



# Revolutionizing early Alzheimer's disease and mild cognitive impairment diagnosis: a deep learning MRI meta-analysis

## *Revolucionando a doença de Alzheimer precoce e o diagnóstico de comprometimento cognitivo leve: uma metanálise de aprendizagem profunda por ressonância magnética*

Li-xue Wang<sup>1,2</sup> Yi-zhe Wang<sup>1,2</sup> Chen-guang Han<sup>2,3</sup> Lei Zhao<sup>1,2</sup> Li He<sup>1,2</sup> Jie Li<sup>1,2</sup>

<sup>1</sup>Beijing Tsinghua Changgung Hospital, Department of Radiology, Beijing, China.

Address for correspondence Jie Li, (email: Jieli\_btch@163.com).

<sup>2</sup>Tsinghua University, School of Clinical Medicine, Beijing, China.

<sup>3</sup>Beijing Tsinghua Changgung Hospital, Department of Information Administration, Beijing, China.

Arq. Neuro-Psiquiatr. 2024;82(8):s00441788657.

### Abstract

**Background** The early diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI) remains a significant challenge in neurology, with conventional methods often limited by subjectivity and variability in interpretation. Integrating deep learning with artificial intelligence (AI) in magnetic resonance imaging (MRI) analysis emerges as a transformative approach, offering the potential for unbiased, highly accurate diagnostic insights.

**Objective** A meta-analysis was designed to analyze the diagnostic accuracy of deep learning of MRI images on AD and MCI models.

**Methods** A meta-analysis was performed across PubMed, Embase, and Cochrane library databases following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, focusing on the diagnostic accuracy of deep learning. Subsequently, methodological quality was assessed using the QUADAS-2 checklist. Diagnostic measures, including sensitivity, specificity, likelihood ratios, diagnostic odds ratio, and area under the receiver operating characteristic curve (AUROC) were analyzed, alongside subgroup analyses for T1-weighted and non-T1-weighted MRI.

**Results** A total of 18 eligible studies were identified. The Spearman correlation coefficient was -0.6506. Meta-analysis showed that the combined sensitivity and specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio were 0.84, 0.86, 6.0, 0.19, and 32, respectively. The AUROC was 0.92. The quiescent point of hierarchical summary of receiver operating characteristic (HSROC)

### Keywords

- ▶ Alzheimer Disease
- ▶ Cognitive Dysfunction
- ▶ Magnetic Resonance Imaging
- ▶ Deep Learning
- ▶ Meta-Analysis

received  
February 22, 2024  
received in its final form  
April 24, 2024  
accepted  
May 6, 2024

DOI <https://doi.org/10.1055/s-0044-1788657>.  
ISSN 0004-282X.

Editor-in-Chief: Hélio A. G. Teive.  
Associate Editor: Sonia Maria Dozzi Brucki.

© 2024. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution 4.0 International License, permitting copying and reproduction so long as the original work is given appropriate credit (<https://creativecommons.org/licenses/by/4.0/>).  
Thieme Revinter Publicações Ltda., Rua do Matoso 170, Rio de Janeiro, RJ, CEP 20270-135, Brazil

was 3.463. Notably, the images of 12 studies were acquired by T1-weighted MRI alone, and those of the other 6 were gathered by non-T1-weighted MRI alone.

**Conclusion** Overall, deep learning of MRI for the diagnosis of AD and MCI showed good sensitivity and specificity and contributed to improving diagnostic accuracy.

## Resumo

**Antecedentes** O diagnóstico precoce da doença de Alzheimer (DA) e do comprometimento cognitivo leve (CCL) continua sendo um desafio significativo na neurologia, com métodos convencionais frequentemente limitados pela subjetividade e variabilidade na interpretação. A integração da aprendizagem profunda com a inteligência artificial (IA) na análise de imagens de ressonância magnética surge como uma abordagem transformadora, oferecendo o potencial para *insights* diagnósticos imparciais e altamente precisos.

**Objetivo** Uma metanálise foi projetada para analisar a precisão diagnóstica do aprendizado profundo de imagens de ressonância magnética em modelos de DA e CCL.

**Métodos** Uma metanálise foi realizada nos bancos de dados das bibliotecas PubMed, Embase e Cochrane seguindo as diretrizes *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA), com foco na precisão diagnóstica do aprendizado profundo. Posteriormente, a qualidade metodológica foi avaliada por meio do *checklist* QUADAS-2. Medidas diagnósticas, incluindo sensibilidade, especificidade, razões de verossimilhança, razão de chances diagnósticas e área sob a curva característica de operação do receptor (*area under the receiver operating characteristic curve* [AUROC]) foram analisadas, juntamente com análises de subgrupo para ressonância magnética ponderada em T1 e não ponderada em T1.

**Resultados** Um total de 18 estudos elegíveis foram identificados. O coeficiente de correlação de Spearman foi de -0,6506. A metanálise mostrou que a sensibilidade e a especificidade combinadas, a razão de verossimilhança positiva, a razão de verossimilhança negativa e a razão de chances de diagnóstico foram 0,84, 0,86, 6,0, 0,19 e 32, respectivamente. A AUROC foi de 0,92. O ponto quiescente do resumo hierárquico da característica de operação do receptor (*hierarchical summary of receiver operating characteristic* [HSROC]) foi 3,463. Notavelmente, as imagens de 12 estudos foram adquiridas apenas por ressonância magnética ponderada em T1, e as dos outros 6 foram obtidas apenas por ressonância magnética não ponderada em T1.

**Conclusão** Em geral, a aprendizagem profunda da ressonância magnética para o diagnóstico de DA e CCL mostrou boa sensibilidade e especificidade e contribuiu para melhorar a precisão diagnóstica.

