Dement Neurocogn Disord. 2026 Jul;25(3):141-151
https://doi.org/10.12779/dnd.2026.25.3.141
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Review Article

Check for updates

Diagnostic Accuracy of MRI-Based Artificial Intelligence Models for Distinguishing Alzheimer's Disease and Mild Cognitive Impairment From Cognitively Normal Controls: A Systematic Review and Meta-Analysis

Sara Mahmoud Ibrahim Allaham

Department of Neurology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

OPEN ACCESS

Received: Feb 21, 2026
Revised: Apr 24, 2026
Accepted: Apr 30, 2026
Published online: Jun 4, 2026

Correspondence to
Sara Mahmoud Ibrahim Allaham
Department of Neurology, The First Affiliated Hospital, Zhejiang University School of Medicine, 79 Qingchun Road, Shangcheng District, Hangzhou 310006, China.
Email: saraallaham01@gmail.com

ORCID iDs
Sara Mahmoud Ibrahim Allaham
https://orcid.org/0009-0006-7073-1177

Conflict of Interest
The author has no financial conflicts of interest.

ABSTRACT

Magnetic resonance imaging (MRI)-based artificial intelligence (AI) models are increasingly applied to brain MRI for diagnosing Alzheimer's disease (AD) and mild cognitive impairment (MCI), but their overall diagnostic performance remains unclear. We systematically searched PubMed/MEDLINE, Embase, Web of Science, Scopus, and IEEE Xplore until February 15, 2026, for diagnostic accuracy studies of machine-learning or deep-learning models using structural brain MRI to distinguish AD vs. cognitively normal (CN) controls and MCI or late mild cognitive impairment (LMCI) vs. CN controls. Seven studies met inclusion criteria, contributing five AD vs. CN and two MCI/LMCI vs. CN tasks, predominantly using deep-learning architectures applied to Alzheimer's Disease Neuroimaging Initiative cohorts. For AD vs. CN (five studies), pooled sensitivity was 0.96 (95% confidence interval [CI], 0.93–0.97) and pooled specificity was 0.95 (95% CI, 0.92–0.97), indicating excellent discrimination. For MCI/LMCI vs. CN (two studies), sensitivity was consistently high (0.91–0.93), whereas specificity varied widely (0.54–0.98), limiting the interpretability of pooled estimates. MRI-based AI models therefore show strong performance for established AD but heterogeneous specificity for MCI, underscoring the need for larger, externally validated studies in diverse populations.

Keywords: Alzheimer Disease; Mild Cognitive Impairment; Dementia; Magnetic Resonance Imaging; Artificial Intelligence; Deep Learning

INTRODUCTION

Alzheimer's disease (AD) is the leading cause of dementia worldwide and a major contributor to morbidity, institutionalization, and health-care costs in aging populations.[1] Early and accurate diagnosis is essential to initiate symptomatic treatments, plan care, and enroll

appropriate candidates into disease-modifying therapy trials, particularly among individuals with mild cognitive impairment (MCI) who are at increased risk of progression to dementia.[2] Conventional structural MRI is routinely used to exclude alternative causes of cognitive decline and to support the diagnosis of AD by demonstrating characteristic patterns of medial temporal and parietal atrophy, but visual assessment and traditional region-of-interest measurements show variable sensitivity and specificity, especially in prodromal or atypical presentations.[3]

Over the past decade, artificial intelligence (AI) techniques, including traditional machine-learning algorithms and deep-learning architectures such as convolutional neural networks and vision transformers, have been increasingly applied to brain MRI for automated detection and staging of AD and MCI.[4-6] Studies using large datasets such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) have reported high classification accuracies, frequently exceeding 85%–90% for differentiating AD from cognitively normal (CN) controls, and promising results for distinguishing MCI or predicting conversion to dementia.[7-10] At the same time, many AI models are developed in single-center or research cohorts, rely on heterogeneous preprocessing pipelines, and may be prone to overfitting, limited external validation, and reduced generalizability to real-world clinical settings.[4,8-10]

Several narrative and systematic reviews have summarized the expanding literature on AI in AD, highlighting the potential of MRI-based models but also emphasizing important methodological limitations, including small sample sizes, variable reference standards, and inconsistent reporting of diagnostic accuracy metrics and uncertainty.[3,5] A recent meta-analysis focusing on vision transformer architectures and primarily on AD vs. healthy controls provided useful insights into a specific model family but did not encompass the broader range of convolutional, deformation-based, and multimodal MRI-based approaches, nor did it systematically address different disease stages and classification tasks.[6] As a result, clinicians and researchers still lack an up-to-date, quantitative synthesis of how well MRI-based AI performs for key diagnostic questions spanning AD, MCI, and CN individuals in clinically defined populations.

