Research Status of Multimodal Medical Image Fusion In AI-Assisted Diagnosis of Alzheimer's Disease

Zhixing Ni *

School of Medical Imaging, Mudanjiang Medical University, Mudanjiang, 157011, China

* Corresponding Author Email: Ivan005768@outlook.com

Abstract. As the global population ages, the socioeconomic burden caused by Alzheimer's disease (AD) is rising. Existing clinical interventions mostly focus on delaying the course of the disease and lack support for the accuracy and timeliness of early diagnosis, which has become a key bottleneck restricting the prevention and treatment of AD. This study aims to explore the application value of multi-modal medical image fusion technology in the early diagnosis of AD under the guidance of artificial intelligence (AI), and to verify its advantages and feasibility compared with single-modal imaging diagnosis. The research will integrate multi-modal imaging data such as structural magnetic resonance imaging (sMRI), positron emission tomography (PET) and diffusion tensor imaging (DTI), and complete the extraction, matching and fusion of image features through AI models such as convolutional neural networks (CNN) or Transformer to build an early diagnosis model for AD; and use clinically confirmed AD high-risk groups (patients with mild cognitive impairment) and healthy controls as the research subjects If this study proves feasible, it can provide efficient and accurate technical support for early AD screening, thereby providing a new practical path for delaying the course of the disease and reducing social costs.

Keywords: Alzheimer's disease; artificial intelligence; multimodal medical imaging; medical image fusion.

1. Introduction

Fundamental pathogenic characteristics of Alzheimer's disease (AD) include tau protein aggregation, β-amyloid (Aβ) deposition, and hippocampus shrinkage. The prevalence of AD is increasing globally as a result of population aging, placing an increasing strain on social economies, healthcare systems, and patients' families [1]. Current medicines simply slow down the disease's progression, and there are still several diagnostic and therapy bottlenecks to be overcome, despite modern medicine's tremendous advancements.

Even though there have been some developments in AD diagnostic technologies recently, including imaging tests like Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI), as well as the identification of biomarkers in cerebrospinal fluid, which can aid in diagnosis [2], these technologies clearly have limitations. High-end equipment such as PET and MRI is expensive and difficult to popularize at the primary healthcare level, greatly restricting the promotion of early accurate diagnosis. Blood biomarker detection shows potential, however, there is currently a lack of sufficiently sensitive, specific, and easily generalizable detection methods, and the evidence for its clinical effectiveness in the general population is relatively limited.

Meanwhile, because in the field of medical diagnosis, with the development of imaging technology, the information provided by single-modal medical images is relatively limited. Moreover, the pathological process of AD (amyloid deposition, neurofibrillary tangles, and neuronal damage) begins before the appearance of clinical symptoms, so single-modal images are unable to fully capture these early pathological changes. Due to the limitations of AD diagnostic technologies and the shortcomings of single-modal medical images, research on multimodal medical image fusion has gradually become a hot field and is favored by researchers for its advantage of integrating more key information. The process of combining several images from various imaging modalities to create a fused image with a lot of information is known as multimodal medical image fusion. This improves the clinical usability of medical images. Through the application of imaging techniques including MRI, fluorodeoxyglucose Positron Emission Tomography (FDG-PET), amyloid Positron Emission

Tomography (Amyloid-PET), and the relatively newly developed Tau Positron Emission Tomography (Tau-PET), a multi-dimensional account of AD pathological characteristics spanning the "structure + function + molecule" aspects can be achieved. This ability boosts the potential of medical imaging in the diagnosis and prognosis of AD. Notably, the 2024 updated clinical guidelines released by the National Institute on Aging and Alzheimer's Association (NIA-AA) have incorporated multimodal imaging into the criteria for AD diagnosis and staging, which further underscores its clinical value [3].

This article aims to clarify the mechanism of action of multimodal medical image fusion in the early diagnosis, prognosis prediction, and treatment evaluation of AD, and analyze the supporting value of AI and deep learning for fusion technologies. It will elaborate on four aspects: the association between AD and medical imaging, technical methods of multimodal image fusion, the collaborative application of AI and image fusion, and current shortcomings and prospects. Ultimately, it intends to provide a theoretical basis for the clinical diagnosis of AD.