## Palavras-chave

- ▶ Doença de Alzheimer
- ▶ Disfunção Cognitiva
- ▶ Imageamento por Ressonância Magnética
- ▶ Aprendizado Profundo
- ▶ Metanálise

## INTRODUCTION

A study has shown that Alzheimer's disease (AD), the most common neurodegenerative cause in patients with dementia, is primarily characterized by a decline in brain function in multiple areas, including memory, reasoning, and language.<sup>1</sup> Patients with AD account for 50–70% of all patients with neurodegenerative dementia. Unfortunately, with the trend towards an aging population, the number of patients with AD is expected to surge. Xia, P. et al. investigated that there were approximately 6.08 million cases of AD in the United States in 2017, but the number is expected to reach 1.5 billion by 2060.<sup>2</sup> Besides, according to Klyucherev, T. O. et al., the healthcare expenditure on caring for dementia patients was estimated to be \$700 million in the United States in

2020, and the economic cost spent on AD patients exceeded the cost of cancer or cardiovascular disease,<sup>3</sup> which brought great trouble to mankind. Based on previous studies, it is known that AD is a complex, heterogeneous, and progressive disease. The predominant molecular mechanism of AD is the formation of toxic amyloid- $\beta$  oligomers and protein aggregates, as well as the formation of neurofibrillary tangles composed of Tau Protein Hyperphosphorylation, thereby leading to region-specific reduction of brain glucose metabolism synaptic dysfunction, and mitochondrial dysfunction.<sup>4</sup> There are 4 main stages in the development of AD, including the presymptomatic stage, the prodromal stage of mild cognitive impairment (MCI), and the clinical form of AD. Although the efficacy of pharmacological treatments for AD is unsatisfactory, cognitive and physical activity treatments

in the early stages of AD, such as the MCI period, play a positive role in reducing cognitive decline.<sup>5</sup> Therefore, researchers have concluded that early detection of brain changes associated with AD is the key to more effective clinical interventions and prevention of disease progression and morbidity.<sup>6</sup> Hence, early diagnosis is particularly important for AD patients.

Currently, imaging modalities are used to identify the early diagnostic and prognostic factors of AD in clinical practice. Among them, magnetic resonance imaging (MRI) is an essential method often used to explore the neuropathological mechanisms and clinical diagnosis of AD and MCI.<sup>7</sup> MRI, as morphometry primarily based on voxel, is a valid and noninvasive method to quantify volumetric brain atrophy caused by severe neuronal loss. Moreover, the integrity of white matter fiber bundles within axonal projections can be assessed *in vivo* using diffusion tensor imaging.<sup>8</sup> MRI can clearly define the pattern of brain injury. On the one hand, MRI measurements of medial temporal lobe atrophy are considered to be a valid marker for clinical AD diagnosis, which can distinguish AD from other brain disorders. On the other hand, MRI can also determine the risk of developing AD or other brain abnormalities from MCI.<sup>9</sup> Due to the high utility of MRI for the diagnosis of AD and MCI, several rating scales have also been established to aid in diagnosis.<sup>10</sup> Besides, a number of studies have begun to develop computer software for automatic MRI assessment to increase diagnostic accuracy and consistency. However, in the process of clinical transformation, the reliability of automatic diagnostic results is not ideal as a wide range of disease characteristics are not specific. Luckily, with the development of artificial intelligence (AI) technology and the popularity of medical applications, diagnostic algorithms can be built through deep learning of clinical data. Such algorithms can be optimized repeatedly to minimize errors.

Moreover, there are many studies today that deep learning—a branch of AI that utilizes layered neural networks to model and analyze vast amounts of data—based on MRI images is employed for the diagnosis of AD. This method is able to reliably capture and quantify a variety of subtle MRI changes throughout the brain, and then amplify the complexity and heterogeneity of AD and brain aging.<sup>11</sup> In addressing the complexities of AD and MCI diagnosis, the role of deep learning cannot be overstated.<sup>12</sup> Specifically, within the domain of MRI analysis, deep learning techniques have the potential to revolutionize diagnostic processes by autonomously identifying and extracting pivotal features from imaging data, a task that traditionally required extensive manual intervention.<sup>13</sup> The application of these advanced AI models to neuroimaging data has been shown to significantly enhance the detection of early and subtle neuroanatomical changes indicative of AD and MCI, thereby offering promising avenues for timely and accurate diagnoses.<sup>14</sup> Nonetheless, no studies have evaluated the accuracy of deep learning in MRI for diagnosing AD and MCI. The objective of this paper was to collect relevant studies on deep learning of MRI for diagnosis of AD and MCI and conduct a meta-analysis to verify the diagnostic accuracy of these

articles. These findings have the potential to significantly enhance the diagnostic process for Alzheimer's disease and mild cognitive impairment, leading to earlier and more accurate identification of these conditions. Such advancements could greatly improve patient outcomes by enabling timely intervention and personalized treatment strategies, ultimately contributing to a better quality of life and slower disease progression.

## METHODS

### Literature search strategy

This study was conducted on the grounds of the regulations in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).<sup>15</sup> Taking MRI, deep learning, AD, MCI, and diagnostics as keywords, the data were searched from PubMed, Embase and Cochrane library database, and the results of different queries were combined using Boolean operator AND. Also, the reference lists of the included studies were manually searched to identify any relevant articles.

### Inclusion and exclusion criteria

Among the collected studies, the research complying with the following inclusion criteria was enrolled for meta-analysis:

- observational project: a meta-analysis of MRI-based deep learning diagnosis of AD and MCI;
- subjects: patients with AD and MCI;
- intervention (subgroups): deep learning modeling group; and
- evaluation metrics: sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio, and area under the receiver operating characteristic curve (AUROC).

Exclusion criteria included magazine publication types (e.g., reviews, letters to the editors, editorials, conference abstracts); and scientific publication types, such as case reports, meta-analyses, literature reviews, and cross-sectional studies; full-text not available; data not extracted; and participants with AD and other mental illnesses. Also, the subjects must not have been the study of MRI-based deep learning for differential diagnosis of AD and MCI.

### Data extraction

A standardized form was developed to capture information including first author, country, year of publication, type of AI model, number of patients, patient characteristics (–mean/median age, gender), and diagnostic efficiency. The above information was extracted independently by two evaluators from each eligible study. In addition, data such as AUROC, sensitivity, specificity, and accuracy were extracted for data processing and forest mapping.

### Quality assessment

The risk of bias was initially assessed independently by two evaluators. Then, a third evaluator was responsible for reviewing each study using the Quality Assessment of

Diagnostic Accuracy Studies-2 (QUADAS-2) guidelines.<sup>16</sup> The QUADAS-2 tool could be used to assign the risk of bias to “low”, “high”, or “uncertain” based on a “yes”, “no”, or “uncertain” response to the relevant marked question contained in each section. For example, if the answer to all landmark questions in the scope was “yes”, then it could be rated as a low risk of bias; if the answer to all information questions was “no”, then the risk of bias was rated as “high”. To ensure the accuracy and reliability of the information gathered, all data were initially extracted independently by two evaluators. In cases where discrepancies arose, a pre-defined protocol was employed to resolve these differences. This involved a detailed discussion between the evaluators to reach a consensus. If consensus could not be achieved, a third, senior evaluator was consulted to make the final decision. This rigorous procedure was designed to minimize subjective interpretation and bias, thereby enhancing the reliability of the data extraction process.

