Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37966

# *Deep Learning-Based Assisted Diagnosis of Alzheimer's Disease: From Monomodal Imaging to Multimodal Fusion*

**Renpu Han**

*Tianjin Yaohua High School, Tianjin, China*
*david_han1145@outlook.com*

***Abstract.*** In the field of auxiliary diagnosis of Alzheimer's disease (AD), deep learning is developing along two paths: single-modal imaging and multi-modal fusion. The single-modal approach utilizes lightweight networks (with efficient receptive fields and multi-axis attention mechanisms) to optimize preprocessing and improve three-dimensional models, achieving high-precision classification while controlling the number of parameters; multi-modal fusion integrates structural magnetic resonance, genetic information, scale data, diffusion tensor imaging, and electroencephalogram, and employs technical means such as Transformer and graph convolutional networks to significantly enhance the performance of early identification and multi-classification. The existing limitations include weak cross-center generalization ability, difficulties in obtaining and aligning multi-modal data, insufficient interpretability, and the lack of longitudinal prediction. Future development directions include lightweight edge deployment, federated and self-supervised learning, incorporating low-cost modalities such as plasma markers, and constructing longitudinal prediction and frameworks with pathological prior interpretability, thereby facilitating the implementation of clinical applications. This article aims to provide a reference for the further optimization and clinical translation of related methods by summarizing the main technical paths, application progress and existing problems of deep learning in the auxiliary diagnosis of Alzheimer's disease.

***Keywords:*** Alzheimer's disease, auxiliary diagnosis, deep learning, single-modal brain imaging, multi-modal fusion

## 1. Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder closely related to aging. Its core clinical features include cognitive decline and abnormal behaviors and mental states. Against the backdrop of the accelerating global aging population, the number of AD patients is increasing, making it one of the major challenges faced by the global public health system.

Alzheimer's Disease is a common neurodegenerative disorder worldwide [1], with its core clinical manifestations being progressive cognitive decline and the emergence of non-cognitive neurological and mental symptoms [2]. This disease primarily affects individuals in the pre-elderly stage, resulting in a considerable number of patients worldwide. Alzheimer's Disease not only severely disrupts patients' ability to take care of themselves but may even cause complete loss of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37966

ability. It also brings unbearable economic and mental pressure to patients and their families. In the current research and clinical practice context, achieving precise diagnosis and dynamic assessment of the disease through brain imaging methods has become a key issue that needs to be urgently addressed. Neurologists can provide core reference information for tailoring treatment strategies for patients by interpreting the pathological features presented in brain imaging data, such as the reduction in specific brain lobe volumes [3].

The clinical manifestations of Alzheimer's Disease mainly include cognitive dysfunction. Patients often experience memory decline, impaired spatial orientation, language expression disorders, reduced calculation ability, and weakened comprehension. The disease has a latent onset and relatively mild early symptoms, but gradually worsens as the disease progresses. In the later stages, patients' daily living abilities significantly decline, and eventually they may completely lose the ability to live independently and self-care.

Currently, there is still no effective method to completely cure or reverse the course of Alzheimer's Disease. This disease not only severely affects patients' cognitive functions, quality of life, and mental state, but also brings long-term physical and mental burdens to family caregivers, and causes huge medical, care, and socioeconomic costs. Therefore, continuously advancing research and related studies on the pathogenesis of Alzheimer's Disease, early screening, and auxiliary diagnosis is of great practical significance for delaying disease progression, improving patient prognosis, and reducing the burden on families and society.

In recent years, with the continuous integration of artificial intelligence technology with medical imaging, clinical diagnosis, and clinical treatment technologies, computer-assisted diagnosis has gradually become an important technical means to improve the efficiency of disease recognition and the objectivity of diagnosis. Compared with the traditional manual interpretation method, computer-assisted diagnosis can extract potential disease features through data analysis and intelligent algorithms, providing new technical paths for the early identification, classification assessment, and clinical decision support of Alzheimer's Disease, and laying the foundation for the development of related intelligent diagnosis research. For example, with the popularization of artificial intelligence, deep learning has demonstrated outstanding data analysis and feature extraction capabilities in medical scenarios. The birth of the Convolutional Neural Network (CNN) provides an effective way for the precise identification of image features [4].

Therefore, this article focuses on the application of artificial intelligence in the auxiliary diagnosis of Alzheimer's Disease, summarizing the research progress of two types of methods: single-modal brain imaging analysis and multi-modal information fusion, and extracting and summarizing the main problems faced by existing technologies and future development directions.