2. AD and Medical Imaging

2.1. Core Pathological Features of AD and Diagnostic Requirement

AD is primarily caused by hippocampal atrophy and related with the buildup of tau protein, which is characterized by molecular abnormalities prior to structural damage. Pathologists' microscopic investigations have indicated the presence of protein fragments and amyloid deposits. A connection between tau protein accumulation and B-amyloid, the last stage of the disease's pathophysiology, has been confirmed by numerous researchers. In the brains of AD patients, tau protein aggregation and β-amyloid deposits progress in a certain geographic pattern: Tau protein progressively travels to the neocortex after preferentially accumulating in the medial temporal lobe (entorhinal cortex hippocampus) as the condition worsens. B-amyloid deposits often start in the prefrontal lobe and entorhinal cortex of the brain before progressively moving to the parietal and temporal lobes. The range and location of such deposits are not only closely related to the severity of AD. For instance, tau aggregation limited to the entorhinal cortex mostly corresponds to the mild cognitive impairment stage, while the spread of tau protein to the entire temporal lobe or even the parietal lobe often indicates progression to the AD dementia stage. They also activate immune cells like microglia in the brain, which triggers an excessive immune response and releases inflammatory factors that further damage surrounding neurons—ultimately forming a vicious cycle of "pathological deposition - immune damage - neuronal death" [4]. These findings require that imaging technologies for AD diagnosis cover both the molecular pathology and structural function dimensions.

2.2. Characteristics and Limitations of Single-Modal Medical Imaging Technologies

Traditionally, AD has been detected using invasive techniques. In recent years, numerous neuroimaging tools have been created to detect AD, including PET visualizes and measures amyloid plaques in the brain using particular radioactive tracers. Electrical activity is recorded via electroencephalography (EEG); by identifying variations in oxygen levels in different brain regions (such as voxels), functional magnetic resonance imaging (fMRI) assesses brain activity. Furthermore, because of its great spatial resolution and capacity to compare soft tissues, MRI is utilized to investigate the anatomical features of the brain.

Although the above imaging methods provide a rich visualization basis for clinical diagnosis, there are still a lot of issues with the aforementioned technologies: By detecting structural alterations like hippocampus atrophy, which can only reveal neurodegenerative lesions and not directly identify the two main clinical characteristics of AD (β-amyloid and tau protein tangles), structural imaging methods like MRI/CT aid in diagnosis. Since structural changes in the early stage of AD are not yet obvious, the rate of missed diagnosis is relatively high. Functional imaging techniques such as PET/ Single-Photon Emission Computed Tomography (SPECT) have significant inconveniences in AD diagnosis due to their high cost and limitations of tracers. Moreover, EEG, which records brain electrical activity, cannot achieve precise localization and is easily affected by external factors.

2.3. Concept and Significance of Multimodal Image Fusion

Due to the obvious shortcomings of single-modal imaging in AD diagnosis—structural imaging cannot directly detect core pathological markers such as Aβ and tau, resulting in a high rate of missed diagnosis in the early stage; functional imaging has high costs and tracer limitations, and neither of them can fully meet the pathological diagnosis requirements of AD. So multimodal medical image fusion becomes particularly important [5]. The other reason is that functional imaging (PET) lacks anatomical features but can provide quantitative metabolic and functional information related to the disease, while structural imaging (CT/MRI) can provide specific information about anatomical structures through high contrast and spatial resolution. The combination of the two (functional and structural imaging) can better describe the disease condition and the characteristics of brain damage. For example, the combination of fMRI and EEG, which are widely used in neuroscience, can provide both high spatial resolution and high temporal resolution of brain dynamics. The core advantage of EEG is its high temporal resolution (capturing brain electrical activity at the millisecond level), but its spatial localization is poor; fMRI has high spatial resolution (displaying structures at the millimeter level) but low temporal resolution. The fusion of the two can meet the dual needs of dynamic activity monitoring and precise localization [6].

Regarding the latest updated 2024 National Institute on Aging and Alzheimer's Association (NIA-AA) clinical guidelines [3], the descriptions of features for AD diagnosis, staging, and prognosis all involve the application of multimodal medical imaging, including PET, MRI, and FDG-PET, along with the recently prevalent Tau-PET. This adjustment marks that multimodal image fusion has become a core technical direction in AD diagnosis. Its advantage of multi-dimensional information complementation can effectively make up for the limitations of single-modal imaging and provide a standardized basis for the early diagnosis and staging of AD.

