

Multimodal AI for Alzheimer's Disease Diagnosis: A Systematic Review of Datasets, Models and Modalities

Ziwen Yu, Anthony Mulholland, Tianyan Huang, Qiang Liu

Submitted to: Journal of Medical Internet Research
on: October 07, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
---------------------------------	----------

Preprint
JMIR Publications

Multimodal AI for Alzheimer's Disease Diagnosis: A Systematic Review of Datasets, Models and Modalities

Ziwen Yu^{1*}; Anthony Mulholland^{1*} Prof Dr; Tianyan Huang²; Qiang Liu¹ Dr

¹School of Engineering Mathematics and Technology University of Bristol BRISTOL GB

²Medical Physics and Biomedical Engineering University College London london GB

*these authors contributed equally

Corresponding Author:

Qiang Liu Dr

School of Engineering Mathematics and Technology

University of Bristol

Tankard's Close

Ada Lovelace Building

BRISTOL

GB

Abstract

Background: Alzheimer's disease (AD) is a major global health challenge, characterised by progressive cognitive decline and rising prevalence due to population ageing. Despite increased awareness, a substantial diagnostic gap remains, underscoring the urgent need for scalable and accurate early detection methods. Artificial intelligence (AI) provides a promising solution by integrating multimodal data to enhance diagnostic accuracy and clinical decision-making. However, current research remains fragmented, with studies often focusing separately on clinical phenotyping or linguistic-based cognitive impairment data.

Objective: This review aims to systematically synthesise recent advances in multimodal AI approaches for AD diagnosis and risk prediction, bridging the gap between clinical phenotyping and linguistic-based cognitive assessment research.

Methods: Following the PRISMA 2020 guidelines, a comprehensive search was conducted across major databases to identify studies published in the past five years that applied multimodal AI models for AD diagnosis or cognitive impairment detection. A total of 97 eligible studies were included for qualitative synthesis. Extracted data covered modality combinations, AI modelling strategies, evaluation metrics, and dataset usage.

Results: Multimodal AI models consistently outperformed single-modality approaches in diagnostic accuracy and generalisability. Deep learning and ensemble learning frameworks were the most frequently adopted methods, offering strong feature-fusion capabilities. However, methodological heterogeneity, data imbalance, and limited dataset accessibility were recurring challenges. The review further summarises representative modality pairings such as imaging + clinical, imaging + genetic, and speech + cognitive features and provides an overview of publicly available multimodal datasets to facilitate reproducibility and benchmarking.

Conclusions: This systematic review provides a comprehensive overview of multimodal AI models for AD diagnosis, uniquely integrating both clinical phenotyping and linguistic-based cognitive assessment perspectives. It highlights key methodological trends, dataset limitations, and opportunities for future research toward standardised, interpretable, and generalisable AI-driven diagnostic frameworks for Alzheimer's disease.

(JMIR Preprints 07/10/2025:85414)

DOI: <https://doi.org/10.2196/preprints.85414>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>



Original Manuscript

Multimodal AI for Alzheimer's Disease Diagnosis: A Systematic Review of Datasets, Models and Modalities

Abstract

Alzheimer's disease (AD) is a significant global health concern, characterised by progressive cognitive decline and increasing prevalence due to population ageing. Despite growing awareness, a considerable diagnostic gap persists, emphasising the urgent need for scalable and effective early detection methods. Artificial intelligence (AI) offers a promising solution by integrating multimodal data to improve diagnostic accuracy and support clinical decision-making. However, existing multimodal AI studies for AD diagnosis remain fragmented, focusing separately on either clinical phenotyping or linguistic-based cognitive impairment data. Following the PRISMA protocol This review paper identified 97 publications for analysis and systematically synthesises multimodal AI models developed over the past five years, aiming to bridge current gaps in research. The primary contributions of this review include: (1) summarising representative modality combinations and evaluating their diagnostic performance; (2) critically assessing a wide array of modelling strategies to identify optimal methods and methodological trade-offs; and (3) compiling a comprehensive overview of publicly available multimodal datasets to support future research. This review provides essential guidance for researchers, facilitating standardised, integrative approaches and advancing multimodal AI-driven diagnostic capabilities in AD.

Keyword: Alzheimer's disease, neurodegenerative disease, Multimodal dataset, Machine learning, Deep learning

1. Introduction

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder and the leading cause of dementia worldwide [1]. With global population ageing, AD has become one of the costliest and deadliest diseases of the 21st century, imposing an enormous emotional, financial, and caregiving burden on patients and their families. According to recent estimates, the global number of people living with AD is expected to double from 55 million in 2020 to around 139 million by 2050 [1]. AD follows a continuous progression that includes several clinical stages [2]: preclinical AD, mild cognitive impairment (MCI), mild AD, moderate AD, and severe AD. Among these, the preclinical AD stage is considered a critical window for early intervention, during which neuropathological changes have begun, but clinical symptoms are not yet apparent [2]. However, despite increasing awareness, a substantial proportion of AD cases remain undiagnosed globally, with rates reaching up to 75% in low- and middle-income countries [3]. This alarming diagnostic gap underscores the urgent need for low-cost, scalable, and effective AD detection models to enable early diagnosis and timely therapeutic intervention, thereby slowing disease progression and improving patients' quality of life [4].

Given this significant diagnostic challenge, artificial intelligence (AI) has emerged as a promising approach for improving the early detection and management of AD. By systematically integrating and analysing vast multimodal data, AI-based diagnostic frameworks offer powerful tools to enhance early detection accuracy and facilitate timely intervention. The workflow of AI-assisted AD diagnosis involves three stages, as illustrated in Figure 1. The initial stage involves comprehensive data acquisition, leveraging diverse modalities and devices to collect neural signals, biomarkers, genetic profiles, and physiological parameters. Given the heterogeneity and potential noise in the raw data, data preprocessing, including normalisation and feature augmentation, plays a vital role in ensuring the integrity and analytical value of the modelling process. The second step is

model development/feature extraction which will discuss in detail in Sections 3. Single modality and 4. Multimodal data. The last step is interpretable and explainable analysis for AI models, which will be elaborated on in Section 5. .

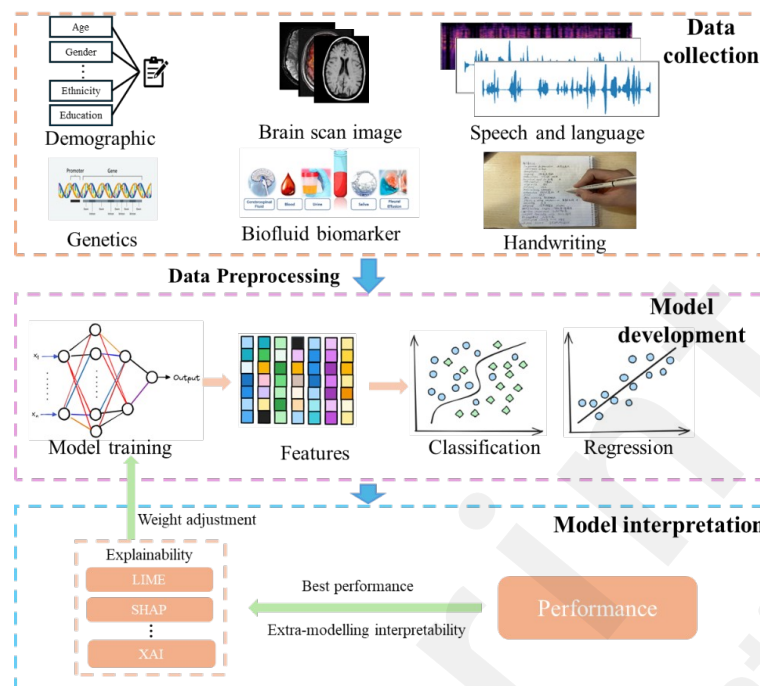


Figure 1 Artificial intelligence for Alzheimer's disease diagnosis

Traditional machine learning [5], ensemble learning [6], deep learning [7], and reinforcement learning (RL) [8] methods have achieved high performance in AI-aided AD diagnosis when applied to single modalities. In contrast, clinical diagnosis inherently relies on integrating both structural and behavioural indicators[9]. This highlights a gap: AI systems that remain unimodal risk diverging from clinical decision-making workflows. Limited multimodal integration restricts diagnostic accuracy and fails to leverage complementary strengths, such as combining MRI for structural changes with speech features for cognitive decline [10]. Without such integration, models risk overfitting to a single modality and underperforming in real-world settings. Consequently, recent research has increasingly emphasised multimodal data integration for AD diagnosis as shown in Figure 3. Yet most studies focus on demonstrating incremental accuracy gains, with limited attention to generalisability, interpretability, or cost-effectiveness factors that determine clinical adoption. Although AI methods have demonstrated significant potential, the rapid expansion and fragmented nature of multimodal studies have led to isolated findings. Furthermore, recent review papers predominantly summarise Multimodal Clinical Phenotyping datasets [11] and Multimodal Linguistic-based Cognitive Impairment datasets [12] in isolation. This separation obscures cross-modal insights, for example how imaging features and speech biomarkers might jointly improve early diagnosis. To date, no comprehensive review has (1) systematically integrated insights on optimal modality combinations, (2) evaluated best modelling strategies, and most importantly (3) provided an overview of the full array of available multimodal datasets. Motivated by these gaps, this review systematically analyses recent multimodal AI models for computer-aided AD diagnosis published over the past five years. The primary contributions of this review are as follows: (1) based on commonly used multimodal datasets, it provides a detailed summary of the most representative modality combinations and their performance in AD diagnosis. (2) it critically evaluates a wide range of modelling strategies including traditional machine learning, ensemble learning, deep learning, and reinforcement learning to identify best practices and methodological trade-offs; and (3) it compiles a comprehensive overview of publicly available multimodal datasets to support future research. Therefore, this review serves as a critical resource for researchers entering or advancing in the field of multimodal AI-driven AD diagnosis. The structure of this paper is as follows: Section 2.

Systematic literature review methodology outlines the methodology employed to retrieve relevant research articles. Section 3. Single modality introduces computer-aided AD diagnosis based on single modalities and presents commonly used performance evaluation metrics. Section 4. Multimodal data offers a detailed review of recent multimodal models developed specifically for AD diagnosis. Section 5. discusses existing challenges and proposes directions for future research. Finally, Section 6. Conclusion concludes the review.

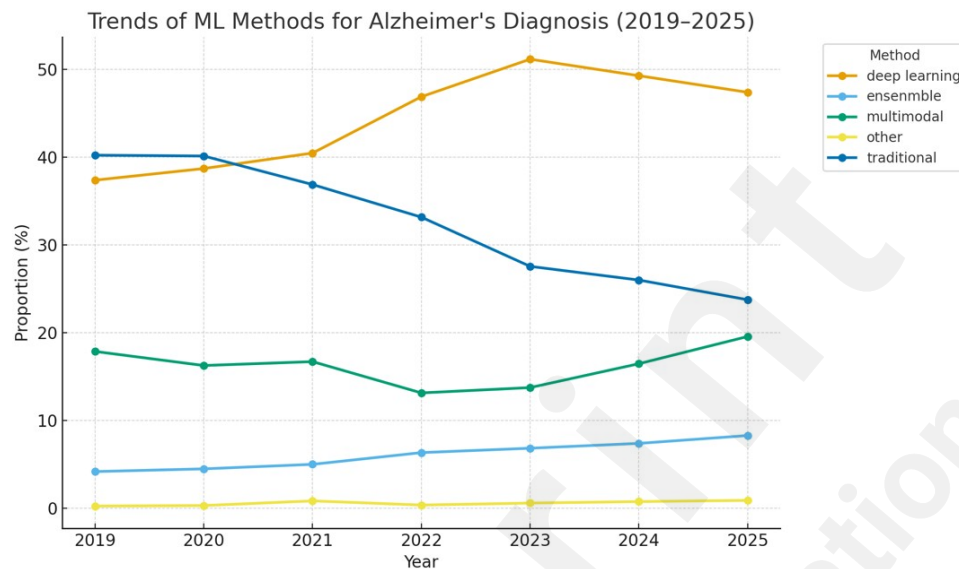


Figure 2 Trends of ML Methods for computer-aided diagnosis AD

2. Systematic literature review methodology

This section outlines the application of the PRISMA [13] methodology to systematically investigate computer-aided approaches for Alzheimer's disease diagnosis, with a particular focus on studies utilising multimodal clinical phenotyping datasets and multimodal linguistic-based cognitive impairment datasets.

2.1 Source of the study and Search criteria

This Literature searches were conducted in the following electronic databases: IEEE Xplore, ACM Digital Library, Scopus, PubMed, arXiv, and Web of Science. A unified search strategy was applied across all databases, combining disease-related, diagnostic, and methodological terms. The core query was ("Alzheimer" OR "dementia") AND ("diagnosis" OR "classification" OR "prediction") AND (specific methodological keywords). Method-specific search terms were then added as follows: for traditional machine learning ("support vector machine" OR "SVM" OR "random forest" OR "decision tree" OR "logistic regression" OR "k-nearest neighbour" OR "Naïve Bayes"); for deep learning ("deep learning" OR "neural network" OR "CNN" OR "convolutional neural network" OR "RNN" OR "LSTM" OR "autoencoder" OR "transformer" OR "BERT" OR "GPT") et al. The complete search strings are provided in the supplementary Materials.

2.2 Eligibility criteria

Inclusion criteria. Studies were considered eligible if they met all the following conditions:

- Focused on Alzheimer's disease (AD) or related dementias as the primary clinical outcome.
- Applied artificial intelligence or machine learning methods for computer-aided diagnosis, classification, or prediction.
- Utilised multimodal datasets, including combinations of clinical phenotyping, neuroimaging, genetic, or linguistic/cognitive data.
- Reported quantitative evaluation metrics (e.g., accuracy, AUC, sensitivity, specificity).

- Written in English.

Exclusion criteria. Studies were excluded if they met any of the following:

- Unimodal approaches using only a single imaging modality, cognitive test, or biomarker, without any multimodal integration.
- Works without reported performance metrics or with insufficient methodological detail.
- works not addressing diagnosis, classification, or prediction (e.g., treatment response, drug trials, lifestyle interventions)
- duplicate publications or overlapping datasets without providing additional methodological contribution.
- non-English publications.

2.3 Selection process

The study selection process followed the PRISMA 2020 guidelines. All records retrieved from the databases were first imported into Zotero, where duplicates were automatically detected and removed.

The initial search identified 825 records. After removing 372 duplicates, 453 records remained for title and abstract screening, 356 studies were excluded for the following main reasons:

- Obviously irrelevant at title/abstract level (n = 132)
- Focused on outcomes unrelated to AD diagnosis, classification, or prediction (e.g., drug trials, treatment response, lifestyle interventions) (n = [115]).
- Used unimodal data without multimodal integration (n = [106]).
- Non-English publications (n = [3]).

Finally, 97 studies were included in the systematic synthesis and all were successfully retrieved (reports not retrieved = 0). The complete selection process is summarised in the PRISMA 2020 flow diagram Figure 3.