### Statistical analysis

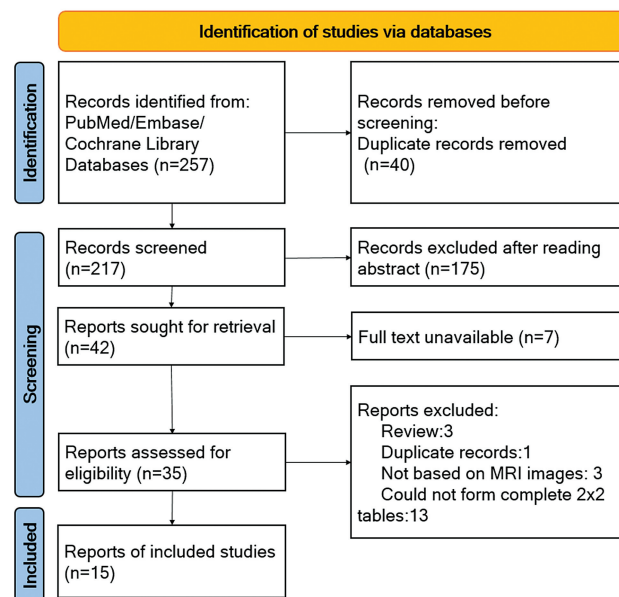
In this study, Stata 16.0 software was used to calculate the combined sensitivity and specificity of each study to assess the accuracy of the meta-analysis, and the combined positive likelihood ratio, combined negative likelihood ratio, and combined diagnostic ratio were also calculated. The total receiver operating characteristic (ROC) curve was also drawn to calculate AUC. Hierarchical summary of receiver operating characteristic (HSROC) curves with credible and predictive regions were simultaneously constructed. The Spearman's correlation coefficient test was calculated to determine whether there was heterogeneity caused by the threshold effect. Deek's funnel plot was plotted to assess the publication bias more intuitively.  $p < 0.05$  indicated a statistically significant difference.

## RESULTS

### Literature screening

The literature search process is shown in the PRISMA flowchart in ► **Figure 1**. There were 257 studies identified, and 40 duplicate studies were removed. The remaining studies were screened. Of them, 175 did not meet the inclusion criteria based on title and abstract. The remaining 42 complete manuscripts were individually assessed, and finally, 35 studies were eligible for inclusion in our systematic review. 15 papers were available for meta-analysis,<sup>17–31</sup> and 20 articles were excluded due to insufficient data.

The included studies were published from 2019 to 2023. Among them, 1 study involved 3 research projects, and 1 study included 2 research projects. Therefore, we collected data from 18 studies. Of them, 6 of these studies were conducted in China, 4 in South Korea, 2 each in Spain, Italy, and Iran, and 1 each in India and Singapore (► **Table 1**). The adopted AI learning models in these studies included CNN (EfficientNet), 3D CNN, CNN + Transfer Learning, Long Short-Term Memory (LSTM) networks + CNN, CNN + iterated Radio Frequency (RF), 2D CNNs, Support Vector Machine (SVM) + Auto-Encoder Neural Networks, CNN + XGBoost, 3D CNN



**Figure 1** PRISMA flowchart outlines the process of studies' selection.

+SVM and 3D CNN + Computer Aided Engineering (CAE) (► **Table 2**). The results of the quality assessment of the included literature are shown in ► **Table 3** and ► **Figure 2**.

### Meta-analysis results

Among the 18 studies included in this research, the Spearman correlation coefficient test showed that the coefficient was  $-0.6506$ , ( $p = 0.0035$ ), suggesting the presence of a threshold effect. A random-effects model was adopted for meta-analysis owing to the heterogeneity of the included studies ( $I^2 = 31\%$ ,  $95\% \text{ CI} = 0-100$ ). The sensitivity and specificity of deep learning in MRI for diagnosing AD and MCI were combined and analyzed. The results suggested that the combined sensitivity, combined specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio were  $0.84$  ( $95\% \text{ CI}: 0.77-0.89$ ),  $0.86$  ( $95\% \text{ CI}: 0.79-0.91$ ), and  $6.0$  ( $95\% \text{ CI}: 3.8-9.4$ ),  $0.19$  ( $95\% \text{ CI}: 0.12-0.28$ ),  $32$  ( $95\% \text{ CI}: 14-72$ ), respectively (► **Figures 3A and B**). In addition, the AUROC curve (► **Figure 3C**) was  $0.92$  ( $95\% \text{ CI}: 0.89-0.94$ ), indicating that MRI-based deep learning had high accuracy in differentiating the diagnosis for AD and MCI in terms of sensitivity and specificity. Then, HSROC analysis was performed to avoid the threshold effect that could adversely affect the results (► **Figure 3D**). The Q-point of the HSROC curve was  $3.463$ . The results of HSROC suggested that MRI-based deep learning exhibited good sensitivity and specificity in diagnosing AD and MCI, as well as a high diagnostic odds ratio.

### Analysis of publication bias in the literature

Deek's funnel plot was adopted to assess whether there was publication bias in the collected literature. The results showed (► **Figure 4**) that the collected studies were approximately distributed along the central axis of symmetry in Deek's funnel plot ( $p = 0.77$ ). This data illustrated the absence of publication bias in our collected literature.