3. Technical Methods of Multimodal Medical Image Fusion and Their Applications in AD

Since the advent of image fusion, it has developed rapidly in various fields: especially in military and civilian applications, including the fusion of infrared and visible light, material analysis, etc. In the medical field, with the development of medical imaging, a variety of imaging devices have emerged, such as Ultrasound (USG), CT, MRI, PET, and SPECT. Different devices have distinct characteristics and specialized fields based on their properties. Multimodal medical image fusion achieves the accurate presentation of AD pathological features by integrating complementary information from different modalities. The core technical methods can be divided into three categories: spatial domain fusion, transform domain fusion, and AI/deep learning-based fusion. Each method has different technical characteristics and application scenarios in AD image fusion [7].

3.1. Spatial Domain Fusion

Spatial domain fusion involves integrating spatial information (such as pixels, saturation, intensity, etc.) from different images to generate a fused image that contains key multimodal information (such as the spatial correspondence between anatomical structures or functional distributions). Its main purpose is to highlight the complementarity of various modalities, such as anatomical details, functional metabolic status, and pathological information. For example, in the multimodal medical image fusion technology system, the fusion method based on the IHS (Intensity-Hue-Saturation) domain is a typical spatial domain fusion technology. The core theoretical basis of this method is the IHS color model proposed by American scientist Munsell. This model is based on the mechanism of human visual perception and clearly explains the ways by which the visual system processes image brightness and color. Its core characteristics can be summarized in two points: First, the Intensity (I) component: it only reflects the brightness information of the image, is completely independent of the color attributes of the image, and the brightness representation will not change due to color variations. Second, the Hue (H) and Saturation (S) components: they are directly related to the human perception logic of color—hue determines the "color type" of the image (such as blue, red), and saturation determines the "shade level" of the color; together, they constitute the core color attributes of the image [1]. These characteristics give it unique advantages in medical image fusion: For instance, the functional metabolic information of AD patients from PET (such as Aβ deposition regions) can be mapped to MRI structural images using the IHS model. By adjusting the Intensity (I) component of the IHS model to retain the functional signals of PET while maintaining the Hue (H) and Saturation (S) to ensure the clarity of the anatomical structure in MRI, the visualized fusion of functional and structural information can be achieved. This can preserve more anatomical information, reduce color distortion, and improve the recognition efficiency of early pathological regions in AD.

3.2. Transform Domain Fusion

Transform domain fusion, also known as frequency domain fusion, differs from spatial domain fusion in that it does not directly perform fusion in the image spatial domain. Instead, it first converts the image from the spatial domain to the transform domain (such as frequency domain, wave domain) using digital algorithms, extracts and reorganizes key multimodal features, and then converts it back to the spatial domain through inverse transformation to form a fused image. It is suitable for AD image analysis scenarios where high-frequency details (such as brain region edges, small pathological lesions) need to be retained. Its characteristic lies in the separation and reorganization of image information. Most transform domain fusion methods utilize the Multiscale Transformation (MST) theory, and based on this theory, there are three commonly used transformations: Nonsubsampled Contourlet Transform (NSCT) fusion, Nonsubsampled Shearlet Transform (NSST) fusion, and Discrete Wavelet Transform (DWT) fusion [7]. With the rapid development of this field, new methods combining the above methods with other algorithms have also emerged. For example, Madanala and Jhansi Rani combined the advantages of wavelet transform in frequency and time localization and the shift-invariance advantage of NSCT, and proposed a cascaded fusion framework in the Discrete Wavelet Transform (DWT) + NSCT domain [8]. By combining the advantages of DWT (good at capturing low-frequency global features) and NSCT (good at retaining high-frequency details such as brain region edges), their cascading can avoid information loss from a single transformation. At the same time, the shift-invariance reduces artifacts in the fused image, improves the contrast for diagnosing AD features (such as hippocampal atrophy), reduces redundancy in the fused image, and enhances the contrast of diagnostic features [7]. A typical application of transform domain fusion in AD diagnosis was researched: they proposed an AD classification method based on fMRI image transformation and canonical correlation analysis. By exploring the correlation between fMRI transformation features and AD pathological labels, they realized the correlation modeling of cross-modal features, which effectively improved the classification accuracy of AD in multiple stages (including Subjective Memory Complaint (SMC) and Mild Cognitive Impairment (MCI)) [9].

