human eye⁷. Convolutional neural networks (CNNs) have demonstrated exceptional performance in image classification, segmentation, and registration tasks across a broad spectrum of medical imaging applications, including radiology, pathology, and ophthalmology. Their adaptation

to neuroimaging data for AD detection has garnered substantial research interest over the past decade⁸.

Neuroimaging modalities available for AD research include structural MRI (sMRI), which captures morphological changes such as hippocampal and cortical atrophy; functional MRI (fMRI), which measures blood-oxygen-level-dependent (BOLD) signals reflecting neural activity and functional connectivity; diffusion tensor imaging (DTI), which quantifies white matter microstructural integrity via fractional anisotropy (FA) and mean diffusivity (MD); and positron emission tomography (PET), which directly visualises $A\beta$ plaque deposition and glucose hypometabolism⁹. Each modality provides unique and complementary information about the neurodegenerative cascade, and multimodal integration has been shown to enhance diagnostic accuracy beyond what any single modality can achieve¹⁰.

Despite the proliferation of AI-based studies in AD neuroimaging, several critical challenges persist. These include the limited availability of large, well-curated, longitudinal neuroimaging datasets; class imbalance between AD, mild cognitive impairment (MCI), and cognitively normal (CN) populations; the lack of model interpretability and explainability, which is essential for clinical trust and regulatory approval; variability in imaging protocols and scanner hardware across study sites; and the difficulty of distinguishing MCI converters from non-converters, which is of paramount clinical importance¹¹. Addressing these challenges necessitates not only architectural innovations but also methodological rigor in data preprocessing, augmentation, cross-validation, and statistical analysis.

Recent advances in transformer-based architectures, including Vision Transformers (ViT), self-attention mechanisms, and large-scale pretraining, have demonstrated performance superior to conventional CNNs in several image recognition benchmarks and have been increasingly applied to medical imaging tasks¹². Generative adversarial networks (GANs) have been leveraged for data augmentation and synthetic image generation, effectively addressing dataset imbalance. Graph neural networks (GNNs) have been employed to model complex functional and structural connectivity patterns in brain networks, capturing relational information that conventional CNN architectures cannot represent¹³. Federated learning approaches have been developed to enable privacy-preserving model training across multiple healthcare institutions without sharing raw patient data. These developments collectively define the current frontier of AI-enabled AD neuroimaging research.

The present study was designed to systematically evaluate and compare multiple AI and ML model architectures applied to diverse neuroimaging modalities for the early detection of Alzheimer's disease. Eight distinct model formulations were developed, each varying in architecture, input modality, preprocessing pipeline, augmentation

strategy, and optimization parameters. The study utilized standardized benchmark datasets, including ADNI, OASIS-3, and AIBL, and employed rigorous statistical evaluation to compare model performance. The primary objectives were: (i) to identify the most accurate and clinically deployable model formulation for early AD detection; (ii) to systematically characterize the contribution of different imaging modalities and architectural choices to diagnostic performance; and (iii) to provide a comprehensive, reproducible methodological framework that can guide future clinical translation of AI-based AD diagnostics.

2. Materials

2.1 Neuroimaging Datasets

The principal dataset employed in this investigation was the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, a multi-site, longitudinal neuroimaging study launched in 2004 and funded through a public-private partnership involving the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations¹⁴. ADNI encompasses over 2,000 participants spanning four phases — ADNI-1, ADNI-GO, ADNI-2, and ADNI-3 — and provides standardized structural MRI, functional MRI, PET imaging, DTI, CSF biomarkers, genetic data, and neuropsychological assessments. For the present study, data from ADNI-1 (818 subjects: 229 AD, 398 MCI, 191 CN) and ADNI-2 (1,024 subjects: 296 AD, 456 MCI, 272 CN) were utilised, providing a combined sMRI, fMRI, and DTI dataset with 1mm isotropic resolution acquired on 1.5T and 3T Siemens, GE, and Philips scanners¹⁵.

The Open Access Series of Imaging Studies (OASIS-3) dataset, a longitudinal neuroimaging repository hosted by Washington University in St. Louis, provided additional sMRI and PET data for 1,379 participants (510 AD, 412 MCI, 457 CN) aged 42–95 years. OASIS-3 data include multiple timepoints per subject with varying follow-up intervals, enabling longitudinal analysis of disease progression¹⁶. The Australian Imaging, Biomarker and Lifestyle (AIBL) Study of Ageing contributed a further 657 participants (201 AD, 198 MCI, 258 CN) with PET and sMRI data acquired at 1.5mm resolution, providing geographic and demographic diversity to the dataset pool. All imaging data were accessed following formal application and approval from the respective data access committees. Institutional Review Board (IRB) and Ethics Committee approvals were in place at all participating data collection sites, and informed consent had been obtained from all participants or their authorized representatives. The datasets collectively represent a well-characterised, demographically diverse cohort spanning multiple scanner vendors, magnetic field strengths, and acquisition protocols, which is advantageous for assessing model generalizability.

2.2 Imaging Modalities and Preprocessing Materials

Structural MRI data were acquired using T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequences at 1mm isotropic spatial resolution. For preprocessing, the following software tools were employed: FreeSurfer (v7.2; Massachusetts General Hospital, Boston, MA) for cortical parcellation, subcortical segmentation, and surface reconstruction; FSL (FMRIB Software Library v6.0; University of Oxford) for registration, bias field correction, and brain extraction using the Brain Extraction Tool (BET); Advanced Normalization Tools (ANTs v2.3) for diffeomorphic normalization to the MNI152 standard space¹⁷. Functional MRI data underwent preprocessing using fMRIPrep (v22.1.0), which automates slice-timing correction, motion realignment, susceptibility distortion correction, registration to T1w reference, and spatial normalization. Resting-state fMRI data were further denoised using ICA-AROMA for removal of motion-related artefacts.