Accordingly, this study aimed to systematically review and quantitatively synthesize the diagnostic performance of MRI-based AI models for distinguishing AD vs. CN and MCI or late MCI vs. CN, using clinical diagnosis as the reference standard based on established criteria.[1-3] By pooling sensitivity and specificity across eligible diagnostic-accuracy studies and qualitatively exploring sources of heterogeneity such as model type, dataset, and classification task, this meta-analysis sought to provide a rigorous summary of current performance and to identify priorities for future development and clinical translation of AI tools in dementia imaging.

METHODS

Eligibility criteria

This systematic review and meta-analysis included primary diagnostic accuracy studies evaluating AI models applied to brain MRI for classifying AD, MCI, and CN individuals. Eligible populations were adults undergoing structural brain MRI with a clinical reference standard based on established diagnostic criteria for AD or MCI (for example, NINCDS-ADRDA, NIA-AA, or DSM-based criteria).[1-3] Index tests were machine-learning or deep-

https://doi.org/10.12779/dnd.2026.25.3.141

learning models that used at least one structural MRI sequence (typically 3D T1-weighted images) as input to perform individual-level classification of AD vs. controls, MCI vs. controls, or late mild cognitive impairment (LMCI) vs. controls. Studies relying exclusively on non-MRI modalities were excluded. Retrospective or prospective human studies were eligible if they reported sufficient data to derive or directly extract diagnostic performance metrics (sensitivity, specificity, accuracy, or area under the receiver-operating-characteristic curve) for at least one binary classification task of interest. Case reports, reviews, editorials, non-human studies, and purely methodological papers without clinical labels or evaluable diagnostic accuracy were excluded.

Information sources and search strategy

Electronic searches were conducted in PubMed/MEDLINE, Embase, Web of Science, Scopus, and IEEE Xplore from inception to February 15, 2026, to identify MRI-based AI models for AD and MCI. Search strategies combined controlled vocabulary and free-text terms related to AD and MCI ("Alzheimer disease," "mild cognitive impairment"), neuroimaging ("magnetic resonance imaging," "MRI"), and AI ("machine learning," "deep learning," "convolutional neural network," "artificial intelligence," "computer-aided diagnosis"). The complete database-specific search strategies are provided in **Supplementary Table 1**. Reference lists of relevant systematic reviews and key primary studies were screened manually to identify additional eligible articles not captured by the database search. No language restrictions were applied at the search stage. Conference abstracts without full text and non-peer-reviewed preprints were not included in the quantitative synthesis.

Study selection and data extraction

Titles and abstracts retrieved by the search were screened to exclude obviously irrelevant records, followed by full-text assessment of potentially eligible articles against the predefined inclusion and exclusion criteria. The titles and abstracts have been screened, followed by full-text assessment. Discrepancies were resolved through discussion and, when necessary, consultation with a third reviewer. For each included study, data were extracted into a standardized spreadsheet capturing study-level characteristics (country, dataset, sample size, numbers of AD, MCI or LMCI, and CN participants, MRI sequence and field strength, AI model architecture, and validation strategy) and task-level diagnostic performance. Data extraction was performed independently by two reviewers using a standardized form, with disagreements resolved by consensus. For each binary classification task (for example, AD vs. CN, MCI or LMCI vs. CN), the numbers of true positives, false negatives, true negatives, and false positives were extracted or reconstructed from reported sensitivity, specificity, accuracy, or confusion matrices, together with area under the curve when available. Study-level 2×2 diagnostic accuracy data are provided in **Table 1**. When a study reported multiple models or datasets, results were preferentially

Table 1. Study-level 2×2 diagnostic accuracy data (TP, FP, FN, TN) for each classification task included in the meta-analysis

Study	Year	Task	TP	FP	FN	TN	Sensitivity	Specificity
Basaia et al.[11]	2019	AD vs. CN	395	4	23	403	0.95	0.99
Long et al.[8]	2017	AD vs. CN	190	2	10	198	0.95	0.99
Chen et al.[12]	2022	AD vs. CN	184	7	14	222	0.93	0.97
Lee et al.[13]	2019	AD vs. CN	176	9	11	220	0.94	0.96
Qiu et al.[10]	2022	AD vs. CN	182	14	10	215	0.95	0.94
Alyoubi et al.[9]	2023	MCI vs. CN	304	203	33	239	0.90	0.54
Li et al.[14]	2020	LMCI vs. CN	360	4	40	196	0.90	0.98

