

Multimodal Biomarker AI for Alzheimer's Diagnosis and Prognosis: A Systematic Review

2026-A-41
10-AAIC

Feliciana Catino^{1,4}, Yalin Zheng², Rodolfo Sardone, PhD, MPH³.

¹Urban Health Center - Local Public Health Authority of Taranto, Taranto, Italy;
²University of Liverpool, Liverpool, UK;
³Research and Innovation, Local Public Health Authority of Taranto, Taranto, Italy;
⁴University of Bari - Department of Computer Science, Italy.

Across 27 studies, multimodal biomarker AI consistently outperformed the best unimodal models for Alzheimer's/MCI classification (diagnostic AUC often ~0.85–0.95) and, more modestly, for MCI-to-AD prognosis (~0.75–0.85); limited external validation and interpretability still constrain clinical translation.

INTRODUCTION

Background

Early diagnosis of Alzheimer's disease (AD) and related dementias remains challenging: no single biomarker captures the complex, multifactorial pathology.

Multimodal artificial intelligence that integrates heterogeneous data — neuroimaging, fluid biomarkers, genetics and cognition — is a promising route to earlier, more accurate detection and risk stratification.

Aim

A PRISMA-guided systematic review (PROSPERO CRD420251049848) of the performance and feasibility of machine-learning models combining ≥2 biomarker modalities for AD/MCI diagnosis and prognosis.

METHODS

Search

PubMed, Scopus, IEEE Xplore and Web of Science Core Collection (Jan 2010 – Sep 2025), plus conference proceedings and preprints.

Eligibility

Peer-reviewed studies applying AI/ML to ≥2 biomarker modalities for AD/MCI diagnosis or prognosis (e.g., MCI-to-AD conversion), with ≥1 minimally invasive or widely deployable modality (cognition, blood biomarkers, APOE/genetics, retinal imaging, routine clinical data).

Synthesis & quality

Narrative synthesis of performance and data-fusion strategy; risk of bias assessed with QUADAS-2. 27 studies met the criteria.

Identification

Screening

Included

Records identified from databases (Jan 2010 - Sep 2025) N=3381: N=872 (PubMed) N=1,096 (Scopus) N=240 (IEEE Xplore) N=458 (Web) N=534 (Google Scholar) N=127 (NeurIPS, ICML, MICCAI) N=54 (arXiv, bioRxiv)

Records removed before screening: • Duplicate records removed (N=1,046) • Records marked as ineligible by automation tools (non-human, not English) (N=230) • Records removed for other reasons (N=155)

Records screened (N=1,950)

Records excluded (N=1,777)

Full-text assessed for eligibility (N=173)

Full-text excluded with reasons (N=146): • Unimodal (did not meet ≥2 modalities) (N=67) • Wrong population/outcome (not AD/MCI, non-diagnostic/prognostic) (N=28) • AI/ML not applicable/no performance metrics (N=23) • Duplicate cohort/no original analysis (N=14) • Case report/insufficient sample (N=14)

Studies included in review (N=27)

RESULTS

In nearly all studies, multimodal models beat the best unimodal baseline. Diagnostic classification often reached AUC ~0.85–0.95 (internal validation); MCI-to-AD prognosis was harder (~0.75–0.85). ~67% used ≥1 deep neural network and ~74% used early (feature-level) fusion.

Modality

MRI

PET

CSF

Plasma

Cognition

APOE

Retina

EHR

19

8

3

9

9

8

3

4

Fig 3. Biomarker modalities across the 27 studies — MRI most frequent, often paired with cognition, plasma/CSF or PET.

Multimodal vs best unimodal AUC — selected studies

Study	Modalities	AUC (multi vs uni)
Ravichandran 2025	Retina + plasma	0.97 vs 0.88
Zhao 2022	MRI + plasma + genetics	0.99 vs 0.96
Zhang 2011	MRI + PET + CSF	0.93 vs 0.85
Qiu 2022	MRI + clinical	~0.95 vs 0.84
Lee 2019	MRI + CSF + cognition	0.86 vs 0.83

A novel retinal-imaging + plasma pairing reached AUC ~0.97 — highlighting scalable, minimally invasive modalities. But over half of studies relied on ADNI, with infrequent external validation.

CONCLUSIONS

Take-home

Multimodal biomarker AI consistently outperforms unimodal approaches for AD/MCI classification and risk prediction — strongest for diagnosis, more modest for MCI-to-AD prognosis.

Emerging minimally invasive pairings (retinal imaging + plasma) are especially promising for scalable early detection.

Limitations & next steps

Translation is constrained by dataset homogeneity (over half used ADNI), infrequent external validation, missing data, the cost of modalities like PET, and limited interpretability; QUADAS-2 flagged frequent Index-Test concerns. Priorities: prospective multi-centre studies, robust external validation, subgroup fairness and transparent, interpretable reporting.

Number of Studies

D1

D2

D3

D4

27

27

27

27

Risk Level

Low

Unclear

High

Fig 2b. QUADAS-2 risk of bias by domain (D1–D4); the Index Test domain (D2) had the most high-risk ratings.

REFERENCES

1. Catino F, et al. Multimodal biomarker AI techniques for early neurocognitive disorder diagnosis: a systematic review. *Artif Intell Med* 2026;177:103389.

2. Page MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.

3. Whiting PF, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–536.

4. Livingston G, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet* 2024;404:572–628.

CONTACT

Feliciana Catino

Urban Health Center, ASL Taranto, Italy

Department of Computer Science, University of Bari, Italy

f.catino5@phd.uniba.it

Rodolfo Sardone, PhD, MPH

Research & Innovation, ASL Taranto

rsardone@liverpool.ac.uk

ALZHEIMER'S ASSOCIATION

AAIC

26

Fig 1. PRISMA selection: 3,381 records screened → 27 studies included.