

Original Research

Received 03 June 2026

Revised 18 July 2026

Accepted 01 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (15s): 102–129

KEYWORDS:

Alzheimer's disease; artificial intelligence; neuroimaging; multimodal learning; graph neural networks; Transformers; longitudinal modelling; explainable AI.

Artificial Intelligence for Early Alzheimer's Disease Diagnosis and Progression Analysis: A Comprehensive Survey of Deep Learning, Multimodal Learning, Graph Networks, and 3D Transformers

Shruti VijayKumar Hegdekar¹, Prathibhavani P. Maruthi¹, Venugopal Kuppanna Rajuk²

¹Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bengaluru, India

²Bangalore University, Bengaluru, India

Corresponding author: Shruti VijayKumar Hegdekar

Email: hegdekarshruti@gmail.com

Co-author emails:

Prathibhavani P. Maruthi: prathibhavani.pm@uvce.ac.in

Venugopal Kuppanna Rajuk: venugopalkr@gmail.com

ABSTRACT: Alzheimer's disease (AD) is a progressive neurodegenerative disorder in which early and reliable diagnosis remains challenging because pathological changes often begin years before severe clinical symptoms become evident. Recent advances in artificial intelligence (AI) have enabled increasingly sophisticated analysis of neuroimaging, cognitive, genetic, and biomarker data for AD diagnosis, staging, and progression prediction. This survey presents a comprehensive review of AI-based approaches for Alzheimer's disease, covering conventional machine-learning methods, 2D and 3D convolutional neural networks, multimodal deep learning, contrastive and metric learning, graph neural networks, Siamese architectures, and Transformer-based models. Particular attention is given to the use of structural MRI, fMRI, PET, clinical and cognitive variables, genetic information, and fluid biomarkers, together with commonly adopted preprocessing, harmonization, and multimodal fusion strategies. The survey also examines explainability, uncertainty estimation, cross-dataset robustness, longitudinal progression modelling, computational efficiency, and clinical generalizability. Comparative analysis of representative studies indicates that while very high performance is frequently reported for AD-versus-cognitively-normal classification, more clinically relevant tasks such as mild cognitive impairment conversion and individualized progression prediction remain substantially more difficult. To address these limitations, a conceptual future architecture is proposed that integrates modality-specific foundation encoders, cross-modal contrastive alignment, a structure-aware multimodal Transformer, dynamic patient-graph reasoning, longitudinal temporal modelling, multi-task prediction, and an explainable clinical layer. Overall, the survey highlights the need to move beyond single-dataset accuracy toward robust, interpretable, uncertainty-aware, longitudinal, and clinically deployable AI systems for early Alzheimer's disease diagnosis and progression assessment.

1. INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder characterized by deterioration in memory, cognition, behaviour, and functional ability. Early identification of the transition from cognitively normal ageing to Mild Cognitive Impairment (MCI) and subsequently to Alzheimer's disease is particularly important because structural and functional abnormalities may begin several years before severe clinical symptoms become evident [1]. Neuroimaging techniques such as structural Magnetic Resonance Imaging (sMRI), functional MRI (fMRI), Positron Emission Tomography (PET), and other biological and clinical measurements have therefore become important sources of diagnostic information. Artificial intelligence has significantly transformed automated Alzheimer's disease diagnosis. Initial approaches relied on handcrafted image features and conventional machine-learning classifiers. Deep Convolutional Neural Networks subsequently enabled automatic representation learning directly from neuroimaging data [2]. More recent developments have introduced three-dimensional convolutional networks, multimodal fusion, attention mechanisms, Vision Transformers, graph neural networks, contrastive learning, metric learning, explainable AI, and longitudinal modelling. Despite substantial progress, several problems remain unresolved. These include subtle differences between neighbouring disease stages, limited labelled datasets, class imbalance, variations among imaging centres and MRI scanners, incomplete multimodal information, high computational requirements for three-dimensional imaging, model interpretability, uncertainty estimation, reproducibility, and limited prospective clinical validation [3]. Therefore, this survey provides a comprehensive analysis of artificial-intelligence techniques for Alzheimer's disease diagnosis, staging, and progression prediction. The survey systematically organizes existing research according to imaging modality, data representation, deep-learning architecture, multimodal fusion strategy, learning objective, explainability, computational efficiency, and clinical applicability [4]. Particular attention is devoted to recent developments involving 3D CNNs, multimodal representation learning, graph neural networks, contrastive learning, Siamese architectures, and three-dimensional Transformers.

1.1 Motivation

The motivation for this survey arises from the rapid diversification of Artificial Intelligence techniques used for Alzheimer's disease diagnosis and progression analysis. Earlier studies primarily focused on conventional machine-learning and CNN-based classification, whereas recent approaches increasingly exploit volumetric 3D learning, multimodal information, contrastive representation learning, graph-based reasoning, metric learning, and Transformer architectures. These developments address different limitations of Alzheimer's diagnosis, including loss of volumetric context, poor discrimination between neighbouring disease stages, heterogeneity between imaging and clinical data, and high computational requirements [5]. Another important motivation is the difficulty of detecting early-stage disease progression. Distinguishing cognitively normal subjects from established Alzheimer's disease is generally less challenging than identifying subtle transitions such as stable MCI to progressive MCI. Recent volumetric approaches have therefore focused on preserving fine anatomical information such as hippocampal shrinkage, cortical thinning, and localized grey-matter changes. For example, the ACSPSF framework employs 3D dilated convolutions and Proxy Anchor Loss to improve volumetric discrimination, while SAR-Swin3D introduces structure-aware routing to preserve subtle anatomical boundaries during hierarchical processing [6]. The increasing availability of multimodal information provides further motivation for this survey. Structural imaging alone may not sufficiently represent the complex progression of Alzheimer's disease. Recent systems therefore integrate structural MRI with fMRI, clinical, demographic, cognitive, and genetic information. NeuroDiagAI demonstrates MRI–fMRI integration, whereas AlzGraph-CL combines structural MRI and clinical information using contrastive alignment and graph-based reasoning. These developments indicate the need for a unified review that explains how different AI paradigms contribute to early diagnosis, staging, and progression prediction [7].

1.2 Need for the Survey

Although numerous studies have reported high classification performance, direct comparison among them remains difficult because they differ in datasets, class definitions, imaging modalities, preprocessing strategies, validation protocols, and diagnostic objectives. A model designed for binary CN-versus-AD classification cannot be directly compared with a model designed for multiclass staging or sMCI-versus-pMCI prediction. Existing research is also distributed across several methodological directions, including CNNs, 3D CNNs, multimodal learning, graph neural networks, contrastive learning, Siamese networks, and Transformers. Consequently, there is a need for a survey that organizes these methods within a common

framework and evaluates them not only in terms of accuracy but also according to representation quality, multimodal integration, computational efficiency, interpretability, robustness, and clinical relevance [8].

1.3 Scope of the Survey

This survey focuses on Artificial Intelligence approaches for Alzheimer's disease detection, classification, staging, and progression prediction using neuroimaging and complementary clinical information. The major areas covered include:

- conventional machine-learning approaches;
- 2D and 3D CNN-based methods;
- transfer-learning architectures;
- multimodal MRI, fMRI, PET, and clinical-data integration;
- contrastive and metric-learning techniques;
- Siamese learning architectures;
- graph neural networks and population-based learning;
- Vision Transformers and 3D Swin Transformers;
- explainable AI and attention-based interpretation;
- longitudinal disease-progression modelling;
- cross-dataset generalization and robustness; and
- computational efficiency and clinical deployment.

The survey primarily emphasizes recent developments while also considering important foundational approaches required to understand the evolution of the field.

1.4 Key Research Challenges

The literature reveals several recurring research challenges that motivate continued investigation. These include limited labelled neuroimaging data, class imbalance, overlapping characteristics between neighbouring disease stages, scanner and acquisition variability, incomplete multimodal information, high computational requirements of 3D models, and limited external validation. The early-stage classification problem is particularly important. SAR-Swin3D, for instance, reports stronger performance for CN-versus-AD classification than for sMCI-versus-pMCI discrimination, illustrating the increased difficulty of detecting subtle progression-related changes [9]. Similarly, existing multimodal research identifies overfitting, interpretability, data scarcity, cross-site generalization, and the practical availability of expensive modalities as continuing limitations. These challenges indicate that future systems must move beyond maximizing benchmark accuracy toward reliable and clinically transferable disease assessment.

1.5 Objectives of the Survey

The major objectives of this survey are:

1. To review AI-based methods for Alzheimer's disease diagnosis and progression prediction.
2. To compare ML, CNN, GNN, Transformer, multimodal, and representation-learning approaches.
3. To analyze neuroimaging modalities, datasets, and evaluation protocols used in existing studies.
4. To examine early-stage and MCI progression prediction, including multimodal fusion and explainability.
5. To identify limitations, challenges, and future research directions in AI-based Alzheimer's diagnosis.

1.6 Major Contributions of the Survey

The principal contributions of this survey are as follows:

- A unified taxonomy of AI-based Alzheimer's disease diagnosis methods is presented.
- Existing methods are categorized according to architecture, imaging modality, learning strategy, and clinical task.
- Major public datasets and preprocessing approaches are comparatively reviewed.
- Multimodal fusion approaches are examined from conventional feature concatenation to contrastive and graph-based alignment.
- Particular emphasis is placed on early-stage classification and MCI progression prediction.
- Explainability, uncertainty, computational complexity, and cross-dataset robustness are considered alongside conventional performance measures.
- Major research gaps involving missing modalities, domain shift, clinical interpretability, external validation, and longitudinal modelling are identified.
- Emerging directions including multimodal foundation models, federated learning, self-supervised learning, longitudinal Transformers, dynamic patient graphs, and uncertainty-aware diagnosis are discussed.

1.7 Organization of the Paper

The remainder of this survey is organized as follows. Section 2 presents the survey methodology, literature-search strategy, inclusion criteria, and study-selection procedure. Section 3 discusses Alzheimer's disease biomarkers and neuroimaging modalities. Section 4 reviews commonly used datasets and preprocessing techniques. Subsequent sections examine conventional machine-learning methods, CNN and 3D CNN architectures, multimodal learning, contrastive and metric learning, graph neural networks, Siamese approaches, and Transformer-based models. Later sections analyze explainability, robustness, longitudinal progression modelling, computational efficiency, research gaps, and future directions. Finally, the survey concludes by summarizing the major findings and opportunities for clinically reliable AI-assisted Alzheimer's disease diagnosis.

2. SURVEY METHODOLOGY

A structured review methodology was adopted to identify, screen, classify, and critically analyse studies related to Artificial Intelligence (AI)-based Alzheimer's disease diagnosis, staging, and progression prediction. The review primarily considered recent developments in machine learning, deep learning, multimodal learning, Graph Neural Networks (GNNs), contrastive and metric learning, Siamese architectures, and Transformer-based neuroimaging. The overall review process consisted of literature identification, duplicate removal, title and abstract screening, full-text eligibility assessment, data extraction, methodological categorization, and comparative synthesis.