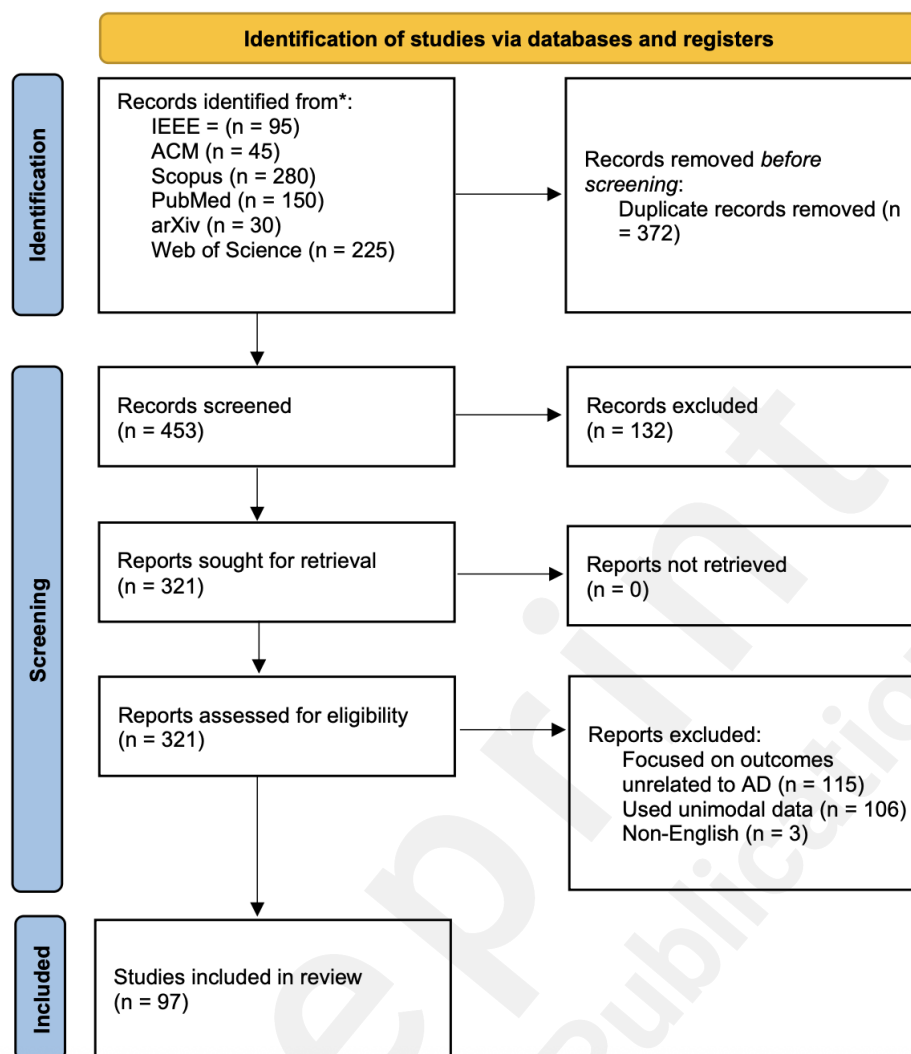


Figure 3 Flow diagram of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses)

3. Single modality

If the reader is already familiar with standard machine learning methods, they may wish to skip directly to Section 4. Multimodal data, where discuss multimodal data in AD diagnosis. There are several machine learning methods in the computer-aided AD diagnosis, these methods can be summarised into traditional machine learning, deep learning, ensemble learning, reinforcement learning and multimodal learning. These methods are summarised below.

3.1 Traditional machine learning models

Support Vector Machine (SVM) is a supervised learning model initially proposed by Cortes [14], widely utilised for both classification and regression tasks. The core principle of SVM involves identifying an optimal hyperplane in a high-dimensional feature space that maximally separates data points from different classes. Due to its robust generalisation capabilities, SVM has been extensively applied in computer-aided diagnosis systems, particularly for medical image classification and disease detection [15], [16], [17], [18], [19]). However, SVM exhibits notable limitations under certain conditions. Specifically, in scenarios where the dimensionality of the feature space exceeds the number of training samples, the model is prone to overfitting. Additionally, conventional SVM does not inherently provide probabilistic outputs, which constrains its interpretability in clinical decision-making processes. These drawbacks underscore the need for hybrid or enhanced models that integrate SVM with dimensionality reduction techniques or probabilistic calibration methods to

improve performance and interpretability.

Decision Tree (DT) is a widely utilised, interpretable machine learning technique applicable to both classification and regression tasks. It operates by recursively partitioning the feature space based on a set of decision rules, which are derived from prior probabilities and the expected value of target outcomes. Due to its inherent simplicity and transparency, DT models are frequently employed in scenarios involving datasets with high proportions of missing values [20], [21], [22], [23]. Their intuitive structure requires minimal domain-specific expertise and is computationally efficient, making them particularly suitable for preliminary data exploration and decision support systems in clinical and diagnostic settings. However, a well-known limitation of decision trees is their tendency to overfit, especially when applied to large datasets with complex, non-linear relationships, unless appropriately regularised through techniques such as pruning or ensemble learning.

Hidden Markov Model (HMM) is a classical machine learning framework widely employed in both classification and regression tasks. In HMMs, the underlying system is modelled as a set of hidden states that are not directly observable, while the observable outputs are probabilistically dependent on these latent states. Each hidden state is associated with a specific probability distribution over the set of possible observed outputs. By generating a sequence of observations, the model provides insights into the underlying sequence of hidden states. Owing to its ability to effectively model temporal dependencies and stochastic processes, HMM is particularly suitable for analysing series data. Consequently, it has been extensively applied to computer-aided diagnosis of AD [24], [5], [25], [26], where temporal patterns in verbal communication can serve as critical biomarkers for early detection and disease progression monitoring.

In addition to HMM-based approaches, various other machine learning algorithms such as k-Nearest Neighbors (KNN), Logistic Regression (LR), and Gaussian Mixture Models (GMMs) have been applied in the field of AI to assist in the diagnosis of AD [27], [28], [29], [30], [31]. Although these traditional models have demonstrated effectiveness in specific tasks, they often struggle to capture the complex and comprehensive features present in large-scale multimodal datasets. To overcome these limitations, researchers have proposed deep learning models that can automatically extract rich and latent feature representations from diverse data sources.

3.2 Deep Learning

Deep learning models possess a superior capability for hierarchical feature extraction compared to traditional machine learning algorithms. This advantage enables them to effectively capture complex and subtle patterns within high-dimensional data. Consequently, deep learning has been extensively employed to process and integrate multimodal data related to AD, including neuroimaging, clinical scores, genetics, and speech signals. The major approaches and findings in this domain are summarised below.

Convolutional Neural Network (CNN) is specifically designed for the analysis of image data, leveraging spatial hierarchies by capturing local patterns through convolutional operations. By exploiting spatial correlations among adjacent pixels, CNNs are particularly effective in extracting discriminative features from medical imaging. Given that neuroimaging modalities play a critical role in the diagnosis and staging of AD, a wide range of CNN architectures and their variants have been proposed. These models have demonstrated high performance in AD classification and prediction tasks, often surpassing traditional approaches in terms of accuracy and robustness [32], [33], [34], [35], [36]. However, despite their widespread adoption, CNNs exhibit several limitations beyond overfitting. Firstly, their inherent inductive bias toward locality and translation invariance, while effective for capturing low-level features, may restrict their capacity to model long-range dependencies and global contextual information that an aspect increasingly relevant in capturing subtle neurodegenerative patterns across distributed brain regions. Secondly, CNNs often require extensive architecture tuning and large labelled datasets to achieve generalisability, making them less flexible in data-scarce medical contexts, and challenges remain in balancing computational

efficiency, interpretability, and performance on small, heterogeneous clinical datasets.

Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN). RNN is well-suited for modelling dynamic temporal datasets, owing to their inherent architecture that allows information to be cyclically propagated through recurrent connections. This design enables RNNs to capture temporal dependencies and contextual relationships within sequential data. As a result, RNNs can handle a wide variety of time-series inputs, in applications involving longitudinal clinical data, speech patterns, and physiological signals relevant to AD progression [37], [38], [39], [40]. Although RNNs perform well in modelling sequential data, they are inherently susceptible to the vanishing gradient problem during training, particularly over long input sequences.

To address this limitation, LSTM was introduced as a specialised variant of RNN. It introduces memory cells and gating mechanisms, including a forget gate, which help regulate the flow of information and preserve long-term dependencies in the data. Compared with traditional RNNs, LSTM offers greater training stability and enhanced performance in capturing temporal patterns. As a result, it has been widely applied in AD research, particularly for processing longitudinal clinical data, speech signals, and other sequential modalities [41], [42], [43], [44].

Transformer model, which leverages attention mechanisms, dynamically assigns different weights to input features based on their relative importance. Each layer of the Transformer consists of multiple attention heads, allowing the model to capture diverse feature representations by attending to various aspects of the input. This parallelised attention structure significantly enhances training efficiency and scalability. Owing to these advantages, Transformer-based architectures have been widely adopted recently in the domain of AD diagnosis [45], [7], [46], [47]. Furthermore, the inherent encoder-decoder framework of Transformers facilitates the integration of heterogeneous data sources, enabling the development of multimodal approaches that effectively fuse features from various modalities. This capability has proven particularly valuable in computer-aided diagnosis of AD, as will be elaborated in the following section.

3.3 Ensemble learning

Ensemble learning is a machine learning paradigm that enhances model generalisation, robustness, and predictive accuracy by aggregating the outputs of multiple base learners. Common ensemble strategies include bagging such as Random Forest, boosting algorithms such as AdaBoost, XGBoost, and LightGBM, which Sequential stacking tech. These methods have been extensively applied to the early detection, subtyping, and progression prediction of AD, owing to its ability not only to optimise performance by combining the strengths of diverse algorithms, but also to improve model reliability in the presence of clinical heterogeneity [6], [48], [49], [50]. Ensemble methods are also known to mitigate the effects of class, however the integration of multiple models increases the risk of redundant feature learning, which may yield only marginal gains in predictive performance, especially when applied to small-scale datasets. Finally, their typically large size and slower inference speed may pose deployment challenges in real-time or resource-constrained environments, such as mobile health or wearable platforms.

3.4 Reinforcement learning

Reinforcement learning (RL), a subfield of machine learning, is centred on the concept of enabling an agent to learn an optimal policy through continuous interaction with its environment, with the objective of maximising cumulative rewards. Given that AD is a chronic, progressive neurodegenerative disorder, its long-term management and the need for personalised intervention strategies align well with the sequential decision-making capabilities inherent in RL. A core component of RL is the value function, which estimates the expected cumulative benefit of a given state or action. In recent applications to AD, researchers have begun using deep neural networks to approximate value functions, enabling the agent to handle complex, high-dimensional patient data.

By simulating the dynamic process of “treatment–feedback–optimisation,” RL offers a more adaptive and flexible framework for devising individualised, long-term therapeutic strategies in AD care [8], [51]).

3.5 Performance evaluation matrices

To ensure the clinical applicability and scientific rigour of computer-aided diagnosis models for AD, it is essential to systematically evaluate their performance using a variety of quantitative metrics. These evaluation metrics not only guide model selection and optimisation but also provide a standardised basis for comparing results across different studies and datasets. Performance metrics help determine how well a model can differentiate between disease stages, predict progression, or generalise to unseen data. Table 1 presents a comprehensive overview of evaluation metrics commonly used in AD diagnosis models, which can be broadly divided into two categories. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Root Mean Squared Logarithmic Error (RMSLE), Normalised MSE (NMSE), Correlation Coefficient (CC), Weighted Correlation, Adjusted Rand Index (ARI), Normalised Mutual Information (NMI), and Area Under the Curve (AUC). These metrics capture model performance in probabilistic, continuous, or unsupervised learning contexts, which play a crucial role in complex, data-rich healthcare applications.

Table 1 Performance evaluation for Alzheimer’s disease diagnosis

Name	Equation	Interpretation
Accuracy (ACC)	$\frac{TP+TN}{TP+FP+TN+FN}$	Overall percentage of correct predictions (range: 0–1; higher is better)
Precision (PRE)	$\frac{TP}{TP+FP}$	The fractions that are positive among predicted positives (range: 0–1; higher is better)
Recall (REC) / Sensitivity (SEN) / True Positive Rate (TPR)	$\frac{TP}{TP+FN}$	Fraction of actual positives correctly identified (range: 0–1; higher is better)
Specificity (SPE)	$\frac{TN}{TN+FP}$	Fraction of actual negatives correctly identified (range: 0–1; higher is better)
F1 score	$\frac{2}{\mathcal{R}C^{-1} + \mathcal{P}R E^{-1}}$	Harmonic means precision and recall (range: 0–1; higher is better)
Matthews Correlation Coefficient (MCC)	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	Correlation of the confusion matrix (range: –1 to 1; closer to 1 is better)
Balanced Accuracy (BAC)	$\frac{1}{2}(REC + SPE) = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right)$	Average of sensitivity and specificity (range: 0–1; higher is better)
Mean Absolute	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	Average magnitude of prediction errors for

Error (MAE)		continuous outputs (non-negative; lower is better)
Mean Squared Error (MSE)	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	Average squared prediction error (non-negative; lower is better)
Normalised Mean Squared Error (NMSE)	$\frac{MSE}{\sigma_y^2} \text{ where } \sigma_y^2 = \text{Var}(y)$	Mean Squared Error normalised by the variance of the target variable (lower is better)
Mean Squared Error (RMSE)	$\sqrt{MSE}$	Square root of Mean Squared Error (non-negative; lower is better)
Root Mean Squared Logarithmic Error (RMSLE)	$\sqrt{\frac{1}{n} \sum_{i=1}^n [\ln(y_i + 1) - \ln(\hat{y}_i + 1)]^2}$	Root Mean Squared Error on a log scale (non-negative; lower is better)
Normalised Mutual Information (NMI)	$\frac{2I(X;Y)}{H(X)+H(Y)}$ I is mutual information	Information-theoretic similarity between predicted and true labels (range: 0–1; higher is better)
Adjusted Rand Index (ARI)	$\frac{\sum_{i,j} \binom{N_{ij}}{2} - \frac{\sum_i \binom{N_i}{2} \sum_j \binom{N_j}{2}}{\binom{N}{2}}}{\frac{1}{2} \left[\sum_i \binom{N_i}{2} + \sum_j \binom{N_j}{2} \right] - \frac{\sum_i \binom{N_i}{2} \sum_j \binom{N_j}{2}}{\binom{N}{2}}}$	Measures agreement between clustering (range: 0–1; higher is better)
Correlation Coefficient (CC)	$\frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$	Linear correlation between two continuous variables (range: –1 to 1; closer to 1 is better)
Weighted Correlation Coefficient	$\text{corr}(x, y; w) = \frac{\text{cov}(x, y; w)}{\sqrt{\text{cov}(x, x; w) \text{cov}(y, y; w)}}$ where w represents a weight vector	Pearson correlation that assigns weights to observations (range: –1 to 1; closer to 1 is better)
Area Under ROC Curve (AUC)	$\frac{1}{ P \cdot N } \sum_{p \in P} \sum_{n \in N} \left[1(s(p) > s(n)) + \frac{1}{2} \cdot 1(s(p) = s(n)) \right]$	The probability that a classifier ranks a random positive classification higher than a random negative classification

		(range: 0–1; higher is better)
--	--	--------------------------------

TP: number of correctly identified positive cases. TN: number of correctly identified negative cases. FP: number of negative cases incorrectly identified as positive. FN: number of positive cases incorrectly identified as negative. y_i : true (actual) value of the i -th observation. $\hat{y}_i$: the predicted value produced by the model for the i -th observation. $H(X)$: the entropy of X, which measures the amount of uncertainty or randomness in variable X. $H(Y)$: the entropy of Y, which measures the amount of uncertainty or randomness in variable Y. N : the total number of data points, N_{ij} : the number of points shared between cluster C_i in partition P and class C_j^* in ground truth partition P^* . $s(p)$: the score (or prediction probability) assigned by the model to a positive sample. $s(n)$: the score assigned to a negative sample.

