

Review

# Multimodal AI for Alzheimer Disease Diagnosis: Systematic Review of Datasets, Models, and Modalities

Ziwen Yu<sup>1\*</sup>, MSc; Anthony Mulholland<sup>1\*</sup>, PhD; Tianyan Huang<sup>2</sup>; Qiang Liu<sup>1</sup>, PhD

<sup>1</sup>School of Engineering Mathematics and Technology, University of Bristol, Bristol, United Kingdom

<sup>2</sup>Medical Physics and Biomedical Engineering, University College London, London, United Kingdom

\*these authors contributed equally

## Corresponding Author:

Qiang Liu, PhD  
School of Engineering Mathematics and Technology  
University of Bristol  
Tankard's Close, Ada Lovelace Building  
Bristol BS8 1TW  
United Kingdom  
Phone: 44 01173746653  
Email: [qiang.liu@bristol.ac.uk](mailto:qiang.liu@bristol.ac.uk)

## Abstract

**Background:** Early detection of Alzheimer disease (AD) is essential for timely intervention; yet, diagnostic performance varies widely across modalities and datasets. Recent multimodal artificial intelligence (AI) models have made significant progress, but the evidence base remains fragmented due to heterogeneous datasets, modeling frameworks, and reporting quality.

**Objective:** This systematic review aimed to analyze studies on multimodal AI models for AD diagnosis, prognosis, and risk prediction over 5 years. We evaluated dataset characteristics, modality combinations, modeling strategies, performance metrics, and methodological limitations. We further discuss real-world implications and translational pathways.

**Methods:** Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, we systematically searched PubMed, IEEE Xplore, Scopus, ACM Digital Library, Cochrane, and arXiv, with the final datasets last searched on November 15, 2025. Studies applying multimodal machine learning or deep learning to AD, mild cognitive impairment, and dementia outcomes were included, whereas studies using a single modality or lacking sufficient methodological detail were excluded. QUADAS-2 (Revised Quality Assessment of Diagnostic Accuracy Studies tool) assessed risk of bias. Extracted performance results were synthesized across 4 major multimodal dataset families.

**Results:** A total of 66 studies met the inclusion criteria. Across datasets, multimodal models consistently outperformed single-modal baselines. Alzheimer's Disease Neuroimaging Initiative–based diagnosis achieved an average accuracy of 92.5% (SD 3.8%), while mild cognitive impairment–conversion models achieved an average area under the curve (AUC) of 0.922 (SD 0.045), and several fusion architectures reported AUCs above 0.95. In contrast, UK Biobank risk-prediction studies reported an average AUC of 0.84 (SD 0.056), and this reflects performance in large, population-based datasets. DementiaBank speech-language studies achieved an average AUC of 0.813 (SD 0.042), and cross-lingual AD detection achieved an accuracy of 77% (SD 6.5%). Self-collected multimodal datasets demonstrated average accuracies around 96% (SD 2.4%), but their generalizability is limited due to small sample sizes and single-center designs.

**Conclusions:** This systematic review demonstrates that multimodal AI models consistently outperform single-modal models for AD diagnosis, prognosis, and risk prediction by integrating complementary biological, clinical, and behavioral information. Unlike prior reviews, this review provides a unified synthesis across heterogeneous clinical, imaging, genetic, and linguistic datasets, enabling cross-domain comparison of modeling strategies and performance. However, the generalizability of reported performance was limited due to substantial heterogeneity in dataset composition, outcome definitions, and validation, and prevalent risks of bias. By evaluating these factors, this review clarifies where current evidence is robust and where caution is warranted. The findings highlight the need for standardized multimodal benchmarks, transparent evaluation protocols, and clinically grounded model design to enable reliable real-world deployment. Overall, this work advances the field by framing multimodal AI not only as a performance-driven tool but also as a translational framework for equitable, interpretable, and scalable AD diagnosis.

**Trial Registration:** PROSPERO CRD420251241895; <https://www.crd.york.ac.uk/PROSPE-RO/view/CRD420251241895>

*J Med Internet Res* 2026;28:e85414; doi: [10.2196/85414](https://doi.org/10.2196/85414)

**Keywords:** alzheimer's disease, neurodegenerative disease; multimodal dataset; machine learning; deep learning; multimodal fusion; computer-aided diagnosis; early diagnosis

## Introduction

Alzheimer disease (AD) is the most prevalent neurodegenerative disorder and the leading cause of dementia worldwide [1]. With an aging global population, AD has become one of the most costly and deadly diseases of the 21st century, imposing profound emotional, financial, and caregiving burdens on patients, families, and health systems. By 2050, the number of people with AD is projected to rise from 55 million in 2020 to approximately 139 million [1].

The progression of AD includes the preclinical stage, mild cognitive impairment (MCI), and symptomatic stages, with varying degrees of symptom severity. The preclinical stage is a key window for intervention, during which neuropathological changes have commenced, but clinical symptoms remain largely undetectable [2]. Despite advances in awareness and screening, up to 75% of dementia cases remain undiagnosed worldwide, particularly in low- and middle-income countries [3]. This persistent diagnostic gap highlights the need for low-cost, scalable, and accurate early detection tools to enable timely intervention and slow disease progression [4].

Artificial intelligence (AI) has emerged as a promising approach for improving the early detection and management of AD. By systematically integrating and analyzing multimodal data, AI-based diagnostic frameworks offer powerful tools to enhance early detection accuracy and facilitate timely intervention.

Recent work has used transformer-based models to integrate imaging, genetic, and linguistic data. Multimodal transformers combining magnetic resonance imaging (MRI) or positron emission tomography (PET) with clinical features and cognitive assessments have reported improved diagnostic accuracy and interpretability [5-7]. In parallel, GPT-style architectures, BERT (Bidirectional Encoder Representations From Transformers) variants, and domain-adapted language models improve extraction of linguistic and semantic markers linked to early cognitive decline [8,9]. Self-supervised speech models also perform strongly for detecting MCI and early AD from spontaneous speech [10]. Together, these advances reflect a shift toward unified, more interpretable, and clinically translatable multimodal systems that capture both biological and behavioral aspects of AD.

Traditional machine learning (ML) [11], ensemble methods [12], deep learning [13], and reinforcement learning (RL) [14] can perform well on unimodal data, but clinical diagnosis integrates structural and behavioral information [15]. Unimodal AI can therefore diverge from clinical workflows and miss complementary signals (eg, MRI for structural change plus speech features for cognitive decline [16]), increasing the risk of modality-specific overfitting and

poorer real-world performance. Accordingly, recent work has shifted toward multimodal integration for AD diagnosis, yet many studies emphasize incremental accuracy gains while underaddressing generalizability, interpretability, and cost-effectiveness needed for adoption. The literature also remains fragmented: recent reviews often cover multimodal clinical phenotyping datasets [17] and multimodal linguistic cognitive-impairment datasets [18] separately, obscuring cross-modal insights such as how imaging and speech biomarkers might jointly improve early detection.

Recent multimodal methods have substantially improved AD detection. However, a comprehensive systematic review that integrates evidence across both clinical and linguistic modalities, fusion strategies, and critically evaluates methodological quality, dataset diversity, and reporting transparency is still lacking. To address these gaps, this review investigates how multimodal models are applied to AD diagnosis, prognosis, and risk prediction and compares performance across different modality combinations and dataset families published between 2019 and 2025. We also examine modeling and fusion strategies alongside validation practices and assess methodological quality and risk of bias using QUADAS-2 (Revised Quality Assessment of Diagnostic Accuracy Studies tool). Furthermore, key multimodal combinations within public datasets are analyzed in relation to their diagnostic performance, and datasets are categorized to evaluate their suitability for AD research and clinical translation. Overall, this review provides a comprehensive synthesis of multimodal AI in AD diagnosis, bridges previously disconnected research streams, and offers practical guidance for future model development and clinical adoption.

## Methods

### Study Design

This review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [19], with the search procedures reported following PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) [20] and developed using the principles outlined in the Cochrane Handbook [21]. These methods were applied to systematically identify and evaluate studies on computer-aided AD diagnosis, with a particular focus on those using multimodal clinical phenotyping datasets and multimodal linguistic-based cognitive impairment datasets.

### Source of the Study and Search Criteria

We developed and internally reviewed independent search strategies (no external peer review). We manually searched

multiple databases to identify AI-driven multimodal approaches for AD diagnosis, rather than using an integrated multidatabase platform. As this review targets methodological advances mainly reported in peer-reviewed computational literature, we did not search trial registries (ClinicalTrials.gov, World Health Organization's International Clinical Trials Registry Platform). We also avoided validated or published filters, instead iteratively refining customized controlled-vocabulary and free-text terms for AD or dementia, multimodal data, and AI through pilot screening to maximize sensitivity.

Searches were performed in PubMed (447 records; January 1, 2019, to November 13, 2025), Scopus (1086 records; all years through November 13, 2025, filtered to PUBYEAR > 2018), IEEE Xplore (2229 records; January 1, 2020 to November 13, 2025), ACM Digital Library (2067 records; 1 January 2020, to 15 November 2025), Cochrane Library (1061 records; all available years through November 15, 2025), and arXiv (1081 records; all available years through November 15, 2025). We included the verbatim search strings for all databases, and because arXiv does not support bulk export, an arXiv search Python (Python Software Foundation) script is provided in [Multimedia Appendix 1](#) [11,14,22-44].

### Eligibility Criteria

The inclusion criteria were if studies were considered eligible if they met all the following conditions: (1) focused on AD, MCI, or related dementias as the primary clinical outcome; (2) applied AI or ML methods for computer-aided diagnosis, classification, or prediction; (3) used multimodal data, defined as any combination of at least two distinct modalities (eg, neuroimaging, clinical phenotyping, genetics, or linguistic features); (4) reported quantitative evaluation metrics; and (5) written in English.

The exclusion criteria were if studies met any of the following conditions: (1) single-modal approaches using only a single imaging modality, cognitive test, or biomarker,

without any multimodal integration; (2) works without reported performance metrics or with insufficient methodological detail; (3) works not addressing diagnosis, classification, or prediction (eg, treatment response, drug trials, and lifestyle interventions); (4) duplicate publications or overlapping datasets without providing additional methodological contribution; and (5) non-English publications.

### Selection Process

The study selection process followed the PRISMA 2020 guidelines, and the protocol was registered. The final search update was conducted in November 2025. All records retrieved from the databases were first imported into Zotero, where duplicates were automatically detected and removed.

The initial search identified 7435 records. After removing 3047 duplicates, 4388 records remained for title and abstract screening. A total of 4021 records were obviously irrelevant at the title and abstract level, 252 studies were excluded for the following main reasons:

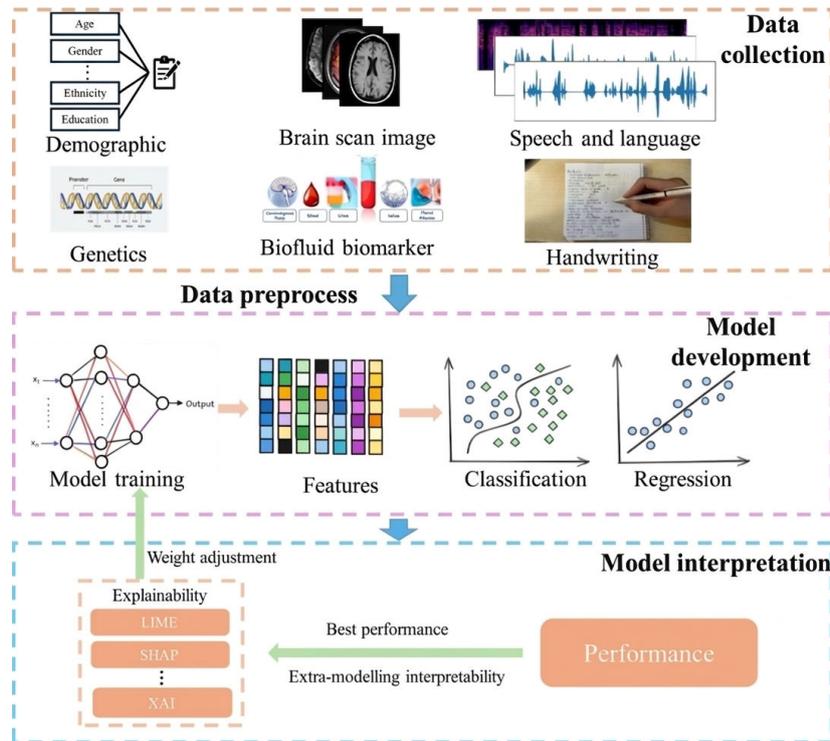
- Focused on outcomes unrelated to AD diagnosis, classification, or prediction (eg, drug trials, treatment response, lifestyle interventions; n=140).
- Used unimodal data without multimodal integration (n=46).
- No sufficient methodological details (n=47).

Finally, 66 studies were included in the systematic synthesis, and all were successfully retrieved (reports not retrieved=0).

### Overview of AI-Assisted AD Diagnosis

The workflow of AI-assisted AD diagnosis involves 3 stages, as illustrated in [Figure 1](#). The initial stage involves comprehensive data acquisition, where information is collected from multiple modalities, including neuroimaging, biomarkers, genetics, and speech or behavioral signals. The second stage involves feature extraction and model development, followed by an interpretable and explainable analysis to ensure that AI models can effectively support clinical decision-making.

**Figure 1.** Overview of the AI pipeline for AD diagnosis. Multimodal inputs undergo preprocessing and feature extraction before model training for classification or regression tasks. Model interpretability supports explanation and performance evaluation. AD: Alzheimer disease; AI: artificial intelligence; LIME: Local Interpretable Model-Agnostic Explanations; SHAP: Shapley Additive Explanations; XAI: explainable artificial intelligence.



### Performance Evaluation Metrics

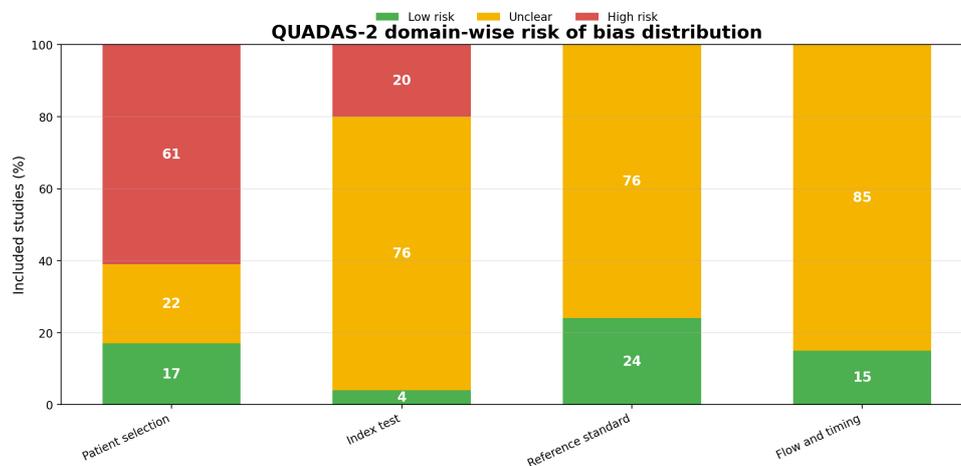
To ensure the clinical applicability and scientific rigor of computer-aided diagnosis models for AD, it is essential to systematically evaluate their performance using a variety of quantitative metrics. We have summarized all performance evaluation metrics in [Multimedia Appendix 2](#).

### Risk of Bias and Quality Assessment

We assessed methodological quality with QUADAS-2 [45], evaluating risk of bias in 4 domains. Patient selection raised

the main concern: 61% (40/66) of outcomes were high risk due to poor reporting or nonrepresentative sampling. For the index test, 76% (50/66) were unclear risk because procedures and decision thresholds were insufficiently described, and 20% (13/66) were high risk. The reference standard showed 76% (50/66) unclear risk from limited methodological detail, with no high-risk ratings. Flow and timing had the greatest uncertainty: 85% (56/66) were unclear owing to missing information on testing intervals and participant flow. [Figure 2](#) summarizes domain-level risk distributions; [Multimedia Appendix 3](#) reports study-level assessments.

**Figure 2.** Summary of the QUADAS-2 plot across the 66 included studies in the domain. QUADAS-2: Revised Quality Assessment of Diagnostic Accuracy Studies tool.



Given frequent unclear and high risk in key domains, we interpreted diagnostic performance cautiously, especially without external validation or a clearly defined reference standard. Future benchmarking should emphasize transparent reporting, prespecified thresholds, and multicenter evaluation to reduce bias and improve reproducibility.

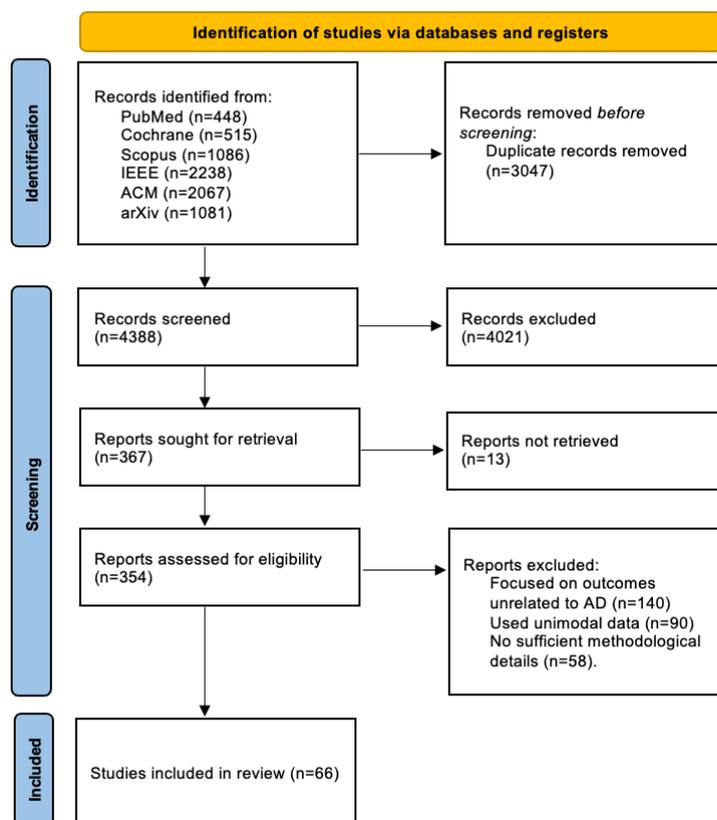
3), we first summarize the overall profile of the included literature to contextualize the subsequent synthesis. [Figure 4](#) provides a temporal overview (2019-2025) of modeling approaches across included studies, illustrating how methodological focus has shifted over time and informing interpretation of the evidence base.