2.1 Literature Search Strategy

Relevant research articles were identified from major scientific databases, including Scopus, Web of Science, IEEE Xplore, PubMed, ScienceDirect, SpringerLink, and the ACM Digital Library. The primary search period covered studies published between 2019 and 2026, while selected earlier studies were also considered when they represented foundational developments in Alzheimer's disease diagnosis or neuroimaging-based machine learning. Search terms were constructed using combinations of keywords such as *Alzheimer's disease*, *Mild Cognitive Impairment*, *MRI*, *fMRI*, *PET*, *deep learning*, *3D CNN*, *multimodal learning*, *graph neural network*, *contrastive learning*, *Siamese network*, *Vision Transformer*, *Swin Transformer*, *explainable AI*, and *progression prediction*. Similar literature-search procedures have been commonly adopted by researchers such as Hegdekar *et al.* in investigating recent AI-based Alzheimer's disease diagnostic approaches [10].

Boolean operators were used to combine disease-related, imaging-related, and AI-related terms. A representative search expression was formulated as ("*Alzheimer's disease*" OR "*Mild Cognitive Impairment*") AND ("*MRI*" OR "*PET*" OR "*neuroimaging*") AND ("*deep learning*" OR "*CNN*" OR "*GNN*" OR "*Transformer*"). Additional targeted searches were performed for multimodal fusion, MCI conversion prediction, explainable AI, graph-based learning, contrastive learning, and three-dimensional Transformer architectures to ensure that emerging methodological directions were adequately represented [11].

2.2 Inclusion and Exclusion Criteria

Studies were considered eligible when they focused on Alzheimer's disease diagnosis, MCI classification, disease staging, or progression prediction and employed machine-learning or deep-learning techniques. Articles using neuroimaging modalities such as MRI, fMRI, and PET, as well as clinical, demographic, cognitive, genetic, electrophysiological, or biological information, were considered [12]. Only studies reporting an identifiable dataset, clearly described methodology, and quantitative experimental evaluation were retained. Peer-reviewed journal articles and relevant conference papers were included to ensure methodological and scientific quality. Studies were excluded when they did not contain an AI-based diagnostic or prognostic component, lacked sufficient experimental information, primarily addressed unrelated neurological disorders without separate Alzheimer's disease results, or failed to report quantitative evaluation metrics. Editorials, abstracts, commentaries, duplicated publications, and studies containing insufficient information regarding datasets or validation procedures were also excluded from the final analysis.

2.3 Study Selection Procedure

The study-selection process followed a PRISMA-oriented screening procedure. Page *et al.* introduced the PRISMA framework to improve transparency in systematic review reporting, and its general principles were adopted to organize the literature-selection process. Initially, all records obtained from the selected databases were combined, after which duplicate, publications were removed using title, author information, publication year, and DOI. The remaining records were screened based on their titles and abstracts to determine their relevance to AI-assisted Alzheimer's disease diagnosis and progression analysis. Full-text assessment was subsequently carried out for potentially eligible studies. During this stage, each article was evaluated according to its clinical objective, dataset, imaging modality, AI architecture, validation methodology, and reported outcomes. Articles satisfying all inclusion requirements were retained for the final qualitative and comparative synthesis. The selection process can therefore be represented as identification → duplicate removal → screening → eligibility assessment → final inclusion. The final PRISMA diagram should report the actual numbers of identified, excluded, assessed, and included studies after completion of the database search.

2.4 Data Extraction and Classification

For every selected publication, information related to authors, publication year, dataset, imaging modality, diagnostic classes, preprocessing technique, model architecture, learning strategy, validation protocol, performance measures, explainability mechanism, computational requirements, and reported limitations was extracted. Studies were then grouped according to their primary methodological characteristics [13]. Conventional machine-learning approaches were considered separately from 2D and 3D CNN-based methods, while more recent approaches were categorized into multimodal learning, contrastive and metric learning, Siamese networks, graph neural networks, Transformer-based models, and longitudinal progression-prediction frameworks.

2.5 Comparative Analysis

The selected studies were comparatively analysed using commonly reported measures such as accuracy, sensitivity, specificity, precision, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Wherever available, additional factors such as model parameters, computational complexity, inference time, statistical significance, explainability, uncertainty estimation, cross-dataset validation, and external clinical evaluation were also considered. This broader comparison was necessary because high classification accuracy alone does not necessarily indicate clinical reliability or generalizability. Studies were further compared according to equivalent diagnostic tasks. Binary CN-versus-AD classification, CN-versus-MCI discrimination, multiclass CN/MCI/AD classification, multistage dementia classification, and stable-MCI-versus-progressive-MCI prediction were considered separately.

3. ALZHEIMER'S DISEASE BIOMARKERS AND NEUROIMAGING MODALITIES

Alzheimer's disease is associated with progressive structural, functional, cognitive, and molecular alterations that can be used as biomarkers for early diagnosis and disease-stage assessment. Because no single biomarker completely represents the heterogeneous pathological process of AD, recent studies increasingly combine neuroimaging with clinical and biological information. Hegdekar *et al.* emphasized that commonly observed neuroimaging biomarkers include hippocampal atrophy, cortical thinning, ventricular enlargement, and alterations in functional brain connectivity, while amyloid-beta and tau deposition

provide complementary pathological evidence. These biomarkers are particularly important for differentiating normal ageing, Mild Cognitive Impairment (MCI), and established AD.

3.1 Structural MRI Biomarkers

Structural Magnetic Resonance Imaging (sMRI) is one of the most widely used neuroimaging modalities for Alzheimer's disease because it provides high-resolution anatomical information without ionizing radiation. T1-weighted MRI enables the assessment of hippocampal atrophy, cortical thinning, ventricular enlargement, and grey-matter loss, which are among the most frequently investigated structural indicators of AD progression. The major structural, functional, molecular, cognitive, demographic, genetic, electrophysiological, and multi-omics biomarkers considered in Alzheimer's disease research are summarized in **Table 1**. Structural MRI remains particularly important because hippocampal shrinkage and cortical-thickness alterations can be quantified directly from volumetric brain images.

Table 1. Major Alzheimer's Disease Biomarkers

Biomarker Category	Representative Biomarkers	Clinical Relevance	Typical Data Source
Structural	Hippocampal atrophy, cortical thinning, ventricular enlargement, grey-matter loss	Reflects anatomical neurodegeneration and disease severity	Structural MRI
Functional	Altered functional connectivity, network disruption	Captures functional abnormalities that may precede visible atrophy	fMRI
Molecular	Amyloid- β , tau accumulation	Indicates core pathological processes of AD	PET, CSF
Cognitive	MMSE, CDR, ADAS-Cog	Quantifies cognitive decline and disease stage	Clinical assessment
Demographic	Age, sex, education	Provides subject-level contextual and risk information	Clinical records
Genetic	APOE $\epsilon 4$ and susceptibility markers	Represents inherited disease risk	Genetic testing
Electrophysiological	EEG rhythm and connectivity alterations	Captures temporal brain-function abnormalities	EEG
Multi-omics	Genomic, transcriptomic, epigenomic, proteomic markers	Supports personalized disease characterization	Omics platforms

3.2 Functional and Molecular Neuroimaging

Functional MRI provides information about brain connectivity and network activity, complementing the anatomical information obtained from structural MRI. Hegdekar *et al.* used MRI-derived structural features together with fMRI connectivity patterns to represent both anatomical degeneration and functional disruption. PET provides a different perspective by capturing metabolic activity and molecular pathology, including amyloid-related abnormalities. However, PET is comparatively expensive, less accessible, involves radiation exposure, and generally requires accurate alignment with structural imaging. The major characteristics, advantages, limitations, and typical AI applications of the principal neuroimaging and complementary modalities are compared in **Table 2**.

Table 2. Comparison of Neuroimaging and Complementary Modalities

Modality	Information Captured	Main Advantages	Major Limitations	Typical Role in AI-Based AD Analysis
sMRI	Brain morphology, hippocampal volume, cortical thickness	Non-invasive, high anatomical resolution, widely available	Early changes may be subtle	Primary modality for detection and staging

fMRI	Functional connectivity and network activity	Captures functional alterations beyond anatomy	Motion-sensitive and acquisition-dependent	Complementary functional representation
PET	Metabolic activity, amyloid and tau pathology	Provides molecular/pathological information	Expensive, radiation exposure, limited availability	Multimodal diagnosis and characterization
EEG	Electrical brain activity	High temporal resolution and comparatively low cost	Artifact-sensitive and protocol-dependent	Functional and temporal biomarker analysis
CSF	Amyloid and tau biomarkers	Strong biological relevance	Invasive sample collection	Molecular confirmation
Clinical/Cognitive	MMSE, CDR and patient characteristics	Clinically meaningful and easy to interpret	Can be incomplete or subjective	Combined with imaging in multimodal models
Genetic	APOE and related markers	Provides inherited disease-risk information	Does not directly measure current degeneration	Risk-aware prediction
Multi-omics	Genomic, epigenomic and proteomic profiles	Rich individualized biological information	Expensive and less widely available	Emerging precision-medicine applications

3.3 Multimodal Biomarker Integration

The comparison in **Tables 1 and 2** highlights the complementary nature of Alzheimer's biomarkers. Structural MRI provides detailed anatomical information, fMRI captures functional network disturbances, PET and CSF provide molecular evidence, while cognitive, demographic, and genetic variables contribute patient-specific clinical context. Hegdekar *et al.* demonstrated this multimodal principle by combining structural MRI with age, sex, MMSE, and APOE $\epsilon 4$ information for Alzheimer's classification [14]. Consequently, recent research is increasingly moving from isolated single-modality analysis toward multimodal learning frameworks capable of jointly exploiting structural, functional, molecular, and clinical biomarkers.

4. COMMONLY USED DATASETS AND PREPROCESSING TECHNIQUES

The reliability of Artificial Intelligence (AI)-based Alzheimer's disease diagnosis depends strongly on the quality, diversity, and preprocessing of the datasets used for model development. Neuroimaging studies frequently rely on publicly available datasets that contain structural MRI, PET, cognitive assessments, clinical variables, and biological biomarkers. Among these, the Alzheimer's Disease Neuroimaging Initiative (ADNI), Open Access Series of Imaging Studies (OASIS), Australian Imaging, Biomarkers and Lifestyle (AIBL), and Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) are widely used resources [15]. These datasets differ in population characteristics, imaging modalities, diagnostic labels, longitudinal follow-up, and acquisition protocols, making standardized preprocessing essential before AI-based analysis.

4.1 Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative is one of the most extensively used datasets for Alzheimer's disease research. ADNI as a collaborative initiative designed to investigate whether neuroimaging and biological biomarkers could support earlier and more reliable characterization of Alzheimer's disease. Subsequently described ADNI as a longitudinal multicentre initiative incorporating clinical assessments, MRI, PET, genetic information, and biochemical biomarkers for detecting and monitoring disease progression. ADNI is particularly useful for AI studies because it supports several diagnostic tasks, including cognitively normal versus AD classification, MCI detection, multiclass disease staging, and longitudinal conversion analysis. Its multimodal nature also enables the development of models combining anatomical, metabolic, cognitive, and molecular information.

However, because ADNI data originate from multiple acquisition centres, scanner and protocol variability must be carefully addressed during preprocessing and validation.

4.2 Open Access Series of Imaging Studies

The Open Access Series of Imaging Studies is another important dataset for neuroimaging-based Alzheimer's research. OASIS-1 as a cross-sectional MRI resource containing young, middle-aged, nondemented older adults and subjects with varying levels of dementia. OASIS-1 contains 416 subjects and multiple T1-weighted MRI acquisitions, including individuals clinically characterized with very mild to moderate Alzheimer's disease [16]. Later OASIS releases expanded the dataset toward longitudinal and multimodal analysis. OASIS-3 includes longitudinal neuroimaging together with clinical, cognitive, and biomarker information for the study of normal ageing and Alzheimer's disease. Consequently, OASIS is frequently used for MRI classification, dementia-stage analysis, cross-dataset testing, and independent validation of models trained on other cohorts.