**Table 1** Basic characteristics of the included studies

Author	Year	Country	MRI	Sample size		Database	Type of research
				AD	MCI		
Li H <sup>26</sup>	2023	China	structural Magnetic Resonance Imaging (sMRI)	151	142	ADNI	Retrospective
Agarwal D <sup>17</sup>	2023	Spain	T1-weighted brain MRI	229	229	ADNI	Retrospective
Tanveer M <sup>30</sup>	2022	India	T1-weighted brain MRI	187	398	ADNI	Retrospective
Gao L <sup>23</sup>	2022	China	T1-weighted brain MRI	154	145	ANDI	Retrospective
Chen X <sup>21</sup>	2022	China	T2-weighted brain MRI	43	97	ADNI	Retrospective
Agarwal D <sup>18</sup>	2022	Spain	T1-weighted brain MRI	245	229	ADNI + IXI	Retrospective
Mehmood A <sup>27</sup>	2021	China	T1-weighted brain MRI	85	70	ADNI	Retrospective
Kang W <sup>25</sup>	2021	China	T1-weighted brain MRI	187	382	ADNI	Retrospective
Hedayati R <sup>24</sup>	2021	Iran	Functional magnetic resonance imaging (fMRI)	100	100	ADNI	Retrospective
Akramifard H <sup>19</sup>	2021	Iran	Not explicitly stated	156	338	ADNI	Retrospective
Suh C H <sup>29</sup>	2020	South Korea	T1-weighted brain MRI	161	363	Asan Medical Center	Retrospective
Suh C H <sup>29</sup>	2020	South Korea	T1-weighted brain MRI	68	63	Kyung Hee University Hospital at Gangdong	Retrospective
Suh C H <sup>29</sup>	2020	South Korea	T1-weighted brain MRI	178	317	ADNI	Retrospective
Feng W <sup>22</sup>	2020	China	T1-weighted brain MRI	130	133	ADNI	Retrospective
Wee C Y <sup>31</sup>	2019	Singapore	T1-weighted brain MRI	592	899	ADNI	Retrospective
Oh K <sup>28</sup>	2019	South Korea	T1-weighted brain MRI	198	101	ADNI	Retrospective
Basaia S <sup>20</sup>	2019	Italy	T1 and T2-weighted brain MRI	418	533	ADNI + Milan dataset	Retrospective
Basaia S <sup>20</sup>	2019	Italy	T1 and T2-weighted MRI	294	510	ADNI	Retrospective

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; MCI, mild cognitive impairment; MRI, magnetic resonance imaging.

### Subgroup analysis

As shown in ► **Table 1**, structural MRI (sMRI) was used in 1 study, T1-weighted brain MR images<sup>17,18,22,23,25,27-31</sup> were used in 12 studies, and T1 and T2-weighted brain MR images were used in 2 studies.<sup>20</sup> In addition, there was 1 study, using T2-weighted brain MR images<sup>21</sup> and fMRI<sup>24</sup> for diagnosis, and 1 study did not specify the diagnostic method.<sup>19</sup> These diagnostic methods could be reduced to two MRI techniques, namely, the T1-weighted MRI alone group (containing 12 studies) and the non-T1-weighted MRI alone group (containing 6 studies).

According to ► **Table 4**, the combined sensitivity of T1-weighted MRI was 0.84 (95% CI: 0.74–0.91), which was similar to that of non-T1-weighted MRI (0.84, 95% CI: 0.72–0.91), but there was no significant difference. Additionally, the combined specificity, negative likelihood ratio, positive likelihood ratio, and combined diagnostic ratio of T1-weighted MRI and non-T1-weighted MRI were not significantly different. However, the positive likelihood ratio of T1-weighted MRI was 5, while the positive likelihood ratio of non-T1-weighted MRI was 8.3, indicating that the positive likelihood of non-T1-weighted MRI was higher. Therefore, when the result was positive, the patients may be diagnosed as positive more accurately by non-T1-weighted MR. Moreover, the AUC for T1-weighted MRI was 0.91, and for non-T1-weighted MRI was 0.93, both

indicating good diagnostic accuracy. These data suggested that the two MRI techniques were similar in terms of sensitivity, negative likelihood ratio, and AUC, but non-T1-weighted MRI had a slight advantage in specificity and positive likelihood ratio. Nevertheless, these differences may be not statistically significant due to the overlapping of CI.

### DISCUSSION

The data of many images, such as MRI, computed tomography, and positron emission tomography, need to be collected and generated during the diagnosis and treatment of AD or MCI.<sup>32</sup> Clinically, medical staff usually evaluate this data subjectively and formulate treatment plans based on experience. Accurate early diagnosis of AD and MCI is essential for treatment. However, the features of the imaging data observed only relying on the naked eye of the medical staff were limited and may make many potential imaging data not fully revealed.<sup>33</sup> In recent years, many researchers have attempted to use sophisticated mathematical and statistical algorithms to extract hard-to-observe quantitative information to increase the diagnostic accuracy of AD and the potential to predict the worsening progression of MCI.<sup>34</sup>

In this study, we collected 15 articles from 2019 to 2023 and analyzed the diagnostic accuracy of deep learning

**Table 2** Characteristics of artificial intelligence learning models of the included literature

Author	Year	Modeling	Test Set	TP	FP	FN	TN
Li H <sup>26</sup>	2023	CNN (EfficientNet)	AD:30,MCI: 30	27	2	3	28
Agarwal D <sup>17</sup>	2023	3D CNN (EfficientNet-B0)	AD:29,MCI: 29	29	4	0	25
Tanveer M <sup>30</sup>	2022	CNN + Transfer Learning	20%	37	1	1	79
Gao L <sup>23</sup>	2022	LSTM networks + CNN	AD:31,MCI:18	25	2	6	16
Chen X <sup>21</sup>	2022	CNN + iterated RF	AD:30,MCI: 30	28	2	2	28
Agarwal D <sup>18</sup>	2022	3D CNN (DenseNet264)	AD:29,MCI: 29	27	10	2	19
Mehmood A <sup>27</sup>	2021	CNN + Transfer Learning	20%	13	2	2	12
Kang W <sup>25</sup>	2021	2D CNNs	20%	26	35	12	42
Hedayati R <sup>24</sup>	2021	CNN	AD:20,MCI: 20	17	1	3	19
Akramifard H <sup>19</sup>	2021	SVM + Auto-Encoder Neural Networks	10%	8	6	8	28
Suh C H <sup>29</sup>	2020	CNN + XGBoost	20%	23	19	10	54
Suh C H <sup>29</sup>	2020	CNN + XGBoost	20%	10	3	4	10
Suh C H <sup>29</sup>	2020	CNN + XGBoost	20%	24	19	12	45
Feng W <sup>22</sup>	2020	3D CNN + SVM	AD:23,MCI: 24	22	1	1	23
Wee C Y <sup>31</sup>	2019	CNN	10%	42	13	18	77
Oh K <sup>28</sup>	2019	3D CNN + CAE	10%	15	3	5	8
Basaia S <sup>20</sup>	2019	3D CNN	10%	35	6	7	48
Basaia S <sup>20</sup>	2019	3D CNN	10%	25	6	5	45