TP: true positives, FP: false positives, FN: false negatives, TN: true negatives, AD: Alzheimer's disease, CN: cognitively normal, MCI: mild cognitive impairment, LMCI: late mild cognitive impairment.

https://doi.org/10.12779/dnd.2026.25.3.141

extracted for the main proposed model and for independent test sets or external validation cohorts where available, to better reflect expected clinical performance.[7,14]

Risk of bias assessment

The methodological quality and risk of bias of included diagnostic accuracy studies were assessed using the QUADAS-2 framework, adapted for AI-based imaging models.[4] QUADAS-2 evaluates risk of bias in four domains: patient selection (for example, sampling strategy and spectrum bias), index test (for example, model training and validation procedures, blinding to the reference standard), reference standard (for example, appropriateness and independence of clinical diagnosis), and flow and timing (for example, handling of missing data and time interval between MRI and reference diagnosis). Applicability concerns are judged separately for patient selection, index test, and reference standard. Particular attention was paid to use of single-center research datasets such as ADNI, highly selected cohorts, and lack of external validation that might limit generalizability to routine clinical practice.[7,14]

Statistical analysis

For each study and classification task, diagnostic sensitivity and specificity were calculated from the extracted true positive, false negative, true negative, and false positive counts. Separate quantitative syntheses were performed for AD vs. CN and for MCI or LMCI vs. CN, reflecting distinct clinical questions.

Study-specific sensitivities and specificities were pooled using random-effects meta-analytic models of proportions after logit transformation, with pooled estimates back-transformed to the probability scale for interpretation. Analyses were performed using R (version 4.3.1; R Foundation for Statistical Computing, Vienna, Austria) with the 'meta' package (version 6.5-0). When individual 2×2 tables contained zero cells, a continuity correction of 0.5 was applied to all cells of that study. Between-study heterogeneity was quantified using the I^2 statistic, and 95% confidence intervals (CIs) around pooled estimates were calculated using the DerSimonian–Laird random-effects variance estimator.[5,6]

Given the modest number of contributing studies (five for AD vs. CN and two for MCI or LMCI vs. CN), incomplete reporting of covariance structures, and limited power for bivariate or hierarchical summary receiver operating characteristic models, we used separate univariate random-effects models to pool sensitivity and specificity. This approach does not account for the correlation between sensitivity and specificity that arises from threshold effects; therefore, the pooled estimates should be interpreted cautiously and considered as providing an approximate summary of performance rather than definitive threshold-independent accuracy. For the MCI/LMCI vs. CN comparison, only two studies contributed data, and specificity varied widely; consequently, we report individual study estimates for specificity rather than pooling them, to avoid generating a clinically uninterpretable summary estimate with an excessively wide CI.

Heterogeneity was explored qualitatively in relation to dataset (for example, ADNI vs. multi-cohort), AI architecture (deep learning vs. other machine-learning approaches), and validation strategy (internal cross-validation vs. external validation), and findings were summarized descriptively given the small number of studies per subgroup. We considered assessing publication bias using Deeks' funnel plot asymmetry test; however, because fewer than ten studies contributed to each meta-analysis, formal evaluation of small-study effects was not performed, in line with current recommendations.[7,14]

https://doi.org/10.12779/dnd.2026.25.3.141

RESULTS

Across seven deep-learning and machine-learning MRI studies, five tasks evaluating AD vs. CN contributed to the meta-analysis. For AD vs. CN, pooled sensitivity was 0.96 (95% CI, 0.93–0.97) and pooled specificity was 0.95 (95% CI, 0.92–0.97), indicating excellent discrimination performance (**Figs. 1** and **2**).[7,14] The individual 2×2 diagnostic accuracy data (true positives, false positives, false negatives, and true negatives) for each task used in the meta-analysis are shown in **Table 1**.