4.3 Australian Imaging, Biomarkers and Lifestyle Dataset

The Australian Imaging, Biomarkers and Lifestyle study of ageing provides another longitudinal resource for Alzheimer's research. AIBL as a longitudinal study initially involving 1,112 participants and designed to investigate relationships among neuroimaging, biomarkers, lifestyle factors, cognitive function, and Alzheimer's disease progression [17]. AIBL is particularly valuable for external validation because it provides a population different from commonly used North American cohorts. It can therefore assist in evaluating whether AI models generalize across demographic and geographical populations rather than being optimized for a single dataset.

4.4 MIRIAD Dataset

The MIRIAD dataset is primarily designed for longitudinal MRI analysis. MIRIAD as a publicly available collection containing repeated volumetric T1-weighted MRI scans from patients with mild-to-moderate Alzheimer's disease and healthy control subjects. The dataset includes 708 scans from 46 AD patients and 23 controls obtained over multiple follow-up intervals. Unlike datasets primarily developed for cross-sectional classification, MIRIAD is particularly useful for analysing longitudinal volume changes, brain atrophy rates, and disease progression [18]. Therefore, it can support AI models aimed at temporal monitoring rather than only single-time-point diagnosis.

Table 3. Commonly Used Datasets for Alzheimer's Disease AI Research

Dataset	Major Modalities	Nature	Typical Applications	Major Strength
ADNI ¹	MRI, PET, clinical, cognitive, genetic, CSF biomarkers	Longitudinal, multicentre	Diagnosis, staging, MCI conversion, multimodal learning	Rich multimodal information
OASIS ²	Mainly MRI with clinical/cognitive information; later releases are multimodal	Cross-sectional and longitudinal	MRI classification, staging, external validation	Open-access neuroimaging resource
AIBL ³	MRI, PET, biomarkers, cognitive and lifestyle information	Longitudinal	Progression analysis and external validation	Diverse longitudinal biomarker information
MIRIAD ⁴	T1-weighted structural MRI and clinical information	Longitudinal	Atrophy measurement and progression modelling	Repeated MRI measurements

¹ Alzheimer's Disease Neuroimaging Initiative (ADNI): <https://adni.loni.usc.edu/>

² Open Access Series of Imaging Studies (OASIS): <https://www.oasis-brains.org/>

³ Australian Imaging, Biomarkers and Lifestyle (AIBL) Study of Ageing: <https://aibl.org.au/>

⁴ Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD): [MIRIAD dataset – UCL](https://www.miriad-dataset.org/)

As summarized in **Table 3**, no single dataset is optimal for every diagnostic task. ADNI provides broad multimodal and longitudinal information, OASIS is particularly useful for accessible MRI benchmarking, AIBL supports independent population validation, and MIRIAD is well suited for longitudinal structural-change analysis.

4.5 Mathematical Formulation of Neuroimaging Preprocessing

Let the raw three-dimensional MRI volume of subject i be represented by

$$X_i \in \mathbb{R}^{H \times W \times D}, \quad (1)$$

where H , W , and D denote the height, width, and depth of the MRI volume, respectively. As shown in Eq. (1), each subject is represented as a volumetric input rather than a collection of independent 2D slices. The complete preprocessing pipeline transforms the raw image X_i into a standardized volume $\tilde{X}_i$, which can be written as

$$\tilde{X}_i = \mathcal{P}(X_i) = \mathcal{R} \circ \mathcal{N} \circ \mathcal{S} \circ \mathcal{B} \circ \mathcal{D}(X_i), \quad (2)$$

where $\mathcal{D}$, $\mathcal{B}$, $\mathcal{S}$, $\mathcal{N}$, and $\mathcal{R}$ denote denoising, bias-field correction, skull stripping, intensity normalization, and spatial registration/resampling, respectively. Equation (2) provides a generalized representation of the preprocessing sequence commonly adopted in Alzheimer's neuroimaging studies.

4.5.1 Denoising

MRI volumes can contain acquisition noise that may interfere with the extraction of subtle disease-related anatomical characteristics. The observed voxel intensity can be modeled as

$$X(r) = S(r) + \eta(r), \quad (3)$$

where $r = (x, y, z)$ denotes the spatial coordinate, $S(r)$ represents the underlying anatomical signal, and $\eta(r)$ represents acquisition noise. According to Eq. (3), the observed image consists of the desired tissue signal together with an unwanted noise component. A generic denoising operation is expressed as

$$X_D(r) = \mathcal{D}[X(r)], \quad (4)$$

where $X_D(r)$ is the denoised image. As indicated in Eq. (4), the objective is to suppress noise while preserving disease-sensitive boundaries and anatomical structures.

4.5.2 Bias-Field Correction

MRI intensity inhomogeneity can arise because of spatial variation in the magnetic and radio-frequency fields. A multiplicative bias-field model can be expressed as

$$X_D(r) = S(r)B(r) + \eta(r), \quad (5)$$

where $B(r)$ denotes the slowly varying bias field. Equation (5) shows that non-pathological intensity variations may alter the appearance of otherwise similar tissues at different spatial locations.

Once the bias field is estimated as $\hat{B}(r)$, the corrected intensity can be obtained using

$$X_B(r) = \frac{X_D(r)}{\hat{B}(r) + \epsilon}, \quad (6)$$

where ϵ is a small constant introduced to prevent numerical instability. Equation (6) reduces scanner-related intensity variations and produces a more homogeneous representation of tissue intensity.

4.5.3 Skull Stripping

Skull stripping removes non-brain regions such as the scalp, skull, and surrounding tissues. Let $M_i(r)$ denote the binary brain mask for subject i , defined as

$$M_i(r) = \begin{cases} 1, & r \in \Omega_{\text{brain}}, \\ 0, & r \notin \Omega_{\text{brain}}. \end{cases} \quad (7)$$

As defined in Eq. (7), voxels belonging to brain tissue are assigned a value of one, while all remaining regions are assigned zero. The brain-extracted MRI is subsequently obtained as

$$X_S(r) = X_B(r) \odot M_i(r), \quad (8)$$

where $\odot$ denotes voxel-wise multiplication. Equation (8) ensures that subsequent feature extraction focuses predominantly on intracranial anatomical structures.

4.5.4 Intensity Normalization

Intensity normalization reduces variations in voxel intensity distributions caused by different scanners and acquisition protocols. Z-score normalization can be expressed as

$$X_N(r) = \frac{X_S(r) - \mu_i}{\sigma_i + \epsilon}, \quad (9)$$

where μ_i and σ_i denote the mean and standard deviation of brain voxel intensities, respectively. These statistics are calculated as

$$\mu_i = \frac{1}{|\Omega_{\text{brain}}|} \sum_{r \in \Omega_{\text{brain}}} X_S(r), \quad (10)$$

$$\sigma_i = \sqrt{\frac{1}{|\Omega_{\text{brain}}|} \sum_{r \in \Omega_{\text{brain}}} (X_S(r) - \mu_i)^2}. \quad (11)$$

Equations (9)–(11) standardize each MRI volume relative to its brain-tissue intensity distribution, thereby improving comparability across subjects. Alternatively, min-max normalization may be represented as

$$X_N(r) = \frac{X_S(r) - X_{\min}}{X_{\max} - X_{\min} + \epsilon}, \quad (12)$$

where $X_{\min}$ and $X_{\max}$ represent the minimum and maximum voxel intensities, respectively. Equation (12) rescales the voxel intensities to a common numerical interval.

4.5.5 Spatial Registration

Spatial registration aligns individual MRI volumes with a common anatomical template. Let

$$r = [x, y, z, 1]^T$$

represent a voxel coordinate in homogeneous form. The affine transformation is defined as

$$r' = Ar, \quad (13)$$

where A is the transformation matrix given by

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_x \\ a_{21} & a_{22} & a_{23} & t_y \\ a_{31} & a_{32} & a_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (14)$$

Equation (13) maps the original subject coordinates to the target anatomical space, while Eq. (14) incorporates translation, rotation, scaling, and shearing parameters.

The registered image is subsequently represented as

$$X_R(r') = X_N(T^{-1}(r')), \quad (15)$$

where $T(\cdot)$ denotes the spatial transformation. Equation (15) ensures that anatomically corresponding regions are represented at approximately consistent spatial locations across subjects.

4.5.6 Resampling

Following registration, MRI scans may still contain different voxel spacings. Let the original voxel spacing be

$$s = (s_x, s_y, s_z)$$

and the target spacing be

$$s' = (s'_x, s'_y, s'_z).$$

The dimensions of the resampled volume can be approximated as

$$H' = H \frac{s_x}{s'_x}, W' = W \frac{s_y}{s'_y}, D' = D \frac{s_z}{s'_z}. \quad (16)$$

Accordingly, the standardized MRI volume becomes

$$X_R \in \mathbb{R}^{H' \times W' \times D'}. \quad (17)$$

Equations (16) and (17) illustrate how resampling produces a consistent spatial resolution and input dimension across scans.

4.6 Tissue Segmentation and ROI Extraction

Tissue segmentation separates the MRI volume into clinically meaningful tissue categories. A generic segmentation function can be represented as

$$L(r) = \mathcal{G}(X_R(r)), \quad (18)$$

where

$$L(r) \in \{\text{GM}, \text{WM}, \text{CSF}\}.$$

Equation (18) assigns each voxel to grey matter, white matter, or cerebrospinal fluid. For tissue class c , the corresponding binary mask is defined as

$$M_c(r) = \mathbb{I}[L(r) = c], \quad (19)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

The physical volume of tissue c can then be calculated as

$$V_c = N_c v_x v_y v_z, \quad (20)$$

where N_c is the number of voxels belonging to tissue class c , while $v_x v_y v_z$ denotes the physical volume represented by each voxel. Equation (20) enables quantitative estimation of tissue loss or enlargement.

For ROI-based analysis, let M_{ROI} represent an anatomical mask corresponding to a disease-sensitive region such as the hippocampus. The extracted region is obtained as

$$X_{\text{ROI}} = X_R \odot M_{\text{ROI}}, \quad (21)$$

while its physical volume is computed as

$$V_{\text{ROI}} = \sum_r M_{\text{ROI}}(r) v_x v_y v_z. \quad (22)$$

Equations (21) and (22) illustrate how ROI extraction can focus the analysis on specific anatomical structures and support quantitative assessment of atrophy.

4.7 Data Augmentation

Data augmentation increases the diversity of the training data and can reduce overfitting. Let T_k represent the k^{th} transformation sampled from an augmentation set. The augmented image is defined as

$$X_i^{(k)} = T_k(\tilde{X}_i). \quad (23)$$

Equation (23) provides a generic formulation applicable to spatial and intensity-based transformations. For example, controlled geometric augmentation can be represented as

$$r' = R(\theta)r + t, \quad (24)$$

where $R(\theta)$ denotes the rotation matrix and t denotes translation.

Intensity augmentation can be expressed as

$$X_{\text{aug}}(r) = \alpha X(r) + \beta, \quad (25)$$

where α controls intensity scaling and β controls intensity shifting. As indicated by Eqs. (24) and (25), augmentation should introduce controlled variability while preserving clinically meaningful anatomical structures.