3.3. AI and Deep Learning-Based Multimodal Image Fusion Technology

With the development of AI technology, deep learning-based fusion has become the mainstream direction of AD image fusion. Through multi-layer neural networks, deep learning automatically extracts high-level multimodal features, greatly promoting the planning and development of AD diagnosis and treatment. Multimodal medical image fusion improves the accuracy of AD diagnosis through complementary data and has inherent advantages in optimizing diagnostic processes and clinical decision-making.

3.3.1 Deep learning

Deep learning is a technology that relies on artificial neural networks to perform multi-level feature extraction and complex pattern learning by simulating the structure of the human brain's nervous system. The Convolutional Neural Network (CNN) is a typical deep learning model [10]. The application of AI and deep learning in medical imaging has transformed this field by enhancing diagnostic capabilities and workflow efficiency. Compared with traditional image analysis methods, deep learning has significant advantages. Traditional image analysis both requires a large amount of professional knowledge and is subject to certain subjective factors. In contrast, deep learning models can learn and summarize relevant features from a large amount of raw image data, which reduces time costs and minimizes subjective differences caused by human factors.

3.3.2 Deep learning and multimodal medical image fusion

The support of deep learning for multimodal image fusion is mainly reflected in two aspects: first, it is widely used in medical image segmentation and medical image registration [11, 12, 14]. Image registration is a prerequisite for multimodal fusion (such as aligning the brain region coordinates of PET and MRI), and image segmentation is a key step in fusion (such as fusing only regions of interest like the hippocampus to reduce interference from irrelevant regions). Deep learning provides technical support for multimodal fusion by improving registration accuracy (such as CNN-based rigid registration) and segmentation efficiency (such as brain region segmentation using U-Net). With its deep network structure for automatic learning and strong capabilities in feature extraction and nonlinear mapping, deep learning enables the efficient fusion of multimodal images (such as MRI and PET, CT and ultrasound), providing more comprehensive image support for clinical diagnosis and analysis. Second, it performs feature fusion. Through deep networks such as DenseNet and Transformer, it automatically learns high-level multimodal features [15, 16]. For example, a spatial-frequency domain fusion deep learning network (SFNet) based on 3D MRI was proposed, which extracts spatial local features through an improved DenseNet to enhance the efficiency and accuracy of AD diagnosis [17].

4. Collaborative Application of AI and Multimodal Medical Image Fusion in AD

4.1. Application of Multimodal Image Fusion in Early Diagnosis and Prognosis Prediction

Early diagnosis of AD (e.g., MCI stage) and prognosis prediction (e.g., risk of MCI conversion to AD) are core clinical needs. Multimodal image fusion based on AI and deep learning significantly improves the accuracy of diagnosis and prediction through the integration of "molecular pathology + structural function" information.

4.1.1 Application in early diagnosis of AD

In terms of diagnosis, according to the NIA-AA criteria, Mild Cognitive Impairment (MCI) caused by AD is used to describe the pre-dementia stage of AD [18]. Its diagnostic criteria emphasize biomarker data, which includes the application of multimodal medical imaging. The medical imaging methods used fall into two categories: one type is imaging for identifying amyloid positivity, such as amyloid-PET; the other type is imaging for identifying neuronal damage, such as MRI and Tau-PET. For example, Brier et al. used amyloid and Tau-PET imaging to study the topological relationship between Aβ and tau proteins in 36 cognitively normal elderly individuals and 10 patients with mild AD. The study found that the PET imaging data of tau protein and Aβ had strong autocorrelation among various regions of interest but exhibited unique topological structures. Moreover, the deposition of tau protein and Aβ in the temporal lobe better reflected the dementia state and was more predictive of cognitive performance than Aβ deposition in other regions [19].