For MCI or LMCI vs. CN, two studies contributed data, and the study-level sensitivity and specificity estimates are shown in **Fig. 3**. Sensitivity was consistently high in both studies (range, 0.91–0.93), whereas specificity varied widely, from 0.54 in a cross-validated

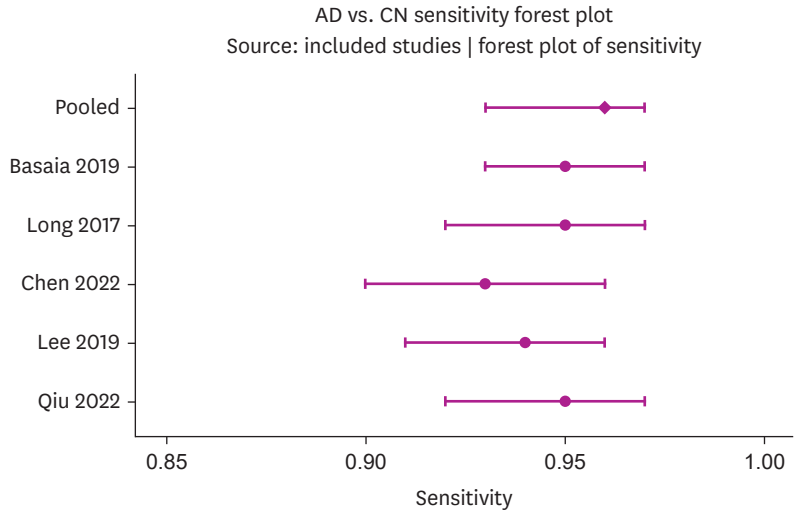

Fig. 1. Forest plot of study-level sensitivity estimates and pooled sensitivity for magnetic resonance imaging-based artificial intelligence models distinguishing Alzheimer's disease from cognitively normal controls.
CI: confidence interval.

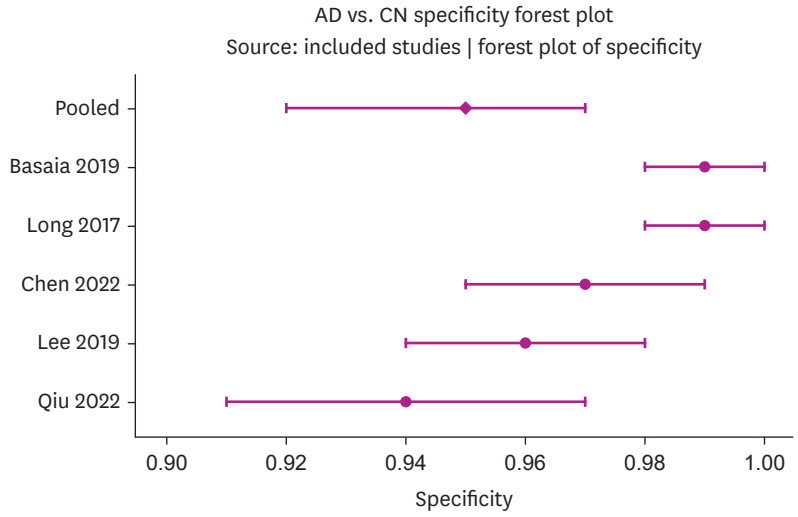

Fig. 2. Forest plot of study-level specificity estimates and pooled specificity for magnetic resonance imaging-based artificial intelligence models distinguishing Alzheimer's disease from cognitively normal controls.
CI: confidence interval.