4.8 Class-Imbalance Handling

Class imbalance is frequently encountered in Alzheimer's datasets because the number of CN, MCI, and AD subjects may differ substantially. Let N_c represent the number of training subjects in class c , N the total number of subjects, and C the number of diagnostic categories. An inverse-frequency class weight can be defined as

$$w_c = \frac{N}{CN_c}. \quad (26)$$

As shown in Eq. (26), minority classes receive larger optimization weights than majority classes. The corresponding weighted cross-entropy loss can be expressed as

$$\mathcal{L}_{WCE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C w_c y_{ic} \log \hat{p}_{ic}, \quad (27)$$

where y_{ic} is the ground-truth class indicator and $\hat{p}_{ic}$ represents the predicted probability for class c . Equation (27) reduces majority-class dominance during training and is particularly useful when early-stage or progressive disease categories are underrepresented.

4.9 Summary of the Preprocessing Pipeline

Using the formulations in Eqs. (1)–(27), a generalized Alzheimer's neuroimaging preprocessing workflow can be represented as

$$X_i \xrightarrow{\mathcal{D}} X_D \xrightarrow{\mathcal{B}} X_B \xrightarrow{\mathcal{S}} X_S \xrightarrow{\mathcal{N}} X_N \xrightarrow{\mathcal{R}} X_R \xrightarrow{\text{ROI/Segmentation}} \tilde{X}_i. \quad (28)$$

Equation (28) summarizes the transformation of a raw volumetric MRI scan into a standardized and anatomically relevant input suitable for subsequent machine-learning or deep-learning analysis. The specific combination of preprocessing operations varies across studies depending on the imaging modality, dataset, and computational architecture.

Table 4. Notations Used in Neuroimaging Preprocessing

Symbol	Description
X_i	Raw MRI volume of subject i
$\tilde{X}_i$	Final preprocessed MRI volume
H, W, D	Height, width, and depth of MRI volume
r	Three-dimensional voxel coordinate

$S(r)$	Underlying anatomical signal
$\eta(r)$	Acquisition noise
$B(r)$	MRI bias field
$\hat{B}(r)$	Estimated bias field
M_i	Brain mask
μ_i, σ_i	Mean and standard deviation of brain voxel intensities
A	Affine registration matrix
$T(\cdot)$	Spatial transformation
M_c	Binary mask of tissue class c
M_{ROI}	Region-of-interest mask
V_{ROI}	Physical volume of selected ROI
T_k	Data-augmentation transformation
N_c	Number of samples in class c
w_c	Class-balancing weight
y_{ic}	Ground-truth class indicator
$\hat{p}_{ic}$	Predicted probability for class

4.6 ROI Extraction and Segmentation

Region-based preprocessing focuses the analysis on brain structures strongly associated with Alzheimer's disease, such as the hippocampus, entorhinal cortex, medial temporal lobe, ventricles, grey matter, and white matter. ROI extraction can reduce computational complexity and improve sensitivity to localized pathological changes. Whole-brain analysis, however, may capture distributed abnormalities that ROI-only methods could overlook. Segmentation is also commonly applied to separate grey matter, white matter, and cerebrospinal fluid or to identify specific anatomical structures. Accurate segmentation can support volumetric biomarker extraction, although segmentation errors may subsequently propagate into the diagnostic model.

4.7 Data Augmentation and Class Imbalance

The limited size of labelled neuroimaging datasets frequently increases the risk of overfitting. Data augmentation techniques such as rotation, translation, flipping, cropping, affine transformation, and controlled intensity variation may therefore be applied to increase training diversity. In three-dimensional MRI analysis, augmentation should preserve anatomically meaningful structures and avoid unrealistic geometric transformations. Class imbalance is another common issue because the number of cognitively normal, MCI, and AD subjects may differ substantially. Strategies such as class weighting, balanced sampling, focal loss, oversampling, and augmentation of minority classes can be used to reduce prediction bias toward dominant categories.

Table 4. Common Preprocessing Techniques for Alzheimer's Neuroimaging

Preprocessing Technique	Main Purpose	Expected Benefit
Bias-field correction	Remove intensity inhomogeneity	Reduces scanner-related intensity variation
Denoising	Suppress acquisition noise	Improves image quality
Skull stripping	Remove non-brain tissue	Focuses analysis on brain structures
Registration	Align scans to a common anatomical space	Improves inter-subject correspondence
Intensity normalization	Standardize intensity distributions	Reduces acquisition variability
Resampling	Standardize voxel size and image dimensions	Provides consistent model inputs
Tissue segmentation	Separate GM, WM and CSF	Supports anatomical and volumetric analysis
ROI extraction	Isolate disease-sensitive regions	Reduces irrelevant information
Data augmentation	Increase training variability	Helps reduce overfitting
Class balancing	Address unequal diagnostic groups	Reduces majority-class prediction bias

As indicated in **Table 4**, preprocessing is not merely an image-cleaning stage but a major component of the overall learning pipeline. Inconsistent preprocessing can significantly influence reported model performance and make comparisons between studies unreliable.

4.8 Data Partitioning and Leakage Prevention

Another critical issue is the separation of subjects during training, validation, and testing. Alzheimer's datasets frequently contain multiple scans from the same individual collected at different visits. If different scans belonging to the same patient are distributed across training and testing sets, the model may indirectly recognize subject-specific characteristics, resulting in overly optimistic performance. Therefore, subject-level splitting should be preferred over image-level random splitting. Longitudinal observations from one patient should remain within the same data partition. Cross-validation should similarly be conducted at the subject level. For stronger evidence of generalization, models should additionally be evaluated on an independent dataset such as training on ADNI and testing on OASIS or AIBL.

5. EVOLUTION OF AI-BASED METHODS FOR ALZHEIMER'S DISEASE DIAGNOSIS

Artificial Intelligence approaches for Alzheimer's disease diagnosis have progressed from conventional machine-learning models based on predefined neuroimaging biomarkers to deep architectures capable of automatically learning spatial, volumetric, multimodal, relational, and long-range representations. The major methodological developments include conventional machine learning, 2D and 3D Convolutional Neural Networks (CNNs), multimodal deep learning, contrastive and metric learning, Graph Neural Networks (GNNs), Siamese architectures, and Transformer-based models [19]. The following subsections review representative studies from each methodological category and highlight how these techniques were applied and the diagnostic results obtained.

5.1 Conventional Machine-Learning Methods

Conventional machine-learning approaches generally involve two stages: extraction of disease-relevant features from neuroimaging or clinical measurements, followed by classification using a supervised learning algorithm. Let X_i denote the neuroimaging information of subject i . Feature extraction can be represented as

$$z_i = \phi(X_i), \quad (29)$$

where $\phi(\cdot)$ denotes the feature extraction operation and z_i represents the resulting feature vector. The diagnostic class is subsequently determined as

$$\hat{y}_i = f(z_i), \quad (30)$$

where $f(\cdot)$ represents a machine-learning classifier.

Moradi *et al.* [20] proposed an MRI-based machine-learning framework for predicting conversion from Mild Cognitive Impairment (MCI) to Alzheimer's disease. Their approach generated an MRI biomarker using feature selection and low-density separation and subsequently combined it with age and cognitive measurements using a Random Forest classifier. Using ADNI data and 10-fold cross-validation, the MRI-only biomarker achieved an AUC of 0.766. When MRI, age, and cognitive information were integrated, the AUC increased to 0.902, with approximately 82% accuracy, 87% sensitivity, and 74% specificity for distinguishing progressive MCI from stable MCI. The results demonstrated that conventional machine-learning methods can provide useful progression prediction when appropriate imaging and cognitive biomarkers are incorporated. Despite their comparatively low computational requirements and relatively interpretable features, conventional methods depend strongly on feature engineering and prior anatomical knowledge. Their performance may therefore deteriorate when handcrafted descriptors fail to capture complex nonlinear neurodegenerative patterns.

5.2 Two-Dimensional CNN Architectures

CNNs introduced automatic hierarchical representation learning and substantially reduced dependence on handcrafted features. In a conventional 2D CNN, the feature map generated at layer l can be expressed as

$$F^{(l)} = \sigma(W^{(l)} * F^{(l-1)} + b^{(l)}), \quad (31)$$

where $W^{(l)}$ and $b^{(l)}$ denote the convolutional kernel and bias, $*$ represents convolution, and $\sigma(\cdot)$ represents a nonlinear activation function. Lower CNN layers generally capture edges and local tissue patterns, whereas deeper layers learn higher-level disease-related representations.

Farooq *et al.* [21] investigated GoogLeNet, ResNet-18, and ResNet-152 for multiclass Alzheimer's disease classification using MRI data. The study considered four diagnostic categories, including AD, MCI, Late MCI (LMCI), and cognitively normal subjects. Their reported classification accuracies were 98.88% using GoogLeNet, 98.01% using ResNet-18, and 98.14% using ResNet-152. These findings demonstrated the potential of deep convolutional representation learning and transfer learning for disease-stage classification.

5.3 Three-Dimensional CNN Architectures

Three-dimensional CNNs directly process complete or cropped MRI volumes and preserve inter-slice anatomical relationships. A 3D convolution can be represented as

$$F_{x,y,z}^{(l)} = \sigma \left[\sum_{u,v,w} W_{u,v,w}^{(l)} F_{x-u,y-v,z-w}^{(l-1)} + b^{(l)} \right]. \quad (32)$$

Unlike the 2D operation in Eq. (31), Eq. (32) performs convolution across the depth, height, and width of the MRI volume. Consequently, 3D CNNs can model volumetric abnormalities including hippocampal shrinkage, cortical atrophy, and ventricular enlargement.

Hosseini-Asl *et al.* [22] developed a 3D deeply supervised adaptable CNN in which a three-dimensional convolutional autoencoder was first pretrained to capture anatomical representations from structural MRI. The resulting network was subsequently fine-tuned for Alzheimer's disease classification. The approach achieved approximately 99.3% accuracy for AD-versus-healthy-control classification and 94.8% accuracy for three-class AD/MCI/healthy-control classification. Basaia *et al.* [23] subsequently applied a 3D CNN directly to T1-weighted MRI volumes for both AD diagnosis and MCI conversion prediction. Their model achieved 99.2% accuracy, 98.9% sensitivity, and 99.5% specificity for AD-versus-healthy-control classification. However, accuracy decreased to approximately 75% for converting-MCI versus stable-MCI classification. This difference illustrates that high performance on clearly separated AD and control groups does not necessarily indicate equivalent capability for identifying subtle prodromal disease progression. The major limitation of 3D CNNs is their computational burden. Processing high-resolution volumetric MRI requires substantially more memory and floating-point operations than 2D methods. Consequently, recent 3D architectures increasingly employ ROI selection, residual learning, dilated convolution, lightweight networks, and attention mechanisms.

5.4 Multimodal Deep Learning

Alzheimer's disease involves structural, functional, metabolic, cognitive, and molecular alterations. Therefore, combining multiple modalities can provide a more complete representation than using structural MRI alone. Let $z_i^{(m)}$ denote the representation obtained from modality m . Conventional feature-level fusion can be expressed as

$$z_i^F = z_i^{(1)} \oplus z_i^{(2)} \oplus \dots \oplus z_i^{(M)}, \quad (33)$$

where $\oplus$ denotes concatenation and M is the total number of modalities.

Attention-based or adaptive multimodal fusion may instead be formulated as

$$z_i^F = \sum_{m=1}^M \alpha_i^{(m)} z_i^{(m)}, \quad \sum_{m=1}^M \alpha_i^{(m)} = 1, \quad (34)$$

where $\alpha_i^{(m)}$ represents the contribution assigned to modality m .