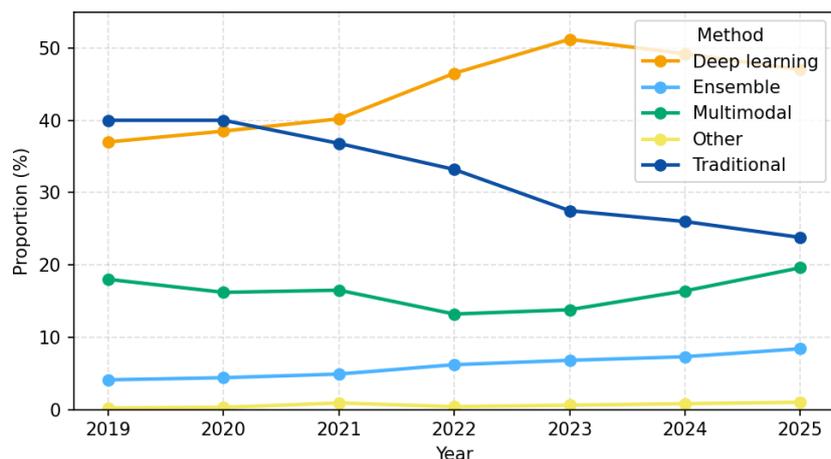
## Results

### Overview

Following study selection (the complete selection process is summarized in the PRISMA 2020 flow diagram, [Figure](#)

**Figure 3.** Flow diagram of PRISMA. AD: Alzheimer disease; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.  
PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only



**Figure 4.** Temporal trends of machine-learning methods used for AD diagnosis (2019-2025). AD: Alzheimer disease.

## Single Modality

### Overview

If readers are already familiar with traditional ML and deep learning approaches, they may wish to proceed directly to the next section, which focuses on multimodal data integration for AD diagnosis.

A concise overview of these baseline methods is provided in [Multimedia Appendix 4](#), which also includes a summary of RL. As most RL studies address sequential decision-making tasks rather than direct diagnostic modeling, their methodological details are presented in [Multimedia Appendix 4](#), to maintain focus on multimodal diagnostic frameworks in the main text.

### Deep Learning

Compared with traditional ML, deep learning enables hierarchical feature extraction, capturing complex patterns in high-dimensional data. It is therefore widely used to process and integrate AD-related multimodal inputs, including neuroimaging, clinical scores, genetics, and speech. Key approaches and findings are summarized below.

Recurrent neural networks are effective for modeling sequential data such as longitudinal clinical records and speech signals, but they are susceptible to vanishing gradients in long sequences [46-49]. Long short-term memory networks address this limitation through gated memory mechanisms, enabling more stable training and improved capture of long-term dependencies. Consequently, long short-term memory models have been widely applied in AD research for analyzing temporal and sequential modalities [50-53].

The 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. Transformer models use attention mechanisms to weight input features and capture diverse representations through multihead attention, enabling efficient and scalable training [54]. Owing to these advantages, they have been widely adopted in AD diagnosis and multimodal

learning, where their encoder-decoder architecture facilitates effective integration of heterogeneous data sources [13,55-57].

### Ensemble Learning

Ensemble learning improves generalization and robustness by combining multiple base models, including bagging and boosting methods such as AdaBoost (Adaptive Boosting), XGBoost (Extreme Gradient Boosting), and LightGBM (Light Gradient-Boosting Machine), and has been widely applied in AD detection and progression prediction [12,58-60]. However, ensemble models may introduce redundant features, offer limited gains on small datasets, and incur higher computational costs, which can restrict real-time or resource-constrained deployment.

### Summarization for Single Modality

Traditional single-modality ML approaches can achieve high performance in AD-related tasks; however, they are constrained by several inherent limitations:

First, regarding information completeness, structural MRI alone has limited sensitivity to functional and molecular changes and cannot fully capture AD-related cognitive and behavioral alterations. Combining MRI with PET, neuropsychological tests, speech, electroencephalography (EEG), and genetic or biomarker data provides a more complete, multidimensional view of disease progression and patient heterogeneity [61].

Second, regarding model robustness, in multimodal data, residual noise in one modality may persist despite denoising, but other modalities can provide complementary signals that improve robustness. Leveraging multisensory-style integration, multimodal models better reflect biological cognition and can yield more reliable decisions [62].

Third, in cross-modal learning, transformer architectures use cross-modal attention to learn associations between modalities. Some studies apply them in weakly supervised or cross-modal guided settings, using one modality to constrain or guide representation learning in another [63].

Fourth, in real-world decision-making, multimodal learning better matches real-world diagnosis, which integrates

multiple information sources. Using diverse modalities aligns models with clinical workflows and improves translational potential in practice [64].

Therefore, this review analyzes the methodological strengths and limitations of multimodal models for computer-assisted AD diagnosis, focusing on how dataset grouping and classification choices affect evaluation. By classifying datasets, we enable model comparisons under a unified setup, allowing for more direct assessment of generalizability and cross-dataset stability.

## Multimodal Data

### 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 generalization ability of models but also

help mitigate the risk of overfitting. Commonly used datasets for AI-assisted diagnosis of AD can be broadly categorized into 2 types. The first type of 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, functional MRI, genetic data, and electronic health records. The second type—multimodal cognitive-linguistic behavioral dataset centers on sequential data modalities, including audio, video, and transcribed language data, such as the Pitt Corpus and ADReSS (Alzheimer’s Dementia Recognition Through Spontaneous Speech). Each dataset type provides unique features that contribute to the comprehensive modeling of AD progression and diagnosis. The commonly used datasets, along with their population demographics and associated modalities, are summarized in Table 1.

**Table 1.** Commonly used dataset.

Dataset	Population demographics		Modalities	Link
	Male	Age (years)		
UK Biobank	23,000 (46) <sup>a</sup>	56.5 (8.1) <sup>b</sup>	MRI <sup>c</sup> , fMRI <sup>d</sup> , genetic, lifestyle scores, activity monitor, and EHR <sup>e</sup>	[65]
ADNI <sup>f</sup>				
ADNI-1	469 (57.3) <sup>a</sup>	75 (6.9) <sup>b</sup>	MRI, PET <sup>g</sup> , genetic, and EHR	[66]
ADNI-GO <sup>h</sup> and 2	473 (53) <sup>a</sup>	72.5 (7.2) <sup>b</sup>	MRI, PET, genetic, and EHR	[66]
ADNI-3	471 (49) <sup>a</sup>	74.9 (8.1) <sup>b</sup>	MRI, PET, genetic, and EHR	[66]
OASIS <sup>i</sup>				
OASIS-1	177 (42.5) <sup>a</sup>	57 (39) <sup>b</sup>	MRI, PET, CT, genetic, and EHR	[67]
OASIS-2	60 (40) <sup>a</sup>	78 (18) <sup>b</sup>	MRI, PET, CT, genetic, and EHR	[67]
OASIS-3	622 (45) <sup>a</sup>	69 (26.56) <sup>b</sup>	MRI, PET, CT, genetic, and EHR	[67]
OASIS-4	663 subjects	57.5 (36.5) <sup>b</sup>	MRI, PET, CT, genetic, and EHR	[67]
NACC <sup>j</sup>	23,625 (44.62) <sup>a</sup>	73.3 (10.5) <sup>b</sup>	MRI, PET, genetics, and EHR	[68]
FHS <sup>k</sup>	718 (42) <sup>a</sup>	80.76 (8.2) <sup>b</sup>	MRI, genetic, and EHR	[69]
AIBL <sup>l</sup>	289 (43.72) <sup>a</sup>	73.5 (7.03) <sup>b</sup>	MRI, PET, genetic, and EHR	[70]
Dementia bank				
Pitt corpus	HC <sup>m</sup> : 104, AD <sup>n</sup> : 208/552	— <sup>o</sup>	Audio and text	[71]
ADReSS <sup>p</sup>	HC: 78, AD: 78/156	—	Audio and text	[71]
ADReSSo <sup>q</sup>	HC: 115, AD: 122/237	—	Audio and text	[71]
ADReSS-M <sup>r</sup>	HC: 143, AD: 148/291	—	Audio and text	[71]
TAUKADIAL <sup>s</sup>	106/507	—	Audio and text	[71]
Multimodal dementia corpus	HC: 10, AD: 12/816	—	Audio, typed, and hand-written	[72]
ADReFV <sup>t</sup>	AD: 25	66.68 (2.08) <sup>b</sup>	Video	[73]
GENCODE <sup>u</sup>	Human: 78,686	—	Genetics	[74]

<sup>a</sup>n (%).

<sup>b</sup>Mean (SD).

<sup>c</sup>MRI: magnetic resonance imaging.

<sup>d</sup>fMRI: functional magnetic resonance imaging.

<sup>e</sup>EHR: electronic health record.

<sup>f</sup>ADNI: Alzheimer’s Disease Neuroimaging Initiative.

<sup>g</sup>PET: positron emission tomography.

<sup>h</sup>ADNI-GO: Alzheimer’s Disease Neuroimaging Initiative – Grand Opportunity.

<sup>i</sup>OASIS: Open Access Series of Imaging Studies.

<sup>j</sup>NACC: National Alzheimer’s Coordinating Centre.

<sup>k</sup>FHS: Framingham Heart Study.

<sup>l</sup>AIBL: Australian Imaging, Biomarkers and Lifestyle Study.

<sup>m</sup>HC: health control.

<sup>n</sup>AD: Alzheimer disease.

<sup>o</sup>Not available.

<sup>p</sup>ADReSS: Alzheimer’s Dementia Recognition Through Spontaneous Speech.

<sup>q</sup>ADReSSo: Alzheimer’s Dementia Recognition Through Spontaneous Speech 2021 Challenge.

<sup>r</sup>ADReSS-M: Multilingual Alzheimer’s Dementia Recognition Through Spontaneous Speech Challenge.

<sup>s</sup>TAUKADIAL: Speech-Based Cognitive Assessment in Chinese and English.

<sup>t</sup>ADReFV: Alzheimer’s Disease Recognition from Face & Voice.

<sup>u</sup>GENCODE: gene code.

To better understand current research trends, this paper reviewed studies from the past 5 years on multimodal AI models for AD diagnosis. The literature was categorized by dataset type. (1) Multimodal clinical phenotyping datasets: ADNI dominates this category, used in about 80% of studies, while others, such as UK Biobank, OASIS, National Alzheimer’s Coordinating Centre (NACC), Framingham Heart Study, and Australian Imaging, Biomarkers and Lifestyle Study, account for the remaining 20%, mainly for supplementary analysis or external validation. (2) Multimodal cognitive-linguistic behavioral datasets: The ADReSS series is most widely used, representing around 70% of studies. The

remaining 30% use datasets often for complementary analysis or benchmarking.

### Multimodal Clinical Phenotyping Datasets

Multimodal clinical phenotyping datasets integrate MRI, PET, or diffusion tensor imaging, biomarkers from blood, cerebrospinal fluid, or genomics, and standardized cognitive assessments. This review summarizes representative resources, highlighting their modalities, distinguishing features, and contributions to diagnostic modeling (Table 2).

**Table 2.** Multimodal clinical phenotyping datasets related paper. The exceptionally high performances reported in some of these studies can be attributed to specific methodological factors: (1) for two studies [75] and [76], the absence of external validation likely inflated the results; (2) for another study [77], 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 yet another study [78], the integration of rich gait features with optimized machine learning techniques in a controlled experimental setting facilitated unusually high performance.

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Xue et al [79], 2024	NACC <sup>a</sup> , ADNI <sup>b</sup> , AIBL <sup>c</sup> , FHS <sup>d</sup> , PPMI <sup>e</sup> , and OASIS <sup>f</sup>	Transformer-based multimodal model	Differential diagnosis of 10 etiologies	MRI <sup>g</sup> (T1, T2, FLAIR <sup>h</sup> ), clinical, neuropsychological tests, and PET <sup>i</sup>	Differential diagnosis probabilities and AD <sup>j</sup> /MCI <sup>k</sup> /NC <sup>l</sup> classification	Internal: NACC held out; external: ADNI and FHS	Etiology classification AUROC <sup>m</sup> 0.96; strong alignment with PET biomarkers and neuropathology	Imbalanced etiologies, training label subjectivity, and limited racial diversity
Shi et al [80], 2018	ADNI	MM-SDPN <sup>n</sup>	Classification (AD vs NC, MCI vs NC)	T1 MRI and FDG-PET <sup>o</sup>	Classification accuracy	Comparative vs single-modality and state-of-the-art models	Outperformed single-modality DPN <sup>p</sup> /SDPN <sup>q</sup> and concatenated models	ADNI-only dataset (limits generalizability); ROI <sup>r</sup> -based features rather than voxel-wise
Allwright et al [81], 2023	UK Biobank	XGBoost <sup>s</sup>	Risk prediction (Incident AD)	Demographics, lifestyle, genetics, and medical history	Prediction of incident AD (2-10 years) and risk factor ranking	Internal: nested 3-fold cross-validation; external: held-out validation set	AUROC 0.77, APOE-ε4 <sup>t</sup> identified as the strongest risk factor, and liver enzymes or frailty as predictors	ICD-10 <sup>u</sup> underascertainment, healthy volunteer bias, and observational design
Gu et al [82], 2025	UK Biobank	LightGBM <sup>v</sup>	Risk prediction (incident dementia in ASCVD <sup>w</sup> patients)	Clinical, biological assays, cognitive tests, and physical measures	All-cause incident dementia, AD, and VD <sup>x</sup> incidence	Temporal: train (2006-2009) and test (2010 cohort)	5-year dementia AUC <sup>y</sup> 0.903, AD AUC 0.775, and accuracy 0.851	Sample mostly European descent, static baseline features, and potential overfitting
You et al [83], 2022	UK Biobank	LightGBM	Risk prediction (5/10-year horizon)	Demographics, lifestyle, blood biomarkers, and genetics	Incident all-cause dementia and AD prediction	Internal: 5-fold cross-validation	AUC 0.848 (all-cause), 0.862 (AD), and outperformed	Limited external validation, population predominantly White, and feature

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Calvo et al [84], 2024	UK Biobank	Multivariable logistic regression	Risk association analysis	Questionnaire, ICD <sup>ab</sup> records, and genotypes	Odds of AD related to menopause type	Single cohort: multivariable adjustment	CAIDE <sup>z</sup> and DRS <sup>aa</sup> Early bilateral oophorectomy associated with 4-fold AD odds (OR <sup>ac</sup> 4.12) and HT <sup>ad</sup> use protective	selection fully data-driven Low case numbers in subgroups; self-reported HT use and healthy volunteer bias
Yi et al [85], 2025	UK Biobank, ADNI, PPMI, and IXI <sup>ae</sup>	3D-ViT <sup>af</sup>	BAG <sup>ag</sup> estimation and GWAS <sup>ah</sup>	T1-weighted MRI, genetics (SNP <sup>ai</sup> and xQTL <sup>aj</sup> )	BAG and drug target prioritization	External: ADNI, PPMI, and IXI	MAE <sup>ak</sup> ≈ 2.6 and identified 7 high-confidence drug targets (eg, MAPT <sup>al</sup> and TNFSF12 <sup>am</sup> )	European-ancestry bias; lack of biological “ground truth” for brain age
Yousefzadeh et al [86], 2024	UK Biobank (retina cohort)	VGG-16 <sup>an</sup> classifier + LAVA <sup>ao</sup> (XAI <sup>ap</sup> )	Binary classification and explainability	Retinal fundus images	AD vs NC classification and neuron-level explanations	Internal: nested 5-fold cross-validation	Accuracy 71.4% and identified 7 latent clusters linking vascular and cognitive decline	Small AD sample size (n=100), cross-sectional design, and UK Biobank volunteer bias
Gong et al [87], 2023	UK Biobank	SuperBigFLICA (semisupervised Bayesian fusion)	Phenotype discovery	Multimodal MRI (47 modalities)	Latent components predictive of nonimaging phenotypes	Internal: train, validation, or test split	Up to 46% improvement over expert IDPs <sup>aq</sup> and interpretable multimodal modes	Linear modeling constraints and UK Biobank population bias
Lian et al [88], 2022	ADNI-1, ADNI-2, and AIBL	Attention-guided HybNet (3D FCN <sup>ar</sup> + hybrid network)	Diagnosis and prognosis	Structural T1 MRI	AD vs NC classification; pMCI <sup>as</sup> vs sMCI <sup>at</sup> prediction	External: trained ADNI-1, and validated ADNI-2 and AIBL	ADNI-2 AD vs NC accuracy 0.919 (AUC 0.965) and outperformed ROI/VBM <sup>au</sup> methods	Heavy preprocessing reliance, limited demographic diversity, and potential overfitting
Lian et al [89], 2022	ADNI-1 and ADNI-2	MWAN <sup>av</sup>	Joint regression of clinical scores	Structural T1 MRI and clinical scores	MMSE <sup>aw</sup> , CDRSB <sup>ax</sup> , and ADAS-Cog <sup>ay</sup> prediction	Cross-validation: across ADNI-1 and ADNI-2	Lower RMSE <sup>az</sup> and higher correlation coefficients than single-task baselines	Restricted to the ADNI cohorts and potential overfitting to the modest sample size
Li et al [90], 2019	ADNI-1, ADNI-GO/2 <sup>ba</sup> , and AIBL	3D CNN <sup>bb</sup> + Cox proportional hazards	Time-to-event prognosis	Hippocampal MRI patches and clinical variables	Progression from MCI to AD and risk stratification	External: trained ADNI-1, and validated ADNI-GO/2 and AIBL	C-index 0.864 (combined model) and significant risk-based stratification of MCI	Focus on the hippocampus only and potential cohort and scanner bias (1.5T vs 3T)
Qiu et al [15], 2022	NACC, ADNI, and ADCP <sup>bc</sup>	Multimodal deep learning (3D CNN + FCN)	Multiclass classification	Structural MRI, demographics, and neuropsychology	Diagnosis (NC, MCI, AD, and nADD <sup>bd</sup> ) and saliency maps	External: trained NACC, and validated ADNI and independent cohorts	Performance comparable to neurologists and saliency aligned with pathology	Retrospective design and heterogeneity in protocols across cohorts
Oh et al [91], 2023	ADNI	LEAR <sup>be</sup> framework (CNN + RL <sup>bf</sup> + XAI)	Diagnosis and interpretation	Structural T1 MRI	AD vs non-AD classification and counterfactual maps	Internal: cross-validation on ADNI	Improved accuracy and generalization, and localized plausible atrophy patterns	Single-cohort (ADNI) and XAI evaluation, partly qualitative