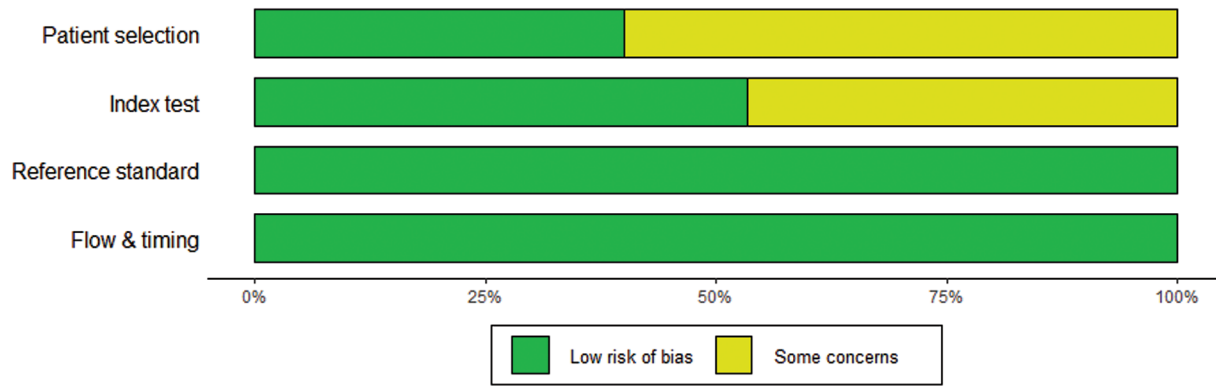
Abbreviations: CAE, Computer Aided Engineering; CNN, Convolutional Neural Networks; LSTM, Long Short-Term Memory; RF, Radio Frequency; SVM, Support Vector Machine.

**Table 3** Quality assessment results of the quality assessment of diagnostic accuracy studies 2 of the included literature

Author	Year	Patient selection	Index test	Reference standard	Flow and timing
Li H <sup>26</sup>	2023	low	unclear	low	low
Agarwal D <sup>17</sup>	2023	unclear	unclear	low	low
Tanveer M <sup>30</sup>	2022	unclear	low	low	low
Gao L <sup>23</sup>	2022	unclear	unclear	low	low
Chen X <sup>21</sup>	2022	unclear	unclear	low	low
Agarwal D <sup>18</sup>	2022	unclear	unclear	low	low
Mehmood A <sup>27</sup>	2021	low	low	low	low
Kang W <sup>25</sup>	2021	unclear	low	low	low
Hedayati R <sup>24</sup>	2021	unclear	unclear	low	low
Akramifard H <sup>19</sup>	2021	low	low	low	low
Suh C H <sup>29</sup>	2020	low	low	low	low
Feng W <sup>22</sup>	2020	unclear	unclear	low	low
Wee C Y <sup>31</sup>	2019	low	low	low	low
Oh K <sup>28</sup>	2019	unclear	low	low	low
Basaia S <sup>20</sup>	2019	low	low	low	low

methods based on brain MRI data in AD and MCI through meta-analysis. These studies were performed based on convolutional neural network (CNN) in conjunction with different algorithms. CNN involves a great many deep learning techniques and has been proven to be effective in diagnosing non-dementia, particularly mild dementia, mild dementia,

and moderate dementia.<sup>35</sup> Jo T et al. collected the deep learning papers on AD published from 2013 to 2018 and performed a meta-analysis. In their study, they concluded that deep learning had 83.7% accuracy for AD classification; the accuracy for predicting progression from MCI to AD was as high as 96.0%.<sup>36</sup> Spasov et al. collected the data of 192



**Figure 2** Results of quality evaluation of the included literature. The quality of the literature was assessed by the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). The color coding represents the assessed risk of bias, with green indicating a low risk of bias and yellow representing some concerns about potential bias.

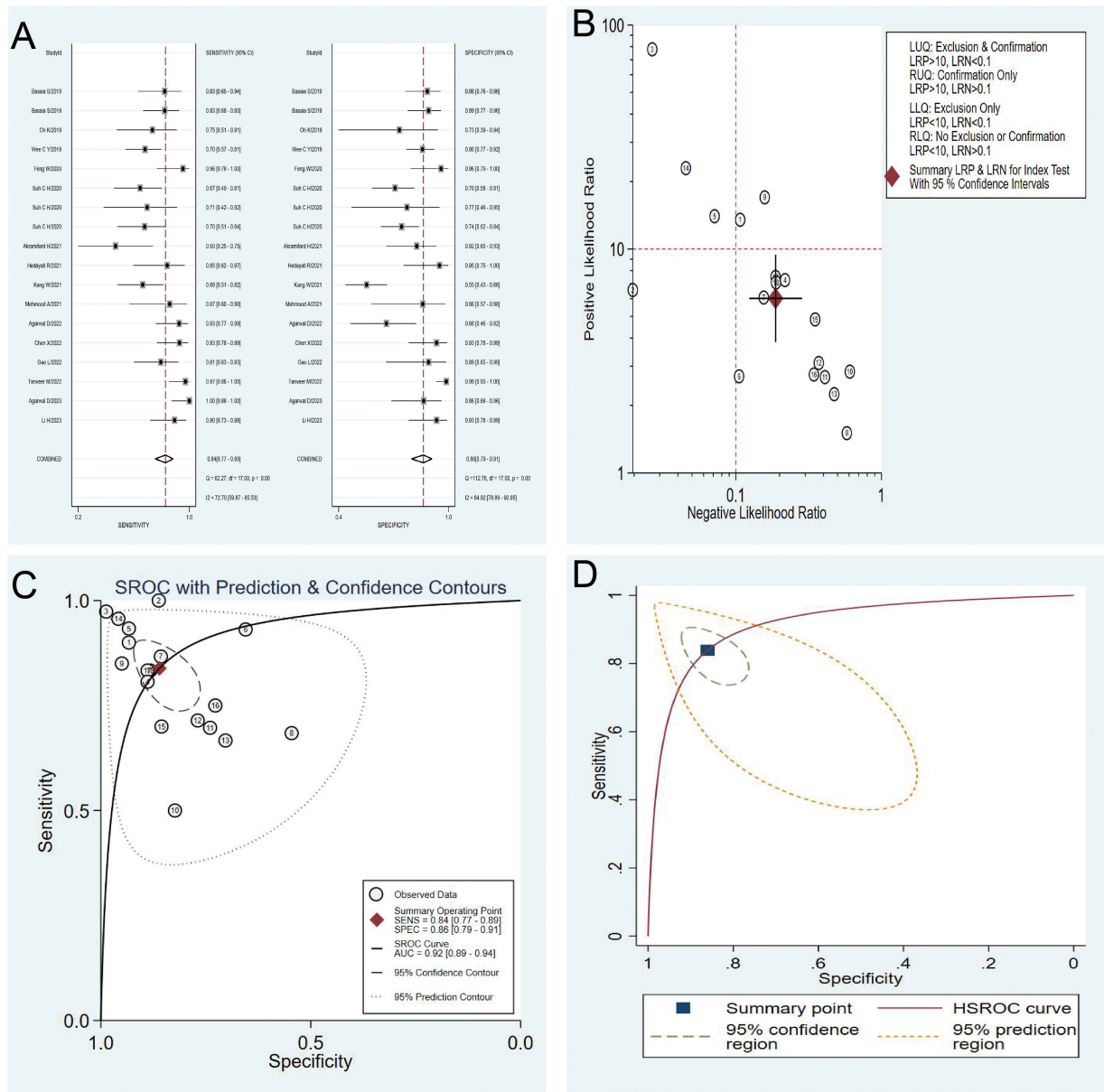
patients with AD and 409 patients with MCI and then built a deep learning algorithm to distinguish the MCI patients developing into AD within 3 years from those MCI patients with stable condition in the same period; the AUC was 0.925, the 10-fold cross-validated accuracy was 86%, the sensitivity was 87.5%, and the specificity of 85%.<sup>37</sup> Wei et al. utilized three-dimensional convolutional neural networks (3D-CNNs) and MRI to build binary and ternary disease classification models. Through these models, they observed that the ternary classification accuracy of the 3D-CNN-support vector machine (3D-CNN-SVM) for the diagnosis of MCI and AD were 96.82% and 96.73%, respectively.<sup>22</sup> The above findings revealed the significance of deep learning in improving the accuracy of MRI-based diagnosis of AD and MCI. The meta-analysis of this study showed that the deep learning of MRI had good sensitivity and specificity in diagnosing AD and MCI overall.