4.1.2 Application in prognosis prediction of AD

In terms of prognosis judgment: Many researchers have focused on the role of different imaging methods in predicting MCI conversion. For example, Jack et al. used key indicators from amyloid-PET: global or regional amyloid (Aβ) deposition load (SUVR value) and a key indicator from MRI: hippocampal volume. Through a series of algorithms, they concluded that MCI patients with positive amyloid were more likely to progress to AD than those with negative amyloid. Among positive patients, hippocampal atrophy indicated a shorter conversion period [20]. In relatively early studies, to predict the conversion of MCI to AD, Multimodal characteristics are typically concatenated into a composite feature set in order to forecast the conversion of MCI to AD. A classifier is built after feature selection is completed to lessen interference from high-dimensional data. For example, Ritter et al. constructed a set of multimodal features including MRI, FDG-PET, CSF, and neuropsychological tests [21]. After feature selection, a Support Vector Machine (SVM) classifier was used to predict MCI conversion within 3 years based on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The 10-fold cross-validation achieved an accuracy of 73% in classifying 86 converters and 151 non-converters. Compared with the accuracy of approximately 65% for a single modality (such as MRI alone) during the same period [22, 23], multimodal fusion increased the accuracy by nearly 8 percentage points, verifying the complementary value of multimodal features.

4.2. Application of Multimodal Image Fusion in AD Treatment Effect Evaluation

In clinical trials and clinical treatment of anti-pathological drugs for AD (such as anti-Aβ drugs), multimodal image fusion can quantify treatment effects through "multi-dimensional pathological monitoring", avoiding the limitations of single-modal evaluation. In anti-Aβ drug clinical trials, the change in Aβ load is detected by amyloid-PET, and the reduction in brain atrophy rate is evaluated by combining with MRI to determine the effectiveness of the drug. For example, the Clarity AD clinical trial of Lecanemab is an important study evaluating the therapeutic effect of early AD. In terms of detecting drug efficacy, amyloid-PET was used to quantitatively evaluate the global amyloid Aβ deposition load, and Tau-PET imaging was used to assess the distribution and load of tau protein tangles in the brain (especially in the medial temporal lobe and neocortex regions) to comprehensively determine the therapeutic effect of the drug [24, 25].

The advantages of using multimodal medical imaging to determine drug efficacy include: (1) Multi-dimensional coverage of the pathological process: Amyloid-PET monitors Aβ deposition (etiological level), Tau-PET monitors neuronal damage (progressive level), and MRI monitors brain atrophy (structural level), avoiding the limitation of a single modality focusing only on a certain pathological stage; (2) Quantitative evaluation of efficacy: It can objectively measure changes in Aβ/tau load, and the rate of brain atrophy can quantify the drug's effect in delaying structural damage, reducing errors from subjective judgment; (3) Prediction of long-term prognosis: Changes in the medial temporal lobe signal on Tau-PET can early predict the protective effect of the drug on cognitive function, providing a basis for clinical adjustment of treatment plans.

5. Shortcomings and Future Prospects of Multimodal Medical Image Fusion in AD

The development of medical image fusion has extended from the spatial domain and transform domain to the field of deep learning. Its rapid development also indicates a strong demand for computer-aided clinical diagnosis. Different researchers have proposed different fusion methods, and each method has its own advantages in terms of different evaluation indicators. Multimodal medical imaging has demonstrated important value in the early diagnosis, prognosis judgment, and clinical guidance of AD. However, limited by the special pathophysiological mechanism of AD and technological development, there are still many bottlenecks to be addressed.

5.1. Existing Core Shortcomings

5.1.1 Insufficiencies in imaging technology and pathological qualification

Firstly, due to the particularity of AD pathophysiology: Aβ deposition and tau aggregation usually occur 5 to 20 years before the appearance of clinical symptoms, which leads to deficiencies in early screening and feature identification. For example, MRI can detect hippocampal atrophy, but hippocampal atrophy usually occurs after Aβ deposition and tau aggregation (i.e., the middle and late pathological stages), so it is impossible to detect abnormalities in the preclinical stage, resulting in missed intervention opportunities. In addition, hippocampal atrophy is not a unique manifestation of AD; conditions such as Lewy body dementia and vascular cognitive impairment also often present with reduced hippocampal volume [5]. Although Amyloid-PET can identify early Aβ deposition, it has false negatives (e.g., familial AD or combined with other neurological diseases leading to abnormal Aβ deposition patterns) and false positives (e.g., some healthy elderly individuals, especially those over 80 years old, may also have physiological Aβ deposition). As for Tau-PET imaging, its sensitivity in the preclinical stage is still insufficient due to current equipment limitations.

