

Review

Comparative Diagnostic Accuracy of AI-Assisted Fluorine-18 Fluorodeoxyglucose Positron Emission Tomography Versus Structural Magnetic Resonance Imaging in Alzheimer Disease: Systematic Review and Meta-Analysis

Bingbing Wang^{1*}, PhD; Tailiang Zhao^{1*}, PhD; Rongrong Ma^{2*}, MSc; Xiaochuan Huo¹, PhD; Xiaoxiao Xiong³, PhD; Minjie Wu¹, PhD; Yuran Wang⁴, MSc; Liu Liu³, PhD; Zhijiang Zhuang³, PhD; Bin Wang^{4,5}, PhD; Jixin Shou^{1*}, PhD

¹Department of Neurosurgery, Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China

²Department of Anesthesiology and Perioperative Medicine, Xuchang Central Hospital, Henan University of Science and Technology, Xuchang, China

³Department of Neurology, Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China

⁴The Fifth Clinical Medical College, Henan Medical College, Zhengzhou University, Zhengzhou, China

⁵Department of Cardiology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

*these authors contributed equally

Corresponding Author:

Tailiang Zhao, PhD
Department of Neurosurgery
Fifth Affiliated Hospital of Zhengzhou University
No. 3 Kangfu-qian Street
Zhengzhou 450052
China
Phone: 86 17630927442
Email: tailiang1996@163.com

Abstract

Background: Neuroimaging is crucial in the diagnosis of Alzheimer disease (AD). In recent years, artificial intelligence (AI)-based neuroimaging technology has rapidly developed, providing new methods for accurate diagnosis of AD, but its performance differences still need to be systematically evaluated.

Objective: This study aims to conduct a systematic review and meta-analysis comparing the diagnostic performance of AI-assisted fluorine-18 fluorodeoxyglucose positron emission tomography (¹⁸F-FDG PET) and structural magnetic resonance imaging (sMRI) for AD.

Methods: Databases including Web of Science, PubMed, and Embase were searched from inception to January 2025 to identify original studies that developed or validated AI models for AD diagnosis using ¹⁸F-FDG PET or sMRI. Methodological quality was assessed using the TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence) checklist. A bivariate mixed-effects model was employed to calculate pooled sensitivity, specificity, and summary receiver operating characteristic curve area (SROC-AUC).

Results: A total of 38 studies were included, with 28 moderate-to-high-quality studies analyzed. Pooled SROC-AUC values were 0.94 (95% CI 0.92-0.96) for sMRI and 0.96 (95% CI 0.94-0.98) for ¹⁸F-FDG PET, demonstrating statistically significant intermodal differences ($P=.02$). Subgroup analyses revealed that for machine learning, pooled SROC-AUCs were 0.89 (95% CI 0.86-0.92) for sMRI and 0.95 (95% CI 0.92-0.96) for ¹⁸F-FDG PET, while for deep learning, these values were 0.96 (95% CI 0.94-0.97) and 0.97 (95% CI 0.96-0.99), respectively. Meta-regression identified heterogeneity arising from study quality stratification, algorithm types, and validation strategies.

Conclusions: Both AI-assisted ¹⁸F-FDG PET and sMRI exhibit high diagnostic accuracy in AD, with ¹⁸F-FDG PET demonstrating superior overall diagnostic performance compared to sMRI.

JMIR Aging 2025;8:e76981; doi: [10.2196/76981](https://doi.org/10.2196/76981)

Keywords: ^{18}F -FDG PET; Alzheimer disease; artificial intelligence; deep learning; machine learning; Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA; sMRI; structural magnetic resonance imaging; systematic review

Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disorder characterized by insidious onset, cognitive decline, and memory impairment. In the United States, approximately 6.7 million adults aged ≥ 65 years are affected by AD, while China faces an even greater burden, with over 13 million cases [1,2]. The prolonged disease course and high comorbidity rates have established AD as one of the most fatal and economically burdensome conditions of the 21st century [3-5]. Accurate diagnosis of AD is critical for therapeutic decision-making and prognostic evaluation, particularly in the context of global population aging and increasing demands for precise patient stratification in clinical trials [6,7].

Neuroimaging modalities, including fluorine-18 fluorodeoxyglucose positron emission tomography (^{18}F -FDG PET) and structural magnetic resonance imaging (sMRI), have become cornerstone technologies in AD diagnostic frameworks (eg, National Institute on Aging-Alzheimer's Association criteria) because of their noninvasive nature and quantitative capabilities [8-10]. Recent advancements in artificial intelligence (AI) have revolutionized medical image analysis: machine learning (ML) enables data-driven predictive modeling beyond traditional rule-based programming, while deep learning (DL), an advanced ML paradigm, employs multilayer neural networks to extract high-level features from complex datasets, demonstrating transformative potential in neuroimaging [11-14]. Despite extensive research on AI-assisted sMRI or ^{18}F -FDG PET for AD diagnosis, two critical challenges hinder evaluation of model generalizability: (1) substantial heterogeneity in algorithm designs and validation frameworks, and (2) significant variability in the quality of individual studies [15]. Furthermore, no high-quality meta-analysis has comprehensively compared the diagnostic performance of these two imaging modalities.

This systematic review and meta-analysis addresses three critical objectives: (1) quantitative evaluation of diagnostic accuracy metrics for AI-assisted ^{18}F -FDG PET and sMRI based on moderate-to-high-quality evidence; (2) direct comparison of diagnostic performance between modalities; and (3) investigation of confounding factors, including study quality (moderate-to-high vs low), algorithm types (ML vs DL), and validation strategies (internal vs external), through meta-regression. The findings are anticipated to inform evidence-based optimization of AD diagnostic pathways.

Methods

Overview

This study adhered to the PRISMA-DTA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy) guidelines and was prospectively registered on PROSPERO (ID: CRD42023449927) to ensure transparency and minimize reporting bias [16,17].

Two reviewers independently conducted all stages of the review process, including title and abstract screening, full-text evaluation, data extraction, adherence assessment to reporting guidelines, and risk-of-bias evaluation. Discrepancies were resolved through group consensus.

Literature Search Strategy

Two investigators (TLZ and BW) systematically searched PubMed, Web of Science, and Embase from inception to January 2025 using a combination of Medical Subject Headings (MeSH) terms and free-text keywords. Additional searches were performed on clinical trial registries and OpenGrey to identify unpublished clinical trials and gray literature. Search terms encompassed four domains: (1) disease terminology (Alzheimer disease, AD, Alzheimer Syndrome); (2) imaging modalities (^{18}F -FDG PET, sMRI); (3) AI methodologies (ML, DL); and (4) diagnostic metrics (sensitivity, specificity, summary receiver operating characteristic curve area [SROC-AUC]). Reference lists of included studies were manually screened to identify additional relevant publications. The specific search strategies are provided in Table S1 in [Multimedia Appendix 1](#).

Literature Inclusion and Exclusion Criteria

The inclusion criteria were as follows: (1) human studies developing or validating AI models using ^{18}F -FDG PET or sMRI to differentiate AD from normal controls; (2) AD diagnosis based on National Institute on Aging-Alzheimer's Association or International Working Group for New Research Criteria for Alzheimer's Disease criteria; (3) availability of diagnostic performance metrics (eg, true positives, false positives, true negatives, false negatives) or explicit reporting of sensitivity and specificity; and (4) full-text availability in English. The exclusion criteria were as follows: (1) case reports, reviews, letters, or conference abstracts; (2) studies lacking sufficient diagnostic performance data; and (3) duplicate publications reporting on the same cohort without novel analyses.

Literature Screening and Data Extraction

Two reviewers (TZ and BW) independently performed title and abstract and full-text screening, followed by cross-verification to ensure accuracy. Two additional investigators (RM and XH) extracted data using predefined forms, including study characteristics (first author, publication year), model specifications (algorithm type, validation strategies), and diagnostic performance metrics (2x2 contingency tables, sensitivity, specificity). Extracted data were cross-checked for completeness and precision.

Quality Assessment of Included Studies

The methodological quality of included studies was evaluated using an adapted TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence) checklist (Table S2 in

[Multimedia Appendix 1](#)) by two independent reviewers (RM and XH) [18]. The instrument assessed four domains: (1) data quality (sample diversity, adequacy, preprocessing standardization), (2) model development (feature extraction rationale, algorithm selection, hyperparameter optimization), (3) validation methods (cross-validation rigor, external validation), and (4) clinical applicability (interpretability, net clinical benefit). Each of the nine items was scored 0-2 (0=unsatisfied; 1=partially satisfied; 2=fully satisfied), with total scores categorized as high quality (≥ 16), moderate quality (10-15), or low quality (≤ 9). Disagreements were resolved through discussion.

Data Analysis

Diagnostic performance metrics were calculated using a bivariate mixed-effects model. SROC-AUC served as the primary outcome due to its threshold independence. Diagnostic accuracy was classified per National Institutes of Health criteria as high ($\text{AUC} \geq 0.90$), moderate (0.70-0.89), or low ($\text{AUC} < 0.70$) [19].

Heterogeneity Assessment and Publication Bias Evaluation

Heterogeneity was quantified using Cochran Q test (significance defined at the .05 level) and I^2 statistics (25%: low; 50%: moderate; 75%: high heterogeneity) [20]. Threshold effects were assessed using Spearman correlation between logit-transformed sensitivity and 1-specificity. Sensitivity analyses evaluated outlier influence by iteratively excluding individual studies. Univariable meta-regression analyses were conducted to assess the influence of potential confounding factors, including study quality (moderate to high vs low quality), algorithm type (ML vs DL), and validation strategy (internal vs external validation). Statistical significance of modifiers of the pooled effect size was defined at the .05 level