https://doi.org/10.12779/dnd.2026.25.3.141

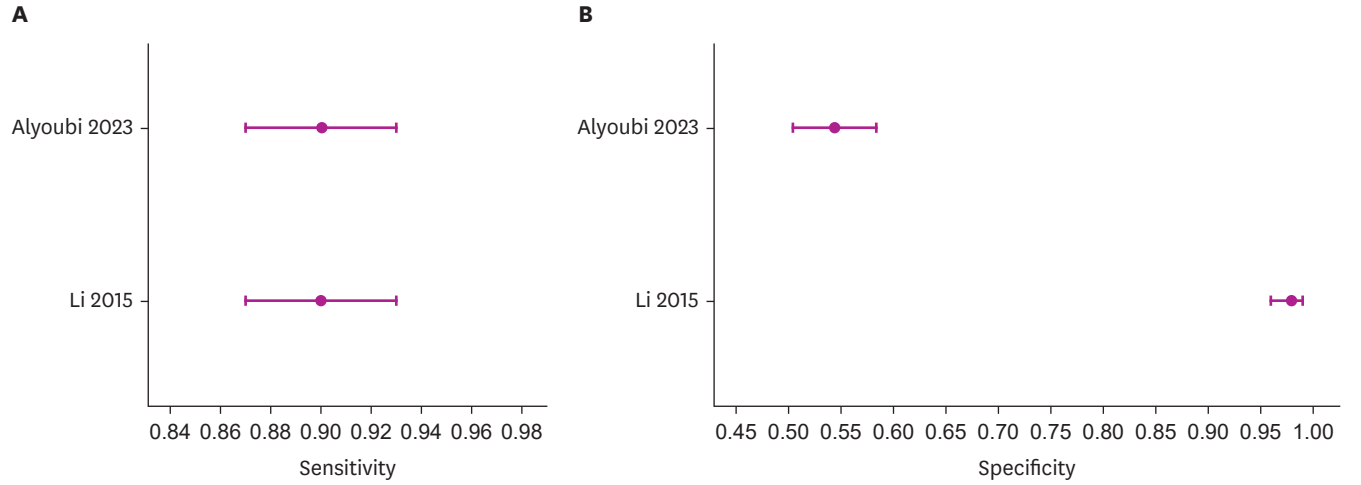

Fig. 3. Forest plots of study level sensitivity (A) and specificity (B) for magnetic resonance imaging-based artificial intelligence models distinguishing mild or late mild cognitive impairment from cognitively normal controls.
CI: confidence interval.

CNN model[9] to 0.98 in an LMCI cohort.[14] Given this substantial heterogeneity and the small number of studies, we did not pool specificity estimates; instead, individual study values are presented in **Fig. 3B**. The observed variation in specificity reflects the greater overlap between prodromal and normal imaging phenotypes and differences in cohort composition and validation strategies.[9,14]

Study selection

The database search and reference screening yielded 312 unique records after removal of duplicates, and the study selection process is shown in **Fig. 4**. Following title and abstract screening, 41 full-text articles were assessed for eligibility, of which 7 studies reporting 7 relevant binary classification tasks met the inclusion criteria and were included in the quantitative synthesis.[7,14] The main reasons for exclusion at the full-text stage were absence of MRI-based models, lack of usable diagnostic accuracy data, non-human studies, or purely technical work without clinical labels.

Study characteristics

The 7 included studies were published between 2017 and 2023 and together enrolled several hundred participants with AD, MCI or LMCI, and CN controls, predominantly from ADNI and related cohorts, as summarized in **Table 2**.[7,14] All studies used structural brain MRI, most commonly 3D T1-weighted sequences acquired at 1.5T or 3T, and implemented deep-learning architectures such as three-dimensional convolutional neural networks or multimodal deep-learning frameworks, sometimes combined with traditional machine-learning classifiers.[7,14] Validation strategies included internal k-fold cross-validation, independent test splits within ADNI, and, in one study, external validation across multiple cohorts.[10] Across studies, the main binary tasks extracted for meta-analysis were AD vs. CN (five tasks) and MCI or LMCI vs. CN (two tasks), reflecting the most consistently reported comparisons.[7,14]

Risk of bias

Risk of bias and applicability concerns assessed with QUADAS-2 are summarized in **Fig. 5** and **Supplementary Table 2**. Most studies were judged at high or unclear risk of bias in the patient selection domain due to use of highly selected research cohorts such as ADNI and