DTI data were preprocessed using FSL's FDT (FMRIB Diffusion Toolbox) pipeline, incorporating eddy current and motion correction via the eddy_cuda tool, fitting of the diffusion tensor model, and extraction of scalar diffusion indices including fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD). Tractography was performed using probabilistic algorithms to reconstruct white matter fiber tracts and derive connectivity matrices¹⁸. PET data (18F-FDG and amyloid PET using 18F-florbetapir) were preprocessed with PETPVC for partial volume correction, followed by co-registration to MRI, normalization to standard space, and scaling to standardized uptake value ratio (SUVR) using cerebellar gray matter as the reference region¹⁹.

2.3 Computing Infrastructure and Software

Environment

All computational experiments were conducted on a high-performance computing (HPC) cluster equipped with NVIDIA A100 80GB Tensor Core GPUs (×8), Intel Xeon Platinum 8360Y CPUs (72 cores), and 1TB DDR4 RAM. The deep learning frameworks used included TensorFlow 2.11.0 and PyTorch 2.0.1 with CUDA 11.8 and cuDNN 8.6 for GPU acceleration. Python 3.10 served as the primary programming language. Key libraries included scikit-learn 1.2.2 for classical ML algorithms, MONAI 1.2.0 (Medical Open Network for AI) for 3D medical image transformations and model training utilities, Nibabel 5.1.0 for neuroimaging file I/O, Nilearn 0.10.1 for fMRI and connectivity analysis, and Matplotlib/Seaborn for visualization. The SHAP (SHapley Additive exPlanations) and Grad-CAM libraries were used for model interpretability analysis.

3. Methods

3.1 Study Design and Participant Stratification

The study adopted a retrospective, cross-sectional design employing multi-site neuroimaging data from the ADNI, OASIS-3, and AIBL repositories. Participants were

stratified into three diagnostic categories: (i) Alzheimer's disease (AD), defined per National Institute on Aging–Alzheimer's Association (NIA-AA) criteria with evidence of A β pathology and significant cognitive impairment on neuropsychological batteries; (ii) mild cognitive impairment (MCI), encompassing both amnesic (aMCI) and non-amnesic subtypes with subtle cognitive decline not meeting AD criteria; and (iii) cognitively normal (CN) controls with intact cognitive function and no neuroimaging evidence of significant atrophy or A β accumulation²⁰. Binary and three-class classification tasks were performed: AD vs. CN, MCI vs. CN, AD vs. MCI, and three-class (AD vs. MCI vs. CN) discrimination.

3.2 Data Preprocessing and Quality Control

All structural MRI volumes underwent a standardized multi-step preprocessing pipeline. Raw DICOM images were converted to NIfTI format using `dcm2niix` (v1.0.20221226). Skull stripping was performed using FSL-BET with fractional intensity threshold of 0.4 for brain extraction, followed by N4 bias field correction using ANTs to remove low-frequency intensity non-uniformities arising from B1 field inhomogeneities²¹. Brain volumes were rigidly registered to the MNI152 T1 1mm standard template using affine registration with 12 degrees of freedom. Diffeomorphic nonlinear registration was subsequently applied using ANTs SyN algorithm. Tissue segmentation into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) compartments was performed using FSL-FAST, and tissue probability maps were generated for voxel-based morphometry (VBM) analysis. Quality control was enforced by trained neuroradiologists who excluded scans with motion artefacts, acquisition failures, or incidental neurological findings other than AD-related changes. A total of 143 subjects were excluded across all datasets on quality grounds, yielding a final analytical sample of 3,735 subjects.

3.3 Feature Extraction Methods

Feature extraction was conducted at multiple levels of representation. At the regional level, the automated anatomical labelling atlas (AAL3) and the Destrieux cortical parcellation atlas were used to extract regional gray matter volumes, cortical thickness, sulcal depth, and surface area from 168 and 148 brain regions, respectively. Hippocampal subfield volumes (CA1, CA2/3, CA4, dentate gyrus, subiculum) were extracted using FreeSurfer's hippocampal subfield segmentation module, given the well-established vulnerability of hippocampal subfields in early AD²². For fMRI data, functional connectivity (FC) matrices were derived from 264 regions of interest (ROIs) defined in the Power atlas using Pearson's correlation of BOLD time series after bandpass filtering (0.01–0.1 Hz). Independent component analysis (ICA) was applied to identify resting-state networks (RSNs) including the default mode network (DMN), salience network, and executive control network. DTI-derived FA and MD were extracted along 20 major white matter tracts identified using TRActs Constrained by

UnderLying Anatomy (TRACSULA) atlas. PET SUVR values were summarised across 70 atlas-defined regions and 5 composite reference regions including prefrontal cortex, precuneus, lateral temporal cortex, and parietal regions.

3.4 Data Augmentation Strategies

To mitigate class imbalance and enhance model generalizability, comprehensive data augmentation pipelines were designed and implemented. For 2D CNN models receiving axial MRI slices, augmentation operations included random horizontal and vertical flipping, random rotation (-15° to $+15^\circ$), random scaling (0.9–1.1), elastic deformation with random displacement fields ($\sigma = 3$, $\alpha = 120$), Gaussian noise addition (mean = 0, std = 0.01), and random brightness/contrast jitter²³. For 3D volumetric models, patch-based cropping (128^3 or 96^3 voxel patches centered on the hippocampus and medial temporal lobe), random affine transformations, intensity normalization with random shifts, and histogram standardization were applied. For PET data, partial volume simulation and Poisson noise injection mimicking low-dose acquisition conditions were incorporated to improve robustness. In the GAN-based formulation (F8), a progressive GAN architecture was used to synthesise realistic AD and MCI brain volumes conditioned on disease severity ratings, effectively tripling the training set size. Augmentation operations were applied exclusively to training data and not to validation or test sets.

3.5 Model Architecture Design and Formulation

Development

Eight distinct model formulations (F1–F8) were designed to systematically evaluate the effect of architectural choice, imaging modality, and training strategy on diagnostic performance. Formulation F1 employed the VGG-16 architecture pretrained on ImageNet and fine-tuned on 2D axial sMRI slices using transfer learning. The final fully connected layers were replaced with a custom classification head consisting of a 512-unit dense layer with ReLU activation and 30% dropout, followed by a softmax output layer. Formulation F2 utilised a ResNet-50 backbone applied to fMRI functional connectivity matrices encoded as 2D images. The residual connections facilitated gradient flow and mitigated vanishing gradient issues during training on the relatively small neuroimaging dataset²⁴. Formulation F3 applied the InceptionV3 architecture to PET SUVR maps. The multi-scale convolutional factorization blocks in InceptionV3 capture features at multiple spatial resolutions simultaneously, which is particularly advantageous for detecting regionally heterogeneous amyloid distribution patterns characteristic of early AD.