## 2. An auxiliary diagnostic method for Alzheimer's Disease based on single-modal brain imaging

Medical imaging is a type of medical detection technology that uses specialized equipment to obtain data on the internal structure, physiological functions, and metabolic levels of the human body, presenting the internal pathological changes and physiological architecture of the body in a direct manner. Currently, this technology encompasses many subtypes such as X-ray photography, computed tomography (CT), structural magnetic resonance imaging (sMRI), ultrasound imaging, and single photon emission computed tomography (SPECT). In the clinical diagnosis of Alzheimer's disease, brain magnetic resonance imaging (MRI) has always been a widely used imaging detection method, with advantages such as good contrast of soft tissues, quantifiable assessment of the degree of brain atrophy, and precise and reliable diagnostic results.

DOI: 10.54254/2755-2721/2026.37966

In response to the common problem of overly complex structures and large amounts of parameters and computations in existing Alzheimer's disease-related models, Li Junqi proposed an Alzheimer's disease classification framework that integrates an efficient receptive field module and a multi-axis attention mechanism. It precisely captures the morphological distortions and atrophy of key brain regions such as the hippocampus. This scheme innovatively introduces a hierarchical residual feedforward network structure and depthwise separable convolution technology, achieving balanced weight distribution within the image and effectively compressing the number of parameters while retaining the core feature expression ability. Through multiple sets of control experiments, this fusion of convolutional neural networks and visual Transformer classification method has achieved a classification accuracy of 98.3% and a parameter size of only 38.13M, with an AUC value of 91.7% in the OASIS public dataset for cross-center verification. For classification models using the fusion strategy of efficient receptive fields and multi-axis attention, the image recognition accuracy can reach 97.29%, and the model parameter size is reduced from 38.13M to 18.3M. This research provides an efficient and automated detection method for the early screening of Alzheimer's disease, which has crucial application significance in reducing the rate of missed diagnoses and promoting the development of precision medicine [5].

Cheng Yeqi developed a method to optimize and upgrade the sMRI brain image skull stripping technology for the problem of interference from irrelevant information such as the skull in the original brain sMRI images. Cheng Yeqi conducted research on Alzheimer's disease classification based on brain nuclear magnetic resonance imaging, addressing issues such as low efficiency of pre-processing for structural magnetic resonance images, insufficient classification accuracy for mild cognitive impairment, and the tendency of deep learning models to overfit. This research first performs skull removal preprocessing on sMRI images to reduce the interference of non-brain tissues such as the skull on the classification results; then, it uses AAL and BN brain maps to extract heterogeneous features of brain regions and integrates them with neuropsychological scale data such as MMSE and CDR-SB, combined with the continuous projection algorithm for feature selection, and finally uses partial least squares discriminant analysis to complete classification. In the three classification tasks of AD vs. normal control, AD vs. MCI, and MCI vs. normal control, the accuracy of this method reaches 98.97%, 92.59%, and 91.40% respectively. Additionally, the study also constructed a lightweight three-dimensional residual convolutional neural network that integrates multi-scale image features and multimodal information to reduce model complexity and overfitting risk, improving the early classification effect of Alzheimer's disease. The image screening is achieved with precise selection, maximizing the retention of brain parenchymal regions with reference values for clinical diagnosis. Experimental results show that this data preprocessing method has a significant advantage in terms of time consumption and relatively less overall computational time. This scheme can conduct secondary processing on images that fail the preprocessing stage, and it takes approximately 9 hours to complete the processing of 730 article samples [6].

Cui Yanan used three-dimensional imaging data, magnetic resonance imaging results, and patient-related clinical indicators as the research objects, and built an automatic recognition model for Alzheimer's disease based on 3D-ResNet. By three-dimensionally processing the residual network (ResNet) as the main feature extraction network, this architecture not only allows for the extension of the network layers but also efficiently avoids the gradient disappearance problem during model training. Experimental results show that this network has an identification accuracy of 82.3% in the binary classification task of AD and NC, and an accuracy of 70.6% in the three-classification task of AD, MCI, and NC. In addition, they also developed an AD auxiliary

DOI: 10.54254/2755-2721/2026.37966

recognition model based on the improved 3D-Mamba. When using the original 3D-Mamba as the image feature extraction network for AD recognition, the residual encoder was embedded in the entire architecture of the original model. Through the analysis and verification of the experimental results, it was found that the proposed method achieved a classification accuracy of 96.1% in the binary classification task of AD/NC and a classification accuracy of 82.7% in the three-classification task of AD/MCI/NC, demonstrating very excellent performance [7].