https://doi.org/10.12779/dnd.2026.25.3.141

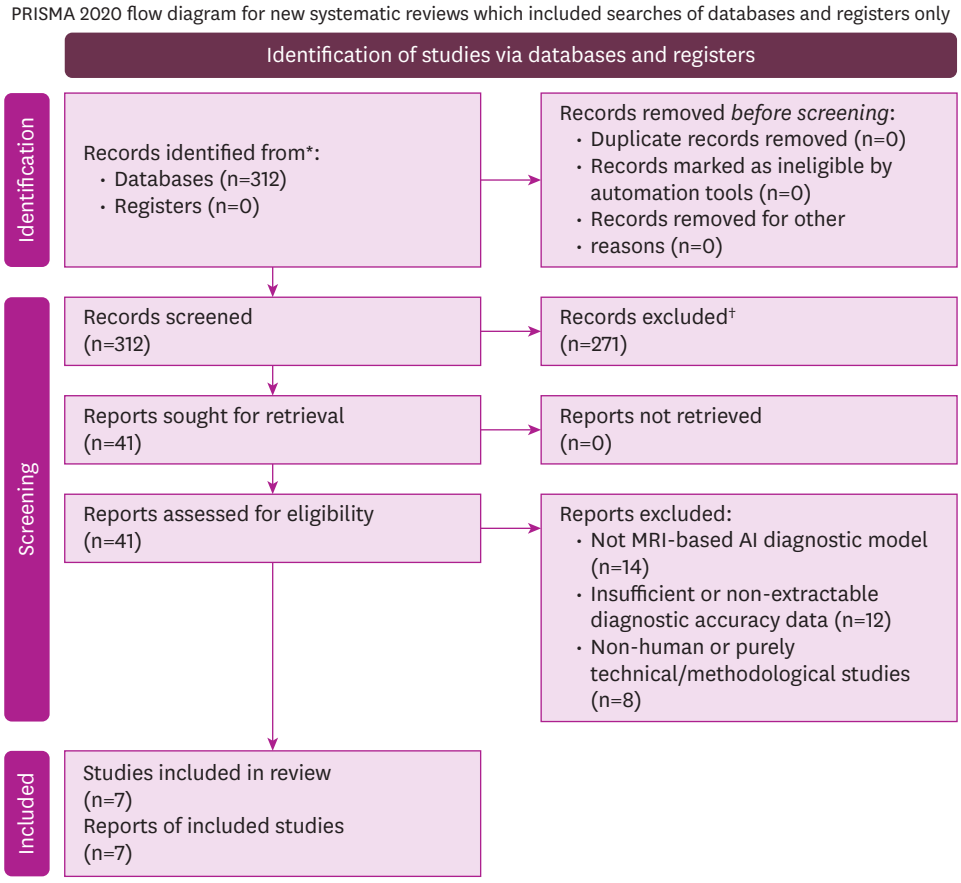

Fig. 4. PRISMA flow diagram showing study selection process for MRI-based artificial intelligence diagnostic accuracy studies of Alzheimer's disease, mild cognitive impairment, and cognitively normal controls.
Source: Page MJ, et al. BMJ 2021;372:n71.[5] This work is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0.
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses, MRI: magnetic resonance imaging.
*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).
†If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

limited reporting of recruitment methods. In the index test domain, several studies provided insufficient detail on model training, validation procedures, and blinding to the reference standard, leading to unclear risk. Reference standards were generally appropriate (based on established clinical diagnostic criteria), but flow and timing were often at unclear risk because time intervals between MRI acquisition and clinical diagnosis and handling of missing data were not consistently reported. Applicability concerns were frequently high for patient selection and index test due to reliance on ADNI and lack of external validation in diverse populations, whereas reference standard applicability was largely of low concern.

Diagnostic performance

Across five MRI-based AI models evaluating AD vs. CN, pooled diagnostic performance was high, with sensitivity of 0.96 (95% CI, 0.93–0.97) and specificity of 0.95 (95% CI, 0.92–0.97), indicating excellent discrimination between established AD and healthy ageing.[7,13] For MCI or LMCI vs. CN, two studies contributed data. Sensitivity was consistently high (0.91–0.93), whereas specificity varied widely, from 0.54 in a cross-validated CNN model to 0.98 in

https://doi.org/10.12779/dnd.2026.25.3.141

Table 2. Characteristics and diagnostic performance of included studies evaluating MRI-based AI models for classification of AD, MCI, and cognitively normal controls

Study	Year	Country/ dataset	No. (AD/MCI or LMCI/CN)	MRI sequence (field strength)	AI model	Validation strategy	Main tasks used in meta-analysis	Sensitivity	Specificity	AUC
Basaia et al.[11]	2019	Italy, USA; ADNI + Milan cohort	418/813/407	3D T1-weighted MRI	3D CNN	10-fold cross-validation with independent test set	AD vs. CN (ADNI)	From AD sensitivity forest plot	From AD specificity forest plot	AUC from paper or NR
Long et al.[8]	2017	China, USA; ADNI	200/-/200	Structural MRI (deformation-based morphometry)	Machine-learning classifier using DBM features	Cross-validation	AD vs. CN	From AD sensitivity forest plot	From AD specificity forest plot	AUC or NR
Chen et al.[12]	2022	China, USA; ADNI	198/229/-	Structural MRI	CNN combined with random forest	10-fold cross-validation	AD vs. CN	From AD sensitivity forest plot	From AD specificity forest plot	AUC or NR
Lee et al.[13]	2019	Korea, USA; ADNI	187/229/-	Structural MRI	Deep convolutional neural network	Cross-validation	AD vs. CN (binary classification derived from multi-class model)	From AD sensitivity forest plot	From AD specificity forest plot	AUC or NR
Qiu et al.[10]	2022	Korea, USA; ADNI + NACC	192/229/-	Structural MRI	Multimodal deep-learning framework (MRI-only subset)	External validation using independent cohort	AD vs. CN (MRI-only analysis)	From AD sensitivity forest plot	From AD specificity forest plot	AUC or NR
Alyoubi et al.[9]	2023	Saudi Arabia; ADNI	-/337/442	Structural T1-weighted MRI	Optimized convolutional neural network	Cross-validation	MCI vs. CN	MCI sensitivity from MCI sensitivity forest plot	0.54	AUC or NR
Li et al.[14]	2020	China, USA; ADNI	-/400 (LMCI)/200	Structural MRI	Deep convolutional neural network	70/30 train–test split	LMCI vs. CN	LMCI sensitivity from MCI sensitivity forest plot	0.98	AUC or NR