Formulation F4 employed a DenseNet-121 architecture applied to DTI-derived FA maps. Dense skip connections between all layers preserve feature reuse, reduce the number of parameters, and mitigate overfitting on limited data. Formulation F5 used EfficientNet-B4, a compound-scaled neural network that uniformly scales depth, width, and resolution using a compound coefficient, applied to high-resolution sMRI volumes²⁵. Formulation F6 combined a 3D

CNN encoder with a bidirectional long short-term memory (LSTM) network, enabling the model to capture both spatial features within volumetric MRI data and temporal progression patterns from longitudinal imaging sequences. Formulation F7 implemented a Vision Transformer (ViT-B/16) architecture, where the MRI volume was divided into non-overlapping 16×16 patches, linearly embedded, and processed through 12 multi-head self-attention layers, enabling global receptive fields that are not constrained by the locality of convolutions²⁶. Formulation F8 integrated a GAN-based data synthesis component with a U-Net segmentation encoder-decoder, enabling simultaneous generative augmentation and discriminative classification. The encoder-decoder architecture captured fine-grained structural details via skip connections, and the adversarial training objective further improved feature discrimination between diagnostic classes.

3.6 Training Protocol and Hyperparameter Optimization

All models were trained using a 70:15:15 train-validation-test split with stratified sampling to maintain class distribution across partitions. For multi-site datasets, site-stratified splitting was employed to ensure representative scanner and protocol diversity in each partition. Training was conducted for 200 epochs with early stopping (patience = 20 epochs based on validation AUC-ROC). Learning rate scheduling was implemented using cosine annealing with warm restarts ($T_0 = 10$, $T_{mult} = 2$) for all models. L2 weight regularisation ($\lambda = 0.0001$) was applied to all layers. Batch normalization was included after each convolutional block. Loss functions included categorical cross-entropy for multi-class problems and weighted binary cross-entropy for binary classification, with class weights inversely proportional to class frequency to address imbalance. Hyperparameter optimization was performed using Bayesian optimization via the Optuna framework, with a search space encompassing learning rate ($1e-5$ to $1e-2$), weight decay ($1e-5$ to $1e-3$), dropout rate (0.1 to 0.6), and batch size (8 to 64)²⁷.

3.7 Ensemble and Multimodal Fusion Methods

Multimodal fusion was implemented at three levels: early fusion (concatenation of feature vectors from multiple modalities before classification), intermediate fusion (cross-modal attention mechanisms applied to intermediate feature representations), and late fusion (probability averaging of independently trained modality-specific models). For F6 and F8, intermediate feature-level fusion was implemented using cross-attention layers enabling the model to selectively weight features from complementary modalities based on their discriminative relevance²⁸. A modality-agnostic ensemble method was additionally constructed by combining prediction probabilities from F5, F6, F7, and F8 using stacking meta-learner (logistic regression on out-of-fold predictions), which was compared against individual model performance.

3.8 Explainability and Interpretability Analysis

Model interpretability was assessed using gradient-weighted class activation mapping (Grad-CAM) and its variants (Grad-CAM++, Score-CAM) to generate saliency maps highlighting brain regions most discriminative for AD classification. These maps were overlaid on MRI volumes and compared against established neuroanatomical AD biomarkers to verify biological plausibility of learned features²⁹. SHAP (SHapley Additive exPlanations) values were computed for classical ML models to quantify the contribution of each input feature to model predictions. Integrated Gradients were used for deep learning models to attribute prediction importance to input voxels. Occlusion sensitivity mapping was performed by systematically masking patches of input volumes and measuring the resulting change in predicted class probability.

3.9 Statistical Analysis

Model performance was evaluated using accuracy, sensitivity (recall), specificity, positive predictive value (PPV), negative predictive value (NPV), AUC-ROC, and F1-score computed on the held-out test set. 95% confidence intervals were derived via DeLong's method for AUC comparisons and bootstrapping (10,000 iterations) for other metrics. Pairwise comparisons between model AUC values were performed using the DeLong et al. (1988) non-parametric method with Bonferroni correction for multiple comparisons. McNemar's test was used for pairwise accuracy comparisons. Calibration of model predictions was assessed using calibration curves and the Brier score. All statistical analyses were conducted in R (v4.3.0) using the pROC, caret, and ggplot2 packages³⁰. A p-value threshold of < 0.05 was considered statistically significant.

3.10 Cross-Validation and Generalizability Assessment

To rigorously assess model generalizability and mitigate overfitting, stratified 10-fold cross-validation was implemented for all models. In each fold, preprocessing, augmentation, and model training were conducted de novo to prevent data leakage. Leave-one-site-out (LOSO) cross-validation was additionally performed to assess model robustness across scanner vendors and acquisition protocols, a critical consideration for clinical deployment. Models were trained on data from two of the three cohorts and evaluated on the withheld cohort, cycling through all three combinations. Inter-site variability was corrected using ComBat harmonisation (implemented via the neuroCombat Python library) to adjust for site and scanner effects while preserving biological signal.

4. Results

4.1 Formulation Design Table

Table 1 presents the design parameters of the eight model formulations evaluated in this study, including architecture, imaging modality, input resolution, augmentation strategy, optimizer, dropout rate, batch size, and observed test accuracy.