Secondly, there are differences in the medical image information obtained by different sensors. Current research focuses on the fusion of two modalities, while the fusion of three modalities is rarely studied. The research bottlenecks of three-modal fusion include: (1) Strong data heterogeneity: For example, the data formats (3D images vs. 1D signals) and resolutions (millimeter-level vs. millisecond-level) of PET (functional), MRI (structural), and EEG (electrophysiological) differ greatly, making registration much more difficult than that of two modalities [7, 25]; (2) High computational cost: The data volume of three modalities is 1.5-2 times that of two modalities. Traditional algorithms are difficult to process efficiently and require more complex deep learning frameworks (such as Transformer), but existing hardware (such as GPUs) cannot support large-scale training [7, 25]; (3) Difficulty in sample acquisition: Patients need to complete examinations of three modalities, which is time-consuming (e.g., PET takes 1 hour, MRI takes 30 minutes) and costly, resulting in a shortage of multi-center datasets and difficulties in repeated verification of research. Research on two-modal fusion mainly focuses on the fusion of MRI/CT, MRI/PET, and MRI/SPECT. In the future, multimodal fusion and algorithm compatibility will be a challenging topic.

5.1.2 Application bottlenecks of deep learning technology

Although deep learning has improved the fusion effect, related research still faces a "triple bottleneck" of "data - model - clinical adaptation": In terms of data: The main current bottlenecks are large differences in data distribution and data scarcity. Firstly, differences in equipment (e.g., GE vs. Siemens PET) and imaging agents (e.g., 18F-florbetapir vs. 18F-florbetaben) in different hospitals lead to inconsistent data distribution. The accuracy of models decreases when applied to cross-center data, and standardized processing is difficult, resulting in large differences in data distribution [26]. Secondly, the quantitative analysis of multimodal images requires manual processing by AI algorithms and radiologists, with each case taking 2-3 hours to annotate, leading to a shortage of high-quality annotated datasets [18].

In terms of models: Acquiring large datasets is a difficulty in medical image research. Deep learning training is time-consuming, with complex frameworks and high requirements for computer hardware configuration. Simplifying training models, proposing new training models, and conducting parallel training are important components of this research field. Some fusion methods rely on accurate image registration and have poor independence [7, 8].

In terms of clinical adaptation: The main problem is the poor interpretability of models. For example, the "black box characteristic" of CNN (only able to input data and obtain results without a clear understanding of the process) makes it impossible to clarify which features (such as tau signal intensity, hippocampal volume) play a key role in the diagnosis during the fusion process. Clinicians find it difficult to trust the model results, which limits clinical translation.

5.1.3 Issues of equipment cost and data standardization

Currently, the cost of equipment and examinations is high. In particular, equipment such as PET has a high unit price and requires imported contrast agents, which greatly increases the cost of a single examination and creates difficulties for screening [27]. Secondly, some equipment such as functional MRI has high requirements for magnetic field strength and long scanning time (30-60 minutes), as well as high site requirements (needing high magnetic field strength of 1.5T or even 3.0T or above), making it difficult for ordinary hospitals to carry out [28]. In addition, there is a lack of "unified standards" for image data from different equipment and imaging agents (e.g., inconsistent SUVR thresholds for Aβ-PET of different brands), making it difficult to integrate data from multi-center studies and affecting the consistency of diagnosis.

5.2. Progress and Future Outlook

In recent years, multimodal medical image fusion has achieved significant results in AD research: With the help of deep learning, AI models, and biomarker analysis, breakthroughs have been made in fields such as early diagnosis and multimodal data integration. Typical technologies include the newly developed GFE-Mamba model (generating high-quality PET images from MRI to achieve dynamic fusion of three modalities, with high accuracy in predicting conversion and solving problems of difficult registration and high computational cost) [29], the AD-Transformer framework (realizing seamless fusion of multimodal data, providing an efficient and unified analysis framework and improving model interpretability) [30], and virtual PET algorithms (simulating PET signals through MRI to reduce examination costs). These technologies have specifically addressed the current core bottlenecks of multimodal fusion, laying a foundation for clinical implementation. Current research focuses on three directions: First, the development of specific imaging biomarkers (such as the PET imaging agent MK-6240 targeting tau protein, which improves the sensitivity of early tau aggregation) [31]. Second, the promotion of AI integration applications (such as integrating imaging, cerebrospinal fluid/blood biomarkers, and clinical information to construct an "imaging - molecular - clinical" combined diagnostic model, improving diagnostic specificity and prognostic accuracy). Third, the development of low-cost technologies (such as virtual PET algorithms based on ordinary MRI, reducing examination costs and improving accessibility).