2.6 Summarisation

Although traditional machine learning approaches based on single-modality feature extraction have demonstrated competitive performance in AD-related tasks, several inherent limitations remain:

1. **Information Completeness:** Structural MRI data alone often fails to capture the full spectrum of cognitive and behavioural alterations associated with AD. While MRI provides valuable information about brain atrophy and structural degeneration, it lacks sensitivity to functional and molecular changes. Therefore, incorporating other data modalities such as FDG-PET (which measures brain metabolism), amyloid or tau PET (which capture pathological burden), neuropsychological assessments, speech and language features, EEG, and genetic or biomarker data can offer a more comprehensive and multidimensional understanding of disease progression and patient heterogeneity. The integration of multimodal data can significantly enhance the completeness of clinical representation (Fatima et al [52]).
2. **Model Robustness:** In multimodal datasets, the presence of noise in one modality, which may not be effectively mitigated by conventional denoising methods can be compensated by complementary information provided by other modalities, thereby improving the overall robustness and reliability of the model. The human cognitive system inherently relies on multisensory integration to perceive and interpret complex stimuli. Multimodal models are thus more aligned with biological mechanisms, enabling a more accurate simulation of human-like cognitive processes [53].
3. **Cross-Modal Learning:** Transformer-based architectures facilitate cross-modal representation learning and support weakly supervised learning paradigms, wherein information from one modality can guide or supervise the learning process of another. This enhances model generalisation and performance under limited supervision [54].
4. **Real-World Decision-Making:** Multimodal learning frameworks are more reflective of real-world clinical decision-making processes, which are inherently multi-dimensional. Incorporating diverse data sources aligns computational models with practical diagnostic workflows and enhances their translational potential in clinical environments [55].

Therefore, this paper provides a comprehensive survey of the advantages and limitations of multimodal models in the context of computer-aided diagnosis of AD, with a particular focus on dataset de-classification strategies. By utilising dataset de-classification, researchers can evaluate and compare the performance of different models under consistent experimental conditions, thereby enabling a more objective and rigorous assessment of model generalisability and robustness.

4. Multimodal data

4.1 Multimodal Dataset overview

High-quality data plays a pivotal role in training AI models for computer-aided diagnosis and detection. Robust datasets not only enhance the generalisation ability of models but also help mitigate the risk of overfitting. Commonly used datasets for AI-assisted diagnosis of AD can be broadly categorised into two types. The first type multimodal clinical phenotyping dataset such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), UK Biobank, and the Open Access Series of Imaging Studies (OASIS) focuses on neuroimaging modalities, including MRI, fMRI, genetic data, and electronic health records (EHRs). The second type - multimodal cognitive-linguistic behavioural dataset, centres on sequential data modalities, including audio, video, and transcribed (typed) language data, such as the Pitt Corpus and ADReSS. Each dataset type provides unique features that contribute to the comprehensive modelling of AD progression and diagnosis. The commonly used datasets, along with their population demographics and associated modalities, are summarised in Table 2.

Table 2 Commonly used dataset

Dataset		Population Demographics		Modalities	Link
		Male, n (%)	Age, mean(y) $\pm$ s.d.		
UK Biobank		23 000, 46%	56.5 $\pm$ 8.1	MRI, fMRI, genetic, lifestyle scores, Activity monitor, EHR	https://ukbiobank.ac.uk
ADNI	ADNI-1	469, 57.3%	75 $\pm$ 6.9	MRI, PET, genetic, EHR	https://adni.loni.usc.edu
	ADNI-GO and 2	473, 53%	72.5 $\pm$ 7.2		
	ADNI-3	471, 49%	74.9 $\pm$ 8.1		
OASIS	OASIS-1	177, 42.5%	57 $\pm$ 39	MRI, PET, CT, genetic, EHR	https://sites.wustl.edu/oasisbrains/
	OASIS-2	60, 40%	78 $\pm$ 18		
	OASIS-3	622, 45%	69 $\pm$ 26.56		
	OASIS-4	663 subjects	57.5 $\pm$ 36.5		
NACC		23 625, 44.62%	73.3 $\pm$ 10.5	MRI, PET, genetics, EHR	https://naccddata.org/
FHS		718, 42%	80.76 $\pm$ 8.2	MRI, genetic, EHR	https://framinghamheartstudy.org
AIBL		289, 43.72%	73.5 $\pm$ 7.03	MRI, PET, genetic, EHR	https://aibl.csiro.au
Dementia Bank	Pitt corpus	HC:104, AD:208/552	-	Audio, text	https://dementia.talkbank.org/
	ADReSS	HC:78, AD: 78/156			
	ADReSSo	HC: 115, AD:122/237			
	ADReSS-M	HC: 143,			

		AD:148/291			
	TAUKADIAL	106/507			
Multimodal Dementia Corpus		HC: 10, AD: 12/816	-	Audio, Typed, Hand-written	[56]
ADReFV		AD:25	66.68 $\pm$ 2.08	Video	[57]
GENCODE		Human:78686	-	Genetics	Gencode nes.org

To better understand current research trends, this paper reviewed recent literature from the past five years focusing on multimodal AI models for AD diagnosis. The studies were categorised based on the type of dataset used: (1) Multimodal clinical phenotyping datasets: Our analysis reveals that the ADNI dataset overwhelmingly dominates this category, being employed in approximately 80% of relevant studies. The remaining 20% of studies utilise other clinical phenotyping datasets, including UK Biobank, OASIS, NACC, FHS, and AIBL, often in supplementary roles or for external validation purposes. (2) Multimodal cognitive-linguistic behavioural datasets: In contrast, studies in this category most frequently adopt the ADReSS series (including ADReSS, ADReSSo, and ADReSS-M), which collectively account for approximately 70% of usage. The remaining 30% of studies incorporate alternative datasets such as TAUKADIAL, the Multimodal Dementia Corpus, and ADReFV. These datasets are typically used in conjunction with the ADReSS series or employed to benchmark model performance across modalities.

4.2 Multimodal clinical phenotyping datasets

Multimodal clinical phenotyping datasets primarily comprise a wide array of data types, including MRI, PET, and DTI neuroimaging data; biological markers derived from blood, cerebrospinal fluid (CSF), and genomics; as well as standardised clinical cognitive assessments. In Section 4.2 Multimodal clinical phenotyping datasets, this paper summarise several representative datasets such as ADNI, UK Biobank, and other multimodal biomedical datasets highlighting their unique characteristics, data modalities, and roles in advancing state-of-the-art diagnostic models. These are presented in Table 3.

4.2.1 UK Biobank dataset

UK Biobank enables population-level association studies and early-risk modelling. It has been widely used in AD diagnosis research, including the following notable studies:

Xue [58] proposed a transformer-based multimodal model to address the challenge of overlapping symptoms across different aetiologies. The AUROC of neurologist assessments augmented by the AI model exceeded that of neurologist-only evaluations by 26.25%, highlighting its potential as a reliable tool for supporting personalised treatment strategies. Allwright [59] developed a custom machine learning model for high-dimensional data and found that the AUC for APOE4 carriers exceeded that of non-APOE4 carriers by 11%, confirming the APOE4 allele as the strongest genetic risk factor for AD. Gu [60] identified predictors from 316 variables and developed a machine learning model to predict the risk of incident dementia, Alzheimer's disease, and vascular dementia over 5-, 10-, and long-term follow-up in patients with ASCVD. The model achieved 88% accuracy, with a mean sensitivity of 63.7%, specificity of 91.4%, precision of 47%, and F1-score of 54.6%. You [61] proposed the UKB-DRP multimodal dementia risk prediction model, which achieved an AUC of 86.2%, clearly outperforming traditional evaluation systems.

Calvo [62] developed a Firth logistic regression-based multimodal model to assess AD risk in women with early bilateral oophorectomy (BO), using large-scale UK Biobank data. The study revealed a causal link between BO and increased AD risk and further identified hormone therapy and

education as modifiable protective factors underscoring the importance of early intervention and cognitive health management in high-risk female populations. Yi [63] established a comprehensive multimodal framework integrating MRI-based brain age estimation, GWAS-based genetic variant screening, Mendelian randomisation analysis, and drug repurposing to systematically identify modifiable genetic targets and candidate therapeutics for brain ageing in the general population. Yousefzadeh [64] introduced LAVA (Granular Neuron-level Explainer), a novel explainable AI framework for assessing AD progression using retinal fundus images. Unlike traditional black-box models, LAVA provides neuron-level interpretability by identifying critical neurons through mutual information-based selection and clustering activation patterns to infer the AD continuum. The model successfully revealed latent subgroups corresponding to different disease stages, correlating with cognitive performance and vascular morphology, thus offering a model-agnostic and clinically relevant diagnostic tool for early detection and monitoring. Gong [65] proposed SuperBigFLICA, a scalable and interpretable semi-supervised learning framework designed to extract low-dimensional image-derived phenotypes from large-scale multimodal brain imaging data. By integrating imaging and non-imaging data within a Bayesian optimisation framework, the model achieved improved predictive performance, reduced manual parameter tuning, and enhanced interpretability compared to both handcrafted features and prior unsupervised methods. It is computationally efficient, generalisable to unseen phenotypes, and particularly well-suited for population-scale datasets such as UK Biobank.

This section outlines the implementation of multimodal models within the UK Biobank framework. As evidenced by the analysis and [Table 3](#), the utilisation of UK Biobank data provides meaningful insights for AD diagnosis and risk prediction. Nevertheless, certain limitations persist, including data imbalance for instance, cognitively healthy individuals greatly outnumber diagnosed AD or MCI cases which can bias model training. Additionally, the need for external validation remains critical to ensure the generalisability of findings beyond the UK Biobank population.

4.2.2 ADNI dataset

ADNI provides a rich and diverse collection of demographic information, multimodal data, and clinical assessments. Owing to its comprehensive scope and longitudinal design, it has become one of the most widely adopted benchmark datasets for computer-aided diagnosis of AD. The following studies exemplify its use:

Lian [66] proposed an attention-guided hybrid deep learning framework to identify AD-related brain regions from structural MRI and perform diagnosis. This method generates individualised lesion attention maps using weakly supervised learning, integrating local and global feature modelling. It addresses limitations of traditional approaches, such as fixed region-of-interest selection and inadequate local detail. Experimental results demonstrated improved AD identification and MCI-to-AD conversion prediction across multiple datasets, while maintaining interpretability and generalisability. Lian [67] introduced a multi-task weakly supervised attention network that leverages whole-brain MRI to jointly predict multiple clinical scores. By applying an attention mechanism to localise subject-specific pathological regions, the model avoids inefficiencies of manual region selection and non-linear registration. Results showed significant improvements in regression accuracy and inference efficiency. Li [68] designed a dual-channel residual neural network within a time-to-event framework using left and right hippocampal MRI to predict both the risk and timing of MCI-to-AD conversion. This overcomes the limitations of binary classification by incorporating temporal prediction, with validation across ADNI and AIBL datasets.

Qiu [9] developed a multimodal deep learning framework integrating MRI, cognitive assessments, and medical history to classify cognitively normal individuals, MCI, AD, and non-AD dementia. Using SHAP, the model enhanced interpretability and mimicked real-world diagnostic workflows, achieving performance comparable to or better than neurologists. Kwanseok [69], introduced the LEAR framework, unifying model training, counterfactual explanation, and

explanation-guided refinement. The model generates multi-class counterfactual maps that optimise attention to clinically relevant brain regions, enhancing both performance and explainability. Lian [70] proposed a Hierarchical Fully Convolutional Network (H-FCN) for localising AD-related atrophy regions at multiple scales in sMRI, achieving superior performance (AUC = 95.9%) in both AD classification and MCI conversion prediction.

Avsec [71] developed Enformer a Transformer-based model for long-range DNA interaction modelling, improving tissue-specific gene expression prediction (correlation increased from 81% to 85%). Yang [72] built a multimodal deep learning model combining MRI, genetic, and cognitive features to predict MCI-to-AD progression, achieving an AUC of 0.921. Lee [73] proposed a prognostic model using low-cost MRI and cognitive tests across real-world UK and Singapore cohorts (AUC = 84%), demonstrating generalisability and outperforming conventional measures. Zhu [74] proposed a DA-MIDL model to identify local/global discriminative features from sMRI for AD/MCI classification. It achieved state-of-the-art performance on ADNI and AIBL datasets, enhancing interpretability and feature visualisation, marking a significant advance in sMRI-based diagnosis.

Zhang [75] proposed a "feature-aware auto-fusion" framework combining SHAP-based feature selection with graph convolutional networks to integrate multimodal data. The novel Automated Model Fusion Strategy (AMFS) adaptively adjusts sub-model weights, achieving 95.9% accuracy on ADNI data while significantly improving interpretability and cross-modal modelling performance. Velazquez [76] developed an ensemble learning model fusing EHR and DTI data, incorporating random forest and CNN sub-models to predict MCI conversion to AD with 98.81% accuracy. The model provides both local and global interpretability, offering valuable insights for clinical decision support and trial participant screening. Zhang [77] introduced an information entropy polynomial expansion method to recover fine-grained information from neuropsychological tests, constructing an extreme learning machine (ELM) ensemble model at the cognitive domain level. The approach achieved 98% accuracy and F1-score without bias in MCI/AD populations, addressing limitations of conventional methods that overlook multimodal cognitive data (e.g., video, speech, text). Bi [78] proposed a Clustering Evolutionary Random Forest (CERF) to integrate fMRI and SNP data, constructing "brain region-gene pair" multimodal features for joint classification, selection, and interpretation. Experiments demonstrated its effectiveness in identifying AD patients and discovering critical brain regions/pathogenic genes even with small-sample high-dimensional data.

Bi [79] presented a Weighted Evolutionary Random Forest for multimodal fMRI-SNP fusion to investigate MCI pathogenesis. The model identifies disease-relevant features strongly associated with brain activity and gene expression, enabling individualised etiological diagnosis with practical implications for early MCI intervention. Hashmi [80] developed a RL framework integrated with convolutional neural networks to address MRI data class imbalance, generating high-resolution lesion probability maps for four-stage dementia classification. Evaluated on OASIS dataset, the method demonstrated performance improvements of 6% (accuracy), 9% (precision), 13% (recall), and 10% (F1-score), showing strong potential for early diagnosis and disease staging. Wang [81] created the DISFC framework integrating sMRI, genetic, and cognitive data while modelling intra- and inter-modal interactions for long-term MCI-to-AD prediction. Achieving 0.962 AUC in multi-phase ADNI cohort validation, the framework demonstrated excellent generalisability, quantifiable biomarker contributions, and robust cross-centre performance with inherent interpretability.