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Lian et al [92], 2020	ADNI-1 and ADNI-2	Hierarchical FCN	Diagnosis and atrophy localization	Structural T1 MRI	AD vs NC, MCI vs NC, and atrophy pattern mapping	External: trained ADNI-1 and tested ADNI-2	Improved accuracy vs conventional features and interpretable atrophy maps	ADNI-only; strong reliance on preprocessing and registration
Avsec et al [93], 2021	Genomic reference datasets	Enformer (transformer)	Genomic prediction	DNA sequence	Gene expression and chromatin state prediction	Internal: held-out chromosomes	Improved capture of long-range regulatory effects vs previous models	Limited to available cell types and assays
Yang et al [94], 2021	ADNI	Deep learning and super learner	Prognosis	MRI, cognitive, and biomarkers	Diagnostic classification and prognostic risk signature	Internal: cross-validation within ADNI	Derived signature distinguished diagnostic groups and progression risk	Limited external validation and restricted to the ADNI research cohort
Lee et al [95], 2024	ADNI and UK or Singapore Clinics	PPM <sup>bg</sup>	Prognosis (MCI to AD)	MRI (gray matter) and cognitive tests	Individualized prognostic index	External: independent real-world memory clinics	Accuracy $\approx$ 81.7%, AUC $\approx$ 0.84; and index predicted conversion better than atrophy alone	Heterogeneity in real-world clinical data and potential site effects
Zhu et al [96], 2021	ADNI and AIBL	DA-MIDL <sup>bh</sup>	Diagnosis	Structural MRI patches	AD vs NC and MCI vs NC	External: trained ADNI and tested AIBL	Higher accuracy and generalizability than baselines, and attention maps aligned with pathology	Reliance on structural MRI and potential dataset-specific overfitting
Zhang et al [97], 2024	ADNI	GCN <sup>bi</sup> , SHAP <sup>bj</sup> , and automatic fusion	Diagnosis	Cognitive, MRI, PET, and risk factors	AD vs non-AD diagnosis, and multimodal feature selection	Internal: two ADNI multimodal cohorts	Accuracies of 95.9% and 91.9%, and efficient selection of clinically important features	Complex model deployment and reliance on ADNI data
Velazquez and Lee [75], 2022	ADNI EMCI <sup>bk</sup>	Ensemble (random forest and CNN)	Prediction of conversion	DTI <sup>bl</sup> (ADC <sup>bm</sup> maps) and EHR <sup>bn</sup>	EMCI to AD conversion prediction	Internal: held-out test set	98.81% accuracy and feature importance explainability provided	Small converter sample size and potential overfitting
Zhang et al [76], 2024	ADNI	Multimodal learning machine (ELM <sup>bo</sup> ensemble)	Diagnosis	MRI features and neuropsychological tests	NC, MCI, and AD classification	Internal: cross-validation on ADNI	>98% accuracy and $F_1$ -score, and no observed bias between MCI and AD	A single research cohort and very high accuracy require external verification
Bi et al [98], 2020	ADNI	Cluster evolutionary random forest	Diagnosis	Resting-state fMRI <sup>bp</sup> and SNP	AD vs control classification and biomarker identification	Comparative vs competing methods	Identified significant brain region-gene pairs and effective classification	Small multimodal sample size and complex hyperparameters
Bi et al [99], 2022	ADNI	Weighted evolutionary random forest	Pathogen detection	Resting-state fMRI and SNP	MCI identification and pathogenic factor extraction	Comparative vs state-of-the-art methods	Superior MCI identification performance, and highlighted key ROIs and SNPs	High-dimensional fusion features, small N, and overfitting risk

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Hashmi and Barukab [100], 2023	OASIS	Deep RL and neural network	Staging	Structural MRI	4-class dementia staging	Internal: augmented vs baseline	RL augmentation improved accuracy by ~6% and recall by ~13%	Single open dataset; focus on MRI only
Wang et al [101], 2024	ADNI-1/2/3	Multimodal DL <sup>bq</sup> with an interaction layer	Prognosis (MCI to AD)	MRI, clinical, and genetics (SNP)	4-year conversion prediction	External: generalized to ADNI-3	AUC 0.962 (cross-validation), 0.939 (test); interaction effects improved accuracy	ADNI-only and potential overfitting despite cross-validation
Hatami et al [102], 2024	ADNI	DNN <sup>br</sup> and RL (data augmentation)	Classification	Structural MRI	AD vs NC classification	Comparative vs baseline augmentation approaches	Precision ~0.95; RL-guided augmentation outperformed baselines	Single research cohort and no external clinical validation
Tabarestani et al [103], 2020	ADNI	Distributed multitask regression	Longitudinal progression	MRI, PET, CSF <sup>bs</sup> , EEG <sup>bt</sup> , and clinical	Prediction of longitudinal cognitive scores	Comparative vs unimodal or multimodal methods	Reduced errors, particularly in sparse or incomplete longitudinal data	Model complexity and potential sensitivity to hyperparameters
Burkhart et al [104], 2024	ADNI and Singapore Memory Clinic	Unsupervised multimodal trajectory modeling	Prognosis	Cognitive, amyloid PET, and MRI	Cognitive health clustering and progression prediction	External: real-world memory clinic data	Better stratification than standard clinical assessments and robust to missing data	Unsupervised complexity and reliance on ADNI for training
El-Sappagh et al [105], 2021	ADNI	Random forest and SHAP (multilayer)	Diagnosis and progression	11 modalities (MRI, PET, CSF, and clinical)	Multiclass diagnosis and MCI progression detection	Internal: cross-validation	Diagnosis accuracy 93.95%, progression accuracy 87.08%, and interpretable	High complexity and challenges for routine care deployment
Lee et al [106], 2024	ADNI and 4 Korean hospitals	GBM <sup>bu</sup>	Conversion prediction	MRI (T1, T2-FLAIR <sup>bv</sup> ), amyloid PET, and clinical	MCI to AD conversion (4-year)	Internal: nested cross-validation with modality combinations	T1 and amyloid PET is the best combination, and T2-FLAIR did not improve prediction	Small multicenter sample and site and scanner heterogeneity
Yuan et al [107], 2021	ADNI	Multimodal cotraining (random forest)	MCI subtype classification	Structural MRI and SNP	sMCI vs pMCI classification	External: ADNI-2 independent test set	Accuracy 85.5% and cotraining outperformed single modality	Dependence on feature selection and ADNI-only
Cirincione et al [108], 2024	TADPOLE <sup>bw</sup> (ADNI)	Ensemble integration	Prediction	MRI, PET, clinical, and cognitive	Future dementia prediction in MCI	Internal: held-out test set	AUC 0.81, and outperformed XGBoost and deep learning baselines	Single research dataset and the complexity of multimodal integration
Cassani and Falk [109], 2020	Clinical EEG	Feature engineering and ML	Diagnosis and severity	Resting-state EEG	AD vs normal, and mild vs moderate AD classification	Internal: cross-validation	Modulation spectral features outperformed traditional EEG features	Small sample size, resting state only, and single center
Cilia et al [110], 2021	Custom (Naples)	Deep transfer learning (CNN)	Diagnosis	Online handwriting (dynamic)	Early AD detection	Internal: cross-validation	Dynamic features (color-encoded) are	Single-center dataset and task-specific protocol

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Kmetzsch et al [111], 2022	PREV-DEMALS <sup>b</sup> <sub>x</sub>	Supervised variational autoencoder	Disease progression modeling	MRI and microRNA	Disease progression score (FTD <sup>by</sup> /ALS <sup>bz</sup> )	Validation: synthetic data and cohort evaluation	superior to shape-only images Outperformed competing models in capturing progression trajectory	Small sample (rare disease) and cross-sectional data used for progression
Mengoudi et al [112], 2020	UCL <sup>ca</sup> and Insight 46	Self-supervised deep neural network	Diagnosis	Eye-tracking (gaze or pupil)	Dementia vs control classification	Comparative vs handcrafted features	Self-supervised features are more sensitive than handcrafted metrics	Modest sample size, mixed dementia subtypes, and specialized hardware
Tsai et al [113], 2024	Taiwan NHI <sup>cb</sup>	MAND <sup>cc</sup>	Incidence prediction	EHR (ICD codes) and demographics	Dementia incidence risk	Internal: held-out test set	AUC 0.901 and outperformed traditional CTR <sup>cd</sup> models	Coding errors in administrative data are specific to the Taiwan NHI
Park et al [22], 2024	Korean memory clinics	SVM <sup>ce</sup>	Diagnosis (MCI vs HC <sup>cf</sup> )	VR <sup>cg</sup> biomarkers, MRI, and neuropsychological tests	MCI vs healthy control classification	Internal: train or test split	VR, MRI AUC 0.89, and VR biomarkers comparable to MRI alone	Small sample (n=54) and VR hardware requirement
Wu et al [114], 2022	Clinical EEG	WiGMM <sup>ch</sup>	Severity detection	Resting-state EEG	Unsupervised dementia degree detection	Internal: latent structure analysis	Captured latent dementia degrees matching clinical status	Unsupervised labeling requires careful interpretation
Zhang et al [115], 2025	Chinese memory clinics	FCRN <sup>ci</sup> and MLP <sup>cj</sup> (patch-based)	Diagnosis	MRI, PET, clinical, and genotype	AD vs normal and MCI vs normal classification	Internal: cross-validation	Accuracy ≈96% (AD), ≈92% (MCI), and interpretable probability maps	Single-country clinical cohorts and limited ethnic diversity
Fabietti et al [77], 2023	Mouse models	Ensemble machine learning	Early detection (animal)	LFP <sup>ck</sup>	AD vs control mouse classification	Internal: channel masking robustness tests	Accuracy 99.4% and robust to artifacts	Preclinical animal model results and small sample size
Seifallahi et al [78], 2022	Single center	SVM	Diagnosis	Kinect V2 (gait or TUG <sup>cl</sup> )	AD vs healthy control classification	Internal: leave-one-out cross-validation	Accuracy 98.68% using 12 skeletal features	Small sample, case-control design may overestimate performance
Fan et al [116], 2024	CVD <sup>cm</sup> patients (Wuhan)	ViT <sup>cn</sup> (MRI) and XGBoost (clinical)	VCI <sup>co</sup> diagnosis	MRI (T1, T2-FLAIR) and clinical	Vascular cognitive impairment diagnosis	External: independent CVD dataset	The hybrid model has an AUC of 0.965 and is comparable to expert neurologists	CVD-specific cohort and complex ViT and XGBoost pipeline
Beebe-Wang et al [117], 2021	Aging cohort (US)	Nonlinear ML and SHAP	Imminent prediction (3 years)	Clinical, neuropsychological	Incident dementia within 3 years	Internal: cross-validation	Sparse model (4 tests) comparable to full battery	Prediction limited to a 3-year horizon and a single health system
Battineni et al [118], 2021	Public MRI dataset	Gradient boosting	Classification	MRI features and demographics	AD vs non-AD classification	Internal: cross-validation	Accuracy 97.58% (gradient boosting performed best)	Small public dataset and lack of external validation

<sup>a</sup>NACC: National Alzheimer’s Coordinating Centre.

<sup>b</sup>ADNI: Alzheimer’s Disease Neuroimaging Initiative.

<sup>c</sup>AIBL: Australian Imaging, Biomarkers, and Lifestyle Study.

<sup>d</sup>FHS: Framingham Heart Study.

<sup>e</sup>PPMI: Parkinson Progression Markers Initiative.

---

<sup>f</sup>OASIS: Open Access Series of Imaging Studies.  
<sup>g</sup>MRI: magnetic resonance imaging.  
<sup>h</sup>FLAIR: fluid-attenuated inversion recovery.  
<sup>i</sup>PET: positron emission tomography.  
<sup>j</sup>AD: Alzheimer disease.  
<sup>k</sup>MCI: mild cognitive impairment.  
<sup>l</sup>NC: normal control.  
<sup>m</sup>AUROC: area under the receiver operating characteristic curve.  
<sup>n</sup>MM-SPDN: multimodal stacked deep polynomial network.  
<sup>o</sup>FDG-PET: fluorodeoxyglucose-positron emission tomography.  
<sup>p</sup>DPN: deep polynomial network.  
<sup>q</sup>SPDN: stacked deep polynomial network.  
<sup>r</sup>ROI: region of interest.  
<sup>s</sup>XGBoost: Extreme Gradient Boosting.  
<sup>t</sup>APOE-ε4: apolipoprotein E epsilon 4 allele.  
<sup>u</sup>ICD-10: *International Statistical Classification of Diseases, Tenth Revision*.  
<sup>v</sup>LightGBM: Light Gradient-Boosting Machine.  
<sup>w</sup>ASCVD: atherosclerotic cardiovascular disease.  
<sup>x</sup>VD: vascular dementia.  
<sup>y</sup>AUC: area under the curve.  
<sup>z</sup>CAIDE: Cardiovascular Risk Factors, Aging, and Incidence of Dementia.  
<sup>aa</sup>DRS: Dementia Risk Score.  
<sup>ab</sup>ICD: *International Classification of Diseases*.  
<sup>ac</sup>OR: odds ratio.  
<sup>ad</sup>HT: hormone therapy.  
<sup>ae</sup>IXI: Information Extraction From Images.  
<sup>af</sup>3D-ViT: 3D vision transformer.  
<sup>ag</sup>BAG: brain age gap.  
<sup>ah</sup>GWAS: genome-wide association study.  
<sup>ai</sup>SNP: single-nucleotide polymorphism.  
<sup>aj</sup>xQTL: molecular quantitative trait locus.  
<sup>ak</sup>MAE: mean absolute error.  
<sup>al</sup>MAPT: microtubule-associated protein tau.  
<sup>am</sup>TNFSF12: Tumor Necrosis Factor (Ligand) Superfamily, Member 12.  
<sup>an</sup>VGG-16: Visual Geometry Group 16-Layer Network.  
<sup>ao</sup>LAVA: Granular Neuron-Level Explainer.  
<sup>ap</sup>XAI: explainable artificial intelligence.  
<sup>aq</sup>IDP: imaging-derived phenotype.  
<sup>ar</sup>FCN: fully convolutional network.  
<sup>as</sup>pMCI: progressive mild cognitive impairment.  
<sup>at</sup>sMCI: stable mild cognitive impairment.  
<sup>au</sup>VBM: voxel-based morphometry.  
<sup>av</sup>MWAN: multi-task weakly-supervised attention.  
<sup>aw</sup>MMSE: Mini-Mental State Examination.  
<sup>ax</sup>CDRSB: Clinical Dementia Rating–Sum of Boxes.  
<sup>ay</sup>ADAS-Cog: Alzheimer Disease Assessment Scale–Cognitive Subscale.  
<sup>az</sup>RMSE: root mean square error.  
<sup>ba</sup>ADNI-GO/2: Alzheimer’s Disease Neuroimaging Initiative – Grand Opportunity / Phase 2.  
<sup>bb</sup>CNN: convolutional neural network.  
<sup>bc</sup>ADPC: Alzheimer Disease Prediction Challenge.  
<sup>bd</sup>nADD: non-Alzheimer disease dementia.  
<sup>be</sup>LEAR: learn-explain-reinforce.  
<sup>bf</sup>RL: reinforcement learning.  
<sup>bg</sup>PPM: Predictive Prognostic Model.  
<sup>bh</sup>DA-MIDL: dual attention multi-instance deep learning.  
<sup>bi</sup>GCN: graph convolutional network.  
<sup>bj</sup>SHAP: Shapley Additive Explanations.  
<sup>bk</sup>EMCI: early mild cognitive impairment.  
<sup>bl</sup>DTI: diffusion tensor imaging.  
<sup>bm</sup>ADC: apparent diffusion coefficient.  
<sup>bn</sup>EHR: electronic health record.  
<sup>bo</sup>ELM: extreme learning machine.  
<sup>bp</sup>fMRI: functional magnetic resonance imaging.  
<sup>bq</sup>DL: deep learning.  
<sup>br</sup>DNN: deep neural network.  
<sup>bs</sup>CSF: cerebrospinal fluid.  
<sup>bt</sup>EEG: electroencephalography.

---

<sup>bu</sup>GBM: Gradient Boosting Machine.

<sup>bv</sup>T2-FLAIR: T2-weighted fluid-attenuated inversion recovery.

<sup>bw</sup>TADPOLE: The Alzheimer Disease Prediction of Longitudinal Evolution.

<sup>bx</sup>PREV-DEMALS: Predict to Prevent Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis.

<sup>by</sup>FTD: frontotemporal dementia.

<sup>bz</sup>ALS: amyotrophic lateral sclerosis.

<sup>ca</sup>UCL: University College London.

<sup>cb</sup>NHI: National Health Insurance.

<sup>cc</sup>MAND: Multimodal Attention Network.

<sup>cd</sup>CTR: clinical trial registration.

<sup>ce</sup>SVM: support vector machine.

<sup>cf</sup>HC: healthy control.

<sup>cg</sup>VR: virtual reality.

<sup>ch</sup>WiGMM: Warped Infinite Gaussian Mixture.

<sup>ci</sup>FCRN: fully convolutional residual network.

<sup>cj</sup>MLP: multilayer perceptron.

<sup>ck</sup>LFP: local field potentials.

<sup>cl</sup>TUG: Timed Up and Go.

<sup>cm</sup>CVD: cardiovascular disease.

<sup>cn</sup>ViT: vision transformer.

<sup>co</sup>VCI: vascular cognitive impairment.

## UK Biobank Dataset

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

Recent UK Biobank-based studies have applied diverse multimodal ML and deep learning approaches for AD risk prediction and diagnosis, integrating neuroimaging, genetic, clinical, and lifestyle data. These models generally achieved moderate to high performance (area under the curve [AUC]  $\approx 0.77$ - $0.90$ ) and demonstrated improved diagnostic utility compared with conventional assessment methods [79, 81-83]. Several studies further emphasized the importance of genetic and hormonal factors in risk stratification [84,85]. In addition, explainable and semisupervised frameworks have enhanced model interpretability and scalability for population-level analysis, facilitating clinically relevant phenotyping and disease monitoring [86,87].

This section describes multimodal model implementation in the UK Biobank. As shown in the analysis and Table 2, UK Biobank data support AD diagnosis and risk prediction, but limitations remain: class imbalance, which may bias training, and the need for external validation to confirm generalizability beyond the UK Biobank cohort.