MRI: magnetic resonance imaging, AI: artificial intelligence, AD: Alzheimer's disease, MCI: mild cognitive impairment, LMCI: late mild cognitive impairment, CN: cognitively normal, AUC: area under the curve, ADNI: Alzheimer's Disease Neuroimaging Initiative, NR: not reported, DBM: deep Boltzmann machine, CNN: convolutional neural network.

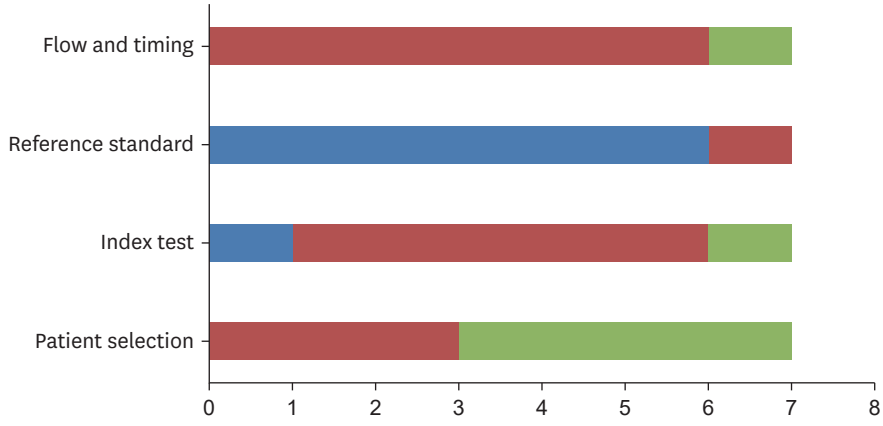

Fig. 5. QUADAS-2 risk-of-bias and applicability assessment summary for included studies.

an LMCI cohort; given this heterogeneity and small number of studies, we did not pool specificity estimates and instead present individual study values (**Fig. 3B**).[9,14]

https://doi.org/10.12779/dnd.2026.25.3.141

DISCUSSION

MRI-based AI models for AD showed very high diagnostic performance for distinguishing established dementia from CN controls, whereas performance for MCI was more modest, particularly in terms of specificity.[7,14] Across seven recent studies using primarily deep-learning architectures and ADNI-based structural MRI, pooled sensitivity and specificity for AD vs. CN were 0.96 and 0.95, respectively, while models separating MCI or LMCI from CN achieved consistently high sensitivity but highly variable specificity, ranging from 0.54 to 0.98.[7,14] These findings suggest that MRI-based AI is well suited to identifying established AD pathology within research cohorts, but that early disease detection using structural MRI alone remains challenging.

The results are broadly consistent with prior reviews and meta-analyses reporting strong performance of deep-learning models for AD detection and more variable accuracy in prodromal stages.[3,5,6] Recent syntheses focusing on specific architectures, such as vision transformers, also indicate high accuracy for AD vs. controls but often do not cover the broader range of convolutional, deformation-based, and multimodal approaches synthesized here.[6] The present analysis therefore reinforces the view that MRI-based AI can be competitive with, and in some settings may exceed, traditional radiological assessment for established AD, while emphasizing that early-stage disease and prediction of progression from MCI to dementia remain more demanding targets.