Hatami [82] introduced a reinforcement learning-based data augmentation method to enhance traditional deep learning models' robustness and generalisation in MRI classification. ADNI experiments confirmed superior performance across accuracy, specificity, sensitivity, and AUC metrics, with the reinforcement feedback mechanism significantly improving model stability. Tabarestani [83] designed a multitask multimodal distributed regression model handling incomplete longitudinal data (MRI/PET/CSF/cognitive assessments). By employing modality-specific regression matrices to capture inter-task relationships and gradient boosting ensemble, the method effectively

addressed missing values and data heterogeneity, demonstrating high robustness and low prediction error in long-term follow-up studies. Burkhart [84] proposed an unsupervised Multimodal Trajectory Modelling method that learns population developmental trajectories from longitudinal cognitive and brain health data without requiring labels. Demonstrating strong generalisation even without costly T2 or PET data, their derived Individual Cognitive Health Index outperformed conventional clinical scales in predicting AD progression, offering a low-cost, non-invasive intelligent early-warning tool. Sappagh [85] developed an interpretable AI-based multilayer multimodal detection model integrating MRI, PET, biomarkers, and clinical scores. The model provides explainable AD progression pathways, delivering transparent diagnostic support information with high potential for early diagnosis and intervention. Min-Woo [86] systematically compared six machine learning models on multimodal MRI/PET data, identifying Gradient Boosting Machine (GBM) as the most stable performer - particularly when excluding T2-FLAIR. This study established a robust prediction tool for MCI-to-AD conversion risk, supporting clinical screening and treatment planning.

Nguyen [87] conducted a systematic review of deep learning in AD diagnosis, highlighting limitations in ADNI and similar datasets including inconsistent evaluation standards and poor generalisability. They emphasised multimodal approaches, standardised evaluation frameworks, and cross-centre reproducibility as critical future directions. Yuan [88] proposed a semi-supervised Co-training framework combining sMRI and SNP data, utilising both labelled and unlabelled MCI samples from ADNI. Achieving 85.5% accuracy and 0.825 AUC on ADNI-2, this work validated the value of joint labelled/unlabelled data training for MCI classification. Cirincione [89] created an Ensemble Integration (EI) multimodal framework using ADNI data that outperformed XGBoost and traditional deep learning (AUC=0.81 vs. 0.68) in predicting dementia conversion from MCI. Key predictive features included structural MRI measures of middle temporal gyrus, posterior cingulate, and lateral ventricles.

While ADNI provides a comprehensive and standardised multimodal resource tailored for AD research and models built on it consistently demonstrate robust performance certain limitations persist. These include data imbalance, such as the disproportionate number of cognitively healthy individuals compared to diagnosed AD or MCI cases, which may bias model training. There is also a notable underrepresentation of racially diverse populations, with most participants being of White European ancestry, while other groups such as Black, South Asian, and mixed-ethnicity individuals are significantly less represented. Furthermore, external validation remains essential to ensure the generalisability of findings across different demographic and clinical settings.

4.2.3 Self-collected datasets

While public datasets like ADNI provide standardised benchmarks, self-collected datasets enable more flexible acquisition of targeted modalities. Representative studies include:

The self-collected dataset also performs multimodal fusion such as: Cassani [90] leveraged EEG power modulation spectral features combined with eLORETA source localisation across frequency bands (α/θ), demonstrating that modulation-domain filtering enhances rsEEG signals for differentiating healthy elderly from AD patients - a crucial advancement for portable, low-cost EEG diagnostics. Wu [91] proposed 14 Hilbert marginal spectrum-based EEG power metrics with a Warped Infinite Gaussian Mixture Model (WiGMM) for unsupervised dementia staging, correlating fast/slow wave power changes with cognitive decline - an innovative quantitative EEG assessment. Kmetzsch [92] designed a supervised Variational Autoencoder (VAE) integrating MRI/PET with microRNA expression for neurodegenerative disease staging, proving that pathological relevance correlates with latent space proximity to ground truth. Zhang [93] built an interpretable patch-level model (FCRN+MLP) fusing MRI/PET/MMSE/ApoE4, achieving 96.2% (AD) and 92.2% (MCI) classification accuracy while generating high-resolution risk maps - bridging AI and pathological mechanisms. Fabietti [94] developed an EXML model analysing local field potentials from animal models, combining time, space and frequency features with 99.4% AD detection accuracy - robust

even with channel occlusion, validating electrophysiological sensitivity to AD progression.

In addition to the multimodal fusion model Behavioural & Digital Biomarkers data was used by Cilia [95] to developed a deep transfer learning approach analysing online handwriting trajectories converted to images, introducing a novel visuomotor fusion paradigm for non-invasive cognitive screening, particularly effective in early AD detection. Mengoudi [96] employed self-supervised learning on instruction-free eye-tracking data to model cognitive differences between dementia patients and controls, revealing naturalistic attention patterns without explicit labelling - demonstrating eye-tracking's potential in cognitive assessment. Park [16] pioneered VR-based behavioural biomarkers integrated with MRI structural data via SVM, achieving 94.4% accuracy (100% sensitivity) for MCI screening - the first validation of VR-morphometry correlations for reproducible, low-cost cognitive assessment. Seifallahi [97] used Kinect V2 to capture 25-joint kinematics during Timed Up and Go tests (47 controls vs 38 AD), achieving 98.68% classification accuracy via SVM - a sensor-free, clinically deployable screening tool.

Tsai [98] created MAND, a Transformer-based model using Taiwan's National Health Insurance data that include 2M patients ×15 years ICD codes, achieving 90.1% AUC for dementia risk prediction via multimodal attention mechanisms. Fan [99] proposed a Vision Transformer-XGBoost hybrid using T1, T2-FLAIR MRI + clinical data, attaining 0.972 AUC for vascular cognitive impairment prediction with CAMs and SHAP revealing hippocampal-prefrontal-amygdala vascular injury linkages. Wang [100] developed a parsimonious ROSMAP-based model requiring just four brief cognitive tests + demographics to predict 3-year dementia onset (AUC = 89.7%) matching 98-minute neuropsychological batteries while offering SHAP-based individualised explanations.

Self-collected datasets offer distinct advantages including targeted modality acquisition, novel biomarker discovery like microRNA, LFP and handwriting, and enhanced real-world clinical utility. However, self-collected datasets typically suffer from limited sample sizes, which increases susceptibility to overfitting and compromises generalisability across diverse populations.

4.2.4 Limitations of datasets

Table 3 Multimodal clinical phenotyping datasets related article

Publication	Dataset	AI Method	performance	Limitation
[58]	AIBL, PPM, ADNI, OASIS, NIFD, UK Biobank	Transformer	-	1. Imbalanced dataset. 2. Training labels can be subjective and inconsistent.
[101]	ADNI, NACC, UK Biobank	Bayesian Monté Carlo dropout	-	Imbalanced dataset
[59]	UK Biobank	XGBoost+SHAP	APOE4 carriers AUC=0.75, Non-APOE4 carriers AUC= 0.64	1.Imbalanced dataset 2. Some deficits may be the earliest preclinical effects of dementia rather than true risk factors.
[60]	UK Biobank	LightGBM	-	1.Risk of overfitting 2.Due to study focused on a specific population of patients with ASCVD, therefore dataset size is small.
[61]	UK Biobank	UKB-DRP	ACC=0.764, SEN=0.774, SPE=0.764 AUC:0.848 for dementia, 0.862 for AD	1.Insufficient data diversity and external validation 2.Ignoring common predictors (e.g., BMI, blood pressure)
[62]	UK Biobank	Firth Logistic regression	-	1. Imbalanced dataset 2. Lower proportion of samples with earlier age of Bilateral oophorectomy onset may underestimate the impact of Bilateral oophorectomy on AD risk

[63]	UK Biobank	VGGNet, ResNet, DenseNet, EfficientNet, SFCN, GLT, 3D-ViT	MAE=2.6, MSE=10.74, $R^2=0.82$, Pearson R=0.91	1.The dataset limits the generalisability of the findings to other races. 2.Absence of Discovery-Replication Design 3. Lack of Ground-Truth Brain Age
[64]	UK Biobank	VGG-16	Acc=0.714	1.Imbalanced dataset 2. Retinal features may have similarities to other neurodegenerative diseases (e.g. Parkinson's disease), affecting model diagnostic performance
[65]	UK Biobank	SuperBigFLICA	-	-
[66]	ADNI, AIBL	HybNet	ADNI-ACC=0.919, SEN=0.887, SPE=0.945, AUC=0.965	Low resolution of attention maps
[67]	ADNI	MWAN	ADNI2-MMSE: CC=0.621 RMSE=1.503, CDRSB: CC=0.648 RMSE=5.701, ADAS-Cog: CC=0.613, RMSE=2.244	1. MWAN still requires High quality data. 2. The model tends to be biased toward the majority class.
[68]	ADNI	CNN+ResNet	Imaging features: C-index=0.762 Clinical measures with imaging: C-index=0.864	1.This study focused on hippocampus-only. 2.Late-MCI bias
[9]	ADNI, NACC, NIFD, AIBL, OASIS, FHS, LBDSU	CNN+CatBoost	NACC-Acc=0.777, F1=0.601, Sen=0.602, Spe=0.914, MCC=0.534	1.Did not contain any confirmed cases of atypical AD 2.Performance was lower for non-parkinsonian dementia
[69]	ADNI	LEAR	mAUC=0.8123, ACC=0.6715	1.Visual explanations need clinic validation 2. CMG relies on the performance of pre-trained diagnostic models
[70]	ADNI	wH-FCN	ACC=0.895, SEN=0.879, SPE=0.910, AUC=0.945	1.Dataset different distribution result in lower performance of model
[71]	GENCODE	Enformer	Pearson=0.85, Z-score=6.9, auROC=0.747,	Cannot generalise to unseen cell types or experimental tasks
[72]	ADNI	VGG16, Super Learner	AUC=0.908, ACC=0.825, SEN=0.755, SPE0.899	1.Limitations of image input and modality
[73]	ADNI, QMIN-MC, MACC	GMLVQ	ACC=0.816, AUC=0.84, SEN=0.823, SPE=0.809	Standardised issues in cognitive testing.
[74]	ADNI, AIBL	DA-MIDL	ACC=0.924, AUC=0.910, SEN=0.938, SPE=0.965	1.The size of input patches is fixed and equivalent. 2. The patch location proposals based on the group comparison are isolated from the subsequent network.
[75]	ADNI	FAMF	Acc=0.931, Sen=0.926, Spe=0.935, AUC=0.974	Using only two modalities (MRI and PET)
[76]	ADNI	MEM	ACC=0.988,	Limitations of the gradient

			AUC=0.992	significance approach
[77]	ADNI	ELMs	Acc=0.981, F1=0.980, MCC=0.969	1. Cognitive domain delineation and diagnostic modelling are not yet well developed 2. Polynomial expansion above a threshold leads to overfitting and memory problems
[78]	ADNI	CERF	Acc=0.913	Limited data modality and small sample size
[79]	ADNI	WERF,	Acc=0.889	The choice of the number of model evolutions may be influenced by the dataset and further validation of generalisability is required.
[80]	OASIS, ADNI	DRL-XGBOOST	Acc=0.902, Pre=0.923, Recall=0.954, F1=0.961	1.Imbalanced dataset 2.Tiny dataset results in overfitting
[102]	ADNI	MM-SDPN	ACC=0.969, SEN=0.942, SPE=0.713	Not combined with other biomarkers
[81]	ADNI	DISFC	AUC=0.962, Acc=0.929, Sen=0.889, Spe=0.953	ADNI dataset is white with high levels of education, which may affect model generalisability
[82]	ADNI	Inception V3+Q-learning+SVM	ACC=0.902, Sen=0.931, Spe=0.908, AUC=0.950	1.Imbalanced dataset 2. Introducing models of RL into Alzheimer classification
[83]	ADNI	Multi-task regression+ Gradient Boosting	-	Models do not fully deal with uncorrelated information between different tasks within the same modality
[84]	ADNI, MACC	MTM	-	MTM assumes that the state space model is linearly dynamic and may not fully capture non-linear disease progression
[85]	ADNI	Random Forest	Acc=0.940, F1=0.939	1.Only using baseline data, but AD is a chronic disease, therefore, time-series data analysis would be of critical value.
[86]	Self-collected dataset+ADNI	Gradient Boosting	BA=0.768, Sen=0.966, Spe=0.881, AUC=0.984	1.Imbalanced dataset 2. The optimal modal combination relies on α PET features, but the test is costly and carries a risk of radiation exposure
[87]	ADNI	CNN+RNN+AE+RB	Acc=0.90	1.Imbalanced dataset 2. Limited ability to generalise the model
[88]	ADNI	SVM+Random Forest	Acc=0.855, Pre=0.722, Spe=0.887, Sen=0.887, F1=0.816, AUC=0.931	Low quality unlabelled samples may affect performance
[89]	ADNI	Ensemble Integration+XGBoost+SDA E	AUC=0.81, F-measure=0.68	Underutilisation of longitudinal data
[90]	Self-collected dataset	SVM+Manually selected features	Acc=0.693, F1=0.682	Smaller sample size may result in lower performance
[95]	Self-collected dataset	CNN+Traditional machine learning	-	1.Difficulty in recognising early patients 2. No integration of personal information
[92]	Self-collected dataset	SMVA	ROC AUC: Control vs. Pre=0.74 Control vs. Patient=0.98 Pre vs. Patient=0.96	1.imbalanced dataset 2. First, there is no ground truth for the progression scores in real datasets
[96]	Self-collected dataset	CNN+SVM+LRP	-	1.Small sample size and high heterogeneity

				2. People who have not been assessed for mild cognitive impairment (MCI) or who are genetically at high risk cannot determine the effectiveness of early detection.
[98]	Taiwan NHI	MAND,	AUC=0.898, Sen=0.737, Spe=0.891	Imbalanced dataset
[16]	Self-collected dataset	SVM	Acc=0.944, Sen=1, Spe=0.909, Pre=0.875, F1=0.933, AUC=0.89	Small sample size
[91]	Self-collected dataset	WiGMM	ACC=0.86	1.Small sample size 2. Modal overlap of traditional indicators
[93]	Self-collected dataset	FCRN+MLP	-	Fixed size image block limit
[94]	Self-collected dataset	EEGNET+GoogLeNet+ Bagged Trees+EXML	Acc=0.994	1.Lack of robustness for channel absence 2. Lower multi-category classification performance
[97]	Self-collected dataset	SVM	Acc=0.986, Sen=0.974, Pre=1 Spe=0.979, F1=0.987,	1.Use of hand-designed features may miss potential information
[99]	Self-collected dataset	ViT+XGBoost	External: AUC=0.972, AP=0.979 Internal: AUC=0.902, AP=0.900	1. Data from central China, no global or national generalisation verified 2.Undifferentiated mixed dementia
[100]	Self-collected dataset	XGBoost+SHAP ROS and MAP	Acc=0.897, AUROC=0.898, AUPRC=0.639	1. Using data from only 3 years to predict dementia risk over the next 3 years may ignore longer-term cognitive changes 2. High rates of missing some cognitive tests
[103]	OASIS	Traditional machine learning method	-	1. No other potentially relevant clinical features were included 2. No external validation, cannot verified robustness of mode

Note: The exceptionally high performances reported in some of these studies can be attributed to specific methodological factors: (1) for [76] and [77], the absence of external validation likely inflated the results; (2) for [94], the use of a small but highly controlled dataset, extensive sample expansion, multimodal feature fusion, and pronounced disease-related electrophysiological signatures contributed to the elevated accuracy; and (3) for [97], the integration of rich gait features with optimised machine learning techniques in a controlled experimental setting facilitated unusually high performance.