## 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:

Recent ADNI-based studies have developed a wide range of multimodal and deep learning frameworks integrating neuroimaging, genetic, cognitive, and clinical data for AD diagnosis and MCI-to-AD progression prediction. Attention-based, multitask, ensemble, and time-to-event models have enabled accurate localization of disease-related regions,

improved prognostic modeling, and enhanced interpretability through explainable artificial intelligence techniques such as SHAP (Shapley Additive Explanations) and counterfactual analysis [15,75,76,88-92,97]. Several approaches further incorporated RL, semisupervised learning, and data augmentation to improve robustness and generalizability in heterogeneous and imbalanced datasets [98-103,107]. These models typically achieved high diagnostic and prognostic performance (AUC up to  $\approx 0.96$ ), with some demonstrating strong external validation and clinical relevance [93-96,101, 104-106,108]. Nevertheless, existing reviews and benchmarking studies have highlighted persistent limitations, including dataset bias, inconsistent evaluation protocols, and limited cross-center validation, underscoring the need for standardized and reproducible multimodal frameworks [119].

While ADNI provides a comprehensive and standardized multimodal resource for AD research and supports robust model performance, several limitations remain. These include class imbalance, underrepresentation of racially diverse populations, and limited external validation, which may bias model training and restrict generalizability across clinical settings.

## Self-Collected Datasets

While public datasets such as ADNI provide standardized benchmarks, self-collected datasets enable more flexible acquisition of targeted modalities. Representative studies include the following.

Studies based on self-collected datasets have explored diverse multimodal fusion strategies. EEG- and local field potentials-based models, as well as hybrid MRI-PET-biomarker frameworks, demonstrated high diagnostic and staging accuracy and supported interpretable risk mapping [77,109,111,114,115]. In parallel, behavioral and digital biomarkers derived from handwriting, eye tracking, virtual reality, and motion capture have enabled noninvasive and low-cost screening with strong classification performance [22,78,110,112]. Large-scale real-world health records and

hybrid deep learning models further facilitated population-level risk prediction and vascular cognitive impairment assessment, achieving robust AUC values above 0.90 [113, 120]. Overall, self-collected datasets have expanded the scope of multimodal AD research by enabling flexible modality integration and novel biomarker discovery, while remaining constrained by limited sample sizes and heterogeneous acquisition protocols.

Self-collected datasets offer distinct advantages, including targeted modality acquisition, novel biomarker discovery, such as microRNA, local field potentials, and handwriting, and enhanced real-world clinical utility. However, self-collected datasets typically endure limited sample sizes, which increases susceptibility to overfitting and compromises generalizability across diverse populations.

### 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 noninvasive 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 semistructured speech and language patterns pushes the development of AI in speech data. Recent work is shown in Table 3.

**Table 3.** Multimodal linguistic-based cognitive impairment datasets related papers.

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Ilias et al [121], 2023	ADReSS <sup>a</sup> and ADReSSo <sup>b</sup>	Multimodal transformer (BERT <sup>c</sup> and DeiT <sup>d</sup> ) with optimal transport	Dementia detection (AD <sup>e</sup> vs non-AD)	Audio (spectrograms) and text (transcripts)	Classification metrics and calibration	Internal: ADReSS or ADReSSo	Accuracy ≈91.25%, F <sub>1</sub> -score ≈91.06%; improved calibration vs baselines	Small, curated datasets, English-only, and potential overfitting
Poor et al [122], 2024	I-CONNECT <sup>f</sup>	Multimodal cross-transformer with coattention	MCI <sup>g</sup> prediction (MCI vs NC <sup>h</sup> )	Audio, text, and vision (facial video)	AUC <sup>i</sup> scores	Internal: cross-validation	Trimodal AUC 85.3%, and outperformed unimodal (60.9%) and bimodal (76.3%) models	Single cohort (I-CONNECT), cross-sectional, and complex architecture
Lin and Washington [123], 2024	DementiaBank (Pitt)	Wav2vec (audio) and Word2Vec (text)	Dementia classification	Audio, text, and timestamps	Accuracy and AUROC <sup>j</sup>	Internal: cross-validation	Text augmentation improved accuracy to ≈80% (AUROC 90%), and timestamps added minimal value	Single corpus: timestamps lacked resolution, and a modest sample size
Ortiz-Perez et al [124], 2023	DementiaBank (Pitt)	Multimodal ensemble (CNN <sup>k</sup> and transformer)	Prediction of dementia signs	Audio and text	Classification accuracy	Internal: held-out test sets	Text-only transformer best (accuracy 90.36%) and audio contributed less than text	Single English dataset, broad diagnosis category, and task constrained to picture description
Ilias and Askounis [125], 2022	ADReSS (DementiaBank)	Transformer (BERT) and Siamese Network	AD identification and severity estimation	Text (transcripts)	Accuracy and interpretability (LIME <sup>l</sup> )	Internal: cross-validation	Single-task accuracy 87.50%, multitask accuracy 86.25%, and distinct linguistic patterns identified	Small dataset, text only, MMSE <sup>m</sup> treated as categorical, and no acoustic information
Wen et al [126], 2023	DementiaBank (Pitt)	Transformer and causal counterfactual XAI <sup>n</sup>	AD detection	Text (part-of-speech tag features)	Accuracy; F <sub>1</sub> -score; feature importance	Internal: cross-validation	Accuracy 92.2%, F <sub>1</sub> -score 0.955, identified 12 key part-of-speech features linked to AD	Text only (part-of-speech), reliance on tagging accuracy, and no acoustic or imaging data
Chen et al [127], 2023	DementiaBank (Pitt)	SpeechFormer ++ (hierarchical transformer)	Paralinguistic AD detection	Audio (acoustic features)	Accuracy; F <sub>1</sub> -score	Internal: held-out test sets	Outperformed standard transformers and CNN/RNN <sup>o</sup>	Single corpus, complex computation, audio only,

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
Zheng et al [128], 2022	DementiaBank (Pitt)	N-gram, AWD-LSTM <sup>q</sup> , or neural models	Dementia detection	Text (context words, stop words, and part-of-speech)	Classification accuracy	Internal: held-out test data	baselines and SOTA <sup>p</sup> performance Combined model (vocabulary and grammar) accuracy 81.54%, and grammar contributes comparably to context	and no cross-lingual evaluation Specific to task or language, and moderate performance vs multimodal approaches
Nambiar et al [129], 2022	DementiaBank (Pitt)	Deep Classifiers (BERT/ALBERT <sup>r</sup> + BiLSTM <sup>s</sup> )	Early dementia detection	Text (transcripts)	Accuracy; $F_1$ -score	Internal: train and test splits	BERT + BiLSTM accuracy 0.812; ALBERT + BiLSTM $F_1$ -score 0.81; contextual embeddings superior	Text only; reliance on manual transcripts; single dataset
Priyadarshinee et al [130], 2023	ADReSSo-2021	ML <sup>t</sup> classifiers (SVM <sup>u</sup> , RF <sup>v</sup> , and NN <sup>w</sup> )	AD detection	Audio and text (transcripts)	Classification accuracy	Internal: held-out test set	Text features (accuracy 88.7%) outperformed audio, and file-level features were superior to frame-level	Benchmarking context, single task, and single language
Liu et al [131], 2023	ADReSS, ADReSSo, and the local Chinese dataset	Ensemble ML (VAD <sup>x</sup> pause and acoustic)	AD detection	Audio (acoustic and VAD pause features)	Accuracy	Internal: cross-validation; cross-lingual (Chinese)	Ensemble improved accuracy by $\approx 8\%$ on ADReSS, and accuracy 80% on the local Chinese dataset	Small local dataset (n=10), handcrafted features, and ensemble complexity
Shah et al [23], 2023	ADReSS-M	Logistic regression and SVR	Cross-lingual AD detection; MMSE regression	Audio (duration, pause, and intelligibility) and metadata	Accuracy and RMSE <sup>y</sup>	External: Greek test set	English cross-validation accuracy 74.7%, Greek Test accuracy 69.57%, and MMSE RMSE 4.77 (Greek)	Small Greek sample, modest accuracy, and simple ML models vs deep learning
Mahajan and Baths [132], 2021	ADReSS	Bimodal framework (CNN-LSTM <sup>z</sup> and Speech-GRU <sup>aa</sup> )	AD detection	Audio and text	Classification accuracy	Internal: cross-validation	Bimodal enriched model improved performance by $\approx 6.25\%$ over acoustic baselines	Small dataset, potential overfitting, and single task (picture description)
Mei et al [133], 2023	ADReSS-M	Bilingual wav2vec 2.0 + XGBoost <sup>ab</sup>	Cross-lingual AD detection and MMSE prediction	Audio (acoustic, silence, and low-frequency bands)	Accuracy and RMSE	External: Greek test set	Accuracy 73.9% (Greek), MMSE RMSE 4.610, and low-frequency speech aided transfer	Very small Greek sample, speech-only, and challenge context
Meerza et al [134], 2022	ADReSS	FL <sup>ac</sup> (LSTM <sup>ak</sup> and feed-forward)	Privacy-preserving AD diagnosis	Audio (Mel-frequency and pause features)	Accuracy and fairness metrics	Internal: simulated FL clients	FL accuracy close to the centralized baseline, and q-FedAvg improved fairness	Simulated clients, single dataset, and relies on feature extraction
Chen et al [135], 2023	ADReSS-M	SVM or NN on pretrained features	Cross-lingual AD detection	Audio (paralinguistic and XLSR-53 <sup>ac</sup> ), and text (ASR <sup>af</sup> )	Accuracy and RMSE	External: Greek test set	Accuracy 69.6% (Greek), RMSE 4.788, and paralinguistic features transferable	Performance below monolingual systems and reliance on ASR quality
Ilias et al [121], 2023	ADReSS	Multimodal transformer (ViT <sup>ag</sup> ,	AD detection	Audio (spectrograms) and text	Accuracy and $F_1$ -score	Internal: cross-validation	High eighties or low nineties accuracy, ViT is best for acoustic,	Small dataset, binary classification focus, and

Study	Datasets	Model type	Type of task	Modalities	Outcomes	Validation	Results	Limitation
		BERT, and GMU <sup>ah</sup> )					and fusion surpassed SOTA	external generalization untested
Tamm et al [136], 2023	ADReSS-M <sup>ai</sup>	Sequence models (transfer learning)	Cross-lingual AD detection and MMSE	Audio features and demographics	Accuracy and RMSE	External: Greek test set	Accuracy 82.6% (Greek), RMSE 4.345, and ranked second in the challenge	Small Greek sample, acoustic only, and transfer limited to English-Greek
Woszczyk et al [137], 2022	ADReSS	Transformers vs traditional ML	AD detection	Audio and text	Classification accuracy	Internal: held-out test data	Data augmentation improved performance and was comparable to SOTA	Augmentations tuned for ADReSS and a single speech task
Jin et al [138], 2023	ADReSS-M	CONSEN <sup>aj</sup> ensemble (acoustic and disfluency)	Multilingual AD detection and MMSE	Audio (acoustic embeddings and disfluency)	Accuracy and RMSE	External: Greek test set	First place in the challenge, accuracy 86.69% (Greek), and RMSE 3.727	Challenge dataset, ensemble complexity, and reliance on diarization quality

<sup>a</sup>ADReSS: Alzheimer Dementia Recognition Through Spontaneous Speech.

<sup>b</sup>ADReSSo: Alzheimer’s Dementia Recognition Through Spontaneous Speech only.

<sup>c</sup>BERT: Bidirectional Encoder Representations From Transformers.

<sup>d</sup>DeiT: Data-Efficient Image Transformers.

<sup>e</sup>AD: Alzheimer disease.

<sup>f</sup>I-CONNECT: Identifying Cognition in the Elderly Through Conversational Engagement.

<sup>g</sup>MCI: mild cognitive impairment.

<sup>h</sup>NC: normal control.

<sup>i</sup>AUC: area under the curve.

<sup>j</sup>AUROC: area under the receiver operating characteristic curve.

<sup>k</sup>CNN: convolutional neural network.

<sup>l</sup>LIME: Local Interpretable Model-Agnostic Explanations.

<sup>m</sup>MMSE: Mini-Mental State Examination.

<sup>n</sup>XAI: explainable artificial intelligence.

<sup>o</sup>RNN: recurrent neural network.

<sup>p</sup>SOTA: state of the art.

<sup>q</sup>AWD-LSTM: Average stochastic gradient descent weight-dropped long short-term memory

<sup>r</sup>ALBERT: A Lite Bidirectional Encoder Representations From Transformers.

<sup>s</sup>BiLSTM: bidirectional long short-term memory.

<sup>t</sup>ML: machine learning.

<sup>u</sup>SVM: support vector machine.

<sup>v</sup>RF: random forest.

<sup>w</sup>NN: neural network.

<sup>x</sup>VAD: voice activity detection.

<sup>y</sup>RMSE: root mean square error.

<sup>z</sup>CNN-LSTM: convolutional neural network long short-term memory.

<sup>aa</sup>Speech-GRU: Speech Gated Recurrent Unit.

<sup>ab</sup>XGBoost: Extreme Gradient Boosting.

<sup>ac</sup>FL: federated learning.

<sup>ad</sup>LSTM: long short-term memory.

<sup>ae</sup>XLSR-53: cross-lingual speech representation-version 53

<sup>af</sup>ASR: automatic speech recognition.

<sup>ag</sup>ViT: vision transformer.

<sup>ah</sup>GMU: gated multimodal unit.

<sup>ai</sup>ADReSS-M: Alzheimer Dementia Recognition through Spontaneous Speech – Multimodal.

<sup>aj</sup>CONSEN: complementary and simultaneous ensemble.

Recent studies have shown that multimodal fusion of speech and text using transformer-based architectures substantially improves AD detection performance, with  $F_1$ -scores above 0.90 on ADReSS and ADReSSo (Alzheimer’s Dementia Recognition Through Spontaneous Speech 2021 Challenge) datasets [121,132,139]. Linguistic feature engineering and interpretable language models further enhanced classification accuracy, achieving up to 92.2% accuracy and  $F_1$ -scores of

0.955 using compact part-of-speech features [124-126,128, 130]. Cross-lingual approaches based on language-agnostic and transfer learning methods enabled moderate generalization, with accuracies ranging from 69% to 73.9% in English-Greek transfer settings [23,127,133,136]. To support real-world deployment, lightweight and hierarchical models achieved around 80% accuracy with reduced computational cost [131,135]. In addition, data augmentation and ensemble

strategies improved robustness in low-resource scenarios, yielding  $F_1$ -score gains of 5%-7% and competitive challenge performance (accuracy 86.69%) [123,137,138].

### Summarization Based on All Multimodal Datasets and Quantitative Analysis

Table 2 and Table 3 summarize the recent state-of-the-art models across the 2 major types of multimodal datasets, extracted according to the Cochrane Handbook. Full QUADAS-2 forms are available in Multimedia Appendix 5. Based on these results, the following quantitative synthesis compares performance trends across all multimodal datasets. Across the 4 major dataset categories, modality choices and model performance show clear dataset-dependent patterns as shown in Table 4. UK Biobank studies mainly combine MRI, clinical variables, and genetic features, with 2 diagnosis studies reporting an average accuracy of 71.4% (SD 5.2%) and 4 risk-prediction studies reaching an average AUC of 0.84 (SD 0.056). ADNI studies use the most comprehensive modality integrations, with 3 diagnosis studies

averaging 92.5% (SD 3.8%) accuracy, 3 MCI-conversion studies achieving a mean AUC of 0.922 (SD 0.045), and risk-prediction studies reaching an average AUC of 0.81 (SD 0.06); these tasks collectively achieve the strongest results, with fusion models frequently reporting AUC values above 0.95. DementiaBank studies differ fundamentally by focusing on speech- and language-based modalities; 9 diagnosis studies report an average AUC of 0.813 (SD 0.042), and 5 cross-lingual AD-detection studies show a mean accuracy of 77% (SD 6.5%), where transformer architectures consistently outperform classical approaches, with models such as BERT + DeiT (Data-Efficient Image Transformers), BERT + ViT (vision transformer), and RoBERTa + (Robustly Optimized Bidirectional Encoder Representations From Transformers Approach) DNN (deep neural network) showing  $F_1$ -scores exceeding 0.90. Self-collected datasets are typically smaller and more heterogeneous; 3 diagnosis studies report an average accuracy of 96% (SD 2.4%), and lightweight models such as EEGNet or ViT-based hybrids demonstrate strong predictive capacity when applied to EEG or structural MRI.

**Table 4.** Summary of representative modality combinations and top-performing models in multimodal AI<sup>a</sup>-aided AD<sup>b</sup> diagnosis.

Dataset and task	Counts	Average performance	Best performance modalities	Related article
<b>UK Biobank</b>				
Diagnosis	2	Accuracy=71.4%	Retinal fundus images	[79,81-87,140]
Risk prediction	4	AUC <sup>c</sup> =84%	Clinical, biological assays, cognitive tests, and physical measures	[79,81-87,140]
Other	3	N/A <sup>d</sup>	Multimodal MRI <sup>e</sup> (T1, T2, MRI, etc)	[79,81-87,140]
<b>ADNI<sup>f</sup></b>				
Diagnosis	3	Accuracy=92.5%	Structural MRI features and neuropsychological tests	[15,75,76,79,80,89-92,94-108,119,141]
MCI <sup>g</sup> conversion	3	AUC=92.2%	Structural MRI, clinical variables, and genetics (SNP <sup>h</sup> )	[15,75,76,79,80,89-92,94-108,119,141]
MMSE <sup>i</sup> regression	2	No integration	Whole-brain T1-weighted MRI and clinical scores	[15,75,76,79,80,89-92,94-108,119,141]
Risk prediction	7	AUC=81%	MRI, PET <sup>j</sup> , clinical, and cognitive	[15,75,76,79,80,89-92,94-108,119,141]
Other	13	N/A	N/A	[15,75,76,79,80,89-92,94-108,119,141]
<b>Dementia bank</b>				
Diagnosis	9	AUC=81.3%	Text transcripts → part-of-speech feature vectors	Table 3
Cross-lingual AD detection	5	Accuracy=77%	Multimodal acoustic fusion	Table 3
Other	6	N/A	N/A	Table 3
<b>Self-collected datasets</b>				
Diagnosis	3	Accuracy=96%	MRI, PET, clinical, and genotype	[22,77,78,106,109-115,117,120,142]
Other	6	No integration	Different task	[22,77,78,106,109-115,117,120,142]

<sup>a</sup>AI: artificial intelligence.

<sup>b</sup>AD: Alzheimer disease.

<sup>c</sup>AUC: area under the curve.

<sup>d</sup>N/A: not available.

<sup>e</sup>MRI: magnetic resonance imaging.

<sup>f</sup>ADNI: Alzheimer Disease Neuroimaging Initiative.