The variability across different deep learning architectures presents both challenges and opportunities for enhancing diagnostic performance in neurodegenerative diseases. Studies have demonstrated that the choice of architecture, from CNNs to more complex models like Recurrent Neural Networks (RNNs) and their hybrids, can significantly affect the model's ability to learn and generalize from neuroimaging data.<sup>38</sup> Furthermore, the diversity in training datasets—spanning various demographics, disease stages, and imaging protocols—introduces additional layers of variability that can influence diagnostic outcomes.<sup>39</sup> Notwithstanding these challenges, adopting strategies such as transfer learning, data augmentation, and ensemble learning models offers promising pathways to achieving more consistent and reliable diagnostic predictions across diverse clinical settings.<sup>40</sup>

In the literature we collected, the uncertainty of heterogeneity was large, but there was no publication bias. Moreover, we did not find any significant difference in the diagnosis of AD and MCI between the T1-weighted MRI alone and non-T1-weighted MRI alone. However, it cannot be ignored that multiple learning models were used in the literature collected in this paper. These learning models inevitably induced data bias and then affected the comparison of the overall diagnostic performance. The literature

collected in this study was also retrospective, and most of the studies had no directly available data and deep learning codes. Moreover, only internal validation or resampling methods were used to judge the accuracy of deep learning. Such validation methods lack generalization, and the internal validation tends to overestimate the AUC, especially for out-of-distribution detection on 3D medical images.<sup>41</sup> Simultaneously, what could not be ignored was that the insufficiency of prospective studies on deep learning for AD and MCI limited the integration of AI models with the clinical setting.<sup>42</sup> Therefore, externally validated predictive models using images from different hospitals are required to create reliable estimates of the performance level at other sites. In addition, reports with incomplete data were removed during the literature screening, which might affect the estimates of diagnostic performance.<sup>43</sup> Besides, the results of the calculations may be geographically biased, since the included studies were from geographically diverse quantitative distributions; moreover, the type of scanner used for diagnosis, the imaging protocol, and the diagnostic criteria for AD and MCI may also affect the accuracy of results.<sup>44</sup> However, deep learning itself has great potential because it can continuously improve the algorithms to increase the accuracy. Therefore, future research should prioritize the development of standardized protocols for data acquisition and preprocessing. Additionally, there's a pressing need for collaborative efforts to establish large, annotated datasets that reflect the diversity of the global population. Moreover, the exploration of federated learning approaches, where AI models are collaboratively trained across multiple institutions while keeping data localized, offers a promising direction.

In summary, our meta-analysis underscores the significant potential of deep learning algorithms applied to MRI images in enhancing the diagnostic accuracy for AD and MCI. The analysis revealed that deep learning models exhibit high sensitivity and specificity, indicating their reliability in identifying AD and MCI from neuroimaging data. These findings highlight the transformative impact of AI in the field of neurology, offering a promising tool for early and accurate disease diagnosis. Moving forward, it is imperative



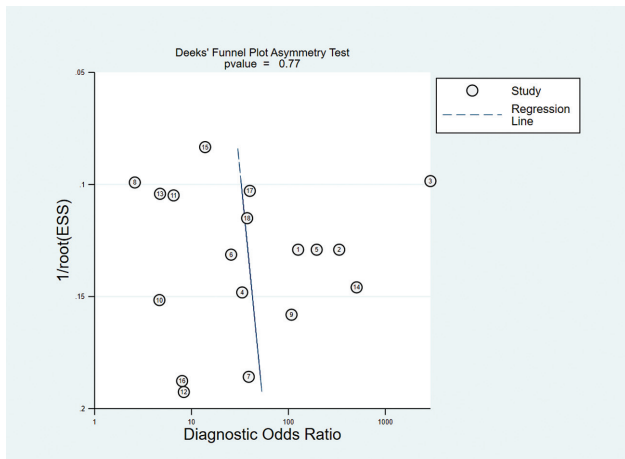
**Figure 3** Meta-analysis results. (A) Combined sensitivity and specificity forest maps for MRI-based deep learning to diagnose Alzheimer's disease and mild cognitive impairment; (B) Combined positive likelihood ratio and negative likelihood ratio for diagnosis of Alzheimer's disease and mild cognitive impairment using MRI-based deep learning; (C) The Receiver Operating Characteristic Curve of MRI-based deep learning for diagnosing Alzheimer's disease and mild cognitive impairment; (D) Hierarchical summary receiver operating characteristic curves for MRI-based deep learning to diagnose Alzheimer's disease and mild cognitive impairment

**Table 4** Subgroup analysis

	T1-weighted MRI alone	Non-T1-weighted MRI alone
Number of studies	12	6
Combined sensitivity (95% CI)	0.84 (0.74, 0.91)	0.84 (0.72, 0.91)
Combined specificity (95% CI)	0.83 (0.73, 0.90)	0.9 (0.84, 0.94)
Positive likelihood ratio (95% CI)	5 (2.8, 8.9)	8.3 (5.0, 14.0)
Negative likelihood ratio (95% CI)	0.19 (0.10, 0.34)	0.18 (0.10, 0.33)
The combined diagnostic odds ratio (95% CI)	26 (9, 79)	46 (17, 127)
AUC	0.91 (0.88, 0.93)	0.93 (0.91, 0.95)

Abbreviations: AUC, Area Under Curve; CI, confidence interval; MRI, magnetic resonance imaging.