4.3 Multimodal linguistic-based Cognitive Impairment Datasets

Beyond multimodal clinical phenotyping datasets, multimodal linguistic-based cognitive impairment datasets represent an equally vital research resource. These datasets offer a non-invasive and cost-effective methodology for detecting cognitive decline, particularly valuable for identifying early-stage or subtle impairments where traditional neuroimaging or biomarker data may yield inconclusive results. Capturing spontaneous or semi-structured speech and language patterns push the development of AI in speech data. Recent work is shown in Table 4:

Speech data and similarly improving model performance through multimodal fusion: Ilias [104] proposed the integration of BERT text encoding with DeiT visual models, combining contextual self-attention mechanisms and optimal transport cross-modal alignment to achieve cross-modal modelling of speech and text. Their model attained an F1-score of 91.06% on ADReSS and ADReSSo datasets, outperforming existing methods in both accuracy and model confidence calibration while demonstrating high robustness and clinical potential. Mahajan [105] achieved 3.5% accuracy gain via dynamic gated fusion of linguistic (72.92%) and audio (66.67%) channels while reducing parameters by 41%, benchmarking five neural architectures for efficient multimodal

integration. Ilias [106] developed a dual-channel framework combining text (BERT embeddings) and audio (wav2vec2 features) Transformers, which outperformed single-channel models in both AD classification and MMSE regression tasks on the ADReSS and Pitt datasets.

Linguistic feature engineering is also one way to improve model performance. Ortiz-Perez [107] demonstrated that text modality alone achieved 90.36% accuracy on Pitt corpus, with CNN-Transformer hybrids further enhancing robustness. Their interpretability analysis revealed characteristic language impairment patterns. Wen [108] achieved 92.2% accuracy (F1=0.955) using only 27 part-of-speech (PoS) features with Transformer models on DementiaBank. Their causal counterfactual analysis identified 12 critical PoS features for AD classification, 75% of which were newly discovered linguistic markers. Zheng [109] systematically evaluated three language models to assess the contributions of syntactic (PoS+stopwords) versus semantic (contextual words) information, finding that combined input improved accuracy to 81.54%, confirming that language decline involves both semantic and syntactic degradation. Ilias [110] developed a multi-task BERT framework with Siamese networks for joint AD diagnosis and MMSE prediction (F1=0.89, RMSE=3.2), identifying dementia-specific semantic degradation patterns through LIME interpretation. Priyadarshinee [111] established file-level text features as most stable (88.7% accuracy) through granularity-modality dual analysis, with audio features showing $\pm 12.3\%$ frame-level volatility in cross-dimensional evaluation.

Cross-language generalisability of high-performance models also performs well. Shah [29] developed language-agnostic features (pause distribution + word duration + word confidence) that achieved 69% accuracy in English to Greek transfer learning (MMSE RMSE=4.77), validating the universality of speech-based cognitive metrics. Mei [112] employed frequency-band clipping and wav2vec2.0 fine-tuning for cross-lingual AD detection, attaining 73.9% accuracy on Greek test sets and MMSE prediction RMSE=4.610 via eGeMAPS-silence feature fusion. Tamm [113] implemented semantic embedding-based transfer learning, achieving 71.6% AD detection accuracy and MMSE RMSE=4.04 in cross-language applications. Xuchu [114] pioneered language-agnostic AD detection combining prosodic features and wav2vec2 embeddings, maintaining smaller 5% performance drop in English to Greek transfer on ADReSS-M, the first validated multilingual solution.

To address computational constraints in real-world deployment, this study developed three efficient methods: Liu [115] developed a lightweight model using only VAD pause features (no transcription required) that achieved 80% accuracy across ADReSS, ADReSSo, and local Chinese-English datasets, outperforming traditional ComParE and eGeMAPS features. Chen [116] proposed SpeechFormer++, which efficiently models hierarchical speech structures (frame-phone-word) through unit encoders and merging blocks, reducing computational overhead while outperforming traditional Transformers.

Data augmentation has also been applied to speech data such as Lin [117] enhanced Pitt corpus data through synonym replacement, improving model accuracy to 80% (AUROC=90%) in low-resource scenarios. Woszczyk [118] systematically compared six text augmentation techniques, finding back-translation and random masking most effective (5-7% F1 improvement for BERT/RoBERTa). Jin [119] proposed CONSEN ensemble that won the ICASSP ADReSS-M challenge (ACC = 86.69%, MMSE RMSE=3.727) through innovative pause-wav2vec fusion.

Table 4 Multimodal linguistic-based cognitive impairment datasets related article

Publication	Dataset	AI Method	performance	Note
[104]	ADReSS, ADReSSo	BERT+DeiT+Context-aware self-attention+optimal transport domain adaptation.	Pre=0.930, Recall=0.891, F1=0.910, ACC=0.912, Spe=0.933	1. Dependent supervised learning 2. Performance degradation when stacking too many context-aware self-attention layers
[120]	I-CONET	Multimodal Cross-Transformer-Based	-	1. Interviewer-participant interactions not analysed

		model		2. Small sample size
[117]	Pitt Cookie Theft	LSTM+ Fully connect layer	Acc=0.845, Pre=0.836, Recall=0.799, F1=0.816, AUROC=0.931	1. Multimodal fusion method is simple 2. Using data from only a single task limits the model's ability to generalise to another dataset
[107]	Pitt Corpus	Mel Spectrogram+ CN+BERT+BiLSTM	ACC Audio=0.734, □ Text=0.903, Multimodal=0.86 6	1.Multimodal models did not significantly improve performance 2.Imbalanced dataset
[110]	ADReSS	Transformer- Based+Siamese+MTL	-	1.No integration of multimodal data 2.small size dataset
[108]	Pitt corpus	CNN-Attention-Feature Network	Acc=0.922, Pre=0.935, Recall=0.971, F1=0.955	The current study focused only on lexical (PoS) features and did not integrate other modal data.
[116]	Pitt corpus	SpeechFormer	Acc=0.813, F1=0.808	Sensitivity to statistical length, the longer the statistical length, the worse the model performance
[109]	Pitt corpus	Attention-based model	Acc=0.815	Small sample size
[121]	Pitt corpus	BERT + BiLSTM	Acc=0.81, Pre=0.79, Recall=0.79, F1=0.79	Imbalanced dataset
[111]	ADReSS	Otter.ai transcription+ RoBERTa+ DNN	Acc=0.887	1.Lack of validation sets 2. Limits to data labelling and quality
[115]	ADReSS	VAD Pause feature and eGeMAPS + Tree bagger	Acc=0.707	1.Small sample size 2. Deep learning models not used
[29]	ADReSS	Logistic regression+L2+SVM	Acc=0.695, RMSE=4.769	1.Cross-language performance degradation 2.Small sample size
[105]	ADReSS	Bi-modal model (GRU + CNN + Bi-LSTM + attention Layer)	Acc=0.729, Pre=0.789, Recall=0.625, F1=0.697	1.Dependence on manual transcription 2. Bimodal models did not improve performance
[112]	ADReSS	SVR+XGBoost	Acc=0.739, RMSE=4.61	Bilingual pre-training relies on internal non-public data
[122]	ADReSS	LSTM + MLP	Acc=0.891, Pre=0.935, Recall=0.923	1.Inbalanced dataset 2. Data heterogeneity leads to performance degradation in non-IID scenarios
[114]	ADReSS	OpenSmile tool + SVM XLSR-53 + fully connected layer	Acc=0.696, Pre=0.692, Recall=0.75, F1=0.72, RMSE=4.788	1. Insufficient ability to generalise across languages 2.Easily overfitting
[106]	ADReSS	BERT + ViT + Crossmodal attention	Acc=0.883, Pre=0.861,	1. The accuracy of BERT+ViT was slightly lower than BERT

			Recall=0.916, F1=0.887	alone. 2. Model robustness is not fully validated
[113]	ADReSS	batch normalisation + down-projected feature + attention pooling + linear projection	Acc=0.826, Pre=0.889, Recall=0.727, F1=0.80, RMSE=4.345	1. Cross-language transfer sample sizes are too small just 8 samples. 2. Model is simple, Limited performance improvement
[118]	ADReSS	Transformer + SVM + BERT	Acc=0.86	Data enhancement by simply increasing the training set tends to cause overfitting
[119]	ADReSS	wav2vec+x-vector feature	Acc=0.869, RMSE=3.727	1. Insufficient validation of cross-language generalisation capabilities 2. Lack of model interpretability

4.4 Summarisation based on all multimodal datasets

Sections 4.2 Multimodal clinical phenotyping datasets and 4.3 Multimodal linguistic-based Cognitive Impairment Datasets reviewed the status of computer-aided AD diagnosis. To enable a more systematic analysis of the included studies, this paper further categorised the literature based on the datasets employed and identified representative modality combinations and high-performance model architectures within each category. This summary covers four major types of datasets: UK Biobank, ADNI, Dementia Bank, and self-collected datasets.

In the UK Biobank, studies predominantly adopted the modality combination of MRI, clinical variables, and genetic features [59] [61] [64]. Frequently used models included interpretable methods such as XGBoost with SHAP [59] and the UKB-DRP framework [61] both achieving strong predictive performance. In addition, some researchers explored Transformer-based architectures [58] which, although not among the top-performing methods, still yielded promising results.

In the ADNI dataset, representative modality combinations included MRI + PET and MRI combined with clinical and cognitive scores [67] [75] [76] [123]. Models such as GCN [123], MEM [76], and FAMF [75] achieved high classification accuracy, with reported AUCs reaching up to 0.999. While standard Transformer architectures have not yet been widely adopted in ADNI studies, several attention-based models (e.g., MEM, DISFC) demonstrate a progressive shift toward Transformer inspired paradigms.

In contrast, studies based on the Dementia Bank datasets primarily focus on speech and language modalities, including combinations such as speech + text and acoustic + part-of-speech (PoS) features [104], [106], [107], [108]. In this domain, Transformer-based models particularly BERT+DeiT[104], BERT+ViT[106], and RoBERTa+DNN [111] have achieved outstanding performance, with F1 scores consistently exceeding 0.91. These results underscore the superior ability of Transformers to model complex semantic and acoustic relationships and indicate that Transformer-based methods have become the dominant paradigm in this subdomain.

In self-collected datasets, common modality combinations include EEG and structural MRI integrated with clinical or demographic information [97], [94], [99]. Among these, models such as ViT+XGBoost [99] and EEGNet with Bagged Trees [94] have demonstrated strong predictive capacity. Notably, ViT and other visual Transformer models have shown considerable adaptability in small-scale and heterogeneous clinical datasets.

In summary, Transformer-based models exhibit outstanding performance in speech- and language-related tasks, whereas their application in multimodal clinical phenotyping datasets remains underexplored. Future research should prioritise the development of cross-modal Transformer frameworks capable of effectively integrating heterogeneous data modalities to further enhance generalisation and computer-aided AD diagnostic performance.

Table 5 Summary of representative modality combinations and top-performing models in multimodal Artificial Intelligence-aided Alzheimer's disease diagnosis

Dataset	Representative Modality Combinations	Representative model	Related article
UK Biobank	MRI + Clinical + Genetic [59] [61] [64]	XGBoost+SHAP [59], UKB-DRP [61]	[58], [101], [59], [60], [61], [62] [63], [64], [65]
ADNI	MRI + PET, MRI + Clinical + Cognitive Scores [67] [75] [76] [123]	GCN[123], MEM [76], FAMF [75], CNN+ResNet [68]	[58], [67], [68], [9], [69] [70], [72], [73] [74], [75], [76], [77] [78], [79] [80], [102], [81], [82], [83], [84], [85], [86], [87], [88], [89], [123]
Dementia bank	Speech + Text, Acoustic features + ASR, PoS + Acoustic Features [104, 106–108]	BERT+DeiT [104], BERT+ViT [106], CNN-Attn-FN [108], RoBERTa+DNN [111]	Table 4
Self-collected datasets	EEG + Deep Learning, MRI + Clinical + Demographic [97] [94] [99]	EEGNet+BT[94], ViT+XGBoost [99]	[86], [90], [95], [92], [96], [98], [124], [16], [91] [97], [93], [94], [99], [100]

5. Challenges and Future Directions

In recent years, multimodal models have demonstrated remarkable potential in computer-aided diagnosis and risk prediction for AD. While these methods have achieved significant successes, a number of challenges remain that warrant careful examination. In this section, this paper summarise the common limitations identified in existing studies and propose directions for future research to advance the field.

5.1 Dataset-related Challenges and Future Directions

5.1.1 Data privacy and data-sharing constraints

Access to multimodal AD data remains severely restricted by privacy regulations and ethical constraints, which limit data sharing and external validation. This restricts the sharing and utilisation of comprehensive datasets needed for robust external validation and generalisability.

Federated learning (FL) provides a technically viable privacy-preserving solution; however, differences in data formats and institutional infrastructures still impede its large-scale deployment. For instance, Meerza [122] pioneered FL for AD speech diagnosis using MFCC and pause features, maintaining model performance while ensuring privacy through q-FedAvg/q-FedSGD optimisation. Nambiar [121] validated an ALBERT + BiLSTM hybrid model on the ADReSS dataset, achieving strong performance without compromising data confidentiality. In parallel, multi-institutional collaborations leveraging publicly available datasets such as ADNI, UK Biobank, and OASIS have enabled richer external validation while adhering to rigorous privacy standards [58], [101], [66], [9], [73], [80], [104].

Despite encouraging demonstrations, FL still lacks harmonised protocols and interoperable platforms. Without such standardisation, results from one centre cannot be reliably reproduced by

another, undermining clinical credibility. Moreover, international data-sharing partnerships remain complex and often limited by varying regulations. Future research should prioritise the development of unified federated frameworks with standardised protocols and privacy-preserving computational methods (e.g., blockchain integration, secure multi-party computation) to further enable secure global data collaboration [125], [126].

5.1.2 Data imbalance

Severe class imbalance, large numbers of healthy controls but few confirmed AD cases, remains a fundamental obstacle. It biases model training towards the majority class and inflates reported accuracy while masking poor sensitivity to early disease. Additionally, datasets such as the UK Biobank predominantly consist of individuals of White European ancestry, which severely limits model generalisability across racially and ethnically diverse populations.

Researchers have applied data-level interventions such as GAN-based augmentation, Diffusion models and Re-sampling [127], [128] [117], [118], [119]; Algorithm-level solutions, including cost-sensitive, loss focal, ensemble and class-weighted training schemes methods [129], [130], [131], [132], [133], [134]; Evaluation-focused remedies [135], have been developed to mitigate biases.

Current methods frequently introduce new challenges such as overfitting or inadequate performance in minority classes. Moreover, efforts to increase diversity remain inadequate. Future directions should focus on novel adaptive resampling methods, generative methods for synthetic minority data creation, and dedicated efforts to include and characterise underrepresented populations to ensure equitable and robust clinical applicability across diverse populations.