<sup>g</sup>MCI: mild cognitive impairment.

<sup>h</sup>SNP: single-nucleotide polymorphism.

<sup>i</sup>MMSE: Mini-Mental State Examination.

<sup>j</sup>PET: positron emission tomography.

To interpret these results and limit metric inflation, note that purely internal cross-validation tends to overestimate performance: AUC is typically  $\approx 5$ -15 points higher than with external validation. Small or tightly controlled datasets also report accuracies  $\approx 10\%$ -20% above those in large, heterogeneous cohorts. Severe class imbalance can further raise accuracy while lowering  $F_1$ -score or sensitivity; without correction, imbalance may inflate results by  $\approx 5\%$ -12%. Cross-sectional models often score higher in single-timepoint evaluations, whereas longitudinal designs usually yield lower but more stable estimates, which are more informative for follow-up and clinical use.

These findings should be interpreted in light of substantial heterogeneity and risk of bias. Variation in sample composition, task definitions, and evaluation procedures across datasets limits direct comparison of performance metrics. QUADAS-2 also indicated frequent unclear and high-risk in-patient selection, reference standards, and flow or timing, especially in studies using only internal validation or selected samples. Reported metrics, therefore, likely represent upper-bound estimates rather than expected real-world performance, and apparent gains often reflect dataset-specific effects rather than generalizable model superiority.

Overall, the evidence shows that modality effectiveness varies substantially across datasets, transformer models deliver the highest gains in speech-language tasks, and large clinical phenotyping datasets such as UK Biobank and ADNI still rely mainly on traditional machine-learning or custom fusion frameworks rather than modern cross-modal transformers. This gap highlights an opportunity to develop transformer-based multimodal integration approaches tailored to large, heterogeneous clinical datasets.

### **Multimodal Fusion Taxonomy**

A structured multimodal fusion taxonomy clarifies the performance of different integration strategies across datasets (Tables 2 and 3). A total of 4 main paradigms are commonly used: early, intermediate, late, and attention- or graph-based fusion.

Early fusion concatenates low-level features and performs well for aligned modalities such as MRI + PET, often achieving  $AUC > 0.95$  in ADNI studies, but is sensitive to missing data and feature-scale heterogeneity. Intermediate fusion combines latent representations from modality-specific encoders and is effective for heterogeneous inputs such as MRI + speech or EEG + clinical data, as demonstrated by high performance in ADRess-based models, although it may be unstable in small datasets. Late fusion aggregates model outputs and is robust to missing modalities, performing well in large datasets such as the UK Biobank, but underuses fine-grained cross-modal interactions.

Across paradigms, limited modality availability and high acquisition costs remain key challenges, underscoring the need for adaptive and clinically feasible fusion strategies.

## **Discussion**

### **Principal Findings**

This review synthesized multimodal AI studies for AD across diverse dataset families, including clinical phenotyping and cognitive-linguistic datasets. Multimodal fusion generally outperformed unimodal baselines, but the gain is dataset-dependent and should be interpreted cautiously. Strong performance in curated cohorts and constrained speech benchmarks may not generalize to population-based or multicenter settings. QUADAS-2 also indicated frequent risk of bias and unclear reporting across domains, likely inflating metrics and limiting comparability. Accordingly, headline accuracy and AUC should be treated as upper-bound estimates unless supported by external validation and transparent reporting.

### **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, several challenges remain that warrant careful examination. In this section, this systematic review summarizes the common limitations identified in existing studies and proposes directions for future research to advance the field.

### **Clinical and Translational Implications**

Multimodal AI could support AD diagnosis through several clinical pathways. In memory clinics, models combining MRI, cognitive scores, and blood biomarkers could triage referrals, prioritizing patients for specialist review or PET. In general practice, speech-based and routine clinical-feature models could be embedded in consultations to flag early cognitive change. In radiology, MRI-clinical fusion could act as a second reader, reducing interobserver variability and supporting less experienced clinicians. Where imaging or specialist access is limited, speech, digital questionnaires, and basic clinical data could enable telemedicine-based screening and follow-up. At the population level, these models could support risk stratification and targeted monitoring. To enable real-world deployment, research should prioritize external multicenter validation, integration with electronic health records, and evaluation of regulatory feasibility, cost-effectiveness, and clinical impact.

### **Ethical and Regulatory Implications**

Deploying multimodal AI for AD diagnosis requires ethical and regulatory safeguards. As datasets often combine imaging, clinical records, genomics, and speech, they fall under strict privacy regimes (eg, General Data Protection Regulation in the European Union; HIPAA [Health Insurance Portability and Accountability Act] in the United States), requiring explicit consent, data minimization, and secure handling, with added complexity for sensitive modalities such as speech and genomic data. Clinical deployment is

also shaped by medical-AI governance frameworks (eg, the European Union AI Act, Food and Drug Administration Software as a Medical Device guidance, and UK Medicines and Healthcare products Regulatory Agency Good Machine Learning Practice), which emphasize transparency, risk management, and postdeployment monitoring. Fairness is essential because demographic imbalance can yield uneven performance across age, ethnicity, and language groups. Interpretability (eg, imaging attention maps and linguistic saliency) supports clinical accountability and aligns with explainability expectations. Future work should incorporate privacy-preserving methods, bias audits, and regulatory-aligned validation pipelines to enable responsible clinical integration.

### **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 usage of comprehensive datasets needed for robust external validation and generalizability.

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 et al [134] pioneered FL for AD speech diagnosis using mel-frequency cepstral coefficients and pause features, maintaining model performance while ensuring privacy through q-FedAvg/q-FedSGD optimization. Nambiar [129] validated an ALBERT (A Lite Bidirectional Encoder Representations From Transformers) + BiLSTM (bidirectional long short-term memory) 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 [15,79,88,95,100,139,140].

Despite encouraging results, FL still lacks harmonized protocols and interoperable platforms. This limits cross-center reproducibility and weakens clinical credibility. International collaboration also remains constrained by regulatory differences. Future work should prioritize unified federated frameworks with standardized protocols and privacy-preserving methods to enable secure global data collaboration [143, 144].

As most datasets lack fully matched modalities per participant, multimodal fusion often relies on representation- or population-level integration rather than early fusion. Early fusion requires paired samples and is therefore infeasible across datasets. By contrast, late fusion and embedding-level integration can train unimodal models separately and combine them via meta-learners, cross-modal transformers, or probabilistic ensembles. Domain adaptation, transfer learning, and harmonization can also combine heterogeneous cohorts at the population level to improve generalizability. A standardized benchmark could further support this by defining

shared preprocessing, label taxonomies, and evaluation metrics, enabling meaningful comparison or representation-stage fusion even without subject-level pairing.

### **Data Imbalance**

Severe class imbalance remains a major obstacle, biasing training toward the majority class and inflating accuracy while masking low sensitivity to early disease. In addition, datasets such as the UK Biobank are dominated by White European ancestry, limiting generalizability across racially and ethnically diverse populations. Addressing this requires both technical mitigation and proactive recruitment of underrepresented groups so models better reflect population heterogeneity.

Researchers have applied data-level interventions such as generative adversarial network-based augmentation, diffusion models, and resampling [123,137,138,145,146]; algorithm-level solutions, including cost-sensitive, loss-focused, ensemble, and class-weighted training schemes [147-152]; and evaluation-focused remedies [153] 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 characterize underrepresented populations to ensure equitable and robust clinical applicability across diverse populations.

### **Lack of Standardized and Longitudinal Data**

Differences in acquisition protocols and diagnostic criteria across datasets limit comparability of imaging, cognitive, and biomarker outcomes. Longitudinal evidence is also constrained: even in relatively standardized resources such as ADNI, limited long-term follow-up hampers modeling the temporal dynamics of disease progression.

Future work should standardize key acquisition elements and diagnostic criteria across longitudinal studies and strengthen coordination across institutions. Building on this, a multimodal benchmark spanning imaging, clinical, biomarker, behavioral, and linguistic modalities would enable cross-dataset validation, improve comparability, and support reproducible evaluation of new models. These steps would strengthen temporal modeling and provide more reliable evidence for clinical translation.

### **Dataset-Specific Limitations**

Data imbalance is prevalent across many AD datasets, but the nature of this issue varies substantially between cohorts. This review, therefore, outlines the dataset-specific limitations of commonly used AD cohorts and corpora.

ADNI participants are generally healthier, with fewer comorbidities and a restricted age range (55-90 y), limiting representativeness. Protocol differences across centers and

evolving diagnostic standards introduce heterogeneity, while frequent reliance on subsets hampers comparability [154].

Of UK Biobank, dementia outcomes are derived mainly from health records, leading to potential misclassification and delayed ascertainment. Participants show strong volunteer bias, and PET or cerebrospinal fluid biomarkers are limited to a small subset, constraining multimodal analyses [155].

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 [156].

Of NACC, data are aggregated from multiple centers with heterogeneous recruitment and diagnostic protocols, making harmonization challenging. The cohort is clinic-based rather than population-representative, and missing biomarker modalities are common [157].

Although high-quality, Australian Imaging, Biomarkers and Lifestyle Study is smaller than ADNI and NACC and is often used only for validation. Regional recruitment and protocol differences reduce ethnic diversity and cross-cohort comparability [158].

Of Pitt Corpus, this is the most widely used speech dataset, but remains small and imbalanced. Tasks are constrained, limiting ecological validity, and cross-linguistic generalizability is poor [159].

Of the ADReSS series, the ADReSS benchmarks provide standardized 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 [18].

Of 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 generalizability [153].

Dataset challenges are compounded by unrepresentative cohorts, incomplete modalities, and poor cross-center consistency, limiting model robustness and cross-dataset generalization in AD diagnosis. Future work should improve data coordination and standardization, enable more practical sharing mechanisms, and adopt cross-cohort validation where feasible. Strengthening data quality and access is essential for translating multimodal AI methods into clinical use.

### ***Model Interpretability and Explainability***

A major limitation of multimodal ML models in clinical AD diagnosis is limited interpretability and transparency. Many high-performing models provide insufficient insight into their decision processes, which can hinder clinical adoption and reduce confidence among end users.

Efforts that have been made toward model interpretability include designing inherently transparent models. For example, some studies demonstrate emerging explainability strategies, including hybrid neuro-symbolic models [160] that generate interpretable reports and post hoc methods such as

SHAP, LIME (Local Interpretable Model-Agnostic Explanations), gradient-based saliency, and graph-masking techniques [161,162], which collectively enhance transparency in multimodal AD diagnosis.

Current interpretability methods often fail to produce explanations that clinicians can use reliably. Future work should prioritize clinically grounded explainability, including interactive visualizations and concise workflow-aligned natural-language summaries. Hybrid designs that combine deep learning with structured reasoning can further improve transparency by making decision logic explicit. For deployment, models should also report prediction uncertainty and demonstrate compatibility with clinical systems and regulatory requirements.

Beyond technical advances, incorporating patient and public involvement can improve multimodal AI development for AD. Patients and caregivers can help shape evaluation and result communication, not just act as end users, aligning explanations with patient priorities and addressing transparency and fairness. Engaging patient and public involvement earlier in model design may therefore support more interpretable and clinically usable diagnostic tools.

### ***Heterogeneous Multiview Learning Problem***

Integrating data across studies is challenging because single datasets rarely cover all modalities, forcing combinations such as ADNI with UK Biobank. However, differences in cohorts, imaging protocols, and cognitive assessment frameworks create substantial heterogeneity that limits direct pooling and comparability.

This heterogeneity hinders building unified models that generalize across nonoverlapping cohorts, so single-dataset models often fail out of domain. Platform-agnostic methods that tolerate missing or inconsistent modalities are therefore needed. Proposed solutions include shared latent-space learning [163], multibranch networks [164], and mixture-of-experts architectures [165] to support partial fusion and cross-dataset adaptation, but most still assume strong cross-domain alignment or require substantial retraining under domain shift.

Despite recent progress, multimodal methods often assume strict cross-domain alignment and require extensive retraining under domain shift or missing modalities. Future work should develop robust, platform-agnostic frameworks that adapt to changing modality availability and distribution shifts with minimal performance loss and advance representation learning to derive stable joint embeddings from heterogeneous data.

### ***Uncertainty Quantification and Clinical Applicability***

Although multimodal AD models have advanced, most studies still omit uncertainty quantification (eg, confidence or prediction intervals). Models typically provide deterministic outputs without communicating reliability, despite clinicians relying on uncertainty to guide management and treatment

decisions. Future work should embed uncertainty metrics into diagnostic models to better align with clinical needs and improve interpretability, reliability, and real-world adoption.

### **Risks of Data Leakage in Multimodal AI Modeling**

Another limitation is data leakage, which can inflate performance. Common forms include subject-level leakage (samples from the same participant in both training and test sets), patch-level overlap in MRI slice and patch models, and transcript or utterance-level leakage in speech datasets when multiple segments come from 1 individual. Many studies did not report whether participant-independent splits were enforced. Clearer reporting of partitioning and rigorous participant-level cross-validation are therefore essential to ensure real-world generalizability.

### **Conclusions**

This review synthesizes evidence on multimodal AI approaches for AD across clinical, neuroimaging, genetic, and linguistic data, systematically comparing modeling strategies, validation practices, and performance trends across heterogeneous datasets. In contrast to prior modality-specific

reviews, the findings show that multimodal models generally outperform unimodal approaches, although performance varies substantially with dataset characteristics, modality availability, and cross-source alignment. High accuracies are often reported in curated or internally validated cohorts, whereas population-based and externally validated studies yield more modest but clinically realistic results, reflecting substantial heterogeneity and risk of bias.

Despite these limitations, the evidence demonstrates that multimodal AI captures complementary biological and behavioral signals relevant to AD, offering clear advantages for diagnosis and risk prediction. Transformer-based architectures and speech- or behavior-derived modalities show promise for scalable and noninvasive early detection. However, meaningful clinical translation will require harmonized benchmarking, transparent reporting, and rigorous external validation. Overall, this review advances the field by contextualizing performance gains within their methodological constraints and by outlining practical directions for developing robust, interpretable, and generalizable multimodal AI systems. These insights support the responsible integration of AI into real-world dementia screening, risk prediction, and early intervention strategies.

### **Acknowledgments**

The authors declare the use of generative artificial intelligence (GenAI) in the research and writing process. According to the GAIDeT (Generative Artificial Intelligence for Digital Twins) taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing. The GenAI tool used was ChatGPT-5.2. Responsibility for the final manuscript lies entirely with the authors. Declaration submitted by: JMIR Publications. We used ChatGPT-5.2 (OpenAI) to conduct a grammatical review of the abstract and conclusion sections.

### **Funding**

This work received no specific financial or nonfinancial support. No funder or sponsor had any role in the design of the review; data collection, analysis, or interpretation; writing of this paper; or the decision to submit for publication.

### **Data Availability**

This systematic review did not generate any new datasets. All data extracted and analyzed in this systematic review were obtained from publicly available publications included in the review. No additional unpublished or proprietary data were used.

### **Conflicts of Interest**

None declared.

### **Multimedia Appendix 1**

Full database search strategies for PubMed, Scopus, IEEE Xplore, and ACM Digital Library, including complete Boolean queries, search fields, filters, and publication date limits used for study identification.

[\[DOCX File \(Microsoft Word File\), 999 KB-Multimedia Appendix 1\]](#)

### **Multimedia Appendix 2**

Performance evaluation for AD diagnosis. AD: Alzheimer disease.

[\[DOCX File \(Microsoft Word File\), 22 KB-Multimedia Appendix 2\]](#)

### **Multimedia Appendix 3**

Complete QUADAS-2 risk-of-bias assessments for all included studies, summarizing judgments across patient selection, index test, reference standard, and flow or timing, with detailed study-level ratings. QUADAS-2: Revised Quality Assessment of Diagnostic Accuracy Studies Tool.

[\[DOCX File \(Microsoft Word File\), 38 KB-Multimedia Appendix 3\]](#)

### **Multimedia Appendix 4**

Overview of traditional machine-learning models applied in Alzheimer disease research, including SVM, decision trees, HMMs, KNN, logistic regression, GMMs, and foundational CNN or RL descriptions, with methodological principles and limitations. CNN: convolutional neural network; GMM: gaussian mixture models; HMM: hidden markov model ; KNN: k-nearest neighbors; RL: reinforcement learning; SVM: support vector machine.

[\[DOCX File \(Microsoft Word File\), 42 KB-Multimedia Appendix 4\]](#)

### Multimedia Appendix 5

Cochrane Handbook 5.3.3–aligned data-extraction tables summarizing study design, datasets, participants, modalities, preprocessing, model architectures, validation schemes, outcomes, and limitations for all included studies.

[\[DOCX File \(Microsoft Word File\), 185 KB-Multimedia Appendix 5\]](#)

### Checklist 1

Completed PRISMA 2020, PRISMA-S checklist, and PRISMA expanded checklist specifying reporting locations for all required items, including eligibility criteria, search methods, extraction procedures, bias assessments, and synthesis reporting. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for literature searches.