**Figure 4** Deek's funnel plot for diagnosis of Alzheimer's disease and mild cognitive impairment based on MRI deep learning. This plot is a tool used to visually inspect for the presence of publication bias in meta-analyses. Each point represents a study included in the meta-analysis, plotted according to its effect size against a measure of precision (the inverse of the standard error). The funnel plot's asymmetry is tested using Deek's test, where a symmetrical distribution suggests the absence of publication bias, and asymmetry may indicate its presence.

for future research to focus on addressing the challenges of model variability and data heterogeneity to further refine AI applications in medical diagnostics. Clinically, integrating AI into diagnostic workflows has the potential to revolutionize patient care by enabling timely and personalized treatment interventions. This research not only contributes to the existing body of knowledge but also lays a foundational path for leveraging AI to improve outcomes for individuals with neurodegenerative disorders.

In conclusion, deep learning models based on MRI images have the potential to improve diagnostic accuracy in AD and MCI, which can not only provide clinicians with individualized preoperative noninvasive auxiliary prediction tools but also increase the early diagnosis rate of the patients with AD and MCI to develop better treatment strategies. Furthermore, with the continuous improvement of deep algorithms, more effective algorithms may be developed to further improve the diagnostic accuracy of AI in AD and MCI.pt

#### Authors' Contributions

LXW, JL: conceived and designed the study, conducted the study, edited the manuscript draft; YZW, CGH: contributed to data acquisition; CGH, LZ: analyzed the data; YZH, LZ: interpreted the data; XW, JL: reviewed and edited the manuscript. All authors have read and approved the manuscript.

#### Conflict of Interest

There is no conflict of interest to declare.

#### References

- Gowda P, Reddy PH, Kumar S. Deregulated mitochondrial microRNAs in Alzheimer's disease: Focus on synapse and mitochondria. *Ageing Res Rev* 2022;73:101529. Doi: 10.1016/j.arr.2021.101529
- Xia P, Chen J, Liu Y, et al. MicroRNA-22-3p ameliorates Alzheimer's disease by targeting SOX9 through the NF- $\kappa$ B signaling pathway in the hippocampus. *J Neuroinflammation* 2022;19(01):180. Doi: 10.1186/s12974-022-02548-1
- Klyucherev TO, Olszewski P, Shalimova AA, et al. Advances in the development of new biomarkers for Alzheimer's disease. *Transl Neurodegener* 2022;11(01):25. Doi: 10.1186/s40035-022-00296-z
- Park JC, Han SH, Mook-Jung I. Peripheral inflammatory biomarkers in Alzheimer's disease: a brief review. *BMB Rep* 2020;53(01):10–19. Doi: 10.5483/BMBRep.2020.53.1.309
- Fonte C, Smania N, Pedrinolla A, et al. Comparison between physical and cognitive treatment in patients with MCI and Alzheimer's disease. *Ageing (Albany NY)* 2019;11(10):3138–3155. Doi: 10.18632/aging.101970
- Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020;396(10248):413–446. Doi: 10.1016/S0140-6736(20)30367-6
- Yin RH, Tan L, Liu Y, et al. Multimodal Voxel-Based Meta-Analysis of White Matter Abnormalities in Alzheimer's Disease. *J Alzheimers Dis* 2015;47(02):495–507. Doi: 10.3233/JAD-150139
- Talwar P, Kushwaha S, Chaturvedi M, Mahajan V. Systematic Review of Different Neuroimaging Correlates in Mild Cognitive Impairment and Alzheimer's Disease. *Clin Neuroradiol* 2021;31(04):953–967. Doi: 10.1007/s00062-021-01057-7
- Zhou Y, Song Z, Han X, Li H, Tang X. Prediction of Alzheimer's Disease Progression Based on Magnetic Resonance Imaging. *ACS Chem Neurosci* 2021;12(22):4209–4223. Doi: 10.1021/acscchem-neuro.1c00472
- Grauskas LA, Guo H, D'Arcy RCN, Song X. Toward MRI-based whole-brain health assessment: The brain atrophy and lesion index (BALI). *Ageing Med (Milton)* 2018;1(01):55–63. Doi: 10.1002%2Fagm.12014
- 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* 2022;77:101614. Doi: 10.1016/j.arr.2022.101614
- Adarsh V, Gangadharan GR, Fiore U, Zanetti P. Multimodal classification of Alzheimer's disease and mild cognitive impairment using custom MKSCDDL kernel over CNN with transparent decision-making for explainable diagnosis. *Sci Rep* 2024;14(01):1774. Doi: 10.1038/s41598-024-52185-2
- Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med* 2021;4(01):65. Doi: 10.1038/s41746-021-00438-z
- Suk HI, Lee SW, Shen D. Alzheimer's Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage* 2014;101:569–582. Doi: 10.1016%2Fj.neuroimage.2014.06.077
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group PPRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151(04):264–269
- Zheng X, He B, Hu Y, et al. Diagnostic Accuracy of Deep Learning and Radiomics in Lung Cancer Staging: A Systematic Review and Meta-Analysis. *Front Public Health* 2022;10:938113. Doi: 10.3389/fpubh.2022.938113
- Agarwal D, Berbís MA, Luna A, Lipari V, Ballester JB, de la Torre-Díez I. Automated Medical Diagnosis of Alzheimer's Disease Using an Efficient Net Convolutional Neural Network. *J Med Syst* 2023;47(01):57. Doi: 10.1007/s10916-023-01941-4
- Agarwal D, Berbís MA, Martín-Noguerol T, Luna A, García SCP, de la Torre-Díez I. End-to-End Deep Learning Architectures Using 3D Neuroimaging Biomarkers for Early Alzheimer's Diagnosis. *Mathematics* 2022;10(15):2575. Doi: 10.3390/math10152575
- Akramifard H, Balafar MA, Razavi SN, Ramli AR. Early Detection of Alzheimer's Disease Based on Clinical Trials, Three-Dimensional Imaging Data, and Personal Information Using Autoencoders. *J*