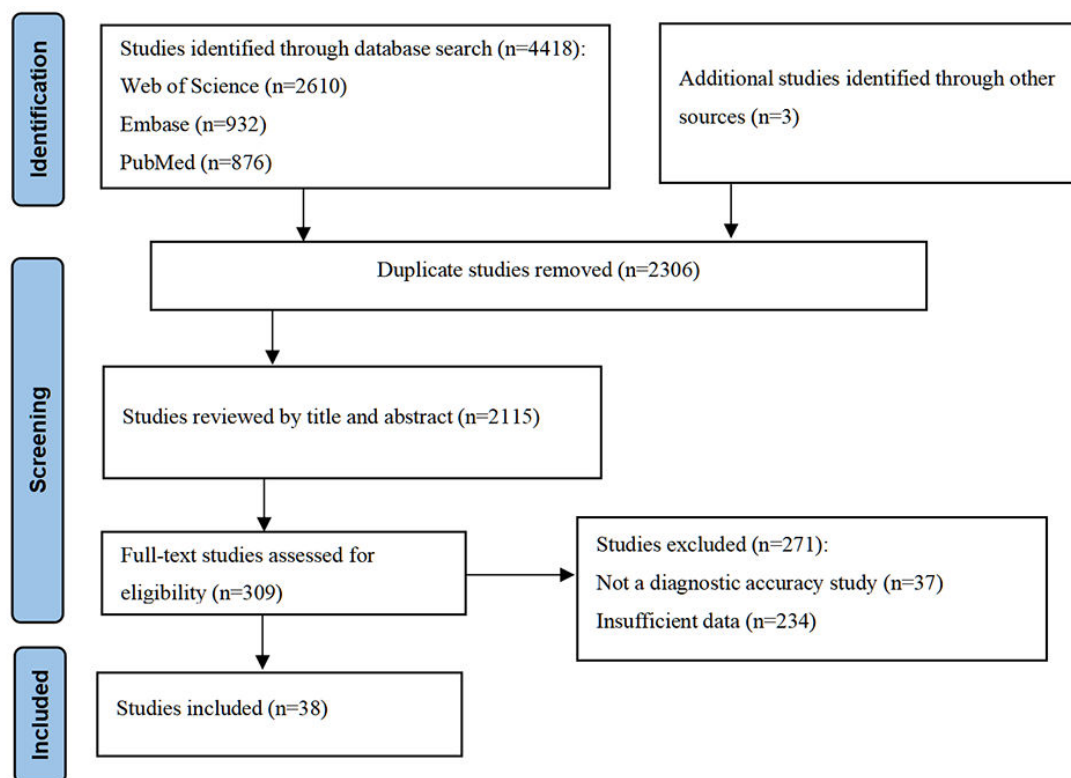
after adjustment via the Holm-Bonferroni method. Publication bias was assessed via funnel plot asymmetry and Egger test, with significance defined at the .05 level. All analyses were conducted in Stata 16.0 using the MIDAS commands.

As this study focused on diagnostic accuracy, we employed the bivariate random-effects model via the MIDAS command in Stata to jointly synthesize sensitivity and specificity. This approach accounts for both within- and between-study variability, as well as the inherent correlation between sensitivity and specificity. In such models, conventional heterogeneity measures like I^2 , although still reported for completeness, may not fully capture the joint variability of test performance measures. Therefore, to better understand heterogeneity, we additionally performed meta-regression and calculated joint likelihood ratio tests to explore potential effect modifiers. These methods align with recommended practices in diagnostic test accuracy meta-analyses [21,22].

Results

Literature Screening Process and Results

Initial searches identified 876 PubMed records ($n=89$ involving ^{18}F -FDG PET; $n=787$ sMRI), 932 Embase records ($n=111$ ^{18}F -FDG PET; $n=821$ sMRI), and 2610 Web of Science records ($n=230$ ^{18}F -FDG PET; $n=2380$ sMRI), supplemented by 3 additional studies. After deduplication in EndNote 20, 2 reviewers independently screened titles and abstracts using the Population, Intervention, Comparison, Outcome, Study Design (PICOS) criteria, followed by full-text evaluation against predefined inclusion/exclusion criteria. This rigorous process yielded 38 studies for systematic review and meta-analysis [23-60]. The literature screening process and results are shown in [Figure 1](#).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram.

Characteristics of Included Studies

Table 1 and Table S3 in [Multimedia Appendix 1](#) detail the characteristics of the included studies. Imaging modalities comprised sMRI (n=15, 39%), ^{18}F -FDG PET (n=15, 39%), and combined sMRI/ ^{18}F -FDG PET (n=8, 22%). Data sources predominantly relied on open-access databases (n=29, 76%), with 28 (97%) from the Alzheimer's Disease Neuroimaging Initiative database and 1 (3%) from the Open Access Series of Imaging Studies database. Four studies (11%) used institutional data, while 5 (13%) combined Alzheimer's Disease Neuroimaging Initiative database with local datasets. Algorithmically, 26 studies (68%) employed ML, predominantly support vector machines (17/26, 65%), whereas 12 (32%) utilized DL, primarily convolutional neural networks (9/12, 75%). Internal validation was implemented in 33 studies (87%), with 10-fold cross-validation in 16 (48%); only 5 studies (13%) incorporated external validation.

Table 1. Characteristics of all included studies (n=38).

Studies	Model development and internal validation				Training/validation			Model external validation	
	Definition	Modality	Data source	Algorithm used	(ratio)	Testing	Type of internal validation	Data source	Definition
Zhang et al [23], 2011	AD ^a ; 51; NC ^b ; 52	sMRI ^c ; PET ^d	ADNI ^e database	SVM ^f	NR ^g	NR	10-fold CV ^h	NR	NR
Yun et al [24], 2015	AD; 71; NC: 85	sMRI; PET	ADNI database	LDA ⁱ	NR	NR	LOOCV ^j	NR	NR
Westman et al [25], 2012	AD; 96; NC: 111	sMRI	ADNI database	OPLS ^k	NR	NR	7-fold CV	NR	NR
Vemuri et al [26], 2008	AD: 190; NC: 190	sMRI	Mayo Clinic	SVM	280	100	4-fold CV; hold-out	NR	NR
Suk et al [27], 2014	AD: 93; NC: 101	sMRI; PET	ADNI database	DBM ^l ; SVM	NR	NR	10-fold CV	NR	NR
Sayeed et al [28], 2002	AD: 18; NC: 10	PET	Hammersmith Hospital	DFA ^m	NR	NR	LOOCV	NR	NR
Pan et al [29], 2019	AD: 247; NC: 246	PET	ADNI database	SVM	NR	NR	10-fold CV	NR	NR
Padilla et al [30], 2012	AD: 53; NC: 52	PET	ADNI database	SVM	NR	NR	LOOCV	NR	NR
Ni et al [31], 2021	AD: 638; NC: 629	PET	ADNI database	CNN ⁿ	1000	267	Hold-out	NR	NR
Magnin et al [32], 2009	AD: 16; NC: 22	sMRI	Pitié-Salpêtrière Hospital	SVM	75%	25%	Bootstrap resampling	NR	NR
Lu et al [33], 2018	AD: 226; NC: 304	PET	ADNI database	MDNN ^o	NR	NR	10-fold CV	NR	NR
Liu et al [34], 2012	AD: 198; NC: 229	sMRI	ADNI database	SRC ^p	NR	NR	10-fold CV	NR	NR
Liu et al [35], 2018	AD: 93; NC: 100	PET	ADNI database	2D-CNN	NR	NR	10-fold CV	NR	NR
Li et al [36], 2015	AD: 44; NC: 45	PET	ADNI database; TUM database	GMM ^q ; SVM	NR	NR	10-fold CV	NR	NR
Lerch et al [37], 2008	AD: 19; NC: 17	sMRI	Ludwig Maximilian University of Munich	QDA ^r	NR	NR	LOOCV	NR	NR
Kim et al [38], 2020	AD: 141; NC: 348	PET	ADNI database	CNN	NR	NR	NR	Severance dataset	AD: 80; NC: 72
Kim et al [39], 2020	AD: 139; NC: 347	PET	ADNI database	BEGAN ^s ; SVM	NR	NR	NR	Severance dataset	AD: 73; NC: 68
Katako et al [40], 2018	AD: 94; NC: 111	PET	ADNI database	SVM	NR	NR	10-fold CV	NR	NR
Ismail et al [41], 2023	AD: 511; NC: 535	sMRI; PET	ADNI database	MultiAz-Net; SVM	NR	NR	10-fold CV	NR	NR

Studies	Model development and internal validation					Training/validation			Model external validation		
	Definition	Modality	Data source	Algorithm used	(ratio)	Testing	Type of internal validation	Data source	Definition		
Illán et al [42], 2011	AD: 95; NC: 97	PET	ADNI database	SVM	NR	NR	LOOCV	NR	NR		
Hinrichs et al [43], 2009	AD: 89; NC: 94	sMRI; PET	ADNI database	LPhoosting [†]	NR	NR	Leave-many-out CV	NR	NR		
Gray et al [44], 2012	AD: 50; NC: 54	PET	ADNI database	SVM	75%	25%	Monte Carlo CV	NR	NR		
Gray et al [45], 2013	AD: 37; NC: 35	sMRI; PET	ADNI database	RF ^u	75%	25%	Monte Carlo CV	NR	NR		
Feng et al [46], 2019	AD: 93; NC: 100	sMRI; PET	ADNI database	FSBi-LSTM ^v ; 3D-CNN	NR	NR	10-fold CV	NR	NR		
Cuingnet et al [47], 2011	AD: 137; NC: 162	sMRI	ADNI database	SVM	50%	50%	Hold-out	NR	NR		
Chen et al [48], 2022	AD: 326; NC: 413	sMRI	ADNI-1 database; ADNI-2 database	2D-CNN; 3D-CNN	NR	NR	NR	ADNI-1 database; ADNI-2 database	AD: 326; NC: 413		
Song et al [49], 2021	AD: 95; NC: 126	sMRI; PET	ADNI database	3D-CNN	NR	NR	10-fold CV	NR	NR		
Li et al [50], 2019	AD: 130; NC: 162	PET	ADNI database	SVM	70%	30%	Monte Carlo CV	Huashan database	AD: 22; NC: 22		
Ahila et al [51], 2022	AD: 220; NC: 635	PET	ADNI database	2D-CNN	90%	10%	Hold-out	NR	NR		
Toussaint et al [52], 2012	AD: 80; NC: 80	PET	ADNI database	SVM	NR	NR	LOOCV	NR	NR		
Tong et al [53], 2014	AD: 198; NC: 231	sMRI	ADNI database	SVM	NR	NR	LOOCV	NR	NR		
Min et al [54], 2014	AD: 97; NC: 128	sMRI	ADNI database	SVM	NR	NR	10-fold CV	NR	NR		
Jin et al [55], 2020	AD: 488; NC: 536	sMRI	In-house database; ADNI database	3D attention network	NR	NR	10-fold CV; leave center out CV	ADNI database; in-house database	AD: 488; NC: 536		
Cho et al [56], 2012	AD: 128; NC: 160	sMRI	ADNI database	LDA	142	146	Hold-out	NR	NR		
Chincarini et al [57], 2011	AD: 144; NC: 189	sMRI	ADNI database	RF; SVM	NR	NR	20-fold CV	NR	NR		
Beheshti et al [58], 2017	AD: 92; NC: 94	sMRI	ADNI database	SVM	NR	NR	10-fold CV	NR	NR		
Anandh et al [59], 2016	AD: 30; NC: 55	sMRI	OASIS ^w database	SVM	NR	NR	10-fold CV	NR	NR		

Studies	Model development and internal validation				Model external validation				
	Definition	Modality	Data source	Algorithm used	Training/validation (ratio)	Testing	Type of internal validation	Data source	Definition
Amaroso et al [60], 2018	AD: 86; NC: 81	sMRI	ADNI database	RF	67	100	10-fold CV; hold-out	NR	NR

^aAD: Alzheimer disease.