5.1.4 Lack of standardised and longitudinal data

A major limitation is the absence of standardised acquisition and diagnostic protocols across datasets. Inconsistent imaging sequences, cognitive scales, and biomarker assays make cross-study comparison unreliable and hinder meta-analysis. Furthermore, the longitudinal nature of AD progression demands sustained and consistent follow-ups, yet comprehensive longitudinal data collection remains scarce and fragmented due to logistic, financial, and participant retention challenges. Although ADNI has provided partially standardised longitudinal data and attempts at harmonising multimodal data have been explored in cross-sectional datasets, significant variability and fragmentation still limit the utility of existing datasets for capturing comprehensive temporal dynamics of AD progression. Future research should prioritise establishing a unified global framework for standardised, multimodal longitudinal data collection, emphasising harmonised protocols across institutions. Such initiatives would enhance data compatibility, enable robust temporal modelling, and facilitate better reproducibility and clinical translation.

5.1.1 Dataset-specific limitations

While data imbalance represents a cross-cutting challenge across most resources, each major cohort also exhibits its own unique shortcomings. To illustrate this, this paper summarises the dataset-specific limitations of widely used Alzheimer's disease cohorts and corpora.

ADNI: ADNI participants are generally healthier, with fewer comorbidities and a restricted age range (55–90 years), limiting representativeness. Protocol differences across centres and evolving diagnostic standards introduce heterogeneity, while frequent reliance on subsets hampers comparability [136].

UK Biobank: Dementia outcomes are derived mainly from health records, leading to potential misclassification and delayed ascertainment. Participants show strong volunteer bias, and PET/CSF biomarkers are limited to a small subset, constraining multimodal analyses [137].

OASIS: OASIS provides open neuroimaging data but with relatively small AD/MCI sample sizes and inconsistent modality coverage. Limited longitudinal depth and cross-scanner variability further reduce reproducibility [138].

NACC: Data are aggregated from multiple centres with heterogeneous recruitment and diagnostic protocols, making harmonisation challenging. The cohort is clinic-based rather than population-representative, and missing biomarker modalities are common [139].

AIBL: Although high-quality, AIBL is smaller than ADNI and NACC and often used only for validation. Regional recruitment and protocol differences reduce ethnic diversity and cross-cohort comparability [140].

Pitt Corpus: This is the most widely used speech dataset but remains small and imbalanced. Tasks are constrained (e.g., picture description), limiting ecological validity, and cross-linguistic generalisability is poor [141].

ADReSS Series: The ADReSS benchmarks provide standardised speech corpora but are modest in size and restricted to English. Narrow task design and small training partitions raise concerns of overfitting and limited external validity [12].

Self-collected cohorts: Locally collected datasets often involve small, single-site samples with heterogeneous acquisition protocols. Missing modalities, limited follow-up, and selection bias further restrict their generalisability [135].

In summary, dataset-related challenges arise both from overarching issues, such as privacy restrictions, limited data sharing, and severe class imbalance, and from the inherent shortcomings of widely used cohorts, including restricted representativeness, incomplete modality coverage, and heterogeneity across sites. These limitations constrain the robustness and generalisability of AI models for AD diagnosis. Overcoming them will require not only harmonisation and standardisation efforts, but also the development of strategies for secure data sharing, augmentation using generative models, and large-scale cross-cohort validation. Ultimately, methodological advances must be tightly coupled with improvements in data quality and accessibility to enable clinically reliable multimodal AI systems.

5.2 Model-related Challenges and Future Directions

5.2.1 Model interpretability and explainability

A major limitation of applying multimodal machine learning models in clinical AD diagnosis is the lack of interpretability and transparency. Complex "black box" models, although accurate, often fail to clearly articulate decision-making processes, which severely limits their acceptance by clinicians and trust from patients.

Efforts have been made towards model interpretability include designing inherently transparent models. For example, He [142] proposed the NeuroSymAD framework, which integrates a deep neural network for MRI image analysis with an LLM-driven symbolic reasoning module, combining patient biomarkers and clinical history. This framework can generate interpretable diagnostic reports, offering strong clinical interpretability and scalability. Explainability methods employ external methods to interpret opaque models. Sadeghi [143] reviewed interpretable graph neural network architectures, leveraging SHAP, gradient saliency, and graph masking for explaining multimodal AD diagnosis outcomes. Additionally, Mahamud [144] utilised ensemble methods (LightGBM + Random Forest) with SHAP and LIME techniques to clearly highlight feature importance at both global and individual patient levels.

Current interpretability frameworks still present gaps in translating machine explanations effectively into clinical decision-making processes. Future work should focus on developing clinically oriented explainability methods integrated with interactive visualisation and natural language explanation tools, specifically tailored to clinician workflows. Additionally, there is considerable scope for enhancing hybrid neuro-symbolic approaches, combining symbolic reasoning with deep learning to ensure more interpretable and trustworthy models.

Beyond purely technical solutions, future research should also integrate Patient and Public

Involvement (PPI) frameworks into the development of multimodal AI systems for AD. PPI ensures that patients, caregivers, and the wider public are not only end-users but active stakeholders in shaping how models are trained, evaluated, and most importantly, communicated in clinical settings. Incorporating PPI can help align algorithmic explanations with patient values, improve trust and adoption by clinicians, and mitigate ethical concerns regarding transparency and fairness. Embedding PPI at early stages of AI design may therefore represent a crucial step towards building interpretable, trustworthy, and clinically actionable diagnostic tools.

5.2.2 Heterogeneous multi-view learning problem

Another major challenge arises from the integration of data originating from diverse studies and sources. Given the limited modality coverage within individual datasets, researchers often resort to combining datasets such as ADNI and the UK Biobank. However, each dataset is typically constructed around distinct cohorts, imaging protocols, or cognitive assessment frameworks, resulting in heterogeneous multi-view learning problem.

This heterogeneity complicates the development of unified models capable of generalising across datasets, particularly when sample populations do not overlap. Consequently, models trained on a single dataset often struggle to generalise beyond their original domain. To address this, there is a pressing need for platform-agnostic models that can accommodate missing or inconsistent modalities. Approaches such as shared latent space learning [145], multi-branch neural networks [146], and mixture-of-experts architectures [147] have been proposed to enable partial modality fusion and flexible adaptation across datasets. Nevertheless, most current methods either rely on strong alignment assumptions or require extensive retraining when faced with domain shifts.

Despite advancements, existing methods frequently rely on rigid alignment assumptions and extensive retraining upon encountering domain shifts or missing data scenarios. A future priority is developing robust, platform-agnostic frameworks that dynamically adapt to varying modality availability and distributional shifts without significant performance loss. Additionally, research should explore advanced representation learning techniques (e.g., contrastive learning, multimodal self-supervision) that can effectively learn robust joint representations from heterogeneous data.

5.2.3 Uncertainty quantification and clinical applicability

Despite notable advancements in multimodal predictive models for AD, most AI-based studies currently overlook uncertainty quantification, such as providing confidence intervals alongside model predictions. Typically, models yield deterministic outputs without clearly communicating the reliability or variability of these predictions. However, clinicians frequently emphasise the importance of such uncertainty measures, as confidence intervals greatly influence their decision-making process regarding patient management and treatment strategies. Future research should therefore explicitly incorporate uncertainty quantification metrics such as confidence intervals or predictive intervals, into diagnostic models. By aligning more closely with clinical needs, these enhancements can substantially improve model interpretability, reliability, and acceptance in real-world healthcare settings.

6. Conclusion

AD has become an increasing global concern in the context of rapid population ageing. Accurate and early diagnosis is critical for timely intervention, patient care, and effective disease management. This paper systematically examined the evolution of computer-aided diagnostic techniques for AD, tracing developments from traditional machine learning methods to more advanced deep learning, ensemble learning, reinforcement learning, and multimodal approaches. This paper also conducted a detailed synthesis of recent studies published over the past five years, summarising the datasets, methods, performance metrics, and limitations associated with multimodal models, as shown in

Table 3 and Table 4. Based on our analysis, several key insights emerge:

1. Multimodal models consistently outperform single-modality models in terms of diagnostic accuracy and generalisation, offering a more comprehensive representation of the disease.
2. Data imbalance remains a pervasive challenge in medical datasets and should be explicitly addressed during model development to avoid biased or misleading predictions.
3. FL offers a promising solution for protecting patient privacy, allowing collaborative model training across institutions without the need for centralised data sharing.
4. Explainability and interpretability are not just technical enhancements but essential components of clinical acceptance, improving transparency and fostering trust among both clinicians and patients.
5. While multimodal biomedical cohort datasets and multimodal linguistic-based cognitive impairment datasets each provide valuable and complementary insights into AD pathology, they are often treated in isolation in existing research. However, in real-world clinical practice, physicians interpret behavioural and structural information simultaneously. Bridging this disconnect through cross-modal fusion strategies represents a critical step toward more holistic, reliable, and patient-centred diagnostic systems.
6. Despite notable advancements in multimodal predictive models for AD, most current models predominantly omit uncertainty quantification such as providing confidence intervals and instead yield deterministic outputs that fail to reflect prediction reliability. Crucially, clinical decision-making heavily relies on such uncertainty assessments. Future research must integrate quantitative metrics like confidence intervals into diagnostic models.
7. Existing benchmark datasets such as ADNI, OASIS, NACC, and ADRess remain indispensable but present important limitations, including small sample sizes, cohort imbalance, restricted diversity, and inconsistent longitudinal follow-up. These constraints hinder model generalisability and highlight the urgent need for more inclusive, standardised, and scalable multimodal datasets.

Reference

1. Scheltens P, De Strooper B, Kivipelto M, et al (2021) Alzheimer's disease. *Lancet* 397:1577–1590. [https://doi.org/10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4)
2. (2024) 2024 alzheimer's disease facts and figures. *Alzheimer's Dement* 20:3708–3821. <https://doi.org/10.1002/alz.13809>
3. World-Alzheimer-Report-2024
4. Kaštelan S, Gverović Antunica A, Puzović V, et al (2025) Non-invasive retinal biomarkers for early diagnosis of alzheimer's disease. *Biomedicines* 13:283. <https://doi.org/10.3390/biomedicines13020283>
5. Wijeratne PA, Alexander DC (2021) Learning transition times in event sequences: the Event-Based Hidden Markov Model of disease progression
6. Huh YJ, Park J, Kim YJ, Kim KG (2025) Ensemble Learning-Based Alzheimer's Disease Classification Using Electroencephalogram Signals and Clock Drawing Test Images. *Sensors* 25:2881. <https://doi.org/10.3390/s25092881>
7. Karasu E, Baytaş İM (2025) Conversion-aware forecasting of Alzheimer's disease via featurewise attention. *Pattern Anal Applic* 28:64. <https://doi.org/10.1007/s10044-025-01447-4>

8. Xiao X, Li Y, Wu Q, et al (2025) Development and validation of a novel predictive model for dementia risk in middle-aged and elderly depression individuals: a large and longitudinal machine learning cohort study. *Alz Res Therapy* 17:103. <https://doi.org/10.1186/s13195-025-01750-6>
9. Qiu S, Miller MI, Joshi PS, et al (2022) Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat Commun* 13:3404. <https://doi.org/10.1038/s41467-022-31037-5>
10. Chakravarthi BA, Shivakanth G (2025) Integrating Multimodal AI Techniques and MRI Preprocessing for Enhanced Diagnosis of Alzheimer's Disease: Clinical Applications and Research Horizons. *IEEE Access* 13:63519–63531. <https://doi.org/10.1109/ACCESS.2025.3557533>
11. Elazab A, Wang C, Abdelaziz M, et al (2024) Alzheimer's disease diagnosis from single and multimodal data using machine and deep learning models: Achievements and future directions. *Expert Systems with Applications* 255:124780. <https://doi.org/10.1016/j.eswa.2024.124780>
12. Chetty M, Noori Hoshyar A, Bhattacharya T, Klein B (2024) Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges. *Artif Intell Rev* 57:325. <https://doi.org/10.1007/s10462-024-10961-6>
13. Page MJ, McKenzie JE, Bossuyt PM, et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71. <https://doi.org/10.1136/bmj.n71>
14. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297. <https://doi.org/10.1007/BF00994018>
15. Sharma A, Kaur S, Memon N, et al (2021) Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis. *Neuroscience Informatics* 1:100012. <https://doi.org/10.1016/j.neuri.2021.100012>
16. Park B, Kim Y, Park J, et al (2024) Integrating Biomarkers From Virtual Reality and Magnetic Resonance Imaging for the Early Detection of Mild Cognitive Impairment Using a Multimodal Learning Approach: Validation Study. *J Med Internet Res* 26:e54538. <https://doi.org/10.2196/54538>
17. Gao X, Liu H, Shi F, et al (2023) Brain Status Transferring Generative Adversarial Network for Decoding Individualized Atrophy in Alzheimer's Disease. *IEEE J Biomed Health Inform* 27:4961–4970. <https://doi.org/10.1109/JBHI.2023.3304388>
18. Romano MS, Tuch SAN, Bustamante LMC Selection of Acoustic, Temporal, and Complexity Features for Machine Learning Classification of Alzheimers Disease on a Spanish Population Through Automatic Analysis of Reading-Elicited Speech

19. Lazli L (2025) Improved Alzheimer Disease Diagnosis With a Machine Learning Approach and Neuroimaging: Case Study Development. *JMIRx Med* 6:e60866–e60866. <https://doi.org/10.2196/60866>
20. Hossain F, Halder RK, Uddin MN (2025) An integrated machine learning based adaptive error minimization framework for Alzheimer's stage identification. *Intelligence-Based Medicine* 11:100243. <https://doi.org/10.1016/j.ibmed.2025.100243>
21. Fulkar B, Dhale T, Pacharaney U, Deshmukh S (2025) Early Detection of Chronic Diseases Using Machine and Deep Learning Algorithms. In: 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL). IEEE, Bhimdatta, Nepal, pp 1656–1661
22. Sathiya A, Basha CH, S Vasanth, et al (2025) Enhancing Alzheimer's Disease Detection Using Optimized Attribute Selection and Random Forest Classifier for Improved Accuracy. In: 2025 International Conference on Visual Analytics and Data Visualization (ICVADV). IEEE, Tirunelveli, India, pp 1174–1179
23. Saleh AW, Gupta G, Khan SB, et al (2023) An Alzheimer's disease classification model using transfer learning Densenet with embedded healthcare decision support system. *Decision Analytics Journal* 9:100348. <https://doi.org/10.1016/j.dajour.2023.100348>
24. Baucum M, Khojandi A, Papamarkou T (2021) Hidden Markov models as recurrent neural networks: An application to Alzheimer's disease. In: 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE). IEEE, Kragujevac, Serbia, pp 1–6
25. Cai Z, Zeng D, Marder KS, et al (2024) Dynamic Classification of Latent Disease Progression with Auxiliary Surrogate Labels
26. Chen Y, Pham TD (2013) Development of a brain MRI-based hidden Markov model for dementia recognition. *BioMed Eng OnLine* 12:S2. <https://doi.org/10.1186/1475-925X-12-S1-S2>
27. Vats NA, Yadavalli A, Gurugubelli K, Vuppala AK (2021) ACOUSTIC FEATURES, BERT Model AND THEIR COMPLEMENTARY NATURE FOR ALZHEIMER'S DEMENTIA DETECTION. In: 2021 Thirteenth International Conference on Contemporary Computing (IC3-2021). ACM, Noida India, pp 267–272
28. Xiao R, Cui X, Qiao H, et al (2021) Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net. *Biomedical Signal Processing and Control* 66:102362. <https://doi.org/10.1016/j.bspc.2020.102362>
29. Shah Z, Qi S-A, Wang F, et al (2023) Exploring Language-Agnostic Speech Representations Using Domain Knowledge for Detecting Alzheimer's