[\[PDF File \(Adobe File\), 11026 KB-Checklist 1\]](#)

### References

- Scheltens P, De Strooper B, Kivipelto M, et al. Alzheimer's disease. *Lancet*. Apr 24, 2021;397(10284):1577-1590. [doi: [10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4)] [Medline: [33667416](https://pubmed.ncbi.nlm.nih.gov/33667416/)]
- 2024 Alzheimer's disease facts and figures. *Alzheimers Dement*. May 2024;20(5):3708-3821. [doi: [10.1002/alz.13809](https://doi.org/10.1002/alz.13809)]
- World alzheimer report 2024. Alzheimer's Disease International; 2024. URL: <https://www.alzint.org/resource/world-alzheimer-report-2024/> [Accessed 2026-02-10]
- Kaštelan S, Gverović Antunica A, Puzović V, et al. Non-invasive retinal biomarkers for early diagnosis of Alzheimer's disease. *Biomedicines*. Jan 24, 2025;13(2):283. [doi: [10.3390/biomedicines13020283](https://doi.org/10.3390/biomedicines13020283)] [Medline: [40002697](https://pubmed.ncbi.nlm.nih.gov/40002697/)]
- Castellano G, Esposito A, Lella E, Montanaro G, Vessio G. Automated detection of Alzheimer's disease: a multi-modal approach with 3D MRI and amyloid PET. *Sci Rep*. Mar 3, 2024;14(1):5210. [doi: [10.1038/s41598-024-56001-9](https://doi.org/10.1038/s41598-024-56001-9)] [Medline: [38433282](https://pubmed.ncbi.nlm.nih.gov/38433282/)]
- Bi Y, Abrol A, Fu Z, Calhoun VD. A multimodal vision transformer for interpretable fusion of functional and structural neuroimaging data. *Hum Brain Mapp*. Dec 1, 2024;45(17):e26783. [doi: [10.1002/hbm.26783](https://doi.org/10.1002/hbm.26783)] [Medline: [39600159](https://pubmed.ncbi.nlm.nih.gov/39600159/)]
- Yu Q, Ma Q, Da L, et al. A transformer-based unified multimodal framework for Alzheimer's disease assessment. *Comput Biol Med*. Sep 2024;180:108979. [doi: [10.1016/j.combiomed.2024.108979](https://doi.org/10.1016/j.combiomed.2024.108979)] [Medline: [39098237](https://pubmed.ncbi.nlm.nih.gov/39098237/)]
- Leng Y, He Y, Amini S, et al. A GPT-4o-powered framework for identifying cognitive impairment stages in electronic health records. *npj Digit Med*. Jul 3, 2025;8(1):401. [doi: [10.1038/s41746-025-01834-5](https://doi.org/10.1038/s41746-025-01834-5)]
- Balabin H, Tamm B, Spruyt L, et al. Natural language processing-based classification of early Alzheimer's disease from connected speech. *Alzheimer's Dement*. Feb 2025;21(2):e14530. [doi: [10.1002/alz.14530](https://doi.org/10.1002/alz.14530)] [Medline: [39868827](https://pubmed.ncbi.nlm.nih.gov/39868827/)]
- Yang X, Hong K, Zhang D, Wang K. Early diagnosis of Alzheimer's disease based on multi-attention mechanism. *Fati SM*, editor. *PLOS ONE*. 2024;19(9):e0310966. [doi: [10.1371/journal.pone.0310966](https://doi.org/10.1371/journal.pone.0310966)]
- Wijeratne PA, Alexander DC. Learning transition times in event sequences: the event-based hidden markov model of disease progression. *Inf Process Med Imaging*. Jun 2021;12729(14):583-595. [doi: [10.1007/978-3-030-78191-0\\_45](https://doi.org/10.1007/978-3-030-78191-0_45)]
- Huh YJ, Park JH, Kim YJ, Kim KG. Ensemble learning-based Alzheimer's disease classification using electroencephalogram signals and clock drawing test images. *Sensors (Basel)*. May 2, 2025;25(9):2881. [doi: [10.3390/s25092881](https://doi.org/10.3390/s25092881)] [Medline: [40363322](https://pubmed.ncbi.nlm.nih.gov/40363322/)]
- Karasu E, Baytaş İM. Conversion-aware forecasting of Alzheimer's disease via featurewise attention. *Pattern Anal Applic*. Jun 2025;28(2):64. [doi: [10.1007/s10044-025-01447-4](https://doi.org/10.1007/s10044-025-01447-4)]
- Xiao X, Li Y, Wu Q, et al. 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*. May 13, 2025;17(1):103. [doi: [10.1186/s13195-025-01750-6](https://doi.org/10.1186/s13195-025-01750-6)]
- Qiu S, Miller MI, Joshi PS, et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat Commun*. Jun 20, 2022;13(1):3404. [doi: [10.1038/s41467-022-31037-5](https://doi.org/10.1038/s41467-022-31037-5)] [Medline: [35725739](https://pubmed.ncbi.nlm.nih.gov/35725739/)]
- Chakravarthi BA, Shivakanth G. Integrating multimodal AI techniques and MRI preprocessing for enhanced diagnosis of Alzheimer's disease: clinical applications and research horizons. *IEEE Access*. 2025;13:63519-63531. [doi: [10.1109/ACCESS.2025.3557533](https://doi.org/10.1109/ACCESS.2025.3557533)]
- Elazab A, Wang C, Abdelaziz M, et al. Alzheimer's disease diagnosis from single and multimodal data using machine and deep learning models: achievements and future directions. *Expert Syst Appl*. Dec 2024;255:124780. [doi: [10.1016/j.eswa.2024.124780](https://doi.org/10.1016/j.eswa.2024.124780)]

18. Ding K, Chetty M, Noori Hoshyar A, Bhattacharya T, Klein B. Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges. *Artif Intell Rev*. Oct 12, 2024;57(12):325. [doi: [10.1007/s10462-024-10961-6](https://doi.org/10.1007/s10462-024-10961-6)]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
20. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. Jan 26, 2021;10(1):39. [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
21. Li T, Higgins JP, Deeks JJ. Collecting data. In: *Cochrane Handbook for Systematic Reviews of Interventions*. 2019:109-141. [doi: [10.1002/9781119536604](https://doi.org/10.1002/9781119536604)] ISBN: 978-1-119-53660-4
22. Park B, Kim Y, Park J, et al. 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*. Apr 17, 2024;26:e54538. [doi: [10.2196/54538](https://doi.org/10.2196/54538)] [Medline: [38631021](https://pubmed.ncbi.nlm.nih.gov/38631021/)]
23. Shah Z, Qi SA, Wang F, et al. Exploring language-agnostic speech representations using domain knowledge for detecting Alzheimer's dementia. *ICASSP 2023 - 2023 IEEE Int Conf Acoustics, Speech Signal Process (ICASSP)*. 2023:1-2. [doi: [10.1109/ICASSP49357.2023.10095593](https://doi.org/10.1109/ICASSP49357.2023.10095593)]
24. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. Sep 1995;20(3):273-297. [doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)]
25. Sharma A, Kaur S, Memon N, Jainul Fathima A, Ray S, Bhatt MW. Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis. *Neurosci Inf*. Nov 2021;1(3):100012. [doi: [10.1016/j.neuri.2021.100012](https://doi.org/10.1016/j.neuri.2021.100012)]
26. Gao X, Liu H, Shi F, Shen D, Liu M. Brain status transferring generative adversarial network for decoding individualized atrophy in Alzheimer's disease. *IEEE J Biomed Health Inform*. Oct 2023;27(10):4961-4970. [doi: [10.1109/JBHI.2023.3304388](https://doi.org/10.1109/JBHI.2023.3304388)] [Medline: [37607152](https://pubmed.ncbi.nlm.nih.gov/37607152/)]
27. Lazli L. Improved Alzheimer disease diagnosis with a machine learning approach and neuroimaging: case study development. *JMIRx Med*. Apr 21, 2025;6:e60866. [doi: [10.2196/60866](https://doi.org/10.2196/60866)] [Medline: [40257754](https://pubmed.ncbi.nlm.nih.gov/40257754/)]
28. Hossain F, Halder RK, Uddin MN. An integrated machine learning based adaptive error minimization framework for Alzheimer's stage identification. *Intell-Based Med*. 2025;11:100243. [doi: [10.1016/j.ibmed.2025.100243](https://doi.org/10.1016/j.ibmed.2025.100243)]
29. Fulkar B, Dhale T, Pacharane U, Deshmukh S. Early detection of chronic diseases using machine and deep learning algorithms. *2025 4th Int Conf Sentiment Anal Deep Learn (ICSADL)*. 2025:1656-1661. [doi: [10.1109/ICSADL65848.2025.10933005](https://doi.org/10.1109/ICSADL65848.2025.10933005)]
30. Sathiy A, Basha CH, S V, Sharmila P JJ, S P, Indhumathi R. Enhancing Alzheimer's disease detection using optimized attribute selection and random forest classifier for improved accuracy. *2025 Int Conf Visual Anal Data Visualization (ICVADV)*. 2025:1174-1179. [doi: [10.1109/ICVADV63329.2025.10961844](https://doi.org/10.1109/ICVADV63329.2025.10961844)]
31. Saleh AW, Gupta G, Khan SB, Alkhalidi NA, Verma A. An Alzheimer's disease classification model using transfer learning Densenet with embedded healthcare decision support system. *Decis Anal J*. Dec 2023;9:100348. [doi: [10.1016/j.dajour.2023.100348](https://doi.org/10.1016/j.dajour.2023.100348)]
32. Baucum M, Khojandi A, Papamarkou T. Hidden markov models as recurrent neural networks: an application to Alzheimer's disease. *2021 IEEE 21st Int Conf Bioinf Bioeng (BIBE)*. 2021:1-6. [doi: [10.1109/BIBE52308.2021.9635256](https://doi.org/10.1109/BIBE52308.2021.9635256)]
33. Cai Z, Zeng D, Marder KS, Honig LS, Wang Y. Dynamic classification of latent disease progression with auxiliary surrogate labels. *arXiv*. Preprint posted online on Dec 11, 2024. [doi: [10.3390/math11102335](https://doi.org/10.3390/math11102335)]
34. Chen Y, Pham TD. Development of a brain MRI-based hidden Markov model for dementia recognition. *Biomed Eng Online*. 2013;12 Suppl 1(Suppl 1):S2. [doi: [10.1186/1475-925X-12-S1-S2](https://doi.org/10.1186/1475-925X-12-S1-S2)] [Medline: [24564961](https://pubmed.ncbi.nlm.nih.gov/24564961/)]
35. Vats NA, Yadavalli A, Gurugubelli K, Vuppala AK. Acoustic features, BERT model and their complementary nature for Alzheimer's dementia detection. *IC3 '21*. Aug 5, 2021:267-272. [doi: [10.1145/3474124.3474162](https://doi.org/10.1145/3474124.3474162)]
36. Xiao R, Cui X, Qiao H, et al. Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net. *Biomed Signal Process Control*. Apr 2021;66:102362. [doi: [10.1016/j.bspc.2020.102362](https://doi.org/10.1016/j.bspc.2020.102362)]
37. Ablimit A, Botelho C, Abad A, Schultz T, Trancoso I. Exploring dementia detection from speech: cross corpus analysis. *ICASSP 2022 - 2022 IEEE Int Conf Acoust, Speech Signal Proc (ICASSP)*. 2022:6472-6476. [doi: [10.1109/ICASSP43922.2022.9747167](https://doi.org/10.1109/ICASSP43922.2022.9747167)]
38. Lahmiri S. Integrating convolutional neural networks, kNN, and Bayesian optimization for efficient diagnosis of Alzheimer's disease in magnetic resonance images. *Biomed Signal Process Control*. Feb 2023;80:104375. [doi: [10.1016/j.bspc.2022.104375](https://doi.org/10.1016/j.bspc.2022.104375)]
39. Suwalka D, Pandita D, Godse S, Patil RR, Salam Khan A, Kumar A. AI applications and simulation-based learning integrating future of nursing education. *2024 Int Conf Intell Innovative Pract Eng Manage (IIPEM)*. 2024:1-6. [doi: [10.1109/IIPEM62726.2024.10925639](https://doi.org/10.1109/IIPEM62726.2024.10925639)]

40. Chaudhari A, Saratkar S, Thute T. AI-enhanced imaging techniques for understanding Alzheimer's progression. 2025 Int Conf Mach Learn Auton Syst (ICMLAS). 2025:1174-1179. [doi: [10.1109/ICMLAS64557.2025.10969042](https://doi.org/10.1109/ICMLAS64557.2025.10969042)]
41. Ango R, C KKR, Fatima S, Nag A. Brain connectivity analysis in Alzheimer's disease using graph convolutional network. 2024 4th Int Conf Soft Comput Secur Appl (ICSCSA). 2024:133-139. [doi: [10.1109/ICSCSA64454.2024.00028](https://doi.org/10.1109/ICSCSA64454.2024.00028)]
42. Chattopadhyay T, Joshy NA, Ozarkar SS, et al. Deep learning algorithms for Alzheimer's disease detection based on diffusion MRI: tests in Indian and North American cohorts. Alzheimer's Dementia. Dec 2024;20(S2):e089294. [doi: [10.1002/alz.089294](https://doi.org/10.1002/alz.089294)] [Medline: [39737627](https://pubmed.ncbi.nlm.nih.gov/39737627/)]
43. Ma D, Zhang H, Wang L. Editorial: deep learning methods and applications in brain imaging for the diagnosis of neurological and psychiatric disorders. Front Neurosci. 2024;18:1497417. [doi: [10.3389/fnins.2024.1497417](https://doi.org/10.3389/fnins.2024.1497417)] [Medline: [39411146](https://pubmed.ncbi.nlm.nih.gov/39411146/)]
44. Williams C, Anik FI, Hasan MM, et al. Advancing brain-computer interface closed-loop systems for neurorehabilitation: A systematic review of AI and machine learning innovations in biomedical engineering (preprint). JMIR Biomed Eng. Nov 5, 2025;10:e72218. [doi: [10.2196/72218](https://doi.org/10.2196/72218)] [Medline: [41191851](https://pubmed.ncbi.nlm.nih.gov/41191851/)]
45. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. Oct 18, 2011;155(8):529-536. [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
46. de Swart WK, Loog M, Krijthe JH. A comparative study of methods for dynamic survival analysis. Front Neurol. 2025;16:1504535. [doi: [10.3389/fneur.2025.1504535](https://doi.org/10.3389/fneur.2025.1504535)] [Medline: [40040908](https://pubmed.ncbi.nlm.nih.gov/40040908/)]
47. Kang MK, Hong KS, Yang D, Kim HK. Multi-scale neural networks classification of mild cognitive impairment using functional near-infrared spectroscopy. Biocybern Biomed Eng. Jan 2025;45(1):11-22. [doi: [10.1016/j.bbe.2024.12.001](https://doi.org/10.1016/j.bbe.2024.12.001)]
48. Abir SI, et al. EEG functional connectivity and deep learning for automated diagnosis of Alzheimer's disease and schizophrenia. JCSTS. Jan 26, 2025;7(1):82-99. [doi: [10.32996/jcsts.2025.7.1.7](https://doi.org/10.32996/jcsts.2025.7.1.7)]
49. Sathish R, Muthukumar R, Dhivya K, Karthikkumar S. Deep learning and IoT-enabled framework for accurate classification and monitoring of alzheimer's disease based on eeg signal analysis. 2025 Fifth Int Conf Adv Electr, Comput, Commun Sustainable Technol (ICAECT). 2025:1-8. [doi: [10.1109/ICAECT63952.2025.10958882](https://doi.org/10.1109/ICAECT63952.2025.10958882)]
50. Dubey AK, Kapoor R, Saraswat M. Optimized machine learning for medical data analysis and disease prediction. 2024 Int Conf Artif Intell Emerging Tech (Global AI Summit. 2024:1282-1286. [doi: [10.1109/GlobalAISummit62156.2024.10948001](https://doi.org/10.1109/GlobalAISummit62156.2024.10948001)]
51. K P, Chitla VB, Aftab A, Kamath S. LSTM-based assistance for people with Alzheimer's disease. 2025 Int Conf Intell Innovative Tech Comput, Electr Electron (IITCEE). 2025:1-5. [doi: [10.1109/IITCEE64140.2025.10915385](https://doi.org/10.1109/IITCEE64140.2025.10915385)]
52. Pan J, Fan Z, Smith GE, Guo Y, Bian J, Xu J. Federated learning with multi-cohort real-world data for predicting the progression from mild cognitive impairment to Alzheimer's disease. Alzheimer's Dement. Apr 2025;21(4):e70128. [doi: [10.1002/alz.70128](https://doi.org/10.1002/alz.70128)] [Medline: [40219846](https://pubmed.ncbi.nlm.nih.gov/40219846/)]
53. Zuo Y, Zhang B, Dong Y, et al. Glypred: lysine glycation site prediction via CCU–LightGBM–BiLSTM framework with multi-head attention mechanism. J Chem Inf Model. Aug 26, 2024;64(16):6699-6711. [doi: [10.1021/acs.jcim.4c01034](https://doi.org/10.1021/acs.jcim.4c01034)]
54. Zhu M, Xu Z, Zhang Q, Liu Y, Gu D, Xu SD. GCSTormer: gated swin transformer with channel weights for image denoising. Expert Syst Appl. Jul 2025;284:127924. [doi: [10.1016/j.eswa.2025.127924](https://doi.org/10.1016/j.eswa.2025.127924)]
55. Han X, Xue R, Feng J, et al. Hypergraph foundation model for brain disease diagnosis. IEEE Trans Neural Netw Learning Syst. 2025;36(10):17702-17716. [doi: [10.1109/TNNLS.2025.3554755](https://doi.org/10.1109/TNNLS.2025.3554755)]
56. Lu SY, Zhang YD, Yao YD. A regularized transformer with adaptive token fusion for Alzheimer's disease diagnosis in brain magnetic resonance images. Eng Appl Artif Intell. Sep 2025;155:111058. [doi: [10.1016/j.engappai.2025.111058](https://doi.org/10.1016/j.engappai.2025.111058)]
57. Li X, Zhu W, Qiu P, Dumitrascu OM, Youssef A, Wang Y. A BERT-style self-supervised learning CNN for disease identification from retinal images. arXiv. Preprint posted online on Apr 25, 2025. [doi: [10.48550/arXiv.2504.18049](https://doi.org/10.48550/arXiv.2504.18049)]
58. Mahapatra C. Exploring advanced applications of artificial intelligence in neuropharmacology: a comprehensive overview. Biol Life Sci. Preprint posted online on May 8, 2025. [doi: [10.20944/preprints202505.0369.v1](https://doi.org/10.20944/preprints202505.0369.v1)]
59. Ren H, Zheng Y, Li C, et al. Using machine learning to predict cognitive decline in older adults from the Chinese longitudinal healthy longevity survey: model development and validation study. JMIR Aging. Apr 30, 2025;8:e67437. [doi: [10.2196/67437](https://doi.org/10.2196/67437)] [Medline: [40305830](https://pubmed.ncbi.nlm.nih.gov/40305830/)]
60. Shah YAR, Qureshi SM, Qureshi HA, Shah SUR, Ahmad A, Shiwani A. Advances in artificial intelligence and machine learning for neurodegenerative disease: a literature review. WJRR. Sep 5, 2024;19(3):4-18. [doi: [10.31871/WJRR.19.3.8](https://doi.org/10.31871/WJRR.19.3.8)]