Zhang *et al.* [24] investigated multimodal classification using MRI, PET, and cerebrospinal-fluid biomarkers. Combining the three modalities achieved approximately 93.2% accuracy, 93.0% sensitivity, and 93.3% specificity for AD-versus-healthy-control classification, outperforming individual modalities and demonstrating the complementary value of structural, metabolic, and biological information. Suk *et al.* [25] introduced hierarchical feature representation and multimodal fusion using Deep Boltzmann Machines for MRI and PET. Modality-specific latent representations were learned and subsequently integrated for

classification. Their approach achieved 95.35% accuracy for AD versus normal controls, 85.67% for MCI versus normal controls, and approximately 75.92% for MCI converters versus MCI non-converters. These studies established the importance of multimodal learning; however, modality availability remains a practical challenge. PET and CSF measurements may not be available for every patient, and different modalities can have heterogeneous dimensions and statistical distributions. Recent studies therefore increasingly emphasize attention-based fusion, missing-modality learning, and cross-modal representation alignment.

5.5 Contrastive and Metric Learning

Contrastive learning aims to structure the representation space so that similar subjects or corresponding multimodal observations remain close while unrelated representations become separated. A generalized contrastive objective can be expressed as

$$\mathcal{L}_{con} = -\log \frac{\exp(\text{sim}(z_i, z_i^+)/\tau)}{\sum_j \exp(\text{sim}(z_i, z_j)/\tau)}, \quad (35)$$

where z_i^+ denotes a positive representation, $\text{sim}(\cdot)$ is a similarity measure, and τ denotes the temperature coefficient. According to Eq. (35), representations from related samples are encouraged to exhibit greater similarity than unrelated samples.

Qiao *et al.* [26] proposed a contrastive-learning framework using paired structural MRI. Their approach introduced a Group Category contrastive objective together with an MMSE-guided ranking mechanism to capture subtle disease-stage differences. The model achieved 93.5% accuracy, 96.6% sensitivity, and 94.9% specificity on the MIRIAD dataset. Kwak *et al.* [27] employed self-supervised contrastive learning for MCI progression prediction using 3D amyloid-PET. Their SMoCo framework utilized both labelled and unlabelled data and achieved an AUROC of 85.17%, accuracy of 81.09%, sensitivity of 77.39%, and specificity of 82.17% for identifying MCI converters. Metric-learning techniques follow a similar principle but explicitly optimize distances among class representations. Triplet loss is commonly formulated as

$$\mathcal{L}_{triplet} = \max[0, d(z_a, z_p) - d(z_a, z_n) + m], \quad (36)$$

where z_a , z_p , and z_n denote anchor, positive, and negative embeddings, $d(\cdot)$ denotes a distance function, and m represents the margin.

Cárdenas-Peña *et al.* [28] investigated kernel metric learning using morphological MRI measurements extracted with FreeSurfer. The learned metric was evaluated with k-NN, SVM, and neural-network classifiers. Their best configuration achieved approximately 92.6% accuracy for healthy-control versus AD classification. Contrastive and metric learning are particularly relevant for early-stage Alzheimer's diagnosis because neighbouring categories frequently exhibit highly overlapping structural characteristics.

5.6 Graph Neural Networks

Conventional CNN models generally analyse patients independently. Graph Neural Networks provide an alternative paradigm in which relationships among subjects are explicitly represented. A patient-population graph can be defined as

$$G = (V, E), \quad (37)$$

where V denotes the set of patient nodes and E represents relationships based on imaging, demographic, cognitive, or genetic similarity.

A generalized graph propagation operation can be represented as

$$h_i^{(l+1)} = \sigma \left(W^{(l)} \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} h_j^{(l)} \right), \quad (38)$$

where $\mathcal{N}(i)$ denotes neighbouring subjects, α_{ij} represents the contribution of neighbour j , and $h_i^{(l)}$ represents the feature representation of subject i .

Parisot *et al.* [29] introduced a population-based Graph Convolutional Network in which subjects were represented as nodes, while phenotypic similarities were incorporated through graph edges. When evaluated using ADNI for Alzheimer's-related prediction, the proposed GCN obtained approximately 80.0% classification accuracy, demonstrating that population-level relationships can complement individual imaging features. The main challenge associated with GNN approaches is graph construction. If connections between subjects are based on inappropriate or noisy similarity criteria, message propagation may introduce irrelevant information and reduce generalization.

5.7 Siamese Learning Approaches

Siamese networks typically contain two parallel branches that map different samples into a shared latent space. For two neuroimaging inputs X_i and X_j , the latent embeddings are represented as

$$z_i = f_\theta(X_i), z_j = f_\theta(X_j), \quad (39)$$

where $f_\theta(\cdot)$ denotes the shared encoder. The similarity between the representations may then be measured using Euclidean distance,

$$d_{ij} = \|z_i - z_j\|_2. \quad (40)$$

Liu *et al.* [30] employed a deep Siamese neural network to analyse inter-hemispheric brain asymmetry associated with Alzheimer's disease and MCI. Paired left-right anatomical structures were represented within the Siamese framework. On ADNI, the method achieved 89.92% sensitivity, 95.23% specificity, 92.72% balanced accuracy, and an F1-score of 93.72%. The approach demonstrated that similarity-based learning can effectively characterize asymmetric structural alterations associated with neurodegeneration. Siamese methods are particularly attractive for limited-data environments because pair generation produces multiple training relationships from a relatively small number of subjects. Nevertheless, their performance depends on the quality of positive and negative pair construction.

5.8 Transformer-Based Models

Transformer architectures have recently become increasingly important in neuroimaging because self-attention can capture long-range dependencies among spatially separated brain regions. For a token representation Z , query, key, and value matrices are computed as

$$Q = ZW_Q, K = ZW_K, V = ZW_V, \quad (41)$$

where W_Q , W_K , and W_V are learnable projection matrices. Scaled dot-product attention is then calculated as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (42)$$

where d_k denotes the dimensionality of the key representation. Equation (42) enables direct interaction between spatially distant image regions and is therefore suitable for modelling distributed neurodegenerative abnormalities.

Hoang *et al.* [31] applied Vision Transformers to predict progression from MCI to AD using mid-sagittal structural MRI. The study involved 598 MCI subjects from ADNI, including converter and non-converter groups. Their model achieved 83.27% accuracy, 81.48% sensitivity, and 85.07% specificity. Attention visualization further identified brain regions that contributed to progression prediction, demonstrating the potential interpretability of Transformer attention. Alp *et al.* [32] proposed a joint Transformer architecture in which a pretrained Vision Transformer extracted MRI slice features, followed by another Transformer for modelling inter-slice dependencies. Their approach reported 99.048% accuracy for binary classification and 99.014% accuracy for multiclass classification using ADNI data. However, comparisons with other studies should consider differences in dataset partitioning, sample composition, and validation strategy.

Transformer models provide strong global-context modelling, but global self-attention can become computationally expensive for complete 3D MRI volumes. Consequently, recent research has moved toward hierarchical Transformers, Swin Transformers, CNN-Transformer hybrids, and local-window attention mechanisms.

5.9 Comparative Analysis of Representative Studies

The representative methods discussed in Sections 5.1–5.8 are summarized in **Table 5**. The table highlights considerable variation in diagnostic tasks and evaluation protocols. Therefore, the reported performance values should be interpreted within their respective experimental settings rather than treated as a direct ranking.

Table 5. Representative AI Techniques for Alzheimer’s Disease Diagnosis

Study	Technique	Dataset / Modality	Diagnostic Task	Major Methodological Feature	Reported Performance
Moradi <i>et al.</i> [20]	Conventional ML	ADNI, MRI + cognitive	pMCI vs sMCI	MRI biomarker + Random Forest	Acc. 82%, AUC 0.902
Farooq <i>et al.</i> [21]	CNN	ADNI, MRI	AD/MCI/LMCI/CN	GoogLeNet/ResNet	Best Acc. 98.88%
Hosseini-Asl <i>et al.</i> [22]	3D CNN	ADNI, sMRI	AD/MCI/HC	3D autoencoder + supervised CNN	94.8% three-class Acc.
Basaia <i>et al.</i> [23]	3D CNN	ADNI, sMRI	AD vs HC	Whole-volume 3D CNN	Acc. 99.2%
Basaia <i>et al.</i> [23]	3D CNN	ADNI, sMRI	cMCI vs sMCI	Whole-volume 3D CNN	Acc. ~ 75%
Zhang <i>et al.</i> [24]	Multimodal ML	ADNI, MRI + PET + CSF	AD vs HC	Multimodal biomarker fusion	Acc. 93.2%
Suk <i>et al.</i> [25]	Multimodal DL	ADNI, MRI + PET	AD vs NC	Deep Boltzmann multimodal fusion	Acc. 95.35%
Qiao <i>et al.</i> [26]	Contrastive learning	ADNI/MIRIAD, sMRI	Early AD classification	Category + MMSE contrastive learning	Acc. 93.5%
Kwak <i>et al.</i> [27]	Self-supervised contrastive	ADNI, amyloid-PET	MCI conversion	SMoCo representation learning	Acc. 81.09%, AUC 0.8517
Cárdenas-Peña <i>et al.</i> [28]	Metric learning	ADNI, MRI	HC vs AD	Kernel metric learning	Acc. 92.6%
Parisot <i>et al.</i> [29]	GCN	ADNI, imaging + phenotypic	AD-related prediction	Population graph learning	Acc. 80.0%
Liu <i>et al.</i> [30]	Siamese network	ADNI/BIOCARD, MRI	AD/MCI-related diagnosis	Inter-hemispheric comparison	BACC 92.72%
Hoang <i>et al.</i> [31]	Vision Transformer	ADNI, sMRI	MCI conversion	ViT + attention interpretation	Acc. 83.27%
Alp <i>et al.</i> [32]	ViT + Transformer	ADNI, MRI	Binary/multiclass AD	ViT + inter-slice Transformer	99.048% / 99.014%

As shown in **Table 5**, binary AD-versus-healthy-control classification generally produces considerably higher performance than progression-oriented tasks such as sMCI-versus-pMCI or converter-versus-nonconverter prediction. For example, Basaia *et al.*

[23] achieved 99.2% accuracy for AD-versus-control classification but approximately 75% for MCI conversion prediction. Similarly, multimodal methods reported by Suk *et al.* [25] showed stronger performance for AD-versus-normal classification than for MCI converter discrimination. These findings suggest that early progression prediction remains a substantially more difficult problem than established-disease classification.

