Research Progress on Classification and Diagnosis of Alzheimer's Disease MRI Images Using Neural Network Methods

Weizhe Xiao *

Interdisciplinary Engineering (Science, Technology and Management), Macau University of Science and Technology, Macao, China

* Corresponding Author Email: 1250015402@student.must.edu.mo

Abstract. With the aging of the population increasing, the dangers of Alzheimer's disease are increasingly aggravated. Meanwhile, an accurate diagnostic method is important for today's society. This review aims to sort the literature about research progress depending on the classification and diagnosis of Alzheimer's disease using neural network methods. This review has discussed challenges about the data heterogeneity, incomplete modal imaging, and modal interpretability, through analyzing the applications of multimodal imaging data in the diagnosis of Alzheimer's disease, and summarized the application of data augmentation, feature fusion, and deep learning algorithms for deep neural network models in this area. Research has found that neural network methods offer significant advantages in enhancing the accuracy and reliability of Alzheimer's disease diagnosis. Therefore, it is necessary to conduct should conduct deeper research on the model's generalisability in the future, to promote clinical application. And merge together the related knowledge about traditional medicine and neural networks to realize a more accurate diagnosis of Alzheimer's disease, to relieve the social burden caused by Alzheimer's disease.

Keywords: neural networks, Alzheimer's disease, MRI imaging, deep learning, classification algorithms.

1. Introduction

Alzheimer's disease is a common neurodegenerative, that impacts the elder's life seriously. With the global ageing intensifies, AD's incidence rate is continuing to increase. It brings a heavy burden of society and family. Therefore, strengthening the early diagnosis and intervention for Alzheimer's disease is crucial to mitigate the harm caused by the condition. However, traditional methods for diagnosing Alzheimer's disease lack accuracy and reliability in their diagnostic assessment. Those methods are overly reliant on clinical symptom