## 3. An auxiliary diagnosis method for Alzheimer's Disease based on multi-modal information fusion

Currently, there is another research path in the academic community that utilizes multi-modal imaging. The key to this research path lies in integrating two or more imaging techniques with different characteristics. That is, to process the imaging information from different modalities, ultimately obtaining data with broader coverage, higher precision, and more relevance to clinical diagnosis. Multi-modal imaging analysis technology breaks through the limitations of a single imaging modality and achieves important goals such as cross-scale visualization.

Gong Lejun developed the multi-modal deep learning classification model MDLCM. This model can extract comprehensive information from the fusion of structural magnetic resonance imaging (sMRI), genetic data (including single nucleotide polymorphisms and gene expression profiles), and data related to clinical diagnosis, achieving effective differentiation of Alzheimer's disease, mild cognitive impairment, and normal control. The model first uses the recursive feature elimination algorithm to screen the key core features in the data, then a lightweight classification network composed of Transformer and convolutional neural networks is specifically used to process the image data, and the gated multi-modal unit (GMU) is introduced to integrate the multi-modal information. Experimental results show that this model has a classification accuracy of 90.5% in the three-classification tasks of Alzheimer's disease, mild cognitive impairment (MCI), and normal control (CN), with a macro average precision of 92.5%, a macro average recall rate of 89.9%, and a macro average F1 value of 90.3% [8].

Liu Yu aggregated structural magnetic resonance imaging (sMRI), brain functional connectivity networks, and diffusion tensor imaging (DTI) fiber structure connection networks, and constructed a multi-modal CNN-Transformer model based on the brain neural connection mechanism and with interpretability. This model achieved efficient fusion of brain structural features and two types of neural connection features through implementing position encoding and attention calculation modules, thereby improving the accuracy of the diagnostic model [9]. In addition, this research team also conducted a study on a multi-modal fusion algorithm related to the clinical diagnosis of Alzheimer's disease. This algorithm fused the brain resting-state electroencephalogram (EEG) data, demographic information, and scale assessment scores. For EEG signals, the research team employed sliding windows and convolutional neural networks to capture the dynamic characteristics of the data, and integrated multiple attention mechanisms and time-domain convolution modules to assist the model in achieving precise learning. At the same time, it focused on the effective information in the data. That is, it combined EEG features with demographic data, neuropsychological scale assessment indicators, and expanded the dimension of input features to comprehensively measure individual differences and cognitive function assessment results, thereby achieving precise discrimination of Alzheimer's disease status. This study conducted a comparative experiment based on an open dataset including Alzheimer's disease, frontotemporal dementia, and healthy controls, and ultimately verified that the proposed scheme has better disease recognition ability compared to many mainstream and cutting-edge technologies. This research achievement

DOI: 10.54254/2755-2721/2026.37966

provides a new practical basis for the diversified development of Alzheimer's disease auxiliary screening technologies [9].

Due to the high difficulty in locating brain lesion areas, the cumbersome screening process, and the negative impact of single-modal clinical data on the accuracy and credibility of the diagnosis results, Sun Haoran utilized the multi-modal imaging information of the subjects and demographic phenotypic characteristics such as age and gender to construct a population population sparse map integrating multi-modal imaging; at the same time, he proposed an innovative residual connection-based graph convolutional network model for fine-grained classification and early warning related research for patients with early mild cognitive impairment (EMCI). This study built a graph convolutional module with multi-dimensional information interaction based on the residual connection architecture to improve the model's recognition effect on potential AD populations. The proposed modeling scheme can simultaneously take into account the unique characteristics of individuals within the potential diseased population and the correlation patterns between multiple data sources. The results of the control experiment show that the constructed model has an identification accuracy rate of 92.86% for EMCI, with an AUC of 0.93, which is 8.34% and 8.14% higher than the baseline model, respectively, fully verifying the superiority of the model [10].

## 4. Current limitations and future prospects

Among the various algorithm models reviewed in the past, there has been a significant improvement in classification accuracy. However, the common drawbacks of these methods remain extremely obvious.

Relying solely on a single imaging modality to obtain information related to pathology has a very obvious limitation in terms of the dimensions covered. Although sMRI has a high sensitivity for brain region atrophy changes, it is almost unable to capture molecular-level pathological changes such as abnormal functional connectivity and tau protein deposition. The efficient receptive field and multi-axis attention fusion model proposed by Li Junqi has successfully reduced the parameter quantity from 38.13M to 18.3M and maintains a recognition accuracy of 97.29% in the AD/NC classification task. However, the input dimension of this model is limited to structural image data, and there are inherent deficiencies in the recognition of non-atrophic cognitive decline [5].