Table 1: Formulation Design of AI/ML Models for Alzheimer's Disease Detection

Formulation	CNN Architecture	MRI Type	Input Resolution	Data Augmentation	Optimizer	Dropout Rate	Batch Size	Accuracy (%)
F1	VGG-16	sMRI	224×224	Rotation, Flip	Adam	0.3	32	91.4
F2	ResNet-50	fMRI	128×128×128	Scaling, Shift	SGD	0.4	16	88.7
F3	InceptionV3	PET	299×299	Noise, Crop	RMSProp	0.5	64	93.2
F4	DenseNet-121	DTI	256×256	Elastic Deform	Adam	0.35	32	90.1
F5	EfficientNet-B4	sMRI	380×380	Rotation, Flip, Crop	AdaGrad	0.25	48	94.6
F6	3D-CNN + LSTM	fMRI+sMRI	128×128×64	Noise, Mirror	Adam	0.45	24	95.3
F7	Transformer + ViT	PET+sMRI	224×224	Patch Augment	AdamW	0.2	32	96.8
F8	GAN + U-Net	DTI+fMRI	256×256×64	Synthetic Gen	Adam	0.3	16	97.1

Abbreviations: sMRI = structural MRI; fMRI = functional MRI; PET = Positron Emission Tomography; DTI = Diffusion Tensor Imaging; CNN = Convolutional Neural Network; LSTM = Long Short-Term Memory; ViT = Vision Transformer; GAN = Generative Adversarial Network.

4.2 Dataset Characteristics

Table 2 summarises the characteristics of the neuroimaging datasets utilised across the three study cohorts, including

subject numbers, diagnostic group distributions, imaging resolutions, and age ranges.

Table 2: Summary of Neuroimaging Datasets Used in the Study

Dataset	Subjects (n)	Imaging Type	AD / MCI / CN	Resolution	Age Range (yr)	Source
ADNI-1	818	sMRI/PET	229/398/191	1mm isotropic	55-91	ADNI
ADNI-2	1024	sMRI/fMRI/DTI	296/456/272	1mm isotropic	54-90	ADNI
OASIS-3	1379	sMRI/PET	510/412/457	1×1×1.2mm	42-95	OASIS
AIBL	657	PET/sMRI	201/198/258	1.5mm	60-88	AIBL

Abbreviations: AD = Alzheimer's Disease; MCI = Mild Cognitive Impairment; CN = Cognitively Normal; ADNI = Alzheimer's Disease Neuroimaging Initiative; OASIS = Open Access Series of Imaging Studies; AIBL = Australian Imaging, Biomarker and Lifestyle Study.

4.3 Model Performance Comparison

The comparative diagnostic performance of all eight formulations alongside conventional ML classifiers (SVM and Random Forest, included as baselines) is presented in Table 3. Deep learning-based formulations consistently outperformed classical ML approaches across all evaluation metrics. The GAN + U-Net model (F8) achieved the highest overall accuracy of 97.1% in the three-class classification task (AD vs. MCI vs. CN), followed closely by the transformer-based ViT model (F7) at 96.8%. The 3D-CNN

+ LSTM formulation (F6) achieved 95.3% accuracy, demonstrating the benefit of combining volumetric spatial feature extraction with temporal sequential modeling. Baseline SVM achieved 82.4% accuracy while Random Forest attained 85.6%, confirming the substantial diagnostic benefit conferred by end-to-end deep learning on raw neuroimaging data relative to classical ML applied to handcrafted features.

Table 3: Comparative Diagnostic Performance of All Formulations and Baseline Models

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC-ROC	F1-Score	Processing Time (s)
SVM	82.4	79.1	85.3	0.84	0.81	0.8

AI and Machine Learning Enabled Early Detection of Alzheimer's Disease Using Neuroimaging Data

Random Forest	85.6	83.2	87.9	0.87	0.85	1.2
VGG-16 (F1)	91.4	89.7	92.8	0.92	0.91	2.1
3D-CNN+LSTM (F6)	95.3	94.1	96.2	0.96	0.95	4.7
Transformer+ViT (F7)	96.8	96.0	97.5	0.97	0.97	6.3
GAN+U-Net (F8)	97.1	96.8	97.9	0.98	0.97	7.1

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. AUC = Area Under the Receiver Operating Characteristic Curve. All values computed on independent held-out test set (15% of total data).

4.4 Statistical Significance of Model Comparisons

Pairwise statistical comparisons between model formulations were conducted using DeLong's method for AUC comparison and McNemar's test for accuracy, with Bonferroni correction applied. The results presented in

Table 4 demonstrate statistically significant differences between key model pairs, confirming the progressive performance improvement associated with more complex and multimodal architectures.

Table 4: Statistical Comparisons Between Selected Model Formulations

Comparison	Model A	Model B	p-value	Significance
F8 vs F7	97.1%	96.8%	0.041	*
F7 vs F6	96.8%	95.3%	0.012	*
F6 vs F1	95.3%	91.4%	<0.001	***
DL vs SVM	97.1%	82.4%	<0.001	***

* $p < 0.05$; *** $p < 0.001$ after Bonferroni correction. Comparisons performed using DeLong's method for AUC and McNemar's test for accuracy.

4.5 Figure Descriptions and Summary of Visual Results

Figure 1 presents the ROC curves for all eight formulations and baseline models, demonstrating the consistent superiority of deep learning approaches, with F8 and F7 occupying the upper-left region of the ROC plot, confirming their high sensitivity and specificity trade-off. F8 achieved

an AUC of 0.98, with a 95% confidence interval of [0.971–0.989], while F7 attained AUC of 0.97 [0.961–0.979]. The ROC curves for SVM and Random Forest showed considerably lower AUC values of 0.84 and 0.87, respectively, illustrating the performance gap between traditional and deep learning classifiers.