- Dementia. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Rhodes Island, Greece, pp 1-2
30. Ablimit A, Botelho C, Abad A, et al (2022) Exploring Dementia Detection from Speech: Cross Corpus Analysis. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Singapore, Singapore, pp 6472-6476
 31. Lahmiri S (2023) Integrating convolutional neural networks, kNN, and Bayesian optimization for efficient diagnosis of Alzheimer's disease in magnetic resonance images. *Biomedical Signal Processing and Control* 80:104375. <https://doi.org/10.1016/j.bspc.2022.104375>
 32. Suwalka D, Pandita D, Godse S, et al (2024) AI Applications and Simulation-Based Learning Integrating Future of Nursing Education. In: 2024 International Conference on Intelligent & Innovative Practices in Engineering & Management (IIPEM). IEEE, Singapore, Singapore, pp 1-6
 33. Chaudhari A, Saratkar S, Thute T (2025) AI-Enhanced Imaging Techniques for Understanding Alzheimer's Progression. In: 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS). IEEE, Prawet, Thailand, pp 1174-1179
 34. Ango R, C KKR, Fatima S, Nag A (2024) Brain Connectivity Analysis in Alzheimer's disease using Graph Convolutional Network. In: 2024 4th International Conference on Soft Computing for Security Applications (ICSCSA). IEEE, Salem, India, pp 133-139
 35. Chattopadhyay T, Joshy NA, Ozarkar SS, et al (2024) Deep Learning Algorithms for Alzheimer's Disease Detection based on Diffusion MRI: Tests in Indian and North American Cohorts. *Alzheimer's & Dementia* 20:e089294. <https://doi.org/10.1002/alz.089294>
 36. Ma D, Zhang H, Wang L (2024) Editorial: Deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders. *Front Neurosci* 18:1497417. <https://doi.org/10.3389/fnins.2024.1497417>
 37. De Swart WK, Loog M, Krijthe JH (2025) A comparative study of methods for dynamic survival analysis. *Front Neurol* 16:1504535. <https://doi.org/10.3389/fneur.2025.1504535>
 38. Kang M-K, Hong K-S, Yang D, Kim HK (2025) Multi-scale neural networks classification of mild cognitive impairment using functional near-infrared spectroscopy. *Biocybernetics and Biomedical Engineering* 45:11-22. <https://doi.org/10.1016/j.bbe.2024.12.001>
 39. Mohammad Hasan Sarwer, Tui Rani Saha, Abir SI, et al (2025) EEG

Functional Connectivity and Deep Learning for Automated Diagnosis of Alzheimer's disease and Schizophrenia. JCSTS 7:82-99. <https://doi.org/10.32996/jcsts.2025.7.1.7>

40. Sathish R, Muthukumar R, Dhivya K, Karthikkumar S (2025) Deep Learning and IoT-Enabled Framework for Accurate Classification and Monitoring of Alzheimer's Disease Based on EEG Signal Analysis. In: 2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE, Bhilai, India, pp 1-8
41. Dubey AK, Kapoor R, Saraswat M (2024) Optimized Machine Learning for Medical Data Analysis and Disease Prediction. In: 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit). IEEE, Greater Noida, India, pp 1282-1286
42. K P, Chitla VB, Aftab A, Kamath S (2025) LSTM-Based Assistance for People with Alzheimer's Disease. In: 2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE). IEEE, Bangalore, India, pp 1-5
43. Pan J, Fan Z, Smith GE, et al (2025) Federated learning with multi-cohort real-world data for predicting the progression from mild cognitive impairment to Alzheimer's disease. *Alzheimer's & Dementia* 21:e70128. <https://doi.org/10.1002/alz.70128>
44. Zuo Y, Zhang B, Dong Y, et al (2024) Glypred: Lysine Glycation Site Prediction via CCU-LightGBM-BiLSTM Framework with Multi-Head Attention Mechanism. *J Chem Inf Model* 64:6699-6711. <https://doi.org/10.1021/acs.jcim.4c01034>
45. Han X, Xue R, Feng J, et al (2025) Hypergraph Foundation Model for Brain Disease Diagnosis. *IEEE Trans Neural Netw Learning Syst* 1-15. <https://doi.org/10.1109/TNNLS.2025.3554755>
46. Lu S-Y, Zhang Y-D, Yao Y-D (2025) A regularized transformer with adaptive token fusion for Alzheimer's disease diagnosis in brain magnetic resonance images. *Engineering Applications of Artificial Intelligence* 155:111058. <https://doi.org/10.1016/j.engappai.2025.111058>
47. Li X, Zhu W, Qiu P, et al (2025) A BERT-Style Self-Supervised Learning CNN for Disease Identification from Retinal Images
48. Mahapatra C (2025) Exploring Advanced Applications of Artificial Intelligence in Neuropharmacology: A Comprehensive Overview
49. Ren H, Zheng Y, Li C, et al (2025) Using Machine Learning to Predict Cognitive Decline in Older Adults From the Chinese Longitudinal Healthy Longevity Survey: Model Development and Validation Study. *JMIR Aging* 8:e67437-e67437. <https://doi.org/10.2196/67437>

50. Shah YAR, Qureshi SM, Qureshi HA, et al (2024) Advances in Artificial Intelligence and Machine Learning for Neurodegenerative Disease: A Literature Review. *WJRR* 19:. <https://doi.org/10.31871/WJRR.19.3.8>
51. Williams C, Anik FI, Hasan MdM, et al (2025) Advancing Brain-Computer Interface Closed-Loop Systems for Neurorehabilitation: A Systematic Review of AI and Machine Learning Innovations in Biomedical Engineering (Preprint)
52. Fatima G, Ashiquzzaman A, Kim SS, et al (2025) Vascular and glymphatic dysfunction as drivers of cognitive impairment in Alzheimer's disease: Insights from computational approaches. *Neurobiology of Disease* 208:106877. <https://doi.org/10.1016/j.nbd.2025.106877>
53. Yang X, Dang X, Cai J, et al (2025) Temporal-multimodal consistency alignment for Alzheimer's cognitive assessment prediction. *Medical Physics* mp.17767. <https://doi.org/10.1002/mp.17767>
54. Sadeghian R, Haider F, Fraser K, et al (2024) Editorial: Methods in artificial intelligence for dementia 2024. *Front Dement* 3:1444825. <https://doi.org/10.3389/frdem.2024.1444825>
55. Kale M, Wankhede N, Pawar R, et al (2024) AI-driven innovations in alzheimer's disease: integrating early diagnosis, personalized treatment, and prognostic modelling. *Ageing Res Rev* 101:102497. <https://doi.org/10.1016/j.arr.2024.102497>
56. Gkoumas D, Wang B, Tsakalidis A, et al (2024) A longitudinal multi-modal dataset for dementia monitoring and diagnosis. *Lang Resour Eval* 58:883–902. <https://doi.org/10.1007/s10579-023-09718-4>
57. Xu T, Wang X, Lun X, et al (2023) ADREFV: face video dataset based on human-computer interaction for alzheimer's disease recognition. *Comput Anim Virtual Worlds* 34:e2127. <https://doi.org/10.1002/cav.2127>
58. Xue C, Kowshik SS, Lteif D, et al (2024) AI-based differential diagnosis of dementia etiologies on multimodal data. *Nat Med* 30:2977–2989. <https://doi.org/10.1038/s41591-024-03118-z>
59. Allwright M, Mundell HD, McCorkindale AN, et al (2023) Ranking the risk factors for alzheimer's disease; findings from the UK biobank study. *Aging Brain* 3:100081. <https://doi.org/10.1016/j.nbas.2023.100081>
60. Gu Z, Liu S, Ma H, et al (2025) Estimation of Machine Learning-Based Models to Predict Dementia Risk in Patients With Atherosclerotic Cardiovascular Diseases: UK Biobank Study. *JMIR Aging* 8:e64148. <https://doi.org/10.2196/64148>
61. You J, Zhang Y-R, Wang H-F, et al (2022) Development of a novel dementia risk prediction model in the general population: A large, longitudinal, population-based machine-learning study. *eClinicalMedicine* 53:101665.

<https://doi.org/10.1016/j.eclinm.2022.101665>

62. Calvo N, McFall GP, Ramana S, et al (2024) Associated risk and resilience factors of Alzheimer's disease in women with early bilateral oophorectomy: Data from the UK Biobank. *Journal of Alzheimer's Disease* 102:119-128. <https://doi.org/10.3233/JAD-240646>
63. Yi F, Yuan J, Somekh J, et al (2025) Genetically supported targets and drug repurposing for brain aging: A systematic study in the UK Biobank. *Sci Adv* 11:eadr3757. <https://doi.org/10.1126/sciadv.adr3757>
64. Yousefzadeh N, Tran C, Ramirez-Zamora A, et al (2024) Neuron-level explainable AI for Alzheimer's Disease assessment from fundus images. *Sci Rep* 14:7710. <https://doi.org/10.1038/s41598-024-58121-8>
65. Gong W, Bai S, Zheng Y-Q, et al (2023) Supervised Phenotype Discovery From Multimodal Brain Imaging. *IEEE Trans Med Imaging* 42:834-849. <https://doi.org/10.1109/TMI.2022.3218720>
66. Lian C, Liu M, Pan Y, Shen D (2022) Attention-Guided Hybrid Network for Dementia Diagnosis With Structural MR Images. *IEEE Trans Cybern* 52:1992-2003. <https://doi.org/10.1109/TCYB.2020.3005859>
67. Lian C, Liu M, Wang L, Shen D (2022) Multi-Task Weakly-Supervised Attention Network for Dementia Status Estimation With Structural MRI. *IEEE Trans Neural Netw Learning Syst* 33:4056-4068. <https://doi.org/10.1109/TNNLS.2021.3055772>
68. Li H, Habes M, Wolk DA, et al (2019) A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's & Dementia* 15:1059-1070. <https://doi.org/10.1016/j.jalz.2019.02.007>
69. Oh K, Yoon JS, Suk H-I (2023) Learn-Explain-Reinforce: Counterfactual Reasoning and its Guidance to Reinforce an Alzheimer's Disease Diagnosis Model. *IEEE Trans Pattern Anal Mach Intell* 45:4843-4857. <https://doi.org/10.1109/TPAMI.2022.3197845>
70. Lian C, Liu M, Zhang J, Shen D (2020) Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural MRI. *IEEE Trans Pattern Anal Mach Intell* 42:880-893. <https://doi.org/10.1109/TPAMI.2018.2889096>
71. Avsec Ž, Agarwal V, Visentin D, et al (2021) Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods* 18:1196-1203. <https://doi.org/10.1038/s41592-021-01252-x>
72. Yang L, Wang X, Guo Q, et al (2021) Deep Learning Based Multimodal Progression Modeling for Alzheimer's Disease. *Statistics in Biopharmaceutical Research* 13:337-343. <https://doi.org/10.1080/19466315.2021.1884129>

73. Lee LY, Vaghari D, Burkhart MC, et al (2024) Robust and interpretable AI-guided marker for early dementia prediction in real-world clinical settings. *eClinicalMedicine* 74:102725. <https://doi.org/10.1016/j.eclinm.2024.102725>
74. Zhu W, Sun L, Huang J, et al (2021) Dual Attention Multi-Instance Deep Learning for Alzheimer's Disease Diagnosis With Structural MRI. *IEEE Trans Med Imaging* 40:2354–2366. <https://doi.org/10.1109/TMI.2021.3077079>
75. Zhang M, Cui Q, Lü Y, Li W (2024) A feature-aware multimodal framework with auto-fusion for Alzheimer's disease diagnosis. *Computers in Biology and Medicine* 178:108740. <https://doi.org/10.1016/j.compbio.2024.108740>
76. Velazquez M, Lee Y (2022) Multimodal ensemble model for Alzheimer's disease conversion prediction from Early Mild Cognitive Impairment subjects. *Computers in Biology and Medicine* 151:106201. <https://doi.org/10.1016/j.compbio.2022.106201>
77. Zhang M, Cui Q, Lü Y, et al (2024) A multimodal learning machine framework for Alzheimer's disease diagnosis based on neuropsychological and neuroimaging data. *Computers & Industrial Engineering* 197:110625. <https://doi.org/10.1016/j.cie.2024.110625>
78. Bi X, Hu X, Wu H, Wang Y (2020) Multimodal Data Analysis of Alzheimer's Disease Based on Clustering Evolutionary Random Forest. *IEEE J Biomed Health Inform* 24:2973–2983. <https://doi.org/10.1109/JBHI.2020.2973324>
79. Bi X, Xing Z, Zhou W, et al (2022) Pathogeny Detection for Mild Cognitive Impairment via Weighted Evolutionary Random Forest With Brain Imaging and Genetic Data. *IEEE J Biomed Health Inform* 26:3068–3079. <https://doi.org/10.1109/JBHI.2022.3151084>
80. Hashmi A, Barukab O (2023) Dementia classification using deep reinforcement learning for early diagnosis. *Appl Sci* 13:1464. <https://doi.org/10.3390/app13031464>
81. Wang Y, Gao R, Wei T, et al (2024) Predicting long-term progression of Alzheimer's disease using a multimodal deep learning model incorporating interaction effects. *J Transl Med* 22:265. <https://doi.org/10.1186/s12967-024-05025-w>
82. Hatami M, Yaghmaee F, Ebrahimpour R (2024) Investigating the potential of reinforcement learning and deep learning in improving alzheimer's disease classification. *Neurocomputing* 597:128119. <https://doi.org/10.1016/j.neucom.2024.128119>
83. Tabarestani S, Aghili M, Eslami M, et al (2020) A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *NeuroImage* 206:116317. <https://doi.org/10.1016/j.neuroimage.2019.116317>