^bNC: normal cognitive.

^csMRI: structural magnetic resonance imaging.

^dPET: positron emission tomography.

^eADNI: Alzheimer’s Disease Neuroimaging Initiative.

^fSVM: support vector machine.

^gNR: no report.

^hCV: cross-validation.

ⁱLDA: linear discriminant analysis.

^jLOOCV: leave-one-out cross-validation.

^kOPLS: orthogonal partial least squares.

^lDBM: deep Boltzmann machine.

^mDFA: discriminant function analysis.

ⁿCNN: convolutional neural network.

^oMDNN: multiscale deep neural network.

^pSRC: sparse representation-based classifier.

^qGMM: Gaussian mixture model.

^rQDA: quadratic discriminant analysis.

^sBEGAN: boundary equilibrium generative adversarial network.

^tLP: linear program.

^uRF: random forest.

^vFSBi-LSTM: fully stacked bidirectional long short-term memory.

^wOASIS: Open Access Series of Imaging Studies.

Quality Assessment Results of Included Studies

Nine (24%) studies were high quality, characterized by adequate sample sizes, rigorous validation (eg, multicenter external validation), and standardized reporting. Moderate-quality studies (n=19, 50%) met basic methodological standards (eg, cross-validation) but had limitations such as single-center data or insufficient clinical correlation analyses. Ten (26%) studies were low quality due to small sample sizes (<100 cases) and inadequate validation strategies, compromising external validity. The results of quality assessments for each study are detailed in Table S4 in [Multimedia Appendix 1](#).

Data Analysis Results

Overall Diagnostic Performance

Analysis of 29 sMRI-related contingency tables demonstrated a pooled sensitivity of 88% (95% CI 86%-89%), a specificity of 92% (95% CI 90%-93%), and an SROC-AUC of 0.94 (95% CI 0.92-0.96). Subgroup analyses stratified by algorithm type revealed that ML (19 tables) achieved a sensitivity of 86% (95% CI 85%-88%), a specificity of 91% (95% CI 88%-93%), and an SROC-AUC of 0.88 (95% CI 0.85-0.90), while DL (10 tables) showed improved performance with a sensitivity of 88% (95% CI 86%-91%), a specificity of 92% (95% CI 90%-94%), and an SROC-AUC of 0.96 (95% CI 0.94-0.97). Stratification by study quality indicated that moderate-to-high-quality studies (24 tables) yielded a sensitivity of 87% (95% CI 85%-89%), a specificity of 91% (95% CI 89%-93%), and an SROC-AUC of 0.94 (95% CI 0.92-0.96), whereas low-quality studies (5 tables) reported elevated metrics with a sensitivity of 91% (95% CI 87%-94%), a specificity of 95% (95% CI 92%-97%),

and an SROC-AUC of 0.98 (95% CI 0.96-0.99). Validation strategy-based analysis showed that internal validation (25 tables) achieved a sensitivity of 88% (95% CI 86%-90%), a specificity of 92% (95% CI 90%-94%), and an SROC-AUC of 0.95 (95% CI 0.92-0.96), while external validation (4 tables) demonstrated marginally lower performance with a sensitivity of 85% (95% CI 81%-89%), a specificity of 91% (95% CI 85%-94%), and an SROC-AUC of 0.93 (95% CI 0.90-0.95).

For ¹⁸F-FDG PET (27 tables), pooled estimates demonstrated a sensitivity of 90% (95% CI 88%-92%), a specificity of 93% (95% CI 91%-94%), and an SROC-AUC of 0.96 (95% CI 0.94-0.98). Subgroup analyses stratified by algorithm type revealed that ML (15 tables) achieved a sensitivity of 89% (95% CI 86%-90%), a specificity of 91% (95% CI 88%-93%), and an SROC-AUC of 0.94 (95% CI 0.91-0.96), while DL (12 tables) exhibited superior performance with a sensitivity of 91% (95% CI 89%-93%), a specificity of 94% (95% CI 93%-96%), and an SROC-AUC of 0.98 (95% CI 0.96-0.99). Moderate-to-high-quality studies (21 tables) demonstrated a sensitivity of 90% (95% CI 88%-92%), a specificity of 93% (95% CI 91%-94%), and an SROC-AUC of 0.96 (95% CI 0.94-0.98), whereas low-quality studies (6 tables) showed comparable metrics with a sensitivity of 91% (95% CI 86%-94%), a specificity of 93% (95% CI 87%-96%), and an SROC-AUC of 0.96 (95% CI 0.94-0.98). Internal validation (24 tables) yielded a sensitivity of 90% (95% CI 89%-92%), a specificity of 93% (95% CI 91%-94%), and an SROC-AUC of 0.96 (95% CI 0.94-0.97), while external validation data (3 tables) were insufficient for SROC-AUC calculation but reported a sensitivity of 87% (95% CI 81%-93%) and a specificity of 95% (95% CI 91%-97%). The data analysis results are shown in [Table 2](#).

Table 2. Summary estimates and meta-regression of pooled performance of all studies.

Main and subordinate directory	Tables, n	Sensitivity (%; 95% CI)	<i>I</i> ²	Specificity (%; 95% CI)	<i>I</i> ²	Joint <i>P</i> value	AUC ^a (95% CI)
sMRI ^b							
Overall	29	88 (86-89)	55.32	92 (90-93)	74.08	— ^j	0.94 (0.92-0.96)
Study quality						.02	
Medium-to-high quality	24	87 (85-89)	57.99	91 (89-93)	75.94		0.94 (0.92-0.96)
Low quality	5	91 (87-94)	10.2	95 (92-97)	0		0.98 (0.96-0.99)
Algorithm type						.40	
ML ^d	19	86 (85-88)	21.23	91 (88-93)	73.82		0.88 (0.85-0.90)
DL ^e	10	88 (86-91)	76.56	92 (90-94)	76.53		0.96 (0.94-0.97)
Validation strategy						.33	
Internal validation	25	88 (86-90)	46.38	92 (90-94)	71.65		0.95 (0.92-0.96)
External validation	4	85 (81-89)	72.26	91 (85-94)	83.65		0.93 (0.90-0.95)
SVM ^f	10	87 (85-90)	17.47	93 (90-95)	57.65		0.93 (0.90-0.95)
CNN ^g	9	88 (85-91)	78.19	92 (90-94)	77.55		0.96 (0.93-0.97)
¹⁸ F-FDG PET ^k							
Overall	27	90 (88-92)	42.04	93 (91-94)	61.49	—	0.96 (0.94-0.98)
Study quality						.96	

Main and subordinate directory	Tables, n	Sensitivity (%; 95% CI)	I^2	Specificity (%; 95% CI)	I^2	Joint P value	AUC ^a (95% CI)
Medium-to-high quality	21	90 (88-92)	44.58	93 (91-94)	58.52	.01	0.96 (0.94-0.98)
Low quality	6	91 (86-94)	42.73	93 (87-96)	72.3		0.96 (0.94-0.98)
Algorithm type							
ML	15	89 (86-90)	1.61	91 (88-93)	45.64		0.94 (0.91-0.96)
DL	12	91 (89-93)	55.23	94 (93-96)	45.92	.26	0.98 (0.96-0.99)
Validation strategy							
Internal validation	24	90 (89-92)	34.61	93 (91-94)	64.9		0.96 (0.94-0.97)
External validation	3	87 (81-93)	—	95 (91-99)	—		—
SVM	11	90 (87-92)	0	92 (89-94)	37.5		0.94 (0.92-0.96)
CNN	8	91 (88-94)	69.76	93 (92-95)	34.63		0.97 (0.95-0.98)

^aAUC: area under the receiver operating characteristic curve.

^bsMRI: structural magnetic resonance imaging.

^cNot applicable.

^dML: machine learning.

^eDL: deep learning.

^fSVM: support vector machine.

^gCNN: convolutional neural network.

^h¹⁸F-FDG PET: fluorine-18 fluorodeoxyglucose positron emission tomograph.

Moderate-to-High-Quality Study Subgroup Analysis

For sMRI, pooled sensitivity, specificity, and SROC-AUC were 87% (95% CI 85%-89%), 91% (95% CI 89%-93%), and 0.94 (95% CI 0.92-0.96), respectively (see [Figures 2 and 3](#)). Stratification by algorithm type indicated that ML models achieved 86% (95% CI 84%-88%) sensitivity, 89% (95% CI 86%-92%) specificity, and an SROC-AUC of 0.89 (95% CI 0.86-0.92), while DL models demonstrated 88% (95% CI 86%-91%) sensitivity, 92% (95% CI 90%-94%) specificity, and an SROC-AUC of 0.96 (95% CI 0.94-0.97).

For ¹⁸F-FDG PET, pooled sensitivity, specificity, and SROC-AUC were 90% (95% CI 88%-92%), 93% (95% CI 91%-94%), and 0.96 (95% CI 0.94-0.98), respectively (see [Figures 4 and 5](#)). Subgroup analyses revealed ML models achieved 89% (95% CI 86-91%) sensitivity, 91% (95% CI 87%-93%) specificity, and an SROC-AUC of 0.95 (95% CI 0.92-0.96), whereas DL models outperformed with 91% (95% CI 89%-93%) sensitivity, 94% (95% CI 93%-96%) specificity, and SROC-AUC 0.97 (95% CI 0.96-0.99). The data analysis results are also shown in [Table 3](#).

Figure 2. The pooled sensitivities and specificities based on moderate-to-high-quality studies for structural magnetic resonance imaging [23-27,34, 41,43,45-49,53,55,57,60].