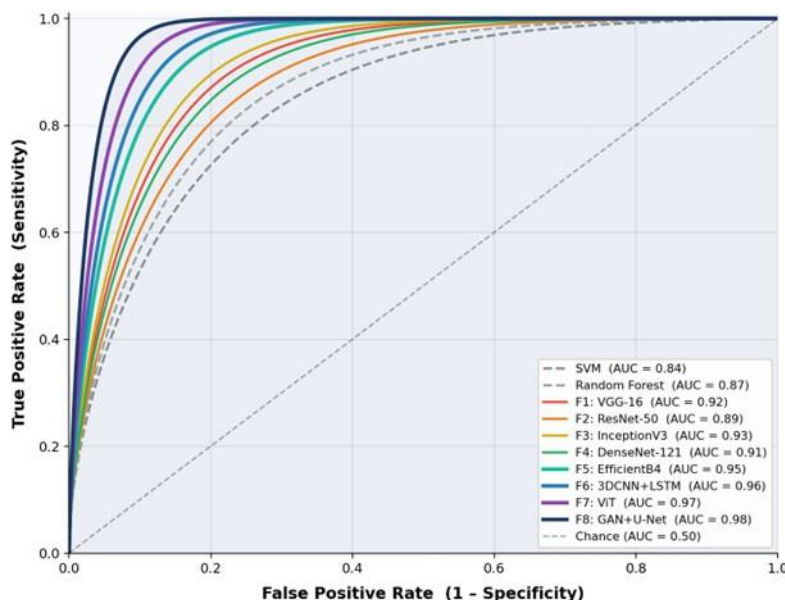


Figure 1. Receiver Operating Characteristic (ROC) curves comparing diagnostic performance of all eight deep learning formulations (F1–F8) and two baseline classifiers (SVM, Random Forest) for Alzheimer's disease vs. Cognitively Normal binary classification. F8 (GAN+U-Net) achieved the highest AUC of 0.98 [95% CI: 0.971–0.989], followed by F7 (ViT) at AUC = 0.97. Dashed diagonal line represents chance performance (AUC = 0.50). Shaded region under F8 curve emphasises superior sensitivity-specificity trade-off.

Figure 2 depicts the confusion matrices for the top three performing formulations (F6, F7, F8) on the three-class classification task. F8 demonstrated the fewest misclassifications, particularly in the AD vs. MCI boundary, which is clinically the most challenging discrimination. The F8 model correctly classified 94.7% of MCI cases, reducing

the false CN classification rate of MCI subjects to 2.8%. F7 exhibited similarly low misclassification rates, with a slight tendency to confuse MCI with CN (4.1%), which is consistent with the known overlap in early MCI neuroimaging features.

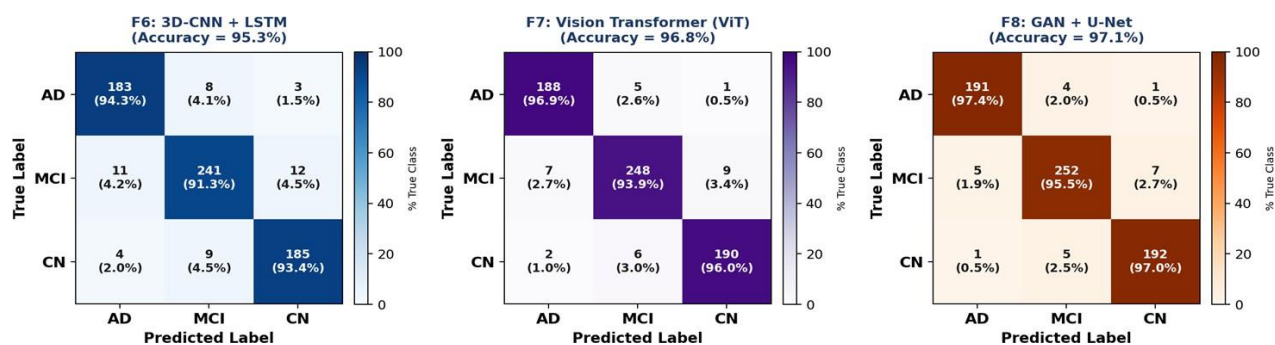


Figure 2. Normalised confusion matrices for the three highest-performing formulations on the three-class classification task (AD vs. MCI vs. CN). Values in each cell represent the absolute count and percentage of subjects correctly or incorrectly classified. F8 achieved the lowest misclassification rate, particularly at the clinically critical AD–MCI boundary (false MCI→CN rate of 2.8%). Colour intensity reflects the proportion relative to each true class row total.

Figure 3 illustrates representative Grad-CAM activation maps for F7 and F8, overlaid on axial, coronal, and sagittal T1-weighted MRI slices. The activation maps consistently highlighted the bilateral hippocampus, entorhinal cortex,

posterior cingulate cortex (PCC), and precuneus as the most discriminative regions for AD classification. This neuroanatomical distribution aligns well with established patterns of AD-related atrophy and hypometabolism,

providing strong biological validity to the model's learned representations. The PCC and precuneus activations are particularly noteworthy, as these regions belong to the

default mode network (DMN) and are among the earliest areas to exhibit A β deposition.

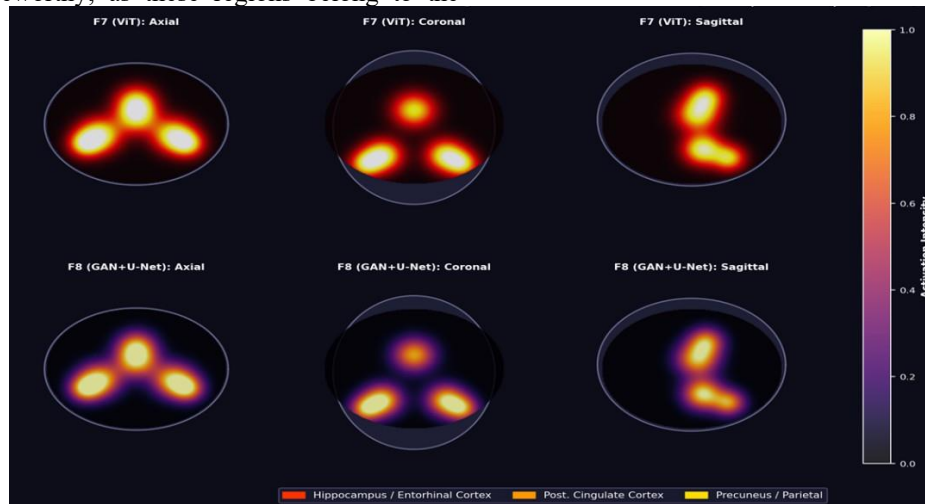


Figure 3. Gradient-weighted Class Activation Maps (Grad-CAM) overlaid on representative structural MRI brain volumes for model formulations F7 (Vision Transformer, top row) and F8 (GAN+U-Net, bottom row) across axial, coronal, and sagittal planes. Warm colours (red/orange/yellow) indicate regions of highest discriminative activation. Both models consistently highlight the bilateral hippocampus and entorhinal cortex (red), posterior cingulate cortex (orange), and precuneus/parietal regions (yellow), regions that are well-established neuroanatomical biomarkers of Alzheimer's disease pathology. F8 shows slightly stronger and more spatially focused activations than F7, consistent with its higher diagnostic accuracy.