84. Burkhart MC, Lee LY, Vaghari D, et al (2024) Unsupervised multimodal modeling of cognitive and brain health trajectories for early dementia prediction. *Sci Rep* 14:10755. <https://doi.org/10.1038/s41598-024-60914-w>
85. El-Sappagh S, Alonso JM, Islam SMR, et al (2021) A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Sci Rep* 11:2660. <https://doi.org/10.1038/s41598-021-82098-3>
86. Lee M-W, Kim HW, Choe YS, et al (2024) A multimodal machine learning model for predicting dementia conversion in Alzheimer's disease. *Sci Rep* 14:12276. <https://doi.org/10.1038/s41598-024-60134-2>
87. Nguyen H, Chu NN (2021) An Introduction to Deep Learning Research for Alzheimer's Disease. *IEEE Consumer Electron Mag* 10:72–75. <https://doi.org/10.1109/MCE.2020.3048254>
88. Yuan S, Li H, Wu J, Sun X (2021) Classification of Mild Cognitive Impairment With Multimodal Data Using Both Labeled and Unlabeled Samples. *IEEE/ACM Trans Comput Biol and Bioinf* 18:2281–2290. <https://doi.org/10.1109/TCBB.2021.3053061>
89. Cirincione A, Lynch K, Bennett J, et al (2024) Prediction of future dementia among patients with mild cognitive impairment (MCI) by integrating multimodal clinical data. *Heliyon* 10:e36728. <https://doi.org/10.1016/j.heliyon.2024.e36728>
90. Cassani R, Falk T (2019) Alzheimer's Disease Diagnosis and Severity Level Detection Based on Electroencephalography Modulation Spectral "Patch" Features. *IEEE J Biomed Health Inform* 1–1. <https://doi.org/10.1109/JBHI.2019.2953475>
91. Wu EQ, Peng X-Y, Chen S-D, et al (2022) Detecting Alzheimer's Dementia Degree. *IEEE Trans Cogn Dev Syst* 14:116–125. <https://doi.org/10.1109/TCDS.2020.3015131>
92. Kmetzsch V, Becker E, Saracino D, et al (2022) Disease Progression Score Estimation From Multimodal Imaging and MicroRNA Data Using Supervised Variational Autoencoders. *IEEE J Biomed Health Inform* 26:6024–6035. <https://doi.org/10.1109/JBHI.2022.3208517>
93. Zhang H, Ni M, Yang Y, et al (2025) Patch-based interpretable deep learning framework for Alzheimer's disease diagnosis using multimodal data. *Biomedical Signal Processing and Control* 100:107085. <https://doi.org/10.1016/j.bspc.2024.107085>
94. Fabietti M, Mahmud M, Lotfi A, et al (2023) Early Detection of Alzheimer's Disease From Cortical and Hippocampal Local Field Potentials Using an Ensembled Machine Learning Model. *IEEE Trans Neural Syst Rehabil Eng* 31:2839–2848. <https://doi.org/10.1109/TNSRE.2023.3288835>

95. Cilia ND, D'Alessandro T, De Stefano C, et al (2021) From Online Handwriting to Synthetic Images for Alzheimer's Disease Detection Using a Deep Transfer Learning Approach. *IEEE J Biomed Health Inform* 25:4243-4254. <https://doi.org/10.1109/JBHI.2021.3101982>
96. Mengoudi K, Ravi D, Yong KXX, et al (2020) Augmenting Dementia Cognitive Assessment With Instruction-Less Eye-Tracking Tests. *IEEE J Biomed Health Inform* 24:3066-3075. <https://doi.org/10.1109/JBHI.2020.3004686>
97. Seifallahi M, Mehraban AH, Galvin JE, Ghoraani B (2022) Alzheimer's Disease Detection Using Comprehensive Analysis of Timed Up and Go Test via Kinect V.2 Camera and Machine Learning. *IEEE Trans Neural Syst Rehabil Eng* 30:1589-1600. <https://doi.org/10.1109/TNSRE.2022.3181252>
98. Tsai H, Yang T-W, Ou K-H, et al (2024) Multimodal Attention Network for Dementia Prediction. *IEEE J Biomed Health Inform* 28:6918-6930. <https://doi.org/10.1109/JBHI.2024.3438885>
99. Fan F, Song H, Jiang J, et al (2024) Development and validation of a multimodal deep learning framework for vascular cognitive impairment diagnosis. *iScience* 27:110945. <https://doi.org/10.1016/j.isci.2024.110945>
100. Beebe-Wang N, Okeson A, Althoff T, Lee S-I (2021) Efficient and Explainable Risk Assessments for Imminent Dementia in an Aging Cohort Study. *IEEE J Biomed Health Inform* 25:2409-2420. <https://doi.org/10.1109/JBHI.2021.3059563>
101. Azevedo T, Bethlehem RAI, Whiteside DJ, et al (2023) Identifying healthy individuals with alzheimer's disease neuroimaging phenotypes in the UK biobank. *Commun Med* 3:100. <https://doi.org/10.1038/s43856-023-00313-w>
102. Shi J, Zheng X, Li Y, et al (2018) Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease. *IEEE J Biomed Health Inform* 22:173-183. <https://doi.org/10.1109/JBHI.2017.2655720>
103. Battineni G, Hossain MA, Chintalapudi N, et al (2021) Improved Alzheimer's Disease Detection by MRI Using Multimodal Machine Learning Algorithms. *Diagnostics* 11:2103. <https://doi.org/10.3390/diagnostics11112103>
104. Ilias L, Askounis D (2023) Context-aware attention layers coupled with optimal transport domain adaptation and multimodal fusion methods for recognizing dementia from spontaneous speech. *Knowledge-Based Systems* 277:110834. <https://doi.org/10.1016/j.knosys.2023.110834>
105. Mahajan P, Baths V (2021) Acoustic and Language Based Deep Learning Approaches for Alzheimer's Dementia Detection From Spontaneous Speech. *Front Aging Neurosci* 13:623607. <https://doi.org/10.3389/fnagi.2021.623607>

106. Ilias L, Askounis D, Psarras J (2023) Detecting dementia from speech and transcripts using transformers. *Computer Speech & Language* 79:101485. <https://doi.org/10.1016/j.csl.2023.101485>
107. Ortiz-Perez D, Ruiz-Ponce P, Tomás D, et al (2023) A Deep Learning-Based Multimodal Architecture to predict Signs of Dementia. *Neurocomputing* 548:126413. <https://doi.org/10.1016/j.neucom.2023.126413>
108. Wen B, Wang N, Subbalakshmi K, Chandramouli R (2023) Revealing the Roles of Part-of-Speech Taggers in Alzheimer Disease Detection: Scientific Discovery Using One-Intervention Causal Explanation. *JMIR Form Res* 7:e36590. <https://doi.org/10.2196/36590>
109. Zheng C, Bouazizi M, Ohtsuki T (2022) An Evaluation on Information Composition in Dementia Detection Based on Speech. *IEEE Access* 10:92294–92306. <https://doi.org/10.1109/ACCESS.2022.3203068>
110. Ilias L, Askounis D (2022) Explainable Identification of Dementia From Transcripts Using Transformer Networks. *IEEE J Biomed Health Inform* 26:4153–4164. <https://doi.org/10.1109/JBHI.2022.3172479>
111. Priyadarshinee P, Clarke CJ, Melechovsky J, et al (2023) Alzheimer's Dementia Speech (Audio vs. Text): Multi-Modal Machine Learning at High vs. Low Resolution. *Applied Sciences* 13:4244. <https://doi.org/10.3390/app13074244>
112. Mei K, Ding X, Liu Y, et al (2023) The Ustc System for Adress-m Challenge. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Rhodes Island, Greece, pp 1–2
113. Tamm B, Vandenberghe R, Van Hamme H (2023) Cross-Lingual Transfer Learning for Alzheimer's Detection from Spontaneous Speech. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Rhodes Island, Greece, pp 1–2
114. Chen X, Pu Y, Li J, Zhang W-Q (2023) Cross-Lingual Alzheimer's Disease Detection Based on Paralinguistic and Pre-Trained Features. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Rhodes Island, Greece, pp 1–2
115. Liu J, Fu F, Li L, et al (2023) Efficient Pause Extraction and Encode Strategy for Alzheimer's Disease Detection Using Only Acoustic Features from Spontaneous Speech. *Brain Sciences* 13:477. <https://doi.org/10.3390/brainsci13030477>
116. Chen W, Xing X, Xu X, et al (2023) SpeechFormer++: A Hierarchical Efficient Framework for Paralinguistic Speech Processing. *IEEE/ACM Trans Audio Speech Lang Process* 31:775–788. <https://doi.org/10.1109/TASLP.2023.3235194>
117. Lin K, Washington PY (2024) Multimodal deep learning for dementia

- classification using text and audio. *Sci Rep* 14:13887. <https://doi.org/10.1038/s41598-024-64438-1>
118. Woszczyk D, Hedlikova A, Akman A, et al (2022) Data Augmentation for Dementia Detection in Spoken Language. In: *Interspeech 2022*. ISCA, pp 2858–2862
 119. Jin L, Oh Y, Kim H, et al (2023) CONSEN: Complementary and Simultaneous Ensemble for Alzheimer's Disease Detection and MMSE Score Prediction. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Rhodes Island, Greece, pp 1–2
 120. Poor FF, Dodge HH, Mahoor MH (2024) A multimodal cross-transformer-based model to predict mild cognitive impairment using speech, language and vision. *Computers in Biology and Medicine* 182:109199. <https://doi.org/10.1016/j.combiomed.2024.109199>
 121. Nambiar AS, Likhita K, Puja KVSS, et al (2022) Comparative study of Deep Classifiers for Early Dementia Detection using Speech Transcripts. In: *2022 IEEE 19th India Council International Conference (INDICON)*. IEEE, Kochi, India, pp 1–6
 122. Ali Meerza SI, Li Z, Liu L, et al (2022) Fair and Privacy-Preserving Alzheimer's Disease Diagnosis Based on Spontaneous Speech Analysis via Federated Learning. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, Glasgow, Scotland, United Kingdom, pp 1362–1365
 123. Liang S, Chen T, Ma J, et al (2024) Identification of mild cognitive impairment using multimodal 3D imaging data and graph convolutional networks. *Phys Med Biol* 69:235002. <https://doi.org/10.1088/1361-6560/ad8c94>
 124. Jahan S, Abu Taher K, Kaiser MS, et al (2023) Explainable AI-based Alzheimer's prediction and management using multimodal data. *PLoS ONE* 18:e0294253. <https://doi.org/10.1371/journal.pone.0294253>
 125. Jahan S, Saif Adib MdR, Huda SM, et al (2025) Federated Explainable AI-Based Alzheimer's Disease Prediction With Multimodal Data. *IEEE Access* 13:43435–43454. <https://doi.org/10.1109/ACCESS.2025.3547343>
 126. Myrzashova R, Alsamhi SH, Shvetsov AV, et al (2025) BCFTL: Blockchain-Enabled Multimodal Federated Transfer Learning for Decentralized Alzheimer's Diagnosis. *IEEE Internet Things J* 1–1. <https://doi.org/10.1109/JIOT.2025.3569652>
 127. Chen K, Weng Y, Huang Y, et al (2025) A multi-view learning approach with diffusion model to synthesize FDG PET from MRI T1WI for diagnosis of alzheimer's disease. *Alzheimer's Dement* 21:e14421. <https://doi.org/10.1002/alz.14421>

128. Lin W, Lin W, Chen G, et al (2021) Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of alzheimer's disease. *Front Neurosci* 15:646013. <https://doi.org/10.3389/fnins.2021.646013>
129. Gupta B, Jegannathan GK, Alam MS, et al (2025) Multimodal lightweight neural network for alzheimer's disease diagnosis integrating neuroimaging and cognitive scores. *Neurosci Inf* 5:100218. <https://doi.org/10.1016/j.neuri.2025.100218>
130. Chen Z, Wang Z, Zhao M, et al (2022) A new classification network for diagnosing alzheimer's disease in class-imbalance MRI datasets. *Front Neurosci* 16:807085. <https://doi.org/10.3389/fnins.2022.807085>
131. Sarma M, Chatterjee DrS (2024) Multistage Diagnosis of Alzheimer's Disease from Clinical Data Using 'Deep Ensemble Learning.' *JAI* 01:122-138. <https://doi.org/10.54364/JAI.2024.1109>
132. Mujahid M, Rehman A, Alam T, et al (2023) An Efficient Ensemble Approach for Alzheimer's Disease Detection Using an Adaptive Synthetic Technique and Deep Learning. *Diagnostics* 13:2489. <https://doi.org/10.3390/diagnostics13152489>
133. Dubey Y, Bhongade A, Palsodkar P, Fulzele P (2024) Efficient Explainable Models for Alzheimer's Disease Classification with Feature Selection and Data Balancing Approach Using Ensemble Learning. *Diagnostics* 14:2770. <https://doi.org/10.3390/diagnostics14242770>
134. Mandawkar U, Diwan T (2024) Hybrid cuttle Fish-Grey wolf optimization tuned weighted ensemble classifier for Alzheimer's disease classification. *Biomedical Signal Processing and Control* 92:106101. <https://doi.org/10.1016/j.bspc.2024.106101>
135. Jasodanand VH, Kowshik SS, Puducheri S, et al (2025) AI-driven fusion of multimodal data for Alzheimer's disease biomarker assessment. *Nat Commun* 16:7407. <https://doi.org/10.1038/s41467-025-62590-4>
136. Weiner MW, Kanoria S, Miller MJ, et al (2025) Overview of alzheimer's disease neuroimaging initiative and future clinical trials. *Alzheimer's Dement* 21:e14321. <https://doi.org/10.1002/alz.14321>
137. on behalf of Dementias Platform UK and UK Biobank, Wilkinson T, Schnier C, et al (2019) Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol* 34:557-565. <https://doi.org/10.1007/s10654-019-00499-1>
138. Thulasimani V, Shanmugavadivel K, Cho J, Veerappampalayam Easwaramoorthy S (2024) A review of datasets, optimization strategies, and learning algorithms for analyzing alzheimer's dementia detection. *Neuropsychiatr Dis Treat* 20:2203-2225. <https://doi.org/10.2147/NDT.S496307>

139. Chan KCG, Xia F, Kukull WA (2025) NACC data: who is represented over time and across centers, and implications for generalizability. *Alzheimer's Dement* 21:e70657. <https://doi.org/10.1002/alz.70657>
140. Fowler C, Rainey-Smith SR, Bird S, et al (2021) Fifteen years of the australian imaging, biomarkers and lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to alzheimer's disease. *J Alzheimer's Dis Rep* 5:443-468. <https://doi.org/10.3233/ADR-210005>
141. Yang Q, Li X, Ding X, et al (2022) Deep learning-based speech analysis for Alzheimer's disease detection: a literature review. *Alzheimer's Res Ther* 14:186. <https://doi.org/10.1186/s13195-022-01131-3>
142. He Y, Wang Z, Zhang Y, et al (2025) NeuroSymAD: A Neuro-Symbolic Framework for Interpretable Alzheimer's Disease Diagnosis
143. Sadeghi A, Hajati F, Argha A, et al Interpretable graph-based models on multimodal biomedical data integration: A technical review and benchmarking
144. Mahamud E, Assaduzzaman M, Islam J, et al (2025) Enhancing Alzheimer's disease detection: An explainable machine learning approach with ensemble techniques. *Intelligence-Based Medicine* 11:100240. <https://doi.org/10.1016/j.ibmed.2025.100240>
145. Zhou T, Liu M, Thung K-H, Shen D (2019) Latent Representation Learning for Alzheimer's Disease Diagnosis With Incomplete Multi-Modality Neuroimaging and Genetic Data. *IEEE Trans Med Imaging* 38:2411-2422. <https://doi.org/10.1109/TMI.2019.2913158>
146. Sharma R, Sibille L, Fahmi R (2023) Multi-branch convolutional neural network for Alzheimer's Disease versus normal control classification using PET images. *Alzheimer's & Dementia* 19:e061092. <https://doi.org/10.1002/alz.061092>
147. Zhang J, Yu X, Chen T, et al (2025) BrainNet-MoE: Brain-Inspired Mixture-of-Experts Learning for Neurological Disease Identification