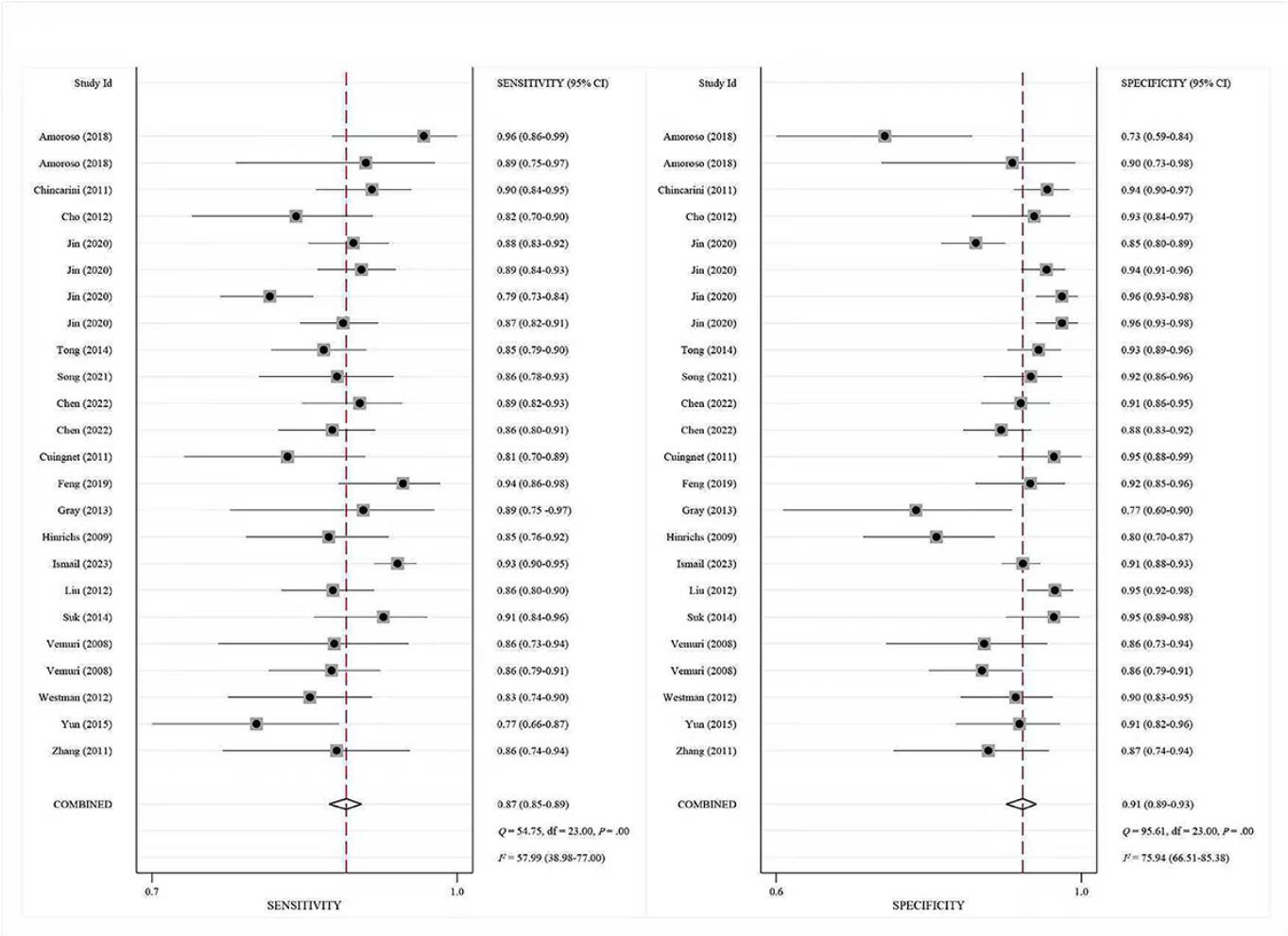


Figure 3. Summary receiver operating characteristic (SROC) based on moderate-to-high-quality studies curves for structural magnetic resonance imaging. AUC: area under the receiver operating characteristic curve area; SENS: sensitivity; SPEC: specificity.

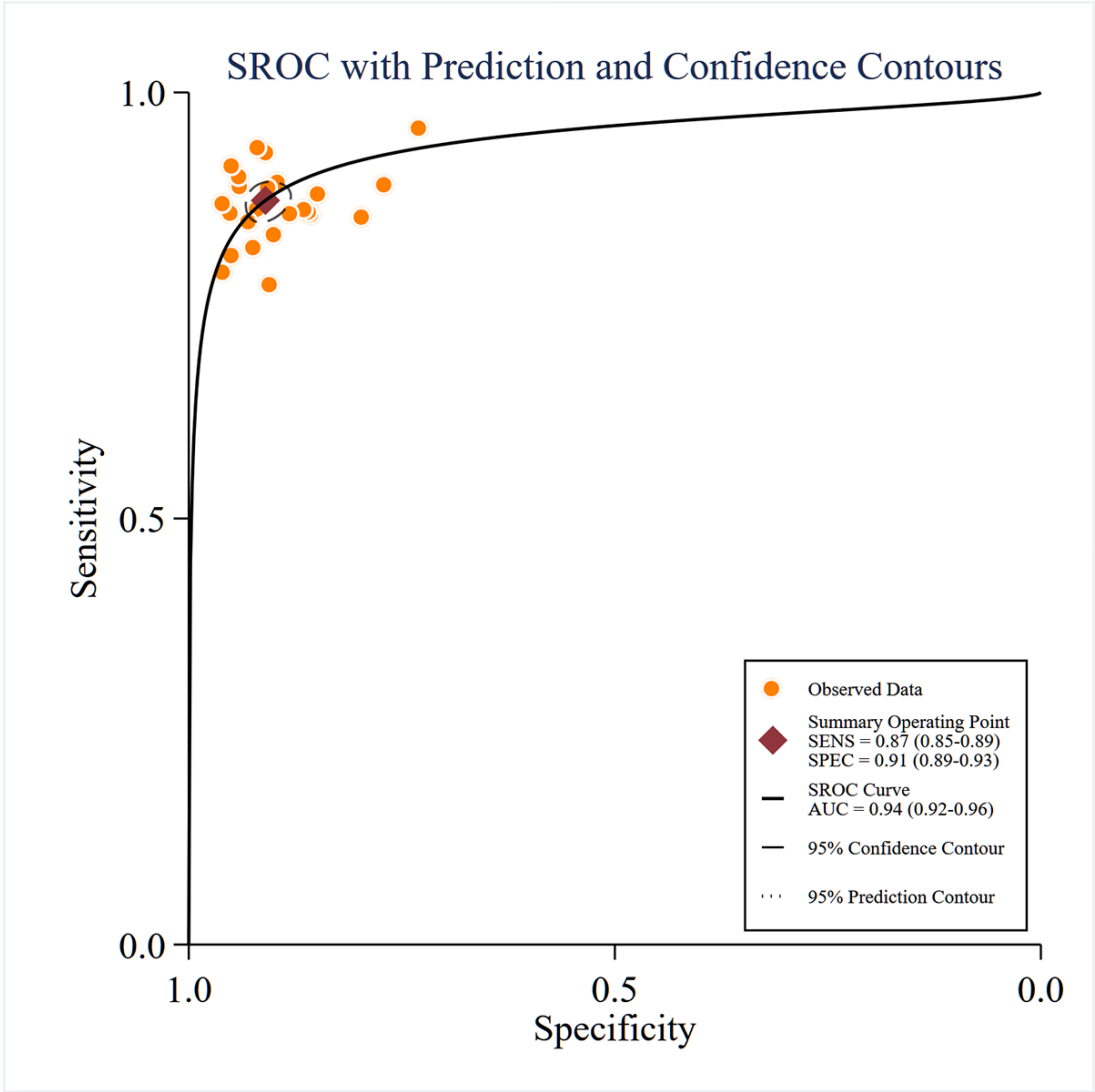


Figure 4. The pooled sensitivities and specificities based on moderate-to-high-quality studies for fluorine-18 fluorodeoxyglucose positron emission tomography (¹⁸F-FDG PET) [23,24,27,29,31,33,35,38,39,41-45,45,46,49,50].

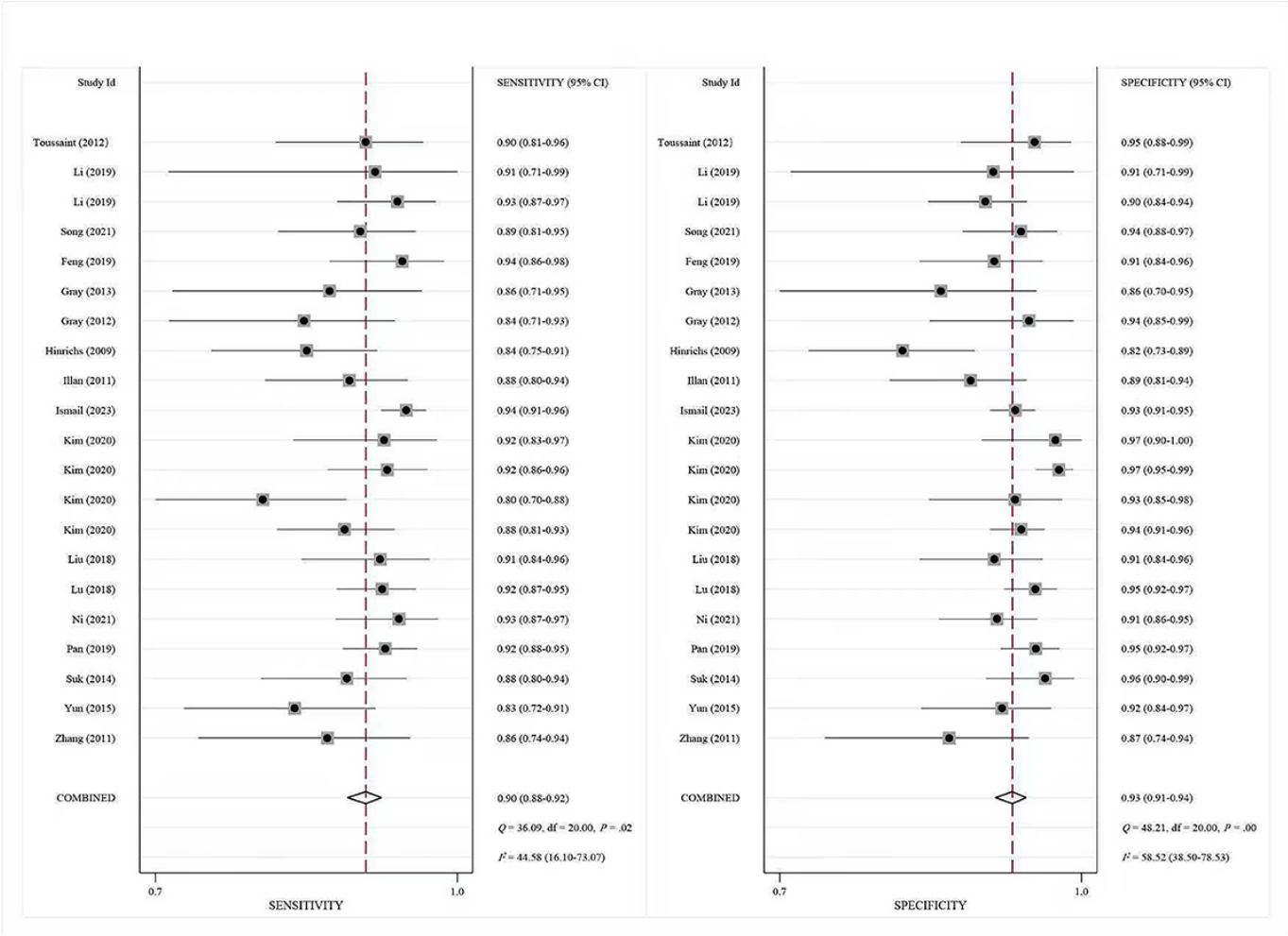


Figure 5. Summary receiver operating characteristic (SROC) based on moderate-to-high-quality studies curves for fluorine-18 fluorodeoxyglucose positron emission tomography (¹⁸F-FDG PET). AUC: area under the receiver operating characteristic curve area; SENS: sensitivity; SPEC: specificity.