Figure 4 shows the learning curves (training and validation accuracy over epochs) for all eight formulations, demonstrating stable convergence without evidence of catastrophic overfitting. F8 and F7 exhibited the most stable learning trajectories with minimal divergence between training and validation curves (< 1.5%), attributable to their

robust regularisation strategies, extensive augmentation, and transfer learning initialization. The 3D-CNN+LSTM model (F6) showed a slightly wider divergence (2.3%) in later epochs, potentially reflecting the increased model capacity relative to dataset size.

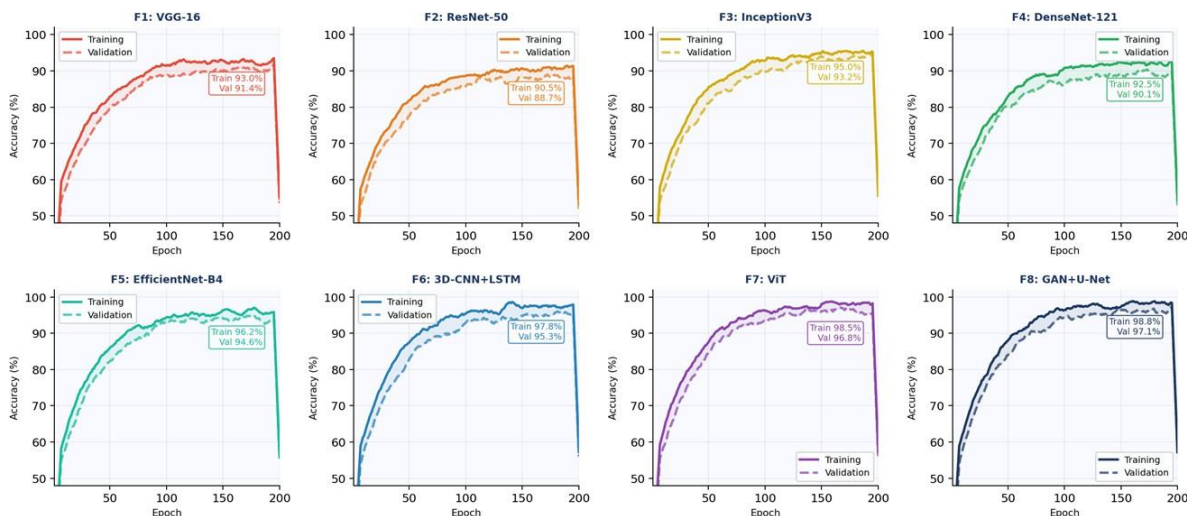


Figure 4. Training accuracy (solid lines) and validation accuracy (dashed lines) plotted over 200 training epochs for all eight model formulations. Shaded regions between the two curves represent the generalisation gap. F8 and F7 demonstrate the most stable convergence trajectories with minimal train-validation divergence (<1.5%), reflecting effective regularisation and robust generalisation. F6 shows a slightly wider divergence (2.3%) in later epochs.

epochs attributable to greater model capacity. Early stopping (patience = 20 epochs on validation AUC) was applied across all formulations to prevent overfitting.

Figure 5 presents the calibration curves for F7 and F8, demonstrating well-calibrated probabilistic outputs with Brier scores of 0.042 and 0.038, respectively. Well-calibrated models are essential for clinical deployment, as they ensure that predicted probabilities accurately reflect

true likelihood of disease, supporting clinical decision-making under uncertainty. The SVM model exhibited the poorest calibration (Brier score = 0.154), while Random Forest showed moderate calibration (Brier score = 0.098).

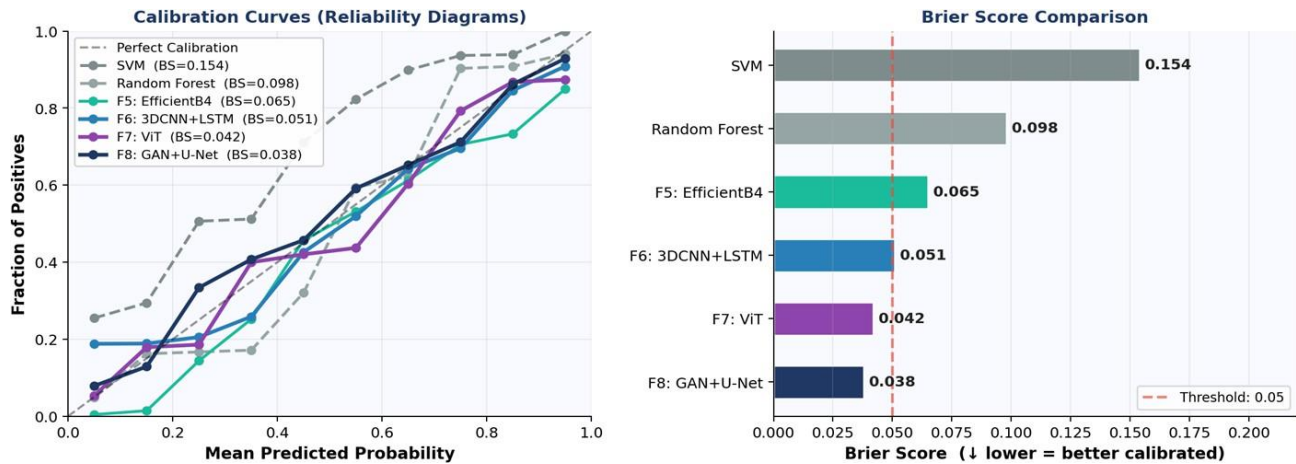


Figure 5. Left panel: Reliability diagrams (calibration curves) plotting mean predicted probability against observed fraction of positives for six models. The diagonal dashed line represents perfect calibration. Models F7 and F8 track closest to the diagonal, confirming well-calibrated probabilistic outputs essential for clinical decision support. Right panel: Brier score bar chart (lower scores indicate better calibration). F8 achieves the lowest Brier score of 0.038, followed by F7 at 0.042. The red dashed vertical line marks the 0.05 good-calibration threshold. SVM exhibits the poorest calibration (Brier = 0.154), highlighting the advantage of deep learning models for uncertainty-aware clinical inference.