Currently, the preprocessing process of medical imaging data in terms of efficiency and robustness still has very obvious flaws. Cheng Yuezhi's skull removal scheme took nearly 9 hours to process 730 samples. Although the performance of this scheme is better than some similar tools on the market, when faced with actual clinical data at a scale of tens of thousands, it may still become the performance bottleneck of the entire process. There is currently insufficient verification to support the stability and adaptability of this method to images collected by different field strengths and equipment from different manufacturers [6].

Although the academic community has noticed the key difficulty of cross-center generalization in the application of models, no practical solution has been constructed yet. Li Junqi's research team achieved an AUC performance value of 91.7% on the OASIS dataset [5], while most related studies currently only use the ADNI single-series dataset as the basis for verification. The current research does not involve the assessment of multi-center, multi-machine, and multi-ethnic populations, which means the applicability of high-performance artificial intelligence models still needs further verification.

Multimodal fusion technology can significantly improve the accuracy of model predictions. For example, the MDLCM developed by Gong Lejun has an accuracy of 90.5% in the three-classification task [8]. Sun Haoran's EMCI recognition model based on graph convolution networks

DOI: 10.54254/2755-2721/2026.37966

has a task accuracy of 92.86%. However, this technology is still constrained by the high threshold of multimodal data acquisition [10]. Not all clinical medical institutions can routinely conduct genetic imaging data, DTI, EEG, etc. special examinations. There are also practical implementation difficulties such as missing modal data, inconsistent time windows, and spatial alignment deviations. The systematic exploration in this field within the academic research domain is still insufficient.

In the field of brain science interpretability research, Liu et al used the gradient-weighted class activation mapping algorithm to complete the positioning of key brain regions and abnormal neural connections, which laid an important foundation for subsequent research in this field. However, from the perspective of clinical diagnosis and treatment practical application, the generated heat maps currently do not have the qualifications to serve as formal diagnostic evidence, and there are still obvious deficiencies in the correlation between the model decision logic and the neural pathological physiological mechanism [9].

Most academic discussions in the current academic circle focus on the horizontal classification dimension of diseases. However, in the longitudinal analysis direction of disease progression paths and the prediction of dementia conversion nodes, there are still obvious gaps in related research. Cui Yanan's improved 3D-Mamba model achieved an accuracy of 82.7% in the three-classification task, but it did not build a dedicated modeling system for the dynamic evolution process of MCI to AD conversion [7].

Future research exploration can be promoted from multiple aspects, such as the development of lightweight technologies and related research on edge-end deployment, which still have broad and in-depth exploration value. The R&D team led by Li Junqi has reduced the model's parameter size to 18.3M. If combined with techniques such as neural architecture search and knowledge distillation, the model size can be further reduced to fit practical application scenarios with limited resources, such as community screening [5].

By introducing a multi-center federated learning architecture, this problem of data silos can be effectively and practically solved. Achieving the collaborative aggregation of model parameters without each party disclosing the original images can enhance adaptability to heterogeneous data and effectively protect the personal privacy information of the research subjects.

Using self-supervised and semi-supervised learning frameworks, a pre-training program can be carried out based on a large number of unlabeled medical images, significantly reducing the reliance on fine-grained labeled tags. This solution demonstrates significant application advantages in clinical scenarios with ambiguous disease stage boundaries such as MCI and EMCI.

In the application scenarios of multimodal fusion, there are obvious research gaps in standardized operation processes and missing value handling mechanisms, and systematic in-depth exploration is urgently needed. One can first attempt to incorporate plasma biomarkers such as p-tau181, Aβ42/Aβ40 and low-cost collected modalities such as digital cognitive assessment into the fusion architecture, thereby filling the gap in the insufficient coverage of high-end imaging modalities and simultaneously developing a multi-modal fusion framework with high robustness for the missing modalities.

Building a longitudinal cohort model and transforming the prediction framework can release the potential capabilities for practical clinical applications. Traditional research usually only conducts classification for AD and NC, while individualized risk prediction for MCI patients developing into dementia in the future can directly provide important key support for early clinical intervention decisions.

It is necessary for the interpretability module to achieve a deeper integration with pathological prior knowledge. For example, embedding the anatomical limitations of known atrophied brain

DOI: 10.54254/2755-2721/2026.37966

regions into the attention mechanism, or making the model output capable of constructing an evidence chain that corroborates autopsy pathological reports. Such practices can make the reasoning process of the algorithm more in line with the inherent operating rules of neuropathology, thereby effectively reducing the acceptance resistance in clinical scenarios.