Table 6. Comparative Characteristics of Major AI Methodologies

Method	Representation	Main Advantage	Major Limitation	Appropriate AD Application
Conventional ML	Handcrafted biomarkers	Low computational cost and greater feature interpretability	Dependence on feature engineering	Baseline diagnosis and biomarker analysis
2D CNN	MRI slices	Efficient automated feature extraction	Loss of inter-slice information	Slice-based classification
3D CNN	Complete MRI volume	Preserves volumetric anatomy	High memory and computation	Whole-brain diagnosis and staging
Multimodal Learning	Imaging + non-imaging features	Captures complementary disease information	Missing modalities and heterogeneous data	Comprehensive AD assessment
Contrastive Learning	Similarity-based latent space	Effective with limited labels and subtle classes	Depends on positive/negative definitions	Early-stage and multimodal learning
Metric Learning	Distance-optimized embedding	Improves class separation	Sensitive to sampling and loss parameters	Fine-grained MCI discrimination
GNN	Patient population graph	Models inter-subject relationships	Sensitive to graph construction	Population-level disease modelling
Siamese Network	Paired embeddings	Efficient similarity learning	Requires appropriate pair generation	Early-stage/few-shot diagnosis
Vision Transformer	Image patches/tokens	Captures long-range dependencies	High computational demand	Global neuroanatomical modelling
3D/Hierarchical Transformer	Volumetric tokens	Integrates local and global 3D information	Large memory and architectural complexity	Advanced volumetric AD diagnosis

As summarized in **Table 6**, no individual architectural family is universally superior. Conventional machine-learning approaches provide interpretable baselines but depend heavily on manual feature design. CNNs automate feature extraction, whereas 3D CNNs improve volumetric representation at increased computational cost. Multimodal methods exploit complementary biomarkers, while contrastive and metric learning improve feature-space discrimination. GNNs extend diagnosis toward population-level relational modelling, Siamese architectures focus on pairwise similarity, and Transformers enable long-range spatial reasoning.

6. CRITICAL EVALUATION, RESEARCH GAPS, AND FUTURE DIRECTIONS

Although artificial-intelligence models have achieved substantial progress in Alzheimer's disease (AD) diagnosis, high classification accuracy alone is insufficient for clinical translation. Reliable systems must additionally provide interpretable decisions, remain robust across populations and acquisition environments, model disease progression over time, and operate

within practical computational constraints. Therefore, this section jointly examines explainability, robustness and generalization, longitudinal progression modelling, computational efficiency, unresolved research gaps, and future research directions.

6.1 Explainability and Clinical Interpretability

The increasing use of deep CNNs, graph networks, and Transformers has improved automated AD classification but has simultaneously increased the complexity of interpreting model decisions. In clinical applications, neurologists and radiologists require evidence indicating why a particular subject has been classified as cognitively normal, MCI, or AD. Consequently, explainable artificial intelligence has become an important component of neuroimaging-based diagnostic systems. Commonly employed approaches include Gradient-weighted Class Activation Mapping (Grad-CAM), saliency maps, SHAP, LIME, attention visualization, and occlusion analysis. These techniques identify image regions or clinical features that contribute strongly to model predictions. For MRI-based AD diagnosis, interpretable systems frequently highlight regions such as the hippocampus, medial temporal lobe, entorhinal cortex, ventricles, and cortical structures associated with neurodegeneration.

For a CNN feature map A^k , the Grad-CAM importance coefficient for class c can be expressed as

$$\alpha_k^c = \frac{1}{Z} \sum_x \sum_y \sum_z \frac{\partial y^c}{\partial A_{xyz}^k}, \quad (43)$$

where y^c represents the prediction score for class c , A^k denotes the k^{th} feature map, and Z represents the number of spatial locations. The resulting activation map is

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right). \quad (44)$$

As indicated by Eqs. (43) and (44), Grad-CAM highlights spatial regions contributing most strongly to a diagnostic decision. However, visualization alone does not guarantee clinically valid reasoning. Future studies should therefore evaluate whether highlighted regions correspond consistently with established neurological biomarkers rather than merely providing visually appealing heat maps.

6.2 Robustness, Generalization, and Reproducibility

A major challenge in Alzheimer's disease AI research is the reduction in performance when a model trained on one dataset is applied to subjects collected at another institution. Differences in MRI scanners, field strengths, acquisition parameters, reconstruction procedures, demographic characteristics, disease prevalence, and preprocessing pipelines can produce substantial domain shift. Let $P_s(X, Y)$ and $P_t(X, Y)$ denote the source- and target-domain distributions, respectively. Cross-dataset variability can generally be expressed as

$$P_s(X, Y) \neq P_t(X, Y). \quad (45)$$

Equation (45) explains why high internal validation accuracy does not necessarily imply reliable external performance. Domain adaptation, domain generalization, intensity harmonization, scanner-invariant feature learning, and multicentre training are therefore increasingly important.

Robust evaluation should additionally investigate sensitivity to noise, motion artefacts, incomplete scans, image-quality degradation, and missing clinical variables. Repeated random seeds, subject-level cross-validation, confidence intervals, and external dataset validation should be reported whenever possible. In particular, repeated scans belonging to the same subject must not be distributed across training and testing partitions, since such leakage can substantially inflate reported diagnostic performance.

6.3 Longitudinal Disease-Progression Modelling

Most early studies treated Alzheimer's diagnosis as a static classification problem. However, AD is inherently progressive, and the clinically important question is often not simply whether a patient currently exhibits MCI, but whether that individual is likely to progress toward AD.

For subject i , a longitudinal sequence can be represented as

$$\mathcal{X}_i = \{X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(T)}\}, \quad (46)$$

where $X_i^{(t)}$ represents the neuroimaging or clinical observations acquired at time point t . A progression model estimates

$$\hat{y}_i^{(T+\Delta)} = f_\theta(X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(T)}), \quad (47)$$

where $\hat{y}_i^{(T+\Delta)}$ denotes the predicted future disease state.

Equation (47) shifts the objective from cross-sectional diagnosis toward prediction of future clinical status. Recurrent Neural Networks, Long Short-Term Memory networks, temporal CNNs, Transformers, survival models, and longitudinal graph architectures have been explored for this purpose. A particularly important task is discrimination between stable MCI (sMCI) and progressive MCI (pMCI). Previous studies consistently show that this task is considerably more difficult than AD-versus-control classification because structural differences between sMCI and pMCI can be subtle. Consequently, longitudinal modelling that combines repeated MRI with cognitive scores, PET biomarkers, and clinical trajectories represents an important direction for early intervention.

6.4 Computational Efficiency and Clinical Deployment

Three-dimensional CNNs, GNNs, and Transformers can contain millions of parameters and require substantial GPU memory and computation. Such resource requirements may restrict their application in hospitals with limited computational infrastructure. Model complexity can be characterized using parameter count,

$$P = \sum_{l=1}^L P_l, \quad (48)$$

where P_l denotes the number of trainable parameters in layer l . Computational performance should additionally be evaluated using FLOPs, inference latency, GPU or CPU memory requirements, and throughput. For clinical deployment, diagnostic performance should therefore be considered jointly with computational cost rather than in isolation. Model compression, pruning, quantization, knowledge distillation, lightweight CNNs, efficient attention mechanisms, sparse Transformers, and ROI-guided processing provide potential strategies for reducing complexity.

A useful evaluation criterion can conceptually be represented as

$$\mathcal{U} = \frac{\mathcal{P}_{\text{diagnostic}}}{\mathcal{C}_{\text{computational}}}, \quad (49)$$

where $\mathcal{P}_{\text{diagnostic}}$ represents diagnostic performance and $\mathcal{C}_{\text{computational}}$ represents computational cost. Equation (49) emphasizes that a marginal improvement in accuracy may not justify a substantial increase in computational complexity.

6.5 Major Research Gaps

Despite extensive research, several important limitations remain unresolved. First, many studies continue to evaluate models primarily using ADNI, resulting in limited evidence of generalization to independent hospitals and geographically diverse populations. Second, extremely high results are frequently reported for relatively simple AD-versus-normal-control classification, whereas clinically important tasks such as CN-to-MCI progression and sMCI-versus-pMCI prediction remain substantially more difficult. Third, multimodal approaches often assume that MRI, PET, CSF, genetic, and cognitive information are simultaneously available for every subject. This assumption does not accurately reflect clinical environments in which modalities are frequently missing. Fourth, current models commonly analyse cross-sectional scans and therefore underutilize longitudinal disease trajectories. Fifth, considerable variation exists in preprocessing, subject selection, class definitions, train-test partitioning, augmentation, and validation protocols, making direct comparison between published accuracies difficult. Sixth, explainability is frequently presented qualitatively without rigorous assessment of whether identified regions correspond to clinically established biomarkers. Seventh, uncertainty estimation and calibration remain less frequently investigated than conventional accuracy and AUC.

Finally, the increasing architectural complexity of multimodal Transformers and graph-based networks introduces challenges relating to computational requirements, reproducibility, and real-time clinical deployment.

6.6 Future Research Directions

Future Alzheimer's disease AI research should increasingly shift from maximizing single-dataset classification accuracy toward developing generalizable, clinically interpretable, longitudinal, multimodal, and computationally efficient systems. Several directions are particularly promising. Multicentre and cross-dataset learning should be emphasized to reduce scanner- and population-specific bias. Domain-generalization and harmonization approaches could enable models trained using ADNI, OASIS, AIBL, and other cohorts to retain performance when transferred to previously unseen hospitals. Longitudinal multimodal modelling represents another important direction. Instead of analysing a single MRI scan, future architectures could jointly model repeated neuroimaging, PET, cognitive scores, demographic factors, and biomarker trajectories to estimate individualized progression risk.

Self-supervised and contrastive learning are also likely to become increasingly important because large quantities of neuroimaging data are available without detailed diagnostic annotations. Pretraining models on unlabelled MRI volumes followed by task-specific fine-tuning may substantially reduce dependence on expensive expert-labelled datasets. Future multimodal systems should additionally support missing-modality learning. Rather than requiring all modalities during inference, adaptive architectures should dynamically exploit whichever clinical and imaging information is available for a particular patient. Graph neural networks may be extended toward dynamic and heterogeneous patient graphs that represent imaging similarity, disease stage, age, genetic risk, cognitive state, and longitudinal progression simultaneously. Similarly, 3D Transformers may increasingly incorporate anatomical priors and hierarchical attention so that global brain relationships can be captured without excessive computational cost. Another important direction concerns uncertainty-aware prediction. Instead of returning only a diagnostic label, future systems should estimate both disease probability and confidence,

$$\hat{y}_i, u_i = f_{\theta}(X_i), \quad (50)$$

where $\hat{y}_i$ represents the predicted diagnostic probabilities and u_i represents predictive uncertainty. As indicated in Eq. (50), uncertain cases could be referred to specialists rather than automatically assigned a definitive diagnosis. Explainable AI should similarly progress beyond qualitative heat maps toward quantitative validation against known neuroanatomical biomarkers. Attention regions, SHAP importance, and saliency patterns should be compared with established measures of hippocampal atrophy, cortical thinning, ventricular enlargement, and other disease-related abnormalities. Federated and privacy-preserving learning also offer an important opportunity for multicentre Alzheimer's research. Hospitals could collaboratively train models without transferring raw patient data, thereby increasing population diversity while maintaining data privacy.

7. PROPOSED FUTURE ARCHITECTURE FOR CLINICALLY RELIABLE AI-ASSISTED ALZHEIMER'S DISEASE DIAGNOSIS

Based on the methodological developments and research gaps identified in the preceding sections, this survey proposes a conceptual future architecture for clinically reliable Alzheimer's disease diagnosis and progression prediction. As illustrated in **Figure 1**, the proposed framework integrates multimodal neuroimaging, clinical and biological information, modality-specific representation learning, cross-modal contrastive alignment, structure-aware Transformer modelling, patient-level graph reasoning, longitudinal analysis, multi-task prediction, and explainable clinical decision support. The framework is intended as a **future research roadmap** rather than a single existing diagnostic model. Its principal objective is to move beyond static AD classification toward an integrated system capable of current-stage assessment, MCI conversion prediction, progression-time estimation, disease-severity assessment, uncertainty quantification, and clinically interpretable decision support.