- Med Signals Sens 2021;11(02):120–130. Doi: 10.4103/jmss.JMSS\_11\_20
- 20 Basaia S, Agosta F, Wagner L, et al; Alzheimer's Disease Neuroimaging Initiative. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin* 2019;21:101645. Doi: 10.1016/j.nicl.2018.101645
  - 21 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(14):765. Doi: 10.21037/atm-22-2961
  - 22 Feng W, Halm-Lutterodt NV, Tang H, et al. Automated MRI-Based Deep Learning Model for Detection of Alzheimer's Disease Process. *Int J Neural Syst* 2020;30(06):2050032. Doi: 10.1142/S012906572050032X
  - 23 Gao L, Hu Z, Li R, et al. Multi-Perspective Feature Extraction and Fusion Based on Deep Latent Space for Diagnosis of Alzheimer's Diseases. *Brain Sci* 2022;12(10):1348. Doi: 10.3390/brainsci12101348
  - 24 Hedayati R, Khedmati M, Taghipour-Gorjilolaie M. Deep feature extraction method based on ensemble of convolutional auto encoders: Application to Alzheimer's disease diagnosis. *Biomed Signal Process Control* 2021;66(03):102397. Doi: 10.1016/j.bspc.2020.102397
  - 25 Kang W, Lin L, Zhang B, Shen X, Wu S Alzheimer's Disease Neuroimaging Initiative. Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis. *Comput Biol Med* 2021;136:104678. Doi: 10.1016/j.combiomed.2021.104678
  - 26 Li H. Attention-based and micro designed EfficientNetB2 for diagnosis of Alzheimer's disease. *Biomed Signal Process Control* 2023;82:104571. Doi: 10.1016/j.bspc.2023.104571
  - 27 Mehmood A, Yang S, Feng Z, et al. A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. *Neuroscience* 2021;460:43–52. Doi: 10.1016/j.neuroscience.2021.01.002
  - 28 Oh K, Chung YC, Kim KW, Kim WS, Oh IS. Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. *Sci Rep* 2019;9(01):18150. Doi: 10.1038/s41598-019-54548-6
  - 29 Suh CH, Shim WH, Kim SJ, et al; Alzheimer's Disease Neuroimaging Initiative. Development and Validation of a Deep Learning-Based Automatic Brain Segmentation and Classification Algorithm for Alzheimer Disease Using 3D T1-Weighted Volumetric Images. *AJNR Am J Neuroradiol* 2020;41(12):2227–2234. Doi: 10.3174/ajnr.A6848
  - 30 Tanveer M, Rashid AH, Ganaie MA, Reza M, Razzak I, Hua KL. Classification of Alzheimer's Disease Using Ensemble of Deep Neural Networks Trained Through Transfer Learning. *IEEE J Biomed Health Inform* 2022;26(04):1453–1463. Doi: 10.1109/JBHI.2021.3083274
  - 31 Wee CY, Liu C, Lee A, Poh JS, Ji H, Qiu A Alzheimer's Disease Neuroimaging Initiative. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. *Neuroimage Clin* 2019;23:101929. Doi: 10.1016/j.nicl.2019.101929
  - 32 Burke AD, Goldfarb D. Diagnosing and Treating Alzheimer Disease During the Early Stage. *J Clin Psychiatry* 2023;84(02):L121019AH3C. Doi: 10.4088/JCP.L121019AH3C
  - 33 Yu Q, Mai Y, Ruan Y, et al; National Alzheimer's Coordinating Center, the Alzheimer's Disease Neuroimaging Initiative Frontotemporal Lobar Degeneration Neuroimaging Initiative. An MRI-based strategy for differentiation of frontotemporal dementia and Alzheimer's disease. *Alzheimers Res Ther* 2021;13(01):23. Doi: 10.1186/s13195-020-00757-5
  - 34 Verma RK, Pandey M, Chawla P, et al. An Insight into the Role of Artificial Intelligence in the Early Diagnosis of Alzheimer's Disease. *CNS Neurol Disord Drug Targets* 2022;21(10):901–912. Doi: 10.2174/1871527320666210512014505
  - 35 Al Shehri W. Alzheimer's disease diagnosis and classification using deep learning techniques. *PeerJ Comput Sci* 2022;8:e1177. Doi: 10.7717/peerj-cs.1177
  - 36 Jo T, Nho K, Saykin AJ. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Front Aging Neurosci* 2019;11:220. Doi: 10.3389/fnagi.2019.00220
  - 37 Spasov S, Passamonti L, Duggento A, Liò P, Toschi N Alzheimer's Disease Neuroimaging Initiative. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* 2019;189:276–287. Doi: 10.1016/j.neuroimage.2019.01.031
  - 38 Song J, Hahm J, Lee J, et al. Comparative validation of AI and non-AI methods in MRI volumetry to diagnose Parkinsonian syndromes. *Sci Rep* 2023;13(01):3439. Doi: 10.1038/s41598-023-30381-w
  - 39 Nazer LH, Zatarah R, Waldrip S, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit Health* 2023;2(06):e0000278. Doi: 10.1371/journal.pdig.0000278
  - 40 Fathi S, Ahmadi A, Dehnad A, Almasi-Dooghaee M, Sadegh M Alzheimer's Disease Neuroimaging Initiative. A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images. *Neuroinformatics* 2024;22(01):89–105. Doi: 10.1007/s12021-023-09646-2
  - 41 Vasiliuk A, Frolova D, Belyaev M, Shirokikh B. Limitations of Out-of-Distribution Detection in 3D Medical Image Segmentation. *J Imaging* 2023;9(09):191. Doi: 10.3390/jimaging9090191
  - 42 Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean J Radiol* 2019;20(03):405–410. Doi: 10.3348/kjr.2019.0025
  - 43 Steyerberg EW, Moons KG, van der Windt DA, et al; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10(02):e1001381. Doi: 10.1371/journal.pmed.1001381
  - 44 Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;1(06):e271–e297. Doi: 10.1016/S2589-7500(19)30123-2