Most similar research methods can only achieve performance verification in standardized experimental data sets with controlled norms, while prospective effectiveness evaluations for clinical real scenarios are largely lacking. Only by integrating an AI model that has undergone multiple meticulous verifications into the actual clinical diagnosis process and conducting systematic tests on its sensitivity, specificity for auxiliary diagnosis and acceptance rate by clinical medical staff in outpatient diagnosis scenarios, can this technology be truly promoted to industrialization and large-scale application.

## 5. Conclusions

This article reviews the research progress of deep learning-assisted diagnosis of Alzheimer's disease from two dimensions: single-modal brain image analysis and multi-modal information fusion.

In the field of single-modal medical image analysis, current research mainly focuses on improving image quality, optimizing feature extraction, and controlling model complexity. Relevant methods have shown certain application potential in the tasks of identifying and classifying brain imaging features of Alzheimer's disease. However, due to the relatively limited information contained in a single imaging modality, the diagnostic performance is still susceptible to factors such as data quality, sample differences, and model generalization ability. Therefore, how to enhance the stability and clinical applicability of the model while improving the recognition performance remains a key issue for single-modal research.

In the multi-modal research field, existing methods mainly integrate multi-source information such as brain imaging, biomarkers, and clinical assessment to depict the pathological features and cognitive changes of Alzheimer's disease from different perspectives. Compared with single-modal analysis, multi-modal fusion can provide a more comprehensive disease characterization and shows great potential in early identification and classification prediction. However, issues such as data heterogeneity, information loss, and feature alignment between different modalities still limit the stability of model performance and its clinical promotion.

The current research results in the academic community have confirmed that integrating scientific and multi-dimensional information, combined with a suitable deep neural network architecture, can truly effectively improve the comprehensive performance of medical diagnosis. However, at the current stage, there are many deficiencies in the main technical approaches: the generalization ability of the model in cross-medical-center scenarios is significantly insufficient, the cost of multi-modal medical data fusion is high and alignment is difficult, the model's interpretability cannot meet the actual needs of clinical applications, and it has not been specifically designed for longitudinal dynamic prediction tasks. If future research can conduct more in-depth exploration in aspects such as model application, unified multi-modal data criteria, and prospective clinical validation, the computer-aided diagnosis technology for Alzheimer's disease has the potential to be truly integrated into the entire process of early screening and precise intervention.

## References

[1] Cohen, J., Mathew, A., Dourvetakis, K. D., et al. (2024). Recent research trends in neuroinflammatory and neurodegenerative disorders. *Cells, 13*(6), 511.

DOI: 10.54254/2755-2721/2026.37966

[2] Teleanu, R. I., Niculescu, A.-G., Roza, E., et al. (2022). Neurotransmitters—key factors in neurological and neurodegenerative disorders of the central nervous system. *International Journal of Molecular Sciences, 23*(11), 5954.
[3] Chu, Y., & Xu, W. (2022). A review of early classification research of Alzheimer's disease based on computer-aided diagnostic technology. *Computer Engineering and Science, 44*(5), 879.
[4] Nan, P. (2025). *Research on brain disease diagnosis algorithm based on multi-information features* [Degree thesis, Yantai University]. https://doi.org/10.27437/d.cnki.gytdu.2025.000252
[5] Li, J. (2025). *Research on Alzheimer's disease classification method based on MRI* [Degree thesis, Shenyang University of Technology]. https://doi.org/10.27322/d.cnki.gsgyu.2025.000136
[6] Cheng, Y. (2023). A study on the classification of Alzheimer's disease based on brain magnetic resonance imaging [Degree thesis, Northeastern University]. https://doi.org/10.27007/d.cnki.gdbeu.2023.001719
[7] Cui, Y. (2025). A study on the identification method of Alzheimer's disease combining MRI images and indicators [Degree thesis, Xi'an Petroleum University]. https://doi.org/10.27400/d.cnki.gxasc.2025.000967
[8] Gong, L., Wei, X., Ji, Y., et al. (2025). Alzheimer's disease classification model based on multimodal deep learning. *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 45*(6), 97–105. https://doi.org/10.14132/j.cnki.1673-5439.2025.06.011
[9] Liu, Y. (2025). Alzheimer's disease classification model based on multimodal data: Research on deep learning-assisted diagnosis of Alzheimer's disease [Degree thesis, Donghua University]. https://doi.org/10.27012/d.cnki.gdhuu.2025.001787
[10] Sun, H. (2024). Research on intelligent diagnosis algorithm for Alzheimer's disease based on multimodal image fusion [Degree thesis, Henan University]. https://doi.org/10.27114/d.cnki.ghnau.2024.001363.