Step 1: Acquisition of Multimodal Patient Data

The first stage of the proposed architecture involves the acquisition of complementary imaging and non-imaging information. As shown at the top of Fig. 1, the imaging branch receives 3D structural MRI, fMRI, and PET, while the parallel non-imaging branch incorporates clinical, cognitive, demographic, genetic, and biomarker information. Structural MRI provides information regarding hippocampal atrophy, cortical thinning, ventricular enlargement, and other morphological alterations associated with neurodegeneration. fMRI provides complementary information regarding functional connectivity and abnormalities in brain-network organization, whereas PET can capture metabolic changes as well as amyloid- and tau-related pathological processes. The complementary branch may incorporate variables such as age, sex, MMSE, CDR, neuropsychological assessments, APOE genotype, CSF biomarkers, blood biomarkers, and relevant clinical history. Combining these heterogeneous sources can provide a more comprehensive representation of disease status than any individual modality.

Step 2: Modality-Specific Foundation Encoders

In the second stage, each imaging modality is processed using a modality-specific foundation encoder, as illustrated in Fig. 1. Instead of forcing structurally different imaging modalities to share the same initial representation mechanism, separate encoders can learn modality-dependent characteristics. For example, a volumetric 3D CNN or 3D Transformer may be employed for structural MRI, whereas fMRI may additionally require connectivity-aware or graph-based representations. PET images can be processed using a separate volumetric encoder designed to capture metabolic or molecular uptake patterns. The output of modality m for subject i can generally be represented as

$$z_i^{(m)} = E_m(X_i^{(m)}), \quad (52)$$

where $X_i^{(m)}$ represents modality m , $E_m(\cdot)$ denotes its corresponding foundation encoder, and $z_i^{(m)}$ represents the learned modality-specific feature embedding. As indicated in Eq. (52), separate encoders allow the framework to preserve modality-specific disease information before multimodal integration.

Step 3: Encoding Clinical, Cognitive, Genetic, and Biomarker Information

In parallel with the imaging branch, structured and textual clinical information is processed using a tabular/language encoder, as shown in Fig. 1. Structured variables such as cognitive scores, demographic factors, genetic characteristics, and laboratory biomarkers can be processed using tabular neural networks or Transformer-based tabular encoders. Unstructured clinical notes, when available, can be processed using language-model-based encoders.

The resulting non-imaging representation can be written as

$$z_i^{(c)} = E_c(x_i^{(c)}), \quad (53)$$

where $x_i^{(c)}$ represents the clinical and biomarker variables and $E_c(\cdot)$ denotes the corresponding tabular or language encoder. Equation (53) transforms heterogeneous clinical information into a representation that can subsequently be aligned with the imaging embeddings.

Step 4: Cross-Modal Contrastive Alignment

Direct concatenation of MRI, PET, fMRI, and clinical representations may be insufficient because these modalities have substantially different feature distributions. Therefore, the third major processing block in Fig. 1 performs cross-modal contrastive alignment. The objective is to place representations belonging to the same patient closer within a shared semantic space while separating representations belonging to unrelated subjects. For paired representations $z_i^{(a)}$ and $z_i^{(b)}$, the cross-modal contrastive objective can be formulated as

$$\mathcal{L}_{CM} = -\log \frac{\exp(\text{sim}(z_i^{(a)}, z_i^{(b)})/\tau)}{\sum_j \exp(\text{sim}(z_i^{(a)}, z_j^{(b)})/\tau)}, \quad (54)$$

where $\text{sim}(\cdot)$ denotes a similarity measure and τ is the temperature parameter. Through Eq. (54), representations of corresponding imaging and clinical observations can be brought into closer agreement before multimodal fusion. Such alignment may also support learning when labelled patient data are limited.

Step 5: Structure-Aware Multimodal Transformer

Following cross-modal alignment, the shared features are supplied to a structure-aware multimodal Transformer, which forms the principal multimodal reasoning component of the proposed architecture in Fig. 1. The Transformer is intended to combine local anatomical information with long-range cross-modal relationships. Rather than treating all volumetric patches identically, structure-aware processing can introduce anatomical priors or region-specific tokens corresponding to clinically relevant areas such as the hippocampus, entorhinal cortex, medial temporal lobe, ventricles, and cortical regions. The fused representation may be defined as

$$Z_i^F = \mathcal{T}_{SA}(z_i^{MRI}, z_i^{fMRI}, z_i^{PET}, z_i^c), \quad (55)$$

where $\mathcal{T}_{SA}(\cdot)$ denotes the structure-aware multimodal Transformer. Equation (55) enables simultaneous modelling of within-modality patterns and cross-modal dependencies while retaining anatomically meaningful information.

Step 6: Dynamic Patient Graph Reasoning

The next stage introduces a dynamic patient graph, as illustrated in Fig. 1, to exploit relationships among patients. Each patient is treated as a graph node, while edges can represent similarity according to multimodal features, demographic characteristics, cognitive profiles, biomarkers, or disease-progression patterns. For time point t , the patient graph can be represented as

$$G^{(t)} = (V, E^{(t)}), \quad (56)$$

where V represents the patient population and $E^{(t)}$ represents time-dependent relationships among patients. Unlike a fixed population graph, the proposed dynamic representation allows patient similarities to evolve as disease status and biomarker profiles change.

The graph-enhanced representation for patient i may be calculated as

$$h_i^{(t+1)} = \sigma \left(W \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(t)} h_j^{(t)} \right), \quad (57)$$

where $\mathcal{N}(i)$ denotes neighbouring patients and $\alpha_{ij}^{(t)}$ indicates their dynamically estimated relationship strength. As shown in Eq. (57), the diagnostic representation of an individual can incorporate information from clinically or biologically similar patients.

Step 7: Longitudinal Temporal Transformer

The dynamic graph representation is subsequently processed by a longitudinal temporal Transformer. This stage is particularly important because Alzheimer's disease evolves over several years and cannot be represented completely using a single clinical visit. As shown in Fig. 1, repeated observations from visits $1, 2, \dots, T$ are provided to the temporal modelling component. For patient i , the longitudinal sequence can be expressed as

$$\mathcal{H}_i = [h_i^{(1)}, h_i^{(2)}, \dots, h_i^{(T)}]. \quad (58)$$

The temporal representation is then obtained as

$$z_i^L = \mathcal{T}_L(\mathcal{H}_i, \Delta t_i), \quad (59)$$

where $\mathcal{T}_L$ denotes the longitudinal Transformer and Δt_i represents the temporal intervals between patient visits. Inclusion of Δt_i is particularly important because clinical follow-up intervals are often irregular. Equations (58) and (59) therefore enable modelling of both temporal disease trajectory and variation in visit timing.

Step 8: Multi-Task Clinical Prediction

Rather than producing only a single AD classification label, the proposed architecture uses a multi-task output layer, as shown in Fig. 1. The shared longitudinal representation can simultaneously support several clinically relevant predictions:

- current AD stage, such as CN, MCI, or AD;
- probability of conversion from MCI to AD;
- estimated time-to-progression;
- continuous or categorical disease-severity assessment; and
- predictive uncertainty.

The combined multi-task learning objective can be expressed as

$$\mathcal{L}_{MT} = \lambda_1 \mathcal{L}_{stage} + \lambda_2 \mathcal{L}_{conversion} + \lambda_3 \mathcal{L}_{progression} + \lambda_4 \mathcal{L}_{severity} + \lambda_5 \mathcal{L}_{uncertainty}, \quad (60)$$

where $\lambda_1, \dots, \lambda_5$ control the relative importance of individual objectives. As indicated in Eq. (60), the model can jointly learn related diagnostic and prognostic tasks instead of training independent models for each clinical endpoint.

Step 9: Explainable Clinical Decision Layer

The final stage of the proposed roadmap is the explainable clinical layer, shown at the bottom of Fig. 1. This component converts complex model outputs into information that can be interpreted by clinicians. Anatomical heatmaps can identify brain regions contributing to the prediction, whereas biomarker-importance analysis can indicate which clinical, cognitive, or genetic variables influenced the result. Counterfactual explanations can indicate which changes in patient characteristics might alter the predicted risk category. Confidence and calibration measures can provide information regarding the reliability of each prediction, while longitudinal explanations can show how the estimated disease risk changes across successive patient visits.

The final clinical output can therefore be represented conceptually as

$$\mathcal{O}_i = \{\hat{y}_i, P_{conv,i}, T_{prog,i}, S_i, U_i, E_i\}, \quad (61)$$

where $\hat{y}_i$ represents the current disease stage, $P_{conv,i}$ is the MCI conversion probability, $T_{prog,i}$ is the estimated time-to-progression, S_i represents disease severity, U_i denotes predictive uncertainty, and E_i represents the associated explanation. Equation (61) emphasizes that future clinical AI systems should provide both predictions and interpretable supporting evidence.

7.1 Overall Research Roadmap

As summarized in Eq. (62) and Fig. 1, the proposed architecture combines the principal strengths of the methodological families discussed throughout this survey. Multimodal foundation encoders provide modality-specific representations, contrastive alignment establishes correspondence across heterogeneous sources, the structure-aware Transformer captures anatomical and cross-modal relationships, dynamic graph reasoning exploits patient-level similarities, and longitudinal modelling captures disease evolution. The multi-task and explainable layers subsequently translate these representations into clinically meaningful diagnostic and prognostic information. The proposed roadmap therefore emphasizes that future Alzheimer's disease AI research should move away from isolated improvements in classification accuracy and toward multimodal, longitudinal, generalizable, uncertainty-aware, explainable, and clinically deployable systems.

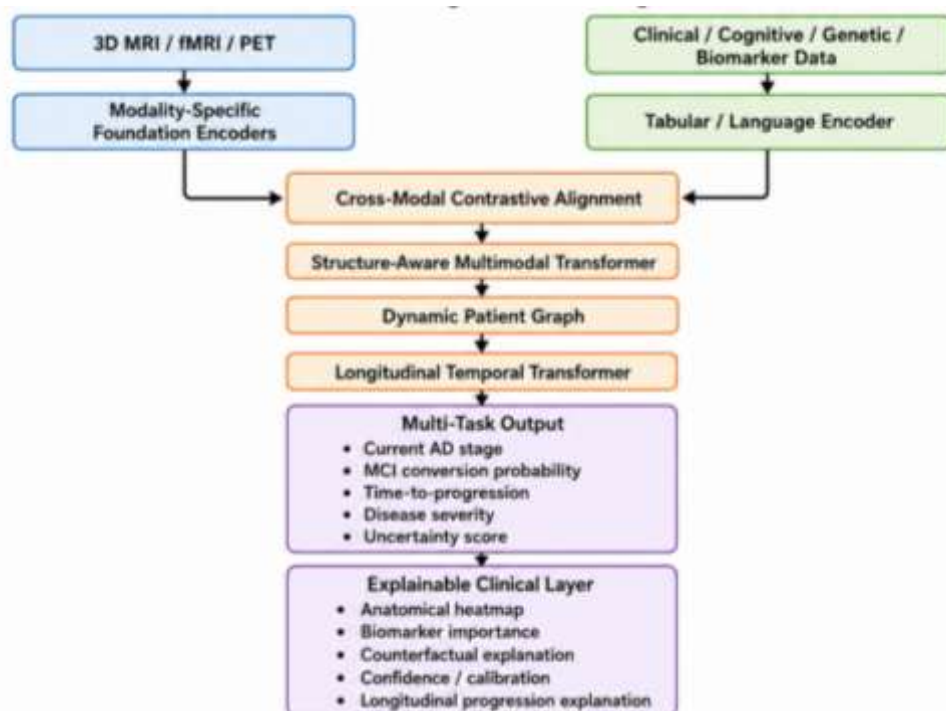


Figure 1. Proposed future architecture and research roadmap for clinically reliable AI-assisted Alzheimer's disease diagnosis and progression prediction.