Although multimodal imaging still has technical limitations, it remains a core tool for AD diagnosis and research. Its shortcomings will be gradually addressed with technological progress and multi-disciplinary integration. With the rapid development of deep learning research in the field of AI, and its effectiveness having been confirmed in computer vision fields such as medical imaging, the application of multimodal imaging in the early diagnosis and prognosis evaluation of AD is expected to further benefit from the iteration of deep learning.

In the future, efforts should focus on three aspects to promote the clinical popularization of multimodal medical image fusion: First, continue to develop specific imaging biomarkers to enhance the ability to capture early pathological features of AD; second, optimize deep learning models to improve the efficiency of multimodal data integration and model interpretability; third, deepen the development of low-cost technologies to reduce equipment and examination costs and improve accessibility in primary medical institutions.

6. Conclusion

This study analyzes the application of multimodal medical image fusion in AI-assisted diagnosis of Alzheimer's Disease (AD). The results show that the characteristic of AD pathology—"molecular abnormalities preceding structural damage"—leads to limitations of single-modal imaging (MRI, PET, EEG, etc.), such as early missed diagnosis, poor localization, or high cost. In contrast, multimodal image fusion (three types of technologies: spatial domain, transform domain, and AI/deep learning-based) can integrate multi-dimensional "structure + function + molecule" information, effectively making up for the shortcomings of single-modal imaging. Specifically, spatial domain fusion (e.g., IHS model) realizes the visual superposition of PET and MRI for function-anatomy; transform domain fusion (e.g., DWT+NSCT cascading) improves the recognition of AD pathological details (e.g., hippocampal atrophy); deep learning-based fusion (e.g., SFNet, AD-Transformer) further enhances the efficiency and accuracy of AD diagnosis through automatic feature extraction and precise registration.

In clinical applications, the combination of multimodal fusion and AI technology can significantly improve the accuracy of early AD diagnosis (e.g., distinguishing MCI from healthy individuals), prognosis prediction (e.g., evaluating the risk of MCI conversion to AD), and treatment effect evaluation (e.g., quantifying the efficacy of anti-Aβ drugs). It has also been included in the 2024 NIA-AA guidelines for AD diagnosis and staging criteria. However, current technologies still face issues such as insufficient matching between imaging and pathology (e.g., low early sensitivity of Tau-PET), deep learning bottlenecks (data scarcity, poor model interpretability), high equipment costs, and lack of unified data standards.

References

[1] Javaid S F, Giebel C, Khan M A B, et al. Epidemiology of Alzheimer’s disease and other dementias: Rising global burden and forecasted trends. F1000Research, 2021, 10: 425.

[2] El-Gamal F, Elmogy M M, Ghazal M, et al. medical imaging diagnosis of early Alzheimer’s disease. Front Biosci, 2018, 23: 671-725.

[3] Jack Jr C R, Andrews J S, Beach T G, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimer's & Dementia, 2024, 20(8): 5143-5169.

[4] Twarowski B, Herbet M. Inflammatory processes in alzheimer’s disease—pathomechanism, diagnosis and treatment: a review. International journal of molecular sciences, 2023, 24(7): 6518.

[5] Nasrallah I M, Wolk D A. Multimodality imaging of Alzheimer disease and other neurodegenerative dementias. Journal of Nuclear Medicine, 2014, 55(12): 2003-2011.

[6] Guo Zhe, Li Xiang, Huang Heng, et al. Deep learning-based image segmentation on multimodal medical imaging. IEEE transactions on radiation and plasma medical sciences, 2019, 3(2): 162-169.

[7] Huang Bing, Yang Feng, Yin Mengxiao, et al. A review of multimodal medical image fusion techniques. Computational and mathematical methods in medicine, 2020, 2020(1): 8279342.