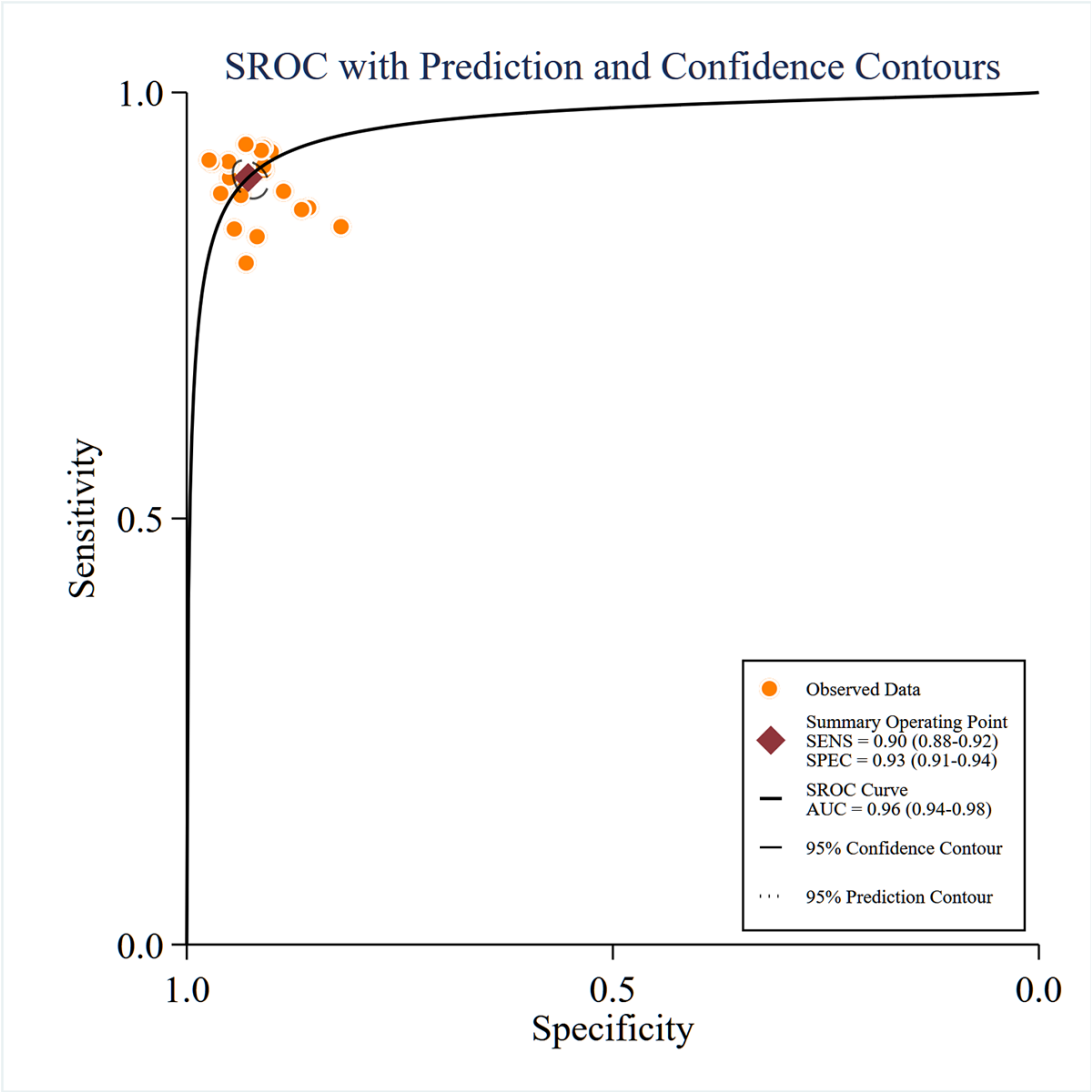


Table 3. Summary estimates and meta-regression of pooled performance of moderate-to-high-quality studies.

Main and subordinate directory	Tables, n	Sensitivity (%; 95% CI)	<i>I</i> ²	Specificity (%; 95% CI)	<i>I</i> ²	Joint <i>P</i> value	AUC ^a (95% CI)
Overall						.02	
sMRI ^b	24	87 (85-97)	57.99	91 (89-93)	75.94		0.94 (0.92-0.96)
¹⁸ F-FDG PET ^c	21	90 (88-92)	44.58	93 (91-94)	58.52		0.96 (0.94-0.98)
ML ^d							
sMRI	14	86 (84-88)	3.72	89 (86-92)	75.58		0.89 (0.86-0.92)
¹⁸ F-FDG PET	10	89 (86-91)	16.51	91 (87-93)	53.95		0.95 (0.92-0.96)
DL ^e							
sMRI	10	88 (86-91)	76.56	92 (90-94)	76.53		0.96 (0.94-0.97)
¹⁸ F-FDG PET	11	91 (89-93)	54.59	94 (93-95)	36.23		0.97 (0.96-0.99)

^aAUC: area under the receiver operating characteristic curve.

^bsMRI: structural magnetic resonance imaging.

^c¹⁸F-FDG PET: fluorine-18 fluorodeoxyglucose positron emission tomography.

^dML: machine learning.

^eDL: deep learning.

Exploration of Statistical Heterogeneity Sources

Full Study Cohort

For sMRI contingency tables, sensitivity exhibited moderate heterogeneity ($P=55.32\%$), while specificity showed high heterogeneity ($P=74.08\%$). Threshold effect testing revealed no significant threshold effect ($r=-0.101$, $P=.60$). Subgroup analyses identified study quality stratification ($P=.02$) as a source of heterogeneity, with moderate-to-high-quality studies demonstrating lower sensitivity (87% vs 91%) and specificity (91% vs 95%) compared to low-quality studies. Algorithm type ($P=.40$) and validation strategy ($P=.33$) were not significant contributors.

For ^{18}F -FDG PET analyses, sensitivity and specificity displayed moderate ($P=42.04\%$) and high heterogeneity ($P=61.49\%$), respectively, with no threshold effect ($r=-0.087$, $P=.67$). Algorithm type ($P=.01$) significantly influenced heterogeneity, as ML models demonstrated lower sensitivity (89% vs 91%) and specificity (91% vs 94%) than DL. Study quality ($P=.96$) and validation strategy ($P=.26$) showed no significant impact. The data analysis results are also shown in [Table 2](#).

Moderate-to-High-Quality Study Subgroups

For sMRI-based models, ML implementations demonstrated minimal sensitivity heterogeneity ($P=3.72\%$) but high specificity heterogeneity ($P=75.58\%$). Exclusion of studies using traditional ensemble algorithms (random forest, boosting) reduced specificity heterogeneity to moderate ($P=45.47\%$), revealing that ensemble methods achieved higher sensitivity (89% vs 87%) but significantly lower specificity (79% vs 93%) compared to nonensemble ML models ($P<.001$). In contrast, DL-based sMRI models exhibited high heterogeneity for both sensitivity ($P=76.56\%$) and specificity ($P=76.53\%$). Removing external validation data mitigated heterogeneity to moderate ($P=54.36\%$) and low levels ($P=38.60\%$), respectively, with external validation studies showing significantly reduced sensitivity (85% vs 90%) and specificity (91% vs 93%) compared to internal validation ($P<.001$).

For ^{18}F -FDG PET-based models, ML implementations showed low sensitivity heterogeneity ($P=16.51\%$) and moderate specificity heterogeneity ($P=53.95\%$). DL models exhibited moderate sensitivity heterogeneity ($P=54.59\%$) and low specificity heterogeneity ($P=36.23\%$). The data analysis results also are shown in [Table 3](#) and [Table S5](#) in [Multimedia Appendix 1](#).

Publication Bias Test

Funnel plot analysis of the full study cohort revealed no significant publication bias for sMRI ($P=.69$), whereas the ^{18}F -FDG PET subgroup exhibited significant bias ($P<.001$). In the moderate-to-high-quality cohort, both sMRI ($P=.03$) and ^{18}F -FDG PET ($P=.01$) demonstrated publication bias. However, subgroup analyses showed no significant bias for

sMRI+ML ($P=.06$), sMRI+DL ($P=.89$), ^{18}F -FDG PET+ML ($P=.08$), or ^{18}F -FDG PET+DL ($P=.28$).

Discussion

Principal Findings

This meta-analysis confirms that AI-enhanced sMRI and ^{18}F -FDG PET achieve high diagnostic accuracy for AD, with pooled SROC-AUCs of 0.94 and 0.96, respectively, outperforming conventional visual assessments [61,62]. ^{18}F -FDG PET demonstrated superior overall performance ($P=.02$), likely attributable to its sensitivity to AD-specific metabolic abnormalities. Notably, ML amplified PET's diagnostic advantage (SROC-AUC: 0.95 vs 0.89), while DL narrowed the gap (SROC-AUC: 0.97 vs 0.96), highlighting DL's capacity to extract complex metabolic features and mitigate structural imaging limitations [63].