61. Fatima G, Ashiquzzaman A, Kim SS, Kim YR, Kwon HS, Chung E. Vascular and glymphatic dysfunction as drivers of cognitive impairment in Alzheimer's disease: insights from computational approaches. *Neurobiol Dis*. May 2025;208:106877. [doi: [10.1016/j.nbd.2025.106877](https://doi.org/10.1016/j.nbd.2025.106877)] [Medline: [40107629](https://pubmed.ncbi.nlm.nih.gov/40107629/)]
62. Yang X, Dang X, Cai J, Li J, Wang X, Heng P. Temporal-multimodal consistency alignment for Alzheimer's cognitive assessment prediction. *Med Phys Mex Symp Med Phys*. Jun 2025;52(6):5064-5080. [doi: [10.1002/mp.17767](https://doi.org/10.1002/mp.17767)]
63. Sadeghian R, Haider F, Fraser K, Tasaki S, Muniz-Terrera G. Editorial: methods in artificial intelligence for dementia 2024. *Front Dement*. 2024;3:1444825. [doi: [10.3389/frdem.2024.1444825](https://doi.org/10.3389/frdem.2024.1444825)] [Medline: [39081595](https://pubmed.ncbi.nlm.nih.gov/39081595/)]
64. Kale M, Wankhede N, Pawar R, et al. AI-driven innovations in Alzheimer's disease: integrating early diagnosis, personalized treatment, and prognostic modelling. *Ageing Res Rev*. Nov 2024;101:102497. [doi: [10.1016/j.arr.2024.102497](https://doi.org/10.1016/j.arr.2024.102497)]
65. UKbiobank. URL: <https://ukbiobank.ac.uk> [Accessed 2026-02-07]
66. ADNI. URL: <https://adni.loni.usc.edu> [Accessed 2026-02-07]
67. Open access series of imaging studies (OASIS). Washington University in St Louis. URL: <https://sites.wustl.edu/oasisbrains/> [Accessed 2026-02-07]
68. NACC. URL: <https://naccdata.org/> [Accessed 2026-02-07]
69. Framingham Heart Study. URL: <https://www.framinghamheartstudy.org/> [Accessed 2026-02-07]
70. aibl. URL: <https://aibl.csiro.au> [Accessed 2026-02-07]
71. TalkBank. URL: <https://dementia.talkbank.org/> [Accessed 2026-02-07]
72. Gkoumas D, Wang B, Tsakalidis A, et al. A longitudinal multi-modal dataset for dementia monitoring and diagnosis. *Lang Resour Eval*. 2024;58(3):883-902. [doi: [10.1007/s10579-023-09718-4](https://doi.org/10.1007/s10579-023-09718-4)] [Medline: [39323983](https://pubmed.ncbi.nlm.nih.gov/39323983/)]
73. Xu T, Wang X, Lun X, Pan H, Wang Z. ADReFV: face video dataset based on human-computer interaction for Alzheimer's disease recognition. *Comput Animation Virtual*. Jan 2023;34(1):e2127. [doi: [10.1002/cav.2127](https://doi.org/10.1002/cav.2127)]
74. GENCODE. URL: <https://www.genencodegenes.org/> [Accessed 2026-02-07]
75. Velazquez M, Lee Y. Multimodal ensemble model for Alzheimer's disease conversion prediction from early mild cognitive impairment subjects. *Comput Biol Med*. Dec 2022;151(Pt A):106201. [doi: [10.1016/j.complbiomed.2022.106201](https://doi.org/10.1016/j.complbiomed.2022.106201)] [Medline: [36370583](https://pubmed.ncbi.nlm.nih.gov/36370583/)]
76. Zhang M, Cui Q, Lü Y, Yu W, Li W. A multimodal learning machine framework for Alzheimer's disease diagnosis based on neuropsychological and neuroimaging data. *Comput Ind Eng*. Nov 2024;197:110625. [doi: [10.1016/j.cie.2024.110625](https://doi.org/10.1016/j.cie.2024.110625)]
77. Fabietti M, Mahmud M, Lotfi A, et al. 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*. 2023;31:2839-2848. [doi: [10.1109/TNSRE.2023.3288835](https://doi.org/10.1109/TNSRE.2023.3288835)]
78. Seifallahi M, Mehraban AH, Galvin JE, Ghoraani B. 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*. 2022;30:1589-1600. [doi: [10.1109/TNSRE.2022.3181252](https://doi.org/10.1109/TNSRE.2022.3181252)] [Medline: [35675251](https://pubmed.ncbi.nlm.nih.gov/35675251/)]
79. Xue C, Kowshik SS, Lteif D, et al. AI-based differential diagnosis of dementia etiologies on multimodal data. *Nat Med*. Oct 2024;30(10):2977-2989. [doi: [10.1038/s41591-024-03118-z](https://doi.org/10.1038/s41591-024-03118-z)] [Medline: [38965435](https://pubmed.ncbi.nlm.nih.gov/38965435/)]
80. Shi J, Zheng X, Li Y, Zhang Q, Ying S. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE J Biomed Health Inform*. Jan 2018;22(1):173-183. [doi: [10.1109/JBHI.2017.2655720](https://doi.org/10.1109/JBHI.2017.2655720)]
81. Allwright M, Mundell HD, McCorkindale AN, et al. Ranking the risk factors for Alzheimer's disease; findings from the UK Biobank study. *Aging Brain*. 2023;3:100081. [doi: [10.1016/j.nbas.2023.100081](https://doi.org/10.1016/j.nbas.2023.100081)] [Medline: [37384134](https://pubmed.ncbi.nlm.nih.gov/37384134/)]
82. Gu Z, Liu S, Ma H, et al. Estimation of machine learning-based models to predict dementia risk in patients with atherosclerotic cardiovascular diseases: UK Biobank study. *JMIR Aging*. Feb 26, 2025;8:e64148. [doi: [10.2196/64148](https://doi.org/10.2196/64148)] [Medline: [40009844](https://pubmed.ncbi.nlm.nih.gov/40009844/)]
83. You J, Zhang YR, Wang HF, et al. Development of a novel dementia risk prediction model in the general population: a large, longitudinal, population-based machine-learning study. *eClinicalMedicine*. Nov 2022;53:101665. [doi: [10.1016/j.eclinm.2022.101665](https://doi.org/10.1016/j.eclinm.2022.101665)]
84. Calvo N, McFall GP, Ramana S, et al. Associated risk and resilience factors of Alzheimer's disease in women with early bilateral oophorectomy: data from the UK Biobank. *J Alzheimers Dis*. Nov 2024;102(1):119-128. [doi: [10.3233/JAD-240646](https://doi.org/10.3233/JAD-240646)] [Medline: [39497303](https://pubmed.ncbi.nlm.nih.gov/39497303/)]
85. Yi F, Yuan J, Somekh J, et al. Genetically supported targets and drug repurposing for brain aging: a systematic study in the UK Biobank. *Sci Adv*. Mar 14, 2025;11(11):eadr3757. [doi: [10.1126/sciadv.adr3757](https://doi.org/10.1126/sciadv.adr3757)] [Medline: [40073132](https://pubmed.ncbi.nlm.nih.gov/40073132/)]

86. Yousefzadeh N, Tran C, Ramirez-Zamora A, Chen J, Fang R, Thai MT. Neuron-level explainable AI for Alzheimer's disease assessment from fundus images. *Sci Rep*. Apr 2, 2024;14(1):7710. [doi: [10.1038/s41598-024-58121-8](https://doi.org/10.1038/s41598-024-58121-8)] [Medline: [38565579](https://pubmed.ncbi.nlm.nih.gov/38565579/)]
87. Gong W, Bai S, Zheng YQ, Smith SM, Beckmann CF. Supervised phenotype discovery from multimodal brain imaging. *IEEE Trans Med Imaging*. Mar 2023;42(3):834-849. [doi: [10.1109/TMI.2022.3218720](https://doi.org/10.1109/TMI.2022.3218720)] [Medline: [36318559](https://pubmed.ncbi.nlm.nih.gov/36318559/)]
88. Lian C, Liu M, Pan Y, Shen D. Attention-guided hybrid network for dementia diagnosis with structural MR images. *IEEE Trans Cybern*. Apr 2022;52(4):1992-2003. [doi: [10.1109/TCYB.2020.3005859](https://doi.org/10.1109/TCYB.2020.3005859)] [Medline: [32721906](https://pubmed.ncbi.nlm.nih.gov/32721906/)]
89. Lian C, Liu M, Wang L, Shen D. Multi-task weakly-supervised attention network for dementia status estimation with structural MRI. *IEEE Trans Neural Netw Learning Syst*. Aug 2022;33(8):4056-4068. [doi: [10.1109/TNNLS.2021.3055772](https://doi.org/10.1109/TNNLS.2021.3055772)]
90. Li H, Habes M, Wolk DA, Fan Y. Alzheimer's disease neuroimaging initiative and the Australian Imaging Biomarkers and Lifestyle Study of Aging. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's Dementia*. Aug 2019;15(8):1059-1070. [doi: [10.1016/j.jalz.2019.02.007](https://doi.org/10.1016/j.jalz.2019.02.007)] [Medline: [31201098](https://pubmed.ncbi.nlm.nih.gov/31201098/)]
91. Oh K, Yoon JS, Suk HI. Learn-explain-reinforce: counterfactual reasoning and its guidance to reinforce an Alzheimer's disease diagnosis model. *IEEE Trans Pattern Anal Mach Intell*. Apr 2023;45(4):4843-4857. [doi: [10.1109/TPAMI.2022.3197845](https://doi.org/10.1109/TPAMI.2022.3197845)] [Medline: [35947563](https://pubmed.ncbi.nlm.nih.gov/35947563/)]
92. Lian C, Liu M, Zhang J, Shen D. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. *IEEE Trans Pattern Anal Mach Intell*. Apr 2020;42(4):880-893. [doi: [10.1109/TPAMI.2018.2889096](https://doi.org/10.1109/TPAMI.2018.2889096)] [Medline: [30582529](https://pubmed.ncbi.nlm.nih.gov/30582529/)]
93. Avsec Ž, Agarwal V, Visentin D, et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods*. Oct 2021;18(10):1196-1203. [doi: [10.1038/s41592-021-01252-x](https://doi.org/10.1038/s41592-021-01252-x)]
94. Yang L, Wang X, Guo Q, et al. For the Alzheimer's disease neuroimaging initiative. deep learning based multimodal progression modeling for Alzheimer's disease. *Stat Biopharm Res*. Jul 3, 2021;13:337-343. [doi: [10.1080/19466315.2021.1884129](https://doi.org/10.1080/19466315.2021.1884129)]
95. Lee LY, Vaghari D, Burkhart MC, et al. Robust and interpretable AI-guided marker for early dementia prediction in real-world clinical settings. *eClinicalMedicine*. Aug 2024;74:102725. [doi: [10.1016/j.eclinm.2024.102725](https://doi.org/10.1016/j.eclinm.2024.102725)]
96. Zhu W, Sun L, Huang J, Han L, Zhang D. Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI. *IEEE Trans Med Imaging*. Sep 2021;40(9):2354-2366. [doi: [10.1109/TMI.2021.3077079](https://doi.org/10.1109/TMI.2021.3077079)]
97. Zhang M, Cui Q, Lü Y, Li W. A feature-aware multimodal framework with auto-fusion for Alzheimer's disease diagnosis. *Comput Biol Med*. Aug 2024;178:108740. [doi: [10.1016/j.combiomed.2024.108740](https://doi.org/10.1016/j.combiomed.2024.108740)] [Medline: [38901184](https://pubmed.ncbi.nlm.nih.gov/38901184/)]
98. Bi XA, Hu X, Wu H, Wang Y. Multimodal data analysis of Alzheimer's disease based on clustering evolutionary random forest. *IEEE J Biomed Health Inform*. Oct 2020;24(10):2973-2983. [doi: [10.1109/JBHI.2020.2973324](https://doi.org/10.1109/JBHI.2020.2973324)] [Medline: [32071013](https://pubmed.ncbi.nlm.nih.gov/32071013/)]
99. Bi XA, Xing Z, Zhou W, Li L, Xu L. Pathogeny detection for mild cognitive impairment via weighted evolutionary random forest with brain imaging and genetic data. *IEEE J Biomed Health Inform*. Jul 2022;26(7):3068-3079. [doi: [10.1109/JBHI.2022.3151084](https://doi.org/10.1109/JBHI.2022.3151084)] [Medline: [35157601](https://pubmed.ncbi.nlm.nih.gov/35157601/)]
100. Hashmi A, Barukab O. Dementia classification using deep reinforcement learning for early diagnosis. *Appl Sci (Basel)*. Jan 22, 2023;13(3):1464. [doi: [10.3390/app13031464](https://doi.org/10.3390/app13031464)]
101. Wang Y, Gao R, Wei T, et al. Predicting long-term progression of Alzheimer's disease using a multimodal deep learning model incorporating interaction effects. *J Transl Med*. Mar 11, 2024;22(1):265. [doi: [10.1186/s12967-024-05025-w](https://doi.org/10.1186/s12967-024-05025-w)]
102. Hatami M, Yaghmaee F, Ebrahimpour R. Investigating the potential of reinforcement learning and deep learning in improving Alzheimer's disease classification. *Neurocomputing*. Sep 2024;597:128119. [doi: [10.1016/j.neucom.2024.128119](https://doi.org/10.1016/j.neucom.2024.128119)]
103. Tabarestani S, Aghili M, Eslami M, et al. A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *Neuroimage*. Feb 1, 2020;206:116317. [doi: [10.1016/j.neuroimage.2019.116317](https://doi.org/10.1016/j.neuroimage.2019.116317)] [Medline: [31678502](https://pubmed.ncbi.nlm.nih.gov/31678502/)]
104. Burkhart MC, Lee LY, Vaghari D, et al. Unsupervised multimodal modeling of cognitive and brain health trajectories for early dementia prediction. *Sci Rep*. May 10, 2024;14(1):10755. [doi: [10.1038/s41598-024-60914-w](https://doi.org/10.1038/s41598-024-60914-w)] [Medline: [38729989](https://pubmed.ncbi.nlm.nih.gov/38729989/)]
105. El-Sappagh S, Alonso JM, Islam SMR, Sultan AM, Kwak KS. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Sci Rep*. Jan 29, 2021;11(1):2660. [doi: [10.1038/s41598-021-82098-3](https://doi.org/10.1038/s41598-021-82098-3)] [Medline: [33514817](https://pubmed.ncbi.nlm.nih.gov/33514817/)]
106. Lee MW, Kim HW, Choe YS, et al. A multimodal machine learning model for predicting dementia conversion in Alzheimer's disease. *Sci Rep*. May 29, 2024;14(1):12276. [doi: [10.1038/s41598-024-60134-2](https://doi.org/10.1038/s41598-024-60134-2)]

107. Yuan S, Li H, Wu J, Sun X. Classification of mild cognitive impairment with multimodal data using both labeled and unlabeled samples. *IEEE/ACM Trans Comput Biol and Bioinf*. Nov 1, 2021;18(6):2281-2290. [doi: [10.1109/TCBB.2021.3053061](https://doi.org/10.1109/TCBB.2021.3053061)]
108. Cirincione A, Lynch K, Bennett J, et al. Prediction of future dementia among patients with mild cognitive impairment (MCI) by integrating multimodal clinical data. *Heliyon*. Sep 15, 2024;10(17):e36728. [doi: [10.1016/j.heliyon.2024.e36728](https://doi.org/10.1016/j.heliyon.2024.e36728)] [Medline: [39281465](https://pubmed.ncbi.nlm.nih.gov/39281465/)]
109. Cassani R, Falk TH. Alzheimer's disease diagnosis and severity level detection based on electroencephalography modulation spectral "patch" features. *IEEE J Biomed Health Inform*. Jul 2020;24(7):1982-1993. [doi: [10.1109/JBHI.2019.2953475](https://doi.org/10.1109/JBHI.2019.2953475)] [Medline: [31725401](https://pubmed.ncbi.nlm.nih.gov/31725401/)]
110. Cilia ND, D'Alessandro T, De Stefano C, Fontanella F, Molinara M. From online handwriting to synthetic images for Alzheimer's disease detection using a deep transfer learning approach. *IEEE J Biomed Health Inform*. Dec 2021;25(12):4243-4254. [doi: [10.1109/JBHI.2021.3101982](https://doi.org/10.1109/JBHI.2021.3101982)]
111. Kmetzsch V, Becker E, Saracino D, et al. Disease progression score estimation from multimodal imaging and MicroRNA data using supervised variational autoencoders. *IEEE J Biomed Health Inform*. Dec 2022;26(12):6024-6035. [doi: [10.1109/JBHI.2022.3208517](https://doi.org/10.1109/JBHI.2022.3208517)]
112. Mengoudi K, Ravi D, Yong KXX, et al. Augmenting dementia cognitive assessment with instruction-less eye-tracking tests. *IEEE J Biomed Health Inform*. Nov 2020;24(11):3066-3075. [doi: [10.1109/JBHI.2020.3004686](https://doi.org/10.1109/JBHI.2020.3004686)] [Medline: [32749977](https://pubmed.ncbi.nlm.nih.gov/32749977/)]
113. Tsai H, Yang TW, Ou KH, Su TH, Lin C, Chou CF. Multimodal attention network for dementia prediction. *IEEE J Biomed Health Inform*. Nov 2024;28(11):6918-6930. [doi: [10.1109/JBHI.2024.3438885](https://doi.org/10.1109/JBHI.2024.3438885)]
114. Wu EQ, Peng XY, Chen SD, Zhao XY, Tang ZR. Detecting Alzheimer's dementia degree. *IEEE Trans Cogn Dev Syst*. Mar 2022;14(1):116-125. [doi: [10.1109/TCDS.2020.3015131](https://doi.org/10.1109/TCDS.2020.3015131)]
115. Zhang H, Ni M, Yang Y, et al. Patch-based interpretable deep learning framework for Alzheimer's disease diagnosis using multimodal data. *Biomed Signal Process Control*. Feb 2025;100:107085. [doi: [10.1016/j.bspc.2024.107085](https://doi.org/10.1016/j.bspc.2024.107085)]
116. Fan CC, Yang H, Zhang C, et al. Graph reasoning module for Alzheimer's disease diagnosis: a plug-and-play method. *IEEE Trans Neural Syst Rehabil Eng*. 2023;31:4773-4780. [doi: [10.1109/TNSRE.2023.3337533](https://doi.org/10.1109/TNSRE.2023.3337533)]
117. Beebe-Wang N, Okeson A, Althoff T, Lee SI. Efficient and explainable risk assessments for imminent dementia in an aging cohort study. *IEEE J Biomed Health Inform*. Jul 2021;25(7):2409-2420. [doi: [10.1109/JBHI.2021.3059563](https://doi.org/10.1109/JBHI.2021.3059563)]
118. Battineni G, Hossain MA, Chintalapudi N, et al. Improved Alzheimer's disease detection by MRI using multimodal machine learning algorithms. *Diagnostics (Basel)*. Nov 13, 2021;11(11):2103. [doi: [10.3390/diagnostics11112103](https://doi.org/10.3390/diagnostics11112103)]
119. Nguyen H, Chu NN. An introduction to deep learning research for Alzheimer's disease. *IEEE Consumer Electron Mag*. May 1, 2021;10(3):72-75. [doi: [10.1109/MCE.2020.3048254](https://doi.org/10.1109/MCE.2020.3048254)]
120. Fan F, Song H, Jiang J, et al. Development and validation of a multimodal deep learning framework for vascular cognitive impairment diagnosis. *iScience*. Oct 2024;27(10):110945. [doi: [10.1016/j.isci.2024.110945](https://doi.org/10.1016/j.isci.2024.110945)]
121. Ilias L, Askounis D, Psarras J. Detecting dementia from speech and transcripts using transformers. *Comput Speech Lang*. Apr 2023;79:101485. [doi: [10.1016/j.csl.2023.101485](https://doi.org/10.1016/j.csl.2023.101485)]
122. Poor FF, Dodge HH, Mahoor MH. A multimodal cross-transformer-based model to predict mild cognitive impairment using speech, language and vision. *Comput Biol Med*. Nov 2024;182:109199. [doi: [10.1016/j.compbimed.2024.109199](https://doi.org/10.1016/j.compbimed.2024.109199)] [Medline: [39332117](https://pubmed.ncbi.nlm.nih.gov/39332117/)]
123. Lin K, Washington PY. Multimodal deep learning for dementia classification using text and audio. *Sci Rep*. Jun 16, 2024;14(1):13887. [doi: [10.1038/s41598-024-64438-1](https://doi.org/10.1038/s41598-024-64438-1)]
124. Ortiz-Perez D, Ruiz-Ponce P, Tomás D, Garcia-Rodriguez J, Vizcaya-Moreno MF, Leo M. A deep learning-based multimodal architecture to predict signs of dementia. *Neurocomputing*. Sep 2023;548:126413. [doi: [10.1016/j.neucom.2023.126413](https://doi.org/10.1016/j.neucom.2023.126413)]
125. Ilias L, Askounis D. Explainable identification of dementia from transcripts using transformer networks. *IEEE J Biomed Health Inform*. Aug 2022;26(8):4153-4164. [doi: [10.1109/JBHI.2022.3172479](https://doi.org/10.1109/JBHI.2022.3172479)] [Medline: [35511841](https://pubmed.ncbi.nlm.nih.gov/35511841/)]
126. Wen B, Wang N, Subbalakshmi K, Chandramouli R. Revealing the roles of part-of-speech taggers in Alzheimer disease detection: scientific discovery using one-intervention causal explanation. *JMIR Form Res*. May 2, 2023;7:e36590. [doi: [10.2196/36590](https://doi.org/10.2196/36590)] [Medline: [37129944](https://pubmed.ncbi.nlm.nih.gov/37129944/)]
127. Chen X, Pu Y, Li J, Zhang WQ. Cross-lingual Alzheimer's disease detection based on paralinguistic and pre-trained features. *ICASSP 2023 - 2023 IEEE Int Conf Acoustics, Speech Signal Proc (ICASSP)*. 2023:1-2. [doi: [10.1109/ICASSP49357.2023.10095522](https://doi.org/10.1109/ICASSP49357.2023.10095522)]
128. Zheng C, Bouazizi M, Ohtsuki T. An evaluation on information composition in dementia detection based on speech. *IEEE Access*. 2022;10:92294-92306. [doi: [10.1109/ACCESS.2022.3203068](https://doi.org/10.1109/ACCESS.2022.3203068)]
129. Nambiar AS, Likhita K, Pujya K, Gupta D, Vekkot S, Lalitha S. Comparative study of deep classifiers for early dementia detection using speech transcripts. *2022 IEEE 19th India Counc Int Conf (INDICON)*. 2022:1-6. [doi: [10.1109/INDICON56171.2022.10039705](https://doi.org/10.1109/INDICON56171.2022.10039705)]