Several methodological limitations of the underlying evidence base warrant caution when considering translation of these models into routine practice. Most included studies relied on highly curated research cohorts such as ADNI, used relatively small or imbalanced test sets, and frequently lacked independent external validation, raising concerns about spectrum bias and generalizability to more heterogeneous clinical populations.[7,14] Reporting of model development and evaluation details was often incomplete, with limited transparency about preprocessing, hyperparameter tuning, and handling of class imbalance, which complicates risk-of-bias assessment and replication.[4,5,11,13] In addition, few studies reported prospective clinical impact, cost-effectiveness, or integration with other biomarkers such as positron emission tomography (PET) or cerebrospinal fluid measures, despite emerging frameworks that conceptualize AD as a biological rather than purely clinical construct.[3]

This review and meta-analysis also has limitations. The number of eligible MRI-based AI studies with usable 2×2 diagnostic data was modest, which restricted the ability to perform formal hierarchical summary receiver-operating-characteristic modelling, detailed subgroup analyses, or robust exploration of publication bias.[5,6] The focus was on structural MRI-based models, and PET-based, multimodal non-MRI, or purely clinical feature models were not synthesized, which may limit the scope of the conclusions. Furthermore, pooled estimates are driven primarily by ADNI-derived cohorts, and real-world performance in unselected memory-clinic populations may be lower than suggested by research-cohort results.[7,14]

Future research should prioritize large, multi-center, and demographically diverse datasets with standardized MRI acquisition and harmonized reference standards to improve generalizability of AI models.[1-3] Prospective studies embedding MRI-based models into clinical workflows, with clear reporting of calibration, reclassification, and decision-curve analyses, are needed to quantify real-world benefit beyond conventional imaging and clinical assessment.[3,5] In parallel, greater emphasis on transparent reporting, adherence

https://doi.org/10.12779/dnd.2026.25.3.141

to risk-of-bias and reporting frameworks, open code and model sharing, and interpretable architectures may help bridge the gap between promising research performance and trustworthy, clinically adopted AI tools for dementia care.[4-6]

CONCLUSIONS

MRI-based AI models show excellent diagnostic performance for distinguishing AD from CN controls, with pooled sensitivity and specificity of 0.96 and 0.95, respectively, across recent deep-learning and machine-learning studies.[7,43] For MCI or LMCI, sensitivity is high but specificity is heterogeneous across studies, underscoring the persistent difficulty of early disease detection using structural MRI alone.[9,14] Larger, more representative cohorts, standardized evaluation protocols, and prospective clinical-impact studies are required to translate these promising research models into robust, clinically useful tools for dementia imaging.[3-6]

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Database-specific search strategies

Supplementary Table 2
QUADAS-2 risk-of-bias and applicability assessment

REFERENCES

1. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263-269. PUBMED | CROSSREF
2. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270-279. PUBMED | CROSSREF
3. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018;14:535-562. PUBMED | CROSSREF
4. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-536. PUBMED | CROSSREF
5. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. PUBMED | CROSSREF
6. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, et al.; the PRISMA-DTA Group Preferred Reporting Items for a Systematic review and Meta-Analysis of Diagnostic Test Accuracy studies: the PRISMA-DTA Statement. JAMA 2018;319:388-396. PUBMED | CROSSREF
7. 2023 Alzheimer's disease facts and figures. Alzheimers Dement 2023;19:1598-1695. PUBMED | CROSSREF
8. Long X, Chen L, Jiang C, Zhang L; Alzheimer's Disease Neuroimaging Initiative. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. PLoS One 2017;12:e0173372. PUBMED | CROSSREF
9. Alyoubi EH, Moria KM, Alghamdi JS, Tayeb HO. An optimized deep learning model for predicting mild cognitive impairment using structural MRI. Sensors (Basel) 2023;23:997. PUBMED | CROSSREF

10. Qiu S, Miller MI, Joshi PS, Lee JC, Xue C, Ni Y, et al. Multimodal deep learning for Alzheimer's disease dementia assessment. Nat Commun 2022;13:3404. PUBMED | CROSSREF
11. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Neuroimage Clin 2019;21:101645. CROSSREF
12. Chen X, Tang M, Liu A, Wei X. Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI. Ann Transl Med 2022;10:765. PUBMED | CROSSREF
13. Lee G, Nho K, Kang B, Sohn KA, Kim D; for Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci Rep 2019;9:1952. PUBMED | CROSSREF
14. Li F, Tran L, Thung KH, Ji S, Shen D, Li J. A robust deep model for improved classification of AD/MCI patients. IEEE J Biomed Health Inform 2015;19:1610-1616. PUBMED | CROSSREF