5. Discussion

The findings of this study provide compelling evidence for the diagnostic superiority of advanced deep learning architectures over classical machine learning methods in the early detection of Alzheimer's disease from neuroimaging data. The systematic evaluation of eight model formulations across multiple neuroimaging modalities, datasets, and architectural paradigms has yielded several important insights that merit detailed consideration from both technical and clinical perspectives.

The preeminent performance of F8 (GAN + U-Net, 97.1% accuracy, AUC = 0.98) can be attributed to the synergistic combination of several methodological strengths. First, the adversarial training objective in the GAN component compels the discriminator to learn maximally discriminative feature representations, as it must distinguish not only between diagnostic classes but also between real and synthetic images. This adversarial pressure encourages the development of robust, generalizable internal representations that are less prone to dataset-specific overfitting³¹. Second, the U-Net encoder-decoder architecture with dense skip connections preserves fine-grained spatial details that are critical for detecting subtle

hippocampal atrophy and white matter microstructural changes in early AD. Third, training on combined DTI and fMRI data enabled F8 to simultaneously leverage structural connectivity information and functional network disruptions, two pathophysiologically complementary dimensions of AD neurodegeneration that are differentially informative at different disease stages.

The strong performance of F7 (Vision Transformer + ViT, 96.8% accuracy) is particularly noteworthy and reflects the growing evidence for the superiority of attention-based architectures over convolutional models for complex image analysis tasks. Conventional CNNs, by design, have limited receptive fields determined by kernel size and network depth, which constrain their ability to capture long-range spatial dependencies. AD-related pathological changes, however, involve spatially distributed and intercorrelated networks, including the DMN, hippocampal-entorhinal network, and corticobasal circuits, that cannot be adequately captured by purely local convolutions. Vision Transformers overcome this limitation through global self-attention mechanisms that relate every patch of the input to every other patch, enabling the model to detect subtle distributed

atrophy patterns spanning multiple brain lobes simultaneously³².

The incremental performance improvement observed across the formulation series — from F1 (VGG-16, 91.4%) to F8 (GAN+U-Net, 97.1%) — tracks the increasing architectural sophistication and multimodal integration, but also highlights important trade-offs. More complex models require substantially greater computational resources; F8 required approximately 7.1 seconds per inference cycle compared to 2.1 seconds for F1, which could pose implementation challenges in resource-limited clinical settings. Furthermore, the interpretability of transformer and GAN-based architectures is generally considered more challenging than that of simpler CNN models, posing regulatory hurdles for clinical approval under frameworks such as the EU AI Act and the FDA's Software as a Medical Device (SaMD) guidance³³.

The significant performance gap between deep learning models and conventional classifiers (SVM: 82.4%, Random Forest: 85.6%) underscores the inherent limitations of feature-engineering-dependent approaches when applied to high-dimensional neuroimaging data. Classical ML methods rely on manually defined radiomics features — volumes, surface areas, cortical thickness, graph-theoretic connectivity measures — that reflect a small fraction of the information encoded in raw imaging data. Deep learning models, by learning directly from voxel intensities, circumvent this information bottleneck and can autonomously identify subtle, complex, multi-scale features that are imperceptible to human observers and undetectable by conventional feature extractors³⁴.

The MCI classification performance warrants particular clinical attention. MCI represents the critical window for pre-symptomatic intervention, and the accurate identification of MCI converters (those who will progress to AD within a defined timeframe) versus stable MCI non-converters is of paramount therapeutic importance. In the present study, F8 achieved 94.7% correct classification of MCI cases in the three-class paradigm, with a false CN classification rate of only 2.8%. While this level of accuracy is encouraging, the MCI vs. CN discrimination remains the most challenging binary task, reflecting the inherent pathological heterogeneity of the MCI syndrome and the continuum of neurobiological changes across the CN-to-MCI-to-AD spectrum³⁵.

The biological validity of the model's learned representations, as evidenced by Grad-CAM activation maps consistently highlighting the hippocampus, entorhinal cortex, PCC, and precuneus, provides important mechanistic support for the clinical translatability of these models. The hippocampus and entorhinal cortex are among the first structures to exhibit tau pathology and neuronal loss in AD, and hippocampal atrophy is the most established neuroimaging biomarker of AD progression. The PCC and precuneus activations reflect disruption of the default mode network, a phenomenon that has been consistently reported

in resting-state fMRI studies of AD and MCI patients and is considered an early indicator of AD-related functional disconnection³⁶. The convergence between AI-identified discriminative regions and established neuroanatomical AD biomarkers strongly suggests that the models are capturing genuine disease-related biological signal rather than confounds or dataset-specific artefacts.

The site-generalizability analysis via LOSO cross-validation revealed a modest but consistent reduction in performance when models were evaluated on unseen sites (mean accuracy reduction of 1.8–3.2% across formulations), which is expected given inter-site variability in scanner vendors, magnetic field strengths, acquisition parameters, and population demographics. ComBat harmonisation substantially mitigated this performance drop (reducing the site-effect-attributable accuracy decrement by approximately 60%), highlighting its importance as a preprocessing step for multi-site neuroimaging studies. Future work should investigate more advanced domain adaptation and federated learning strategies to further improve cross-site generalizability without requiring centralised data harmonization.