130. Priyadarshinee P, Clarke CJ, Melechovsky J, Lin CMY, B. T. B, Chen JM. Alzheimer's dementia speech (audio vs. text): multi-modal machine learning at high vs. low resolution. *Appl Sci (Basel)*. 2023;13(7):4244. [doi: [10.3390/app13074244](https://doi.org/10.3390/app13074244)]
131. Liu J, Fu F, Li L, et al. Efficient pause extraction and encode strategy for Alzheimer's disease detection using only acoustic features from spontaneous speech. *Brain Sci*. Mar 11, 2023;13(3):477. [doi: [10.3390/brainsci13030477](https://doi.org/10.3390/brainsci13030477)] [Medline: [36979287](https://pubmed.ncbi.nlm.nih.gov/36979287/)]
132. Mahajan P, Baths V. Acoustic and language based deep learning approaches for Alzheimer's dementia detection from spontaneous speech. *Front Aging Neurosci*. 2021;13:623607. [doi: [10.3389/fnagi.2021.623607](https://doi.org/10.3389/fnagi.2021.623607)] [Medline: [33613269](https://pubmed.ncbi.nlm.nih.gov/33613269/)]
133. Mei K, Ding X, Liu Y, et al. The USTC system for ADReSS-M challenge. *ICASSP 2023 - 2023 IEEE Int Conf Acoustics, Speech Signal Proc (ICASSP)*. 2023:1-2. [doi: [10.1109/ICASSP49357.2023.10094714](https://doi.org/10.1109/ICASSP49357.2023.10094714)]
134. Ali Meerza SI, Li Z, Liu L, Zhang J, Liu J. Fair and privacy-preserving Alzheimer's disease diagnosis based on spontaneous speech analysis via federated learning. 2022 44th Ann Int Conf IEEE Eng Med Biol Soc (EMBC). 2022:1362-1365. [doi: [10.1109/EMBC48229.2022.9871204](https://doi.org/10.1109/EMBC48229.2022.9871204)]
135. Chen W, Xing X, Xu X, Pang J, Du L. SpeechFormer++: a hierarchical efficient framework for paralinguistic speech processing. *IEEE/ACM Trans Audio Speech Lang Process*. 2023;31:775-788. [doi: [10.1109/TASLP.2023.3235194](https://doi.org/10.1109/TASLP.2023.3235194)]
136. Tamm B, Vandenberghe R, Van Hamme H. Cross-lingual transfer learning for alzheimer's detection from spontaneous speech. *ICASSP 2023 - 2023 IEEE Int Conf Acoust, Speech Signal Process (ICASSP)*. 2023:1-2. [doi: [10.1109/ICASSP49357.2023.10096770](https://doi.org/10.1109/ICASSP49357.2023.10096770)]
137. Wozczyk D, Hedlikova A, Akman A, Demetriou S, Schuller B. Data augmentation for dementia detection in spoken language. *Proc Interspeech 2022*. 2022:2858-2862. [doi: [10.21437/Interspeech.2022-10210](https://doi.org/10.21437/Interspeech.2022-10210)]
138. Jin L, Oh Y, Kim H, et al. CONSEN: complementary and simultaneous ensemble for Alzheimer's disease detection and MMSE score prediction. *ICASSP 2023 - 2023 IEEE Int Conf Acoustics, Speech Signal Proc (ICASSP)*. 2023:1-2. [doi: [10.1109/ICASSP49357.2023.10096253](https://doi.org/10.1109/ICASSP49357.2023.10096253)]
139. Ilias L, Askounis D. Context-aware attention layers coupled with optimal transport domain adaptation and multimodal fusion methods for recognizing dementia from spontaneous speech. *Knowl Based Syst*. Oct 2023;277:110834. [doi: [10.1016/j.knsys.2023.110834](https://doi.org/10.1016/j.knsys.2023.110834)]
140. Azevedo T, Bethlehem RAI, Whiteside DJ, et al. Identifying healthy individuals with Alzheimer's disease neuroimaging phenotypes in the UK Biobank. *Commun Med*. Jul 20, 2023;3(1):100. [doi: [10.1038/s43856-023-00313-w](https://doi.org/10.1038/s43856-023-00313-w)]
141. Liang S, Chen T, Ma J, Ren S, Lu X, Du W. Identification of mild cognitive impairment using multimodal 3D imaging data and graph convolutional networks. *Phys Med Biol*. Dec 7, 2024;69(23):235002. [doi: [10.1088/1361-6560/ad8c94](https://doi.org/10.1088/1361-6560/ad8c94)]
142. Jahan S, Abu Taher K, Kaiser MS, et al. Explainable AI-based Alzheimer's prediction and management using multimodal data. *PLOS ONE*. 2023;18(11):e0294253. [doi: [10.1371/journal.pone.0294253](https://doi.org/10.1371/journal.pone.0294253)] [Medline: [37972072](https://pubmed.ncbi.nlm.nih.gov/37972072/)]
143. Jahan S, Saif Adib MdR, Huda SM, et al. Federated explainable AI-based Alzheimer's disease prediction with multimodal data. *IEEE Access*. 2025;13:43435-43454. [doi: [10.1109/ACCESS.2025.3547343](https://doi.org/10.1109/ACCESS.2025.3547343)]
144. Myrzashova R, Alsamhi SH, Shvetsov AV, Hawbani A, Guizani M, Wei X. BCFTL: blockchain-enabled multimodal federated transfer learning for decentralized Alzheimer's diagnosis. *IEEE Internet Things J*. 2025;12(15):29656-29669. [doi: [10.1109/JIOT.2025.3569652](https://doi.org/10.1109/JIOT.2025.3569652)]
145. Chen K, Weng Y, Huang Y, et al. A multi-view learning approach with diffusion model to synthesize FDG PET from MRI T1WI for diagnosis of Alzheimer's disease. *Alzheimers Dement*. Feb 2025;21(2):e14421. [doi: [10.1002/alz.14421](https://doi.org/10.1002/alz.14421)]
146. Lin W, Lin W, Chen G, et al. Bidirectional mapping of brain MRI and PET with 3D Reversible GAN for the diagnosis of Alzheimer's disease. *Front Neurosci*. 2021;15:646013. [doi: [10.3389/fnins.2021.646013](https://doi.org/10.3389/fnins.2021.646013)] [Medline: [33935634](https://pubmed.ncbi.nlm.nih.gov/33935634/)]
147. Gupta B, Jegannathan GK, Alam MS, et al. Multimodal lightweight neural network for Alzheimer's disease diagnosis integrating neuroimaging and cognitive scores. *Neurosci Inf*. Sep 2025;5(3):100218. [doi: [10.1016/j.neuri.2025.100218](https://doi.org/10.1016/j.neuri.2025.100218)]
148. Chen Z, Wang Z, Zhao M, et al. A new classification network for diagnosing Alzheimer's disease in class-imbalance MRI datasets. *Front Neurosci*. Aug 25, 2022;16:807085. [doi: [10.3389/fnins.2022.807085](https://doi.org/10.3389/fnins.2022.807085)]
149. Sarma M, Chatterjee D. Multistage diagnosis of Alzheimer's disease from clinical data using 'deep ensemble learning'. *JAI*. 2024;01(1):122-138. [doi: [10.54364/JAI.2024.1109](https://doi.org/10.54364/JAI.2024.1109)]
150. Mujahid M, Rehman A, Alam T, Alamri FS, Fati SM, Saba T. An efficient ensemble approach for Alzheimer's disease detection using an adaptive synthetic technique and deep learning. *Diagnostics (Basel)*. Jul 26, 2023;13(15):2489. [doi: [10.3390/diagnostics13152489](https://doi.org/10.3390/diagnostics13152489)] [Medline: [37568852](https://pubmed.ncbi.nlm.nih.gov/37568852/)]
151. Dubey Y, Bhongade A, Palsodkar P, Fulzele P. Efficient explainable models for Alzheimer's disease classification with feature selection and data balancing approach using ensemble learning. *Diagnostics (Basel)*. Dec 10, 2024;14(24):2770. [doi: [10.3390/diagnostics14242770](https://doi.org/10.3390/diagnostics14242770)] [Medline: [39767131](https://pubmed.ncbi.nlm.nih.gov/39767131/)]
152. Mandawkar U, Diwan T. Hybrid cuttle Fish-Grey wolf optimization tuned weighted ensemble classifier for Alzheimer's disease classification. *Biomed Signal Process Control*. Jun 2024;92:106101. [doi: [10.1016/j.bspc.2024.106101](https://doi.org/10.1016/j.bspc.2024.106101)]
153. Jasodanand VH, Kowshik SS, Puducheri S, et al. AI-driven fusion of multimodal data for Alzheimer's disease biomarker assessment. *Nat Commun*. Aug 11, 2025;16(1):7407. [doi: [10.1038/s41467-025-62590-4](https://doi.org/10.1038/s41467-025-62590-4)] [Medline: [40789853](https://pubmed.ncbi.nlm.nih.gov/40789853/)]

154. Weiner MW, Kanoria S, Miller MJ, et al. Overview of Alzheimer's disease neuroimaging Initiative and future clinical trials. *Alzheimer's Dement*. Jan 2025;21(1):e14321. [doi: [10.1002/alz.14321](https://doi.org/10.1002/alz.14321)] [Medline: [39711072](https://pubmed.ncbi.nlm.nih.gov/39711072/)]
155. Wilkinson T, Schnier C, Bush K, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. *Eur J Epidemiol*. Jun 2019;34(6):557-565. [doi: [10.1007/s10654-019-00499-1](https://doi.org/10.1007/s10654-019-00499-1)]
156. Thulasimani V, Shanmugavadivel K, Cho J, Easwaramoorthy SV. A review of datasets, optimization strategies, and learning algorithms for analyzing Alzheimer's dementia detection. *Neuropsychiatr Dis Treat*. 2024;20:2203-2225. [doi: [10.2147/NDT.S496307](https://doi.org/10.2147/NDT.S496307)] [Medline: [39588176](https://pubmed.ncbi.nlm.nih.gov/39588176/)]
157. Chan KCG, Xia F, Kukull WA. NACC data: who is represented over time and across centers, and implications for generalizability. *Alzheimer's Dement*. Sep 2025;21(9):e70657. [doi: [10.1002/alz.70657](https://doi.org/10.1002/alz.70657)] [Medline: [40968249](https://pubmed.ncbi.nlm.nih.gov/40968249/)]
158. Fowler C, Rainey-Smith SR, Bird S, et al. 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*. Mar 11, 2021;5(1):443-468. [doi: [10.3233/ADR-210005](https://doi.org/10.3233/ADR-210005)]
159. Yang Q, Li X, Ding X, Xu F, Ling Z. Deep learning-based speech analysis for Alzheimer's disease detection: a literature review. *Alz Res Therapy*. Dec 14, 2022;14(1):186. [doi: [10.1186/s13195-022-01131-3](https://doi.org/10.1186/s13195-022-01131-3)]
160. He Y, Wang Z, Zhang Y, et al. NeuroSymAD: a neuro-symbolic framework for interpretable Alzheimer's disease diagnosis. Preprint posted online on Mar 1, 2025. [doi: [10.48550/arXiv.2503.00510](https://doi.org/10.48550/arXiv.2503.00510)]
161. Sadeghi A, Hajati F, Argha A, Lovell NH, Yang M. Interpretable graph-based models on multimodal biomedical data integration: a technical review and benchmarking. arXiv. Preprint posted online on May 3, 2025. [doi: [10.48550/arXiv.2505.01696](https://doi.org/10.48550/arXiv.2505.01696)]
162. Mahamud E, Assaduzzaman M, Islam J, Fahad N, Hossen MJ, Ramanathan TT. Enhancing Alzheimer's disease detection: an explainable machine learning approach with ensemble techniques. *Intell-Based Med*. 2025;11:100240. [doi: [10.1016/j.ibmed.2025.100240](https://doi.org/10.1016/j.ibmed.2025.100240)]
163. Zhou T, Liu M, Thung KH, Shen D. Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. *IEEE Trans Med Imaging*. Oct 2019;38(10):2411-2422. [doi: [10.1109/TMI.2019.2913158](https://doi.org/10.1109/TMI.2019.2913158)]
164. Sharma R, Sibille L, Fahmi R. Multi-branch convolutional neural network for Alzheimer's disease versus normal control classification using PET images. *Alzheimer's Dementia*. Jun 2023;19(S3):e061092. [doi: [10.1002/alz.061092](https://doi.org/10.1002/alz.061092)]
165. Zhang J, Yu X, Chen T, et al. BrainNet-moe: brain-inspired mixture-of-experts learning for neurological disease identification. Preprint posted online on Mar 5, 2025. [doi: [10.48550/ARXIV.2503.07640](https://doi.org/10.48550/ARXIV.2503.07640)]

## Abbreviations

- AD:** Alzheimer disease
- AdaBoost:** Adaptive Boosting
- ADNI:** Alzheimer's Disease Neuroimaging Initiative
- ADReSS:** Alzheimer's Dementia Recognition Through Spontaneous Speech
- ADReSSo:** Alzheimer's Dementia Recognition Through Spontaneous Speech 2021 Challenge
- AI:** artificial intelligence
- ALBERT:** A Lite Bidirectional Encoder Representations From Transformers
- AUC:** area under the curve
- BERT:** Bidirectional Encoder Representations From Transformers
- BiLSTM:** bidirectional long short-term memory
- DeiT:** Data-Efficient Image Transformers
- DNN:** DNN
- EEG:** electroencephalography
- FL:** federated learning
- HIPAA :** Health Insurance Portability and Accountability Act
- LightGBM:** Light Gradient-Boosting Machine
- LIME:** Local Interpretable Model-Agnostic Explanations
- MCI:** mild cognitive impairment
- ML:** machine learning
- MRI:** magnetic resonance imaging
- NACC:** National Alzheimer's Coordinating Centre
- OASIS:** Open Access Series of Imaging Studies
- PET:** positron emission tomography
- PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- PRISMA-S:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
- QUADAS-2:** Revised Quality Assessment of Diagnostic Accuracy Studies tool
- RL:** reinforcement learning

**RoBERTa+**: Robustly Optimized Bidirectional Encoder Representations From Transformers Approach

**SHAP**: Shapley Additive Explanations

**ViT**: vision transformer

**XGBoost**: Extreme Gradient Boosting

*Edited by Stefano Brini; peer-reviewed by Farah Elkourdi, Mohammad Mamun Sikder, Oladayo Oyetunji, Shan Jiang; submitted 07.Oct.2025; final revised version received 09.Jan.2026; accepted 09.Jan.2026; published 25.Mar.2026*

*Please cite as:*

*Yu Z, Mulholland A, Huang T, Liu Q*

*Multimodal AI for Alzheimer Disease Diagnosis: Systematic Review of Datasets, Models, and Modalities*

*J Med Internet Res 2026;28:e85414*

*URL: <https://www.jmir.org/2026/1/e85414>*

*doi: [10.2196/85414](https://doi.org/10.2196/85414)*

© Ziwen Yu, Anthony Mulholland, Tianyan Huang, Qiang Liu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 25.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.