[8] Lu Tang, Yu Hui, Hang Yang, et al. Medical image fusion quality assessment based on conditional generative adversarial network. Frontiers in Neuroscience, 2022, 16: 986153.

[9] Jia Hongfei, Wang Yu, Duan Yifan, et al. Alzheimer’s disease classification based on image transformation and features fusion. Computational and mathematical methods in medicine, 2021, 2021(1): 9624269.

[10] LeCun Yann, Bengio Yoshua, Hinton Geoffrey. Deep learning. nature, 2015, 521(7553): 436-444.

[11] Wang Risheng, Lei Tao, Cui Ruixia, et al. Medical image segmentation using deep learning: A survey. IET image processing, 2022, 16(5): 1243-1267.

[12] Balakrishnan G, Zhao A, Sabuncu M R, et al. Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging, 2019, 38(8): 1788-1800.

[13] Yang X, Kwitt R, Styner M, et al. Quicksilver: Fast predictive image registration–a deep learning approach. NeuroImage, 2017, 158: 378-396.

[14] Hu Y, Modat M, Gibson E, et al. Weakly-supervised convolutional neural networks for multimodal image registration. Medical image analysis, 2018, 49: 1-13.

[15] Huang Gao, Liu Zhuang, Van Der Maaten Laurens, et al. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[16] Xu Peng, Zhu Xiatian, David A.Clifton. Multimodal learning with transformers: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 12113-12132.

[17] Yang Xinyue, Liu Meiliang, Xu Yunfang, et al. SFNet: A Spatial- Frequency Domain Deep Learning Network for Efficient Alzheimer's Disease Diagnosis. arxiv preprint arxiv:2507.16267, 2025.

[18]] Liu Xiaonan, Chen Kewei, Teresa Wu, et al. Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Translational Research, 2018, 194: 56-67.

[19] Brier M R, Gordon B, Friedrichsen K, et al. Tau and Aβ imaging, CSF measures, and cognition in Alzheimer’s disease. Science translational medicine, 2016, 8(338): 338ra66-338ra66.

[20] Jack Jr C R, Wiste H J, Vemuri P, et al. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer’s disease. Brain, 2010, 133(11): 3336-3348.

[21] Ritter K, Schumacher J, Weygandt M, et al. Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2015, 1(2): 206-215.

[22] Retico A, Bosco P, Cerello P, et al. Predictive models based on support vector machines: whole-brain versus regional analysis of structural MRI in the Alzheimer's disease. Journal of Neuroimaging, 2015, 25(4): 552-563.

[23] Guerrero R, Wolz R, Rao A W, et al. Manifold population modeling as a neuro-imaging biomarker: application to ADNI and ADNI-GO. NeuroImage, 2014, 94: 275-286.

[24] Van Dyck C H, Swanson C J, Aisen P, et al. Lecanemab in early Alzheimer’s disease. New England Journal of Medicine, 2023, 388(1): 9-21.

[25] Alam F, Rahman S U. Challenges and solutions in multimodal medical image subregion detection and registration. Journal of medical imaging and radiation sciences, 2019, 50(1): 24-30.

[26] Kasban H, El-Bendary M A M, Salama D H. A comparative study of medical imaging techniques. International Journal of Information Science and Intelligent System, 2015, 4(2): 37-58.

[27] McMahon P M, Araki S S, Sandberg E A, et al. Cost-effectiveness of PET in the diagnosis of Alzheimer disease. Radiology, 2003, 228(2): 515-522.

[28] Marchewka A, Kherif F, Krueger G, et al. Influence of magnetic field strength and image registration strategy on voxel-based morphometry in a study of Alzheimer's disease. Human brain map, 2014, 35(5): 1865-1874.

[29] Fang Zhaojie, Zhu Shenghao, Chen Yifei, et al. GFE-Mamba: Mamba-based AD Multi-modal Progression Assessment via Generative Feature Extraction from MCI[J]. arxiv preprint arxiv:2407.15719, 2024.

[30] Yu Qi, Ma Qian, Da Lijuan, et al. A transformer-based unified multimodal framework for Alzheimer's disease assessment. Computers in Biology and Medicine, 2024, 180: 108979.

[31] Pascoal T A, Therriault J, Benedet A L, et al. 18F-MK-6240 PET for early and late detection of neurofibrillary tangles. Brain, 2020, 143(9): 2818-2830.