8. CONCLUSION

This survey presented a comprehensive review of artificial-intelligence techniques for Alzheimer's disease diagnosis, staging, and progression analysis, covering conventional machine learning, 2D and 3D CNNs, multimodal learning, contrastive and metric learning, graph neural networks, Siamese architectures, and Transformer-based models. The reviewed literature demonstrates a clear methodological progression from handcrafted neuroimaging biomarkers and conventional classifiers toward end-to-end representation learning capable of modelling volumetric anatomy, complementary modalities, inter-subject relationships, and long-range dependencies. While conventional machine-learning approaches remain computationally efficient and comparatively interpretable, deep-learning architectures have substantially improved automated feature extraction and disease-stage discrimination. Among deep-learning approaches, 3D CNNs are particularly effective for structural MRI because they preserve volumetric anatomical information that may be lost in slice-based 2D analysis. Multimodal frameworks further improve disease characterization by combining complementary information from MRI, PET, cognitive assessments, CSF biomarkers, demographic variables, and genetic information. Contrastive and metric-learning approaches provide promising mechanisms for distinguishing closely related disease stages by learning more discriminative embedding spaces, whereas Siamese architectures explicitly model similarities and differences between subjects. Graph neural networks extend this capability by incorporating relationships among patients, while Transformer-based models enable long-range spatial interactions and increasingly provide an effective framework for whole-brain and multimodal neuroimaging analysis.

Despite the encouraging performance reported in the literature, several challenges continue to restrict clinical translation. Many studies achieve very high accuracy for relatively easier AD-versus-cognitively-normal classification, whereas clinically important problems such as early MCI detection, stable-MCI versus progressive-MCI discrimination, and individualized progression prediction remain considerably more challenging. Differences in scanner characteristics, preprocessing procedures, class definitions, dataset partitioning, population characteristics, and evaluation protocols also make direct comparison between studies difficult. Limited external validation, missing multimodal information, class imbalance, insufficient longitudinal modelling, high computational requirements, and inadequate assessment of model uncertainty remain important unresolved problems. The review further highlights that diagnostic accuracy alone should no longer be considered sufficient for evaluating Alzheimer's disease AI systems. Future models should demonstrate robustness across institutions and populations, provide clinically meaningful explanations, quantify prediction uncertainty, prevent subject-level data leakage, and report computational requirements alongside conventional diagnostic metrics. Greater emphasis is also required on independent cross-dataset evaluation and prospective clinical validation to determine whether promising experimental results can be reproduced under realistic healthcare conditions. Future research is therefore expected to increasingly integrate 3D neuroimaging, multimodal biomarkers, self-supervised and contrastive representation learning, graph-based reasoning, longitudinal modelling, and efficient Transformer architectures. Missing-modality learning, domain generalization, uncertainty-aware decision support, federated learning, privacy-preserving training, and anatomically guided explainable AI represent particularly promising directions. Longitudinal analysis of repeated neuroimaging and clinical observations may further enable a transition from static disease classification toward personalized prediction of disease progression.

REFERENCES

- [1] Jack CR Jr, Andrews JS, Beach TG et al (2024) Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimers Dement* 20(8):5143–5169. <https://doi.org/10.1002/alz.13859>
- [2] Arslan B, Zetterberg H, Ashton NJ (2024) Blood-based biomarkers in Alzheimer's disease—moving towards a new era of diagnostics. *Clin Chem Lab Med* 62(6):1063–1069. <https://doi.org/10.1515/cclm-2023-1434>
- [3] Lee J, Renslo J, Wong K, Clifford TG, Beutler BD, Kim PE, Gholamrezanezhad A (2024) Current trends and applications of PET/MRI hybrid imaging in neurodegenerative diseases and normal aging. *Diagnostics* 14(6):585. <https://doi.org/10.3390/diagnostics14060585>
- [4] Chetty CA, Bhardwaj H, Kumar GP et al (2024) EEG biomarkers in Alzheimer's and prodromal Alzheimer's: a comprehensive analysis of spectral and connectivity features. *Alzheimers Res Ther* 16:236. <https://doi.org/10.1186/s13195-024-01582-w>
- [5] Alarjani M, Almarri B (2024) fMRI-based Alzheimer's disease detection via functional connectivity analysis: a systematic review. *PeerJ Comput Sci* 10:e2302. <https://doi.org/10.7717/peerj-cs.2302>
- [6] Kaur A, Mittal M, Bhatti JS, Thareja S, Singh S (2024) A systematic literature review on the significance of deep learning and machine learning in predicting Alzheimer's disease. *Artif Intell Med* 154:102928. <https://doi.org/10.1016/j.artmed.2024.102928>

- [7] Seoni S, Shahini A, Meiburger KM et al (2024) All you need is data preparation: a systematic review of image harmonization techniques in multi-center/device studies for medical support systems. *Comput Methods Programs Biomed* 250:108200. <https://doi.org/10.1016/j.cmpb.2024.108200>
- [8] Okazawa H, Nogami M, Ishida S et al (2024) PET/MRI multimodality imaging to evaluate changes in glymphatic system function and biomarkers of Alzheimer's disease. *Sci Rep* 14:12310. <https://doi.org/10.1038/s41598-024-62806-5>
- [9] Jack CR Jr, Graf A, Burnham SC et al (2024) Application of the revised criteria for diagnosis and staging of Alzheimer's disease: drug development and clinical practice. *Alzheimers Dement (N Y)* 10(4):e70013. <https://doi.org/10.1002/trc2.70013>
- [10] Shaw LM, Korecka M, Lee EB et al (2025) ADNI Biomarker Core: a review of progress since 2004 and future challenges. *Alzheimers Dement* 21(1):e14264. <https://doi.org/10.1002/alz.14264>
- [11] Landau SM, Harrison TM, Baker SL et al (2025) Positron emission tomography harmonization in the Alzheimer's Disease Neuroimaging Initiative: a scalable and rigorous approach to multisite amyloid and tau quantification. *Alzheimers Dement* 21(1):e14378. <https://doi.org/10.1002/alz.14378>
- [12] Pahlke S, Kahale LA, Mahinrad S et al (2025) Blood-based biomarkers for detecting Alzheimer's disease pathology in cognitively impaired individuals within specialized care settings: a systematic review and meta-analysis. *Alzheimers Dement* 21(11):e70828. <https://doi.org/10.1002/alz.70828>
- [13] Palmqvist S, Whitson HE, Allen LA et al (2025) Alzheimer's Association Clinical Practice Guideline on the use of blood-based biomarkers in the diagnostic workup of suspected Alzheimer's disease within specialized care settings. *Alzheimers Dement* 21(7):e70535. <https://doi.org/10.1002/alz.70535>
- [14] Marongiu A, Spanu A, Palumbo B et al (2025) Artificial intelligence in PET imaging for Alzheimer's disease: a narrative review. *Brain Sci* 15(10):1038. <https://doi.org/10.3390/brainsci15101038>
- [15] Strelnikova PA, Zakharova NV, Kononikhin AS et al (2025) Blood plasma proteomic markers of Alzheimer's disease, current status and application prospects. *Expert Rev Proteomics* 22(1):11–18. <https://doi.org/10.1080/14789450.2025.2450804>
- [16] Basanta-Torres S, Rivas-Fernández MA, Galdo-Alvarez S (2025) Artificial intelligence for Alzheimer's disease diagnosis through T1-weighted MRI: a systematic review. *Comput Biol Med* 197(Pt A):111028. <https://doi.org/10.1016/j.combiomed.2025.111028>
- [17] Batool Z, Hu S, Kamal MA, Greig NH, Shen B (2025) Advancing Alzheimer's diagnosis with AI-enhanced MRI: a review of challenges and implications. *Curr Neuropharmacol* 23(14):1860–1877. <https://doi.org/10.2174/011570159X353595250303064846>
- [18] Shi Q, Hou J, Peng X, Xu Z, Wang Y, Cao D (2026) Magnetic resonance imaging analysis for Alzheimer's disease diagnosis using artificial intelligence: methods, challenges, and opportunities. *Ageing Res Rev* 113:102943. <https://doi.org/10.1016/j.arr.2025.102943>
- [19] Resen RD, Alzubaidi LS, Alwzway HA, Capel MI (2026) Review of deep learning models for Alzheimer's disease detection: MRI-centric approaches and multimodal extensions. *Front Artif Intell* 9:1707043. <https://doi.org/10.3389/frai.2026.1707043>
- [20] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Alzheimer's Disease Neuroimaging Initiative (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104:398–412. <https://doi.org/10.1016/j.neuroimage.2014.10.002>
- [21] Farooq A, Anwar S, Awais M, Rehman S (2017) A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, Beijing, China, pp 1–6. <https://doi.org/10.1109/IST.2017.8261460>
- [22] Hosseini-Asl E, Ghazal M, Mahmoud A et al (2018) Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci (Landmark Ed)* 23(3):584–596. <https://doi.org/10.2741/4606>
- [23] Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M, Alzheimer's Disease Neuroimaging Initiative (2019) Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin* 21:101645. <https://doi.org/10.1016/j.nicl.2018.101645>
- [24] Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Alzheimer's Disease Neuroimaging Initiative (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage* 55(3):856–867. <https://doi.org/10.1016/j.neuroimage.2011.01.008>

- [25] Suk HI, Lee SW, Shen D, Alzheimer's Disease Neuroimaging Initiative (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage* 101:569–582. <https://doi.org/10.1016/j.neuroimage.2014.06.077>
- [26] Qiao H, Chen L, Ye Z, Zhu F (2021) Early Alzheimer's disease diagnosis with the contrastive loss using paired structural MRIs. *Comput Methods Programs Biomed* 208:106282. <https://doi.org/10.1016/j.cmpb.2021.106282>
- [27] Kwak MG, Su Y, Chen K, Weidman D, Wu T, Lure F, Li J, Alzheimer's Disease Neuroimaging Initiative (2023) Self-supervised contrastive learning to predict the progression of Alzheimer's disease with 3D amyloid-PET. *Bioengineering* 10(10):1141. <https://doi.org/10.3390/bioengineering10101141>
- [28] Cárdenas-Peña D, Collazos-Huertas D, Castellanos-Dominguez G (2017) Enhanced data representation by kernel metric learning for dementia diagnosis. *Front Neurosci* 11:413. <https://doi.org/10.3389/fnins.2017.00413>
- [29] Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B, Rueckert D (2018) Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. *Med Image Anal* 48:117–130. <https://doi.org/10.1016/j.media.2018.06.001>
- [30] Liu CF, Padhy S, Ramachandran S et al (2019) Using deep Siamese neural networks for detection of brain asymmetries associated with Alzheimer's disease and mild cognitive impairment. *Magn Reson Imaging* 64:190–199. <https://doi.org/10.1016/j.mri.2019.07.003>
- [31] Hoang GM, Kim UH, Kim JG (2023) Vision transformers for the prediction of mild cognitive impairment to Alzheimer's disease progression using mid-sagittal sMRI. *Front Aging Neurosci* 15:1102869. <https://doi.org/10.3389/fnagi.2023.1102869>
- [32] Alp S, Akan T, Bhuiyan MS et al (2024) Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification. *Sci Rep* 14:8996. <https://doi.org/10.1038/s41598-024-59578-3>