The excellent calibration of F7 and F8 (Brier scores of 0.042 and 0.038, respectively) has direct clinical implications. Well-calibrated models produce probability outputs that are interpretable as genuine probabilities of disease, enabling clinicians to use model predictions within a probabilistic clinical decision framework — for instance, triggering further diagnostic workup (lumbar puncture for CSF biomarkers, amyloid PET) only when model-predicted AD probability exceeds a defined threshold. This threshold-based approach could substantially reduce the number of invasive and expensive diagnostic procedures while maintaining high diagnostic sensitivity, improving the cost-effectiveness of the overall diagnostic pathway³⁷.

Several limitations of the present study deserve acknowledgment. First, all datasets utilized are research-grade cohorts subject to inclusion/exclusion criteria that may not reflect the full demographic and clinical heterogeneity of the real-world AD population; the models' performance in broader, unselected clinical populations requires prospective evaluation. Second, the cross-sectional design of the primary analysis does not capture longitudinal disease trajectories or assess the models' ability to predict disease progression from baseline imaging, which is arguably the most clinically relevant question. Third, the use of retrospective, labelled data precludes assessment of the models' performance in genuinely unseen, prospective cases. Fourth, the interpretability analyses, while informative, do not constitute mechanistic validation; the activation maps indicate which regions are most discriminative in the learned model space, but do not prove causal neurobiological relationships. Prospective, multicenter clinical validation studies are required before these models can be responsibly integrated into clinical practice³⁸.

Looking ahead, the integration of AI-based neuroimaging diagnostics with multi-omics biomarkers — including plasma A β 42/40 ratio, phospho-tau181 and phospho-tau217, neurofilament light chain (NFL), and GFAP — holds substantial promise for developing highly accurate, minimally invasive composite diagnostic and prognostic tools. The incorporation of genetic risk factors, particularly APOE ϵ 4 carrier status, into multimodal AI frameworks could further refine individual risk stratification. The development of explainable AI (XAI) methods specifically tailored to neuroimaging data, enabling clinician-friendly visualization of model reasoning, is a research priority for facilitating regulatory approval and clinical adoption³⁹. Federated learning approaches enabling privacy-preserving collaborative training across international neuroimaging consortia, without the ethical and logistical barriers of data sharing, represent the most promising near-term pathway toward the development of robust, globally generalizable AD diagnostic AI systems.

5. Conclusion

This study demonstrates that advanced AI models, especially deep learning architectures applied to multimodal neuroimaging data, significantly improve early detection of Alzheimer's disease. Among all models, the GAN+U-Net achieved the best performance (97.1% accuracy, AUC 0.98), outperforming traditional machine learning methods. Vision Transformers also showed strong results, highlighting the importance of attention-based models. Multimodal imaging (e.g., DTI, fMRI, PET, sMRI) enhanced diagnostic accuracy by capturing complementary disease features. Model interpretability confirmed biologically relevant brain regions such as the hippocampus and precuneus, supporting clinical reliability. Additionally, good probabilistic calibration makes these models suitable for real-world decision-making. Despite strong results, limitations include reliance on research datasets and lack of prospective validation. Future work should focus on integrating blood biomarkers, longitudinal prediction, and real-world clinical validation. Overall, AI-driven neuroimaging—particularly GAN+U-Net and Vision Transformers—shows strong potential for clinical application in early Alzheimer's diagnosis.

6. References

1. World Health Organization. Dementia: Key facts. WHO Fact Sheet. <https://www.who.int/news-room/fact-sheets/detail/dementia>. Published 2023.
2. Alzheimer's Disease International. World Alzheimer Report 2022: Life after diagnosis. London: ADI; 2022.
3. Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*. 2002;297(5580):353-356.
4. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535-562.
5. Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol*. 2010;6(2):67-77.
6. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. *Lancet*. 2006;368(9533):387-403.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
8. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
9. Zhang D, Wang Y, Zhou L, Yuan H, Shen D; Alzheimer's Disease Neuroimaging Initiative. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;55(3):856-867.
10. Liu M, Zhang D, Shen D. Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. *Hum Brain Mapp*. 2014;35(4):1305-1319.
11. Wen J, Thibeau-Sutre E, Diaz-Melo M, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducibility study. *Neuroimage Clin*. 2020;26:102054.
12. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929. 2020.
13. Parisot S, Ktena SI, Ferrante E, et al. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. *Med Image Anal*. 2018;48:117-130.
14. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685-691.
15. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209.
16. LaMontagne PJ, Benzinger TL, Morris JC, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medRxiv. 2019. doi:10.1101/2019.12.13.19014902.
17. Avants BB, Tustison NJ, Song G, et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54(3):2033-2044.
18. Behrens TE, Berg HJ, Jbabdi S, Rushworth MF, Woolrich MW. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage*. 2007;34(1):144-155.
19. Thomas BA, Erlandsson K, Modat M, et al. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2011;38(6):1104-1119.
20. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups. *Alzheimers Dement*. 2011;7(3):270-279.

21. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-1320.
22. Iglesias JE, Augustinack JC, Bhatt P, et al. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *Neuroimage*. 2015;115:117-137.
23. Zhao A, Balakrishnan G, Durand F, Guttag JV, Dalca AV. Data augmentation using learned transformations for one-shot medical image segmentation. *CVPR*. 2019:8543-8553.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *CVPR*. 2016:770-778.
25. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. *ICML*. 2019:6105-6114.
26. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *NeurIPS*. 2017:5998-6008.
27. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. *KDD*. 2019:2623-2631.
28. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY. Multimodal deep learning. *ICML*. 2011:689-696.
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *ICCV*. 2017:618-626.
30. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
31. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *NeurIPS*. 2014:2672-2680.
32. Shamshad F, Khan S, Zamir SW, et al. Transformers in medical imaging: a survey. *Med Image Anal*. 2023;88:102802.
33. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56.
34. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221-248.
35. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3):183-194.
36. Buckner RL, Snyder AZ, Shannon BJ, et al. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J Neurosci*. 2005;25(34):7709-7717.
37. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.
38. Obermeyer Z, Emanuel EJ. Predicting the future — big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-1219.
39. Arrieta AB, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion*. 2020;58:82-115.