Subgroup analyses identified three key determinants of heterogeneity and performance variation. First, methodological quality significantly influenced sMRI models: moderate-to-high-quality studies reported lower sensitivity (87% vs 91%) and specificity (91% vs 95%) compared to low-quality studies ($P=.02$), suggesting inflated performance in the latter due to small sample sizes, single-center data, or nonstandardized preprocessing. Adherence to TRIPOD-AI guidelines and transparent reporting of preprocessing workflows are critical for future studies [18]. Second, algorithm type drove technical divergence in PET models: DL outperformed ML (sensitivity: 91% vs 89%; specificity: 94% vs 91%; $P=.01$) through end-to-end feature learning. However, ML's interpretability better aligns with clinical demands for transparency, whereas DL's reliance on large annotated datasets and computing resources limits its deployment in resource-constrained regions [64]. Third, validation strategies exposed generalizability limitations: external validation of sMRI+DL models showed reduced sensitivity (85% vs 90%) and specificity (91% vs 93%) compared to internal validation ($P<.001$).

The reduced diagnostic accuracy observed in externally validated models may stem from (1) data distribution shift—differences in feature distributions, class balance, and temporal trends between training and external datasets; (2) overfitting—models may have overfitted to training-specific noise or spurious patterns, limiting generalization; and (3) implementation and annotation inconsistencies—variations in data preprocessing, feature scaling, and labeling protocols across datasets [65].

Future Directions and Study Limitations

This study has two methodological constraints: potential selection bias from English-only inclusion and possible overestimation by prioritizing optimal contingency tables. Future research should focus on enhancing evidence robustness through multinational, multiethnic cohorts; improving transparency via open-source preprocessing codes, findable, accessible, interoperable, and reusable-compliant data sharing, and independent validation; and

balancing performance and interpretability via explainable DL frameworks to meet clinical ethical standards [66].

We also acknowledge the risk of publication bias, as studies with positive results are more likely to be published, particularly in AI-related research where rapid progress and selective reporting may skew the literature. Additionally, language and geographic biases may exist since only English-language articles were included and most studies originated from a limited number of regions (eg, China, the United States, and Europe). This may limit the generalizability of our findings to underrepresented regions.

Despite ^{18}F -FDG PET's higher accuracy, its radiation exposure and cost hinder widespread screening [67]. Conversely, sMRI's cost-effectiveness and noninvasiveness position it as a first-line screening tool, with PET reserved for confirmatory testing in complex cases—a tiered diagnostic strategy. The 2023 Responsible AI for Social and Ethical Healthcare consensus implementation priorities, including transparent cost-benefit frameworks cross-modal standardization, and dynamic performance monitoring [68].

Comparison With Existing Reviews

Although previous reviews have explored the diagnostic accuracy of AI in AD, studies involving meta-analyses remain scarce [69-71]. Conducting such meta-analyses faces significant challenges, particularly due to substantial methodological heterogeneity across studies. This heterogeneity manifests in multiple dimensions, including variations in neuroimaging modalities, disparities in model validation strategies, and differences in algorithm types—all of which influence AD diagnostic performance and complicate the synthesis of evidence.

Borchert et al [70] conducted a comprehensive systematic review of 255 neuroimaging studies utilizing AI for dementia diagnosis and prognosis. Their findings demonstrated that discriminative models, particularly DL approaches, outperformed algorithmic classifiers in distinguishing AD patients from healthy controls. However, they emphasized critical methodological limitations, with conclusions primarily relying on qualitative synthesis rather than quantitative evidence.

In a systematic review and meta-analysis by Sun et al [71], the diagnostic accuracy of DL models based on ^{18}F -FDG PET for AD was investigated. While the study reported excellent diagnostic performance, notable heterogeneity was observed during meta-analysis, raising concerns about the reliability of the findings. Furthermore, the study focused exclusively on DL, overlooking the widespread application of traditional ML methods in current clinical research for AD diagnostic modeling.

In contrast, our meta-analysis incorporates a broader range of studies, rigorously controls methodological heterogeneity through stringent quality assessment and detailed subgroup analyses, and systematically evaluates the diagnostic accuracy of both ML and DL in AD. By emphasizing methodological rigor and the importance of external validation in AI-assisted neuroimaging for AD diagnosis, this study addresses critical gaps in the existing literature.

Although our subgroup comparisons between ML and DL models provide a useful overview of broad methodological trends, it must be noted that algorithm complexity, training data size, and model optimization procedures vary considerably within each group. The observed performance differences may, therefore, reflect not only model class but also differences in dataset size, feature representation, and implementation quality. Future studies should aim to compare individual algorithms under standardized conditions.

Conclusions

In conclusion, AI can effectively support the diagnosis of AD using sMRI and ^{18}F -FDG PET imaging. Among these approaches, combining PET imaging with DL techniques yields the highest diagnostic accuracy. These findings suggest that a future direction lies in integrating precision neuroimaging with AI tools. To bring such systems into routine clinical use—helping doctors detect AD earlier, personalize treatment, and improve patient outcomes—future studies should focus on repeated validation with high-quality clinical datasets and the development of standardized implementation protocols.

Acknowledgments

This research was funded by the Henan Province Medical Science and Technology Public Relations Plan Joint Construction Project (LHGJ20230402).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed database search process, characteristics of included studies, quality assessment results, and subgroup analysis findings.

[\[DOCX File \(Microsoft Word File\), 60 KB-Multimedia Appendix 1\]](#)

Checklist 1

PRISMA 2020 checklist.

[\[PDF File \(Adobe File\), 101 KB-Checklist 1\]](#)

References

1. 2023 Alzheimer's disease facts and figures. *Alzheimers Dement*. Apr 2023;19(4):1598-1695. [doi: [10.1002/alz.13016](https://doi.org/10.1002/alz.13016)]
2. Ren R, Qi J, Lin S, et al. The China Alzheimer Report 2022. *Gen Psychiatr*. 2022;35(1):e100751. [doi: [10.1136/gpsych-2022-100751](https://doi.org/10.1136/gpsych-2022-100751)] [Medline: [35372787](https://pubmed.ncbi.nlm.nih.gov/35372787/)]
3. Jia J, Wei C, Chen S, et al. The cost of Alzheimer's disease in China and re-estimation of costs worldwide. *Alzheimers Dement*. Apr 2018;14(4):483-491. [doi: [10.1016/j.jalz.2017.12.006](https://doi.org/10.1016/j.jalz.2017.12.006)] [Medline: [29433981](https://pubmed.ncbi.nlm.nih.gov/29433981/)]
4. Scheltens P, Blennow K, Breteler MMB, et al. Alzheimer's disease. *Lancet*. Jul 30, 2016;388(10043):505-517. [doi: [10.1016/S0140-6736\(15\)01124-1](https://doi.org/10.1016/S0140-6736(15)01124-1)] [Medline: [26921134](https://pubmed.ncbi.nlm.nih.gov/26921134/)]
5. Pickett AC, Valdez D, White LA, et al. The CareVirtue digital journal for family and friend caregivers of people living with Alzheimer disease and related dementias: exploratory topic modeling and user engagement study. *JMIR Aging*. Dec 24, 2024;7:e67992. [doi: [10.2196/67992](https://doi.org/10.2196/67992)] [Medline: [39719081](https://pubmed.ncbi.nlm.nih.gov/39719081/)]
6. Cahill S. WHO's global action plan on the public health response to dementia: some challenges and opportunities. *Aging Ment Health*. Feb 2020;24(2):197-199. [doi: [10.1080/13607863.2018.1544213](https://doi.org/10.1080/13607863.2018.1544213)] [Medline: [30600688](https://pubmed.ncbi.nlm.nih.gov/30600688/)]
7. Wu Y, Fu L, Li Q, et al. Recent advancements in the early diagnosis and treatment of Alzheimer's disease. *Adv Therap*. Nov 2023;6(11). URL: <https://onlinelibrary.wiley.com/toc/23663987/6/11> [doi: [10.1002/adtp.202300181](https://doi.org/10.1002/adtp.202300181)] [Medline: [36818419](https://pubmed.ncbi.nlm.nih.gov/36818419/)]
8. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*. Apr 2018;14(4):535-562. [doi: [10.1016/j.jalz.2018.02.018](https://doi.org/10.1016/j.jalz.2018.02.018)]
9. Jack CR, Andrews JS, Beach TG, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*. Aug 2024;20(8):5143-5169. [doi: [10.1002/alz.13859](https://doi.org/10.1002/alz.13859)]
10. By S, Kahl A, Cogswell PM. Alzheimer's disease clinical trials: what have we learned from magnetic resonance imaging. *J Magn Reson Imaging*. Feb 2025;61(2):579-594. [doi: [10.1002/jmri.29462](https://doi.org/10.1002/jmri.29462)] [Medline: [39031716](https://pubmed.ncbi.nlm.nih.gov/39031716/)]
11. Kuo RYL, Harrison C, Curran TA, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. *Radiology*. Jul 2022;304(1):50-62. [doi: [10.1148/radiol.211785](https://doi.org/10.1148/radiol.211785)] [Medline: [35348381](https://pubmed.ncbi.nlm.nih.gov/35348381/)]
12. Wang J, Xue L, Jiang J, et al. Diagnostic performance of artificial intelligence-assisted PET imaging for Parkinson's disease: a systematic review and meta-analysis. *NPJ Digit Med*. Jan 22, 2024;7(1):17. [doi: [10.1038/s41746-024-01012-z](https://doi.org/10.1038/s41746-024-01012-z)] [Medline: [38253738](https://pubmed.ncbi.nlm.nih.gov/38253738/)]
13. Cheng Y, Malekar M, He Y, et al. High-throughput phenotyping of the symptoms of Alzheimer disease and related dementias using large language models: cross-sectional study. *JMIR AI*. Jun 3, 2025;4:e66926. [doi: [10.2196/66926](https://doi.org/10.2196/66926)] [Medline: [40460418](https://pubmed.ncbi.nlm.nih.gov/40460418/)]
14. Bosco C, Shojaei F, Theisz AA, et al. Testing 3 modalities (voice assistant, chatbot, and mobile app) to assist older African American and Black adults in seeking information on Alzheimer disease and related dementias: wizard of Oz usability study. *JMIR Form Res*. Dec 9, 2024;8:e60650. [doi: [10.2196/60650](https://doi.org/10.2196/60650)] [Medline: [39653372](https://pubmed.ncbi.nlm.nih.gov/39653372/)]
15. Rudroff T, Rainio O, Klén R. AI for the prediction of early stages of Alzheimer's disease from neuroimaging biomarkers —a narrative review of a growing field. *Neurol Sci*. Nov 2024;45(11):5117-5127. [doi: [10.1007/s10072-024-07649-8](https://doi.org/10.1007/s10072-024-07649-8)] [Medline: [38866971](https://pubmed.ncbi.nlm.nih.gov/38866971/)]
16. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods*. Sep 2017;8(3):290-302. [doi: [10.1002/jrsm.1240](https://doi.org/10.1002/jrsm.1240)] [Medline: [28378395](https://pubmed.ncbi.nlm.nih.gov/28378395/)]
17. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. Jan 23, 2018;319(4):388-396. [doi: [10.1001/jama.2017.19163](https://doi.org/10.1001/jama.2017.19163)] [Medline: [29362800](https://pubmed.ncbi.nlm.nih.gov/29362800/)]
18. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Apr 2024;q902. [doi: [10.1136/bmj.q902](https://doi.org/10.1136/bmj.q902)]
19. Singh S, Chang SM, Matchar DB, Bass EB. Chapter 7: grading a body of evidence on diagnostic tests. *J Gen Intern Med*. Jun 2012;27 Suppl 1(Suppl 1):S47-55. [doi: [10.1007/s11606-012-2021-9](https://doi.org/10.1007/s11606-012-2021-9)] [Medline: [22648675](https://pubmed.ncbi.nlm.nih.gov/22648675/)]
20. Ruppert T. Meta-analysis: how to quantify and explain heterogeneity? *Eur J Cardiovasc Nurs*. Oct 2020;19(7):646-652. [doi: [10.1177/1474515120944014](https://doi.org/10.1177/1474515120944014)] [Medline: [32757621](https://pubmed.ncbi.nlm.nih.gov/32757621/)]
21. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. Oct 15, 2001;20(19):2865-2884. [doi: [10.1002/sim.942](https://doi.org/10.1002/sim.942)] [Medline: [11568945](https://pubmed.ncbi.nlm.nih.gov/11568945/)]
22. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. Oct 2005;58(10):982-990. [doi: [10.1016/j.jclinepi.2005.02.022](https://doi.org/10.1016/j.jclinepi.2005.02.022)] [Medline: [16168343](https://pubmed.ncbi.nlm.nih.gov/16168343/)]
23. Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. Apr 1, 2011;55(3):856-867. [doi: [10.1016/j.neuroimage.2011.01.008](https://doi.org/10.1016/j.neuroimage.2011.01.008)] [Medline: [21236349](https://pubmed.ncbi.nlm.nih.gov/21236349/)]

24. Yun HJ, Kwak K, Lee JM, Alzheimer's Disease Neuroimaging Initiative. Multimodal discrimination of Alzheimer's disease based on regional cortical atrophy and hypometabolism. PLoS ONE. 2015;10(6):e0129250. [doi: [10.1371/journal.pone.0129250](https://doi.org/10.1371/journal.pone.0129250)] [Medline: [26061669](https://pubmed.ncbi.nlm.nih.gov/26061669/)]
25. Westman E, Muehlboeck JS, Simmons A. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage. Aug 1, 2012;62(1):229-238. [doi: [10.1016/j.neuroimage.2012.04.056](https://doi.org/10.1016/j.neuroimage.2012.04.056)] [Medline: [22580170](https://pubmed.ncbi.nlm.nih.gov/22580170/)]
26. Vemuri P, Gunter JL, Senjem ML, et al. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage. Feb 1, 2008;39(3):1186-1197. [doi: [10.1016/j.neuroimage.2007.09.073](https://doi.org/10.1016/j.neuroimage.2007.09.073)] [Medline: [18054253](https://pubmed.ncbi.nlm.nih.gov/18054253/)]
27. Suk HI, Lee SW, Shen D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage. Nov 1, 2014;101:569-582. [doi: [10.1016/j.neuroimage.2014.06.077](https://doi.org/10.1016/j.neuroimage.2014.06.077)] [Medline: [25042445](https://pubmed.ncbi.nlm.nih.gov/25042445/)]
28. Sayeed A, Petrou M, Spyrou N, Kadyrov A, Spinks T. Diagnostic features of Alzheimer's disease extracted from PET sinograms. Phys Med Biol. Jan 7, 2002;47(1):137-148. [doi: [10.1088/0031-9155/47/1/310](https://doi.org/10.1088/0031-9155/47/1/310)] [Medline: [11814222](https://pubmed.ncbi.nlm.nih.gov/11814222/)]
29. Pan X, Adel M, Fossati C, et al. Multiscale spatial gradient features for ¹⁸F-FDG PET image-guided diagnosis of Alzheimer's disease. Comput Methods Programs Biomed. Oct 2019;180:105027. [doi: [10.1016/j.cmpb.2019.105027](https://doi.org/10.1016/j.cmpb.2019.105027)] [Medline: [31430595](https://pubmed.ncbi.nlm.nih.gov/31430595/)]
30. Padilla P, López M, Górriz JM, et al. NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease. IEEE Trans Med Imaging. Feb 2012;31(2):207-216. [doi: [10.1109/TMI.2011.2167628](https://doi.org/10.1109/TMI.2011.2167628)] [Medline: [21914569](https://pubmed.ncbi.nlm.nih.gov/21914569/)]
31. Ni YC, Tseng FP, Pai MC, et al. Detection of Alzheimer's disease using ECD SPECT images by transfer learning from FDG PET. Ann Nucl Med. Aug 2021;35(8):889-899. [doi: [10.1007/s12149-021-01626-3](https://doi.org/10.1007/s12149-021-01626-3)] [Medline: [34076857](https://pubmed.ncbi.nlm.nih.gov/34076857/)]
32. Magnin B, Mesrob L, Kinkingnéhun S, et al. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology. Feb 2009;51(2):73-83. [doi: [10.1007/s00234-008-0463-x](https://doi.org/10.1007/s00234-008-0463-x)] [Medline: [18846369](https://pubmed.ncbi.nlm.nih.gov/18846369/)]
33. Lu DH, Popuri K, Ding GW, Balachandar R, Beg MF. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Med Image Anal. May 2018;46:26-34. [doi: [10.1016/j.media.2018.02.002](https://doi.org/10.1016/j.media.2018.02.002)]
34. Liu MH, Zhang DQ, Shen DG. Ensemble sparse classification of Alzheimer's disease. Neuroimage. Apr 2012;60(2):1106-1116. [doi: [10.1016/j.neuroimage.2012.01.055](https://doi.org/10.1016/j.neuroimage.2012.01.055)]
35. Liu M, Cheng D, Yan W, Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front Neuroinform. 2018;12:35. [doi: [10.3389/fninf.2018.00035](https://doi.org/10.3389/fninf.2018.00035)] [Medline: [29970996](https://pubmed.ncbi.nlm.nih.gov/29970996/)]
36. Li R, Perneczky R, Yakushev I, et al. Gaussian mixture models and model selection for [18F] fluorodeoxyglucose positron emission tomography classification in Alzheimer's disease. PLoS ONE. Apr 2015;10(4):e0122731. [doi: [10.1371/journal.pone.0122731](https://doi.org/10.1371/journal.pone.0122731)]
37. Lerch JP, Pruessner J, Zijdenbos AP, et al. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging. Jan 2008;29(1):23-30. [doi: [10.1016/j.neurobiolaging.2006.09.013](https://doi.org/10.1016/j.neurobiolaging.2006.09.013)] [Medline: [17097767](https://pubmed.ncbi.nlm.nih.gov/17097767/)]
38. Kim HW, Lee HE, Oh K, Lee S, Yun M, Yoo SK. Multi-slice representational learning of convolutional neural network for Alzheimer's disease classification using positron emission tomography. Biomed Eng Online. Sep 7, 2020;19(1):70. [doi: [10.1186/s12938-020-00813-z](https://doi.org/10.1186/s12938-020-00813-z)] [Medline: [32894137](https://pubmed.ncbi.nlm.nih.gov/32894137/)]
39. Kim HW, Lee HE, Lee S, Oh KT, Yun M, Yoo SK. Slice-selective learning for Alzheimer's disease classification using a generative adversarial network: a feasibility study of external validation. Eur J Nucl Med Mol Imaging. Aug 2020;47(9):2197-2206. [doi: [10.1007/s00259-019-04676-y](https://doi.org/10.1007/s00259-019-04676-y)] [Medline: [31980910](https://pubmed.ncbi.nlm.nih.gov/31980910/)]
40. Katako A, Shelton P, Goertzen AL, et al. Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in Lewy body dementia and Parkinson's disease dementia. Sci Rep. Sep 5, 2018;8(1):13236. [doi: [10.1038/s41598-018-31653-6](https://doi.org/10.1038/s41598-018-31653-6)] [Medline: [30185806](https://pubmed.ncbi.nlm.nih.gov/30185806/)]
41. Ismail WN, P. P. FR, Ali MAS. A meta-heuristic multi-objective optimization method for Alzheimer's disease detection based on multi-modal data. Mathematics. Feb 2023;11(4):957. [doi: [10.3390/math11040957](https://doi.org/10.3390/math11040957)]
42. Illán IA, Górriz JM, Ramírez J, et al. 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. Inf Sci (Ny). Feb 2011;181(4):903-916. [doi: [10.1016/j.ins.2010.10.027](https://doi.org/10.1016/j.ins.2010.10.027)]
43. Hinrichs C, Singh V, Mukherjee L, et al. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage. Oct 15, 2009;48(1):138-149. [doi: [10.1016/j.neuroimage.2009.05.056](https://doi.org/10.1016/j.neuroimage.2009.05.056)] [Medline: [19481161](https://pubmed.ncbi.nlm.nih.gov/19481161/)]
44. Gray KR, Wolz R, Heckemann RA, et al. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage. Mar 2012;60(1):221-229. [doi: [10.1016/j.neuroimage.2011.12.071](https://doi.org/10.1016/j.neuroimage.2011.12.071)] [Medline: [22236449](https://pubmed.ncbi.nlm.nih.gov/22236449/)]

45. Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage*. Jan 15, 2013;65:167-175. [doi: [10.1016/j.neuroimage.2012.09.065](https://doi.org/10.1016/j.neuroimage.2012.09.065)] [Medline: [23041336](https://pubmed.ncbi.nlm.nih.gov/23041336/)]
46. Feng C, Elazab A, Yang P, et al. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. *IEEE Access*. 2019;7:63605-63618. [doi: [10.1109/ACCESS.2019.2913847](https://doi.org/10.1109/ACCESS.2019.2913847)]
47. Cuingnet R, Gerardin E, Tessieras J, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*. May 15, 2011;56(2):766-781. [doi: [10.1016/j.neuroimage.2010.06.013](https://doi.org/10.1016/j.neuroimage.2010.06.013)] [Medline: [20542124](https://pubmed.ncbi.nlm.nih.gov/20542124/)]
48. Chen L, Qiao HZ, Zhu F. Alzheimer's disease diagnosis with brain structural MRI using multiview-slice attention and 3D convolution neural network. *Front Aging Neurosci*. 2022;14:871706. [doi: [10.3389/fnagi.2022.871706](https://doi.org/10.3389/fnagi.2022.871706)] [Medline: [35557839](https://pubmed.ncbi.nlm.nih.gov/35557839/)]
49. Song J, Zheng J, Li P, Lu X, Zhu G, Shen P. An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. *Front Digit Health*. 2021;3:637386. [doi: [10.3389/fdgth.2021.637386](https://doi.org/10.3389/fdgth.2021.637386)] [Medline: [34713109](https://pubmed.ncbi.nlm.nih.gov/34713109/)]
50. Li Y, Jiang J, Lu J, Jiang J, Zhang H, Zuo C. Radiomics: a novel feature extraction method for brain neuron degeneration disease using ¹⁸F-FDG PET imaging and its implementation for Alzheimer's disease and mild cognitive impairment. *Ther Adv Neurol Disord*. 2019;12:1756286419838682. [doi: [10.1177/1756286419838682](https://doi.org/10.1177/1756286419838682)] [Medline: [30956687](https://pubmed.ncbi.nlm.nih.gov/30956687/)]
51. A A, M P, Hamdi M, Bourouis S, Rastislav K, Mohamed F. Evaluation of neuro images for the diagnosis of Alzheimer's disease using deep learning neural network. *Front Public Health*. 2022;10:834032. [doi: [10.3389/fpubh.2022.834032](https://doi.org/10.3389/fpubh.2022.834032)] [Medline: [35198526](https://pubmed.ncbi.nlm.nih.gov/35198526/)]
52. Toussaint PJ, Perlberg V, Bellec P, et al. Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. *Neuroimage*. Nov 1, 2012;63(2):936-946. [doi: [10.1016/j.neuroimage.2012.03.091](https://doi.org/10.1016/j.neuroimage.2012.03.091)] [Medline: [22510256](https://pubmed.ncbi.nlm.nih.gov/22510256/)]
53. Tong T, Wolz R, Gao Q, et al. Multiple instance learning for classification of dementia in brain MRI. *Med Image Anal*. Jul 2014;18(5):808-818. [doi: [10.1016/j.media.2014.04.006](https://doi.org/10.1016/j.media.2014.04.006)] [Medline: [24858570](https://pubmed.ncbi.nlm.nih.gov/24858570/)]
54. Min R, Wu G, Cheng J, Wang Q, Shen D. Multi-atlas based representations for Alzheimer's disease diagnosis. *Hum Brain Mapp*. Oct 2014;35(10):5052-5070. [doi: [10.1002/hbm.22531](https://doi.org/10.1002/hbm.22531)] [Medline: [24753060](https://pubmed.ncbi.nlm.nih.gov/24753060/)]
55. Jin D, Zhou B, Han Y, et al. Generalizable, reproducible, and neuroscientifically interpretable imaging biomarkers for Alzheimer's disease. *Adv Sci (Weinh)*. Jul 2020;7(14):2000675. [doi: [10.1002/advs.202000675](https://doi.org/10.1002/advs.202000675)] [Medline: [32714766](https://pubmed.ncbi.nlm.nih.gov/32714766/)]
56. Cho Y, Seong JK, Jeong Y, Shin SY. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage*. Feb 1, 2012;59(3):2217-2230. [doi: [10.1016/j.neuroimage.2011.09.085](https://doi.org/10.1016/j.neuroimage.2011.09.085)] [Medline: [22008371](https://pubmed.ncbi.nlm.nih.gov/22008371/)]
57. Chincarini A, Bosco P, Calvini P, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage*. Sep 15, 2011;58(2):469-480. [doi: [10.1016/j.neuroimage.2011.05.083](https://doi.org/10.1016/j.neuroimage.2011.05.083)] [Medline: [21718788](https://pubmed.ncbi.nlm.nih.gov/21718788/)]
58. Beheshti I, Demirel H, Matsuda H. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput Biol Med*. Apr 1, 2017;83(109-19):109-119. [doi: [10.1016/j.combiomed.2017.02.011](https://doi.org/10.1016/j.combiomed.2017.02.011)] [Medline: [28260614](https://pubmed.ncbi.nlm.nih.gov/28260614/)]
59. Anandh KR, Sujatha CM, Ramakrishnan S. A method to differentiate mild cognitive impairment and Alzheimer in MR images using eigen value descriptors. *J Med Syst*. Jan 2016;40(1):26547845. [doi: [10.1007/s10916-015-0396-y](https://doi.org/10.1007/s10916-015-0396-y)] [Medline: [26547845](https://pubmed.ncbi.nlm.nih.gov/26547845/)]
60. Amoroso N, La Rocca M, Bruno S, et al. Multiplex networks for early diagnosis of Alzheimer's disease. *Front Aging Neurosci*. 2018;10:365. [doi: [10.3389/fnagi.2018.00365](https://doi.org/10.3389/fnagi.2018.00365)] [Medline: [30487745](https://pubmed.ncbi.nlm.nih.gov/30487745/)]
61. Yamane T, Ikari Y, Nishio T, et al. Visual-statistical interpretation of (18)F-FDG-PET images for characteristic Alzheimer patterns in a multicenter study: inter-rater concordance and relationship to automated quantitative evaluation. *AJNR Am J Neuroradiol*. Feb 2014;35(2):244-249. [doi: [10.3174/ajnr.A3665](https://doi.org/10.3174/ajnr.A3665)] [Medline: [23907243](https://pubmed.ncbi.nlm.nih.gov/23907243/)]
62. Harper L, Fumagalli GG, Barkhof F, et al. MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain (Bacau)*. Apr 2016;139(Pt 4):1211-1225. [doi: [10.1093/brain/aww005](https://doi.org/10.1093/brain/aww005)] [Medline: [26936938](https://pubmed.ncbi.nlm.nih.gov/26936938/)]
63. Loddo A, Buttau S, Di Ruberto C. Deep learning based pipelines for Alzheimer's disease diagnosis: a comparative study and a novel deep-ensemble method. *Comput Biol Med*. Feb 2022;141:105032. [doi: [10.1016/j.combiomed.2021.105032](https://doi.org/10.1016/j.combiomed.2021.105032)] [Medline: [34838263](https://pubmed.ncbi.nlm.nih.gov/34838263/)]
64. Pawar U, O'Shea D, Rea S, O'Reilly R. Explainable AI in healthcare. Presented at: 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA); Jun 15-19, 2020; Dublin, Ireland. [doi: [10.1109/CyberSA49311.2020.9139655](https://doi.org/10.1109/CyberSA49311.2020.9139655)]

65. Birkenbihl C, Emon MA, Vrooman H, et al. Differences in cohort study data affect external validation of artificial intelligence models for predictive diagnostics of dementia - lessons for translation into clinical practice. *EPMA J.* Sep 2020;11(3):367-376. [doi: [10.1007/s13167-020-00216-z](https://doi.org/10.1007/s13167-020-00216-z)] [Medline: [32843907](https://pubmed.ncbi.nlm.nih.gov/32843907/)]
66. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion.* Jun 2020;58:82-115. [doi: [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012)]
67. Mayerhoefer ME, Prosch H, Beer L, et al. PET/MRI versus PET/CT in oncology: a prospective single-center study of 330 examinations focusing on implications for patient management and cost considerations. *Eur J Nucl Med Mol Imaging.* Jan 2020;47(1):51-60. [doi: [10.1007/s00259-019-04452-y](https://doi.org/10.1007/s00259-019-04452-y)] [Medline: [31410538](https://pubmed.ncbi.nlm.nih.gov/31410538/)]
68. Goldberg CB, Adams L, Blumenthal D, et al. To do no harm - and the most good - with AI in health care. *Nat Med.* Mar 2024;30(3):623-627. [doi: [10.1038/s41591-024-02853-7](https://doi.org/10.1038/s41591-024-02853-7)] [Medline: [38388841](https://pubmed.ncbi.nlm.nih.gov/38388841/)]
69. Frizzell TO, Glashutter M, Liu CC, et al. Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: A systematic review. *Ageing Res Rev.* May 2022;77:101614. [doi: [10.1016/j.arr.2022.101614](https://doi.org/10.1016/j.arr.2022.101614)] [Medline: [35358720](https://pubmed.ncbi.nlm.nih.gov/35358720/)]
70. Borchert RJ, Azevedo T, Badhwar A, et al. Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. *Alzheimers Dement.* Dec 2023;19(12):5885-5904. [doi: [10.1002/alz.13412](https://doi.org/10.1002/alz.13412)] [Medline: [37563912](https://pubmed.ncbi.nlm.nih.gov/37563912/)]
71. Sun Y, Chen Y, Dong L, et al. Diagnostic performance of deep learning-assisted [¹⁸F]FDG PET imaging for Alzheimer's disease: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* Aug 2025;52(10):3600-3612. [doi: [10.1007/s00259-025-07228-9](https://doi.org/10.1007/s00259-025-07228-9)] [Medline: [40159544](https://pubmed.ncbi.nlm.nih.gov/40159544/)]

Abbreviations

¹⁸FDG-PET: fluorine-18 fluorodeoxyglucose positron emission tomography

AD: Alzheimer disease

AI: artificial intelligence

DL: deep learning

MeSH: Medical Subject Headings

ML: machine learning

PICOS: Population, Intervention, Comparison, Outcome, Study Design

PRISMA-DTA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy

sMRI: structural magnetic resonance imaging

SROC-AUC: summary receiver operating characteristic curve area

TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence

Edited by Megan O'Connell; peer-reviewed by Ravi Teja Potla, Sadhasivam Mohanadas, Seyed Ali Mirshahvalad; submitted 05.05.2025; final revised version received 03.07.2025; accepted 20.07.2025; published 08.10.2025

Please cite as:

Wang B, Zhao T, Ma R, Huo X, Xiong X, Wu M, Wang Y, Liu L, Zhuang Z, Wang B, Shou J

Comparative Diagnostic Accuracy of AI-Assisted Fluorine-18 Fluorodeoxyglucose Positron Emission Tomography Versus Structural Magnetic Resonance Imaging in Alzheimer Disease: Systematic Review and Meta-Analysis

JMIR Aging 2025;8:e76981

URL: <https://aging.jmir.org/2025/1/e76981>

doi: [10.2196/76981](https://doi.org/10.2196/76981)

© Bingbing Wang, Tailiang Zhao, Rongrong Ma, Xiaochuan Huo, Xiaoxiao Xiong, Minjie Wu, Yuran Wang, Liu Liu, Zhijiang Zhuang, Bin Wang, Jixin Shou. Originally published in *JMIR Aging* (<https://aging.jmir.org>), 08.10.2025. 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 *JMIR Aging*, is properly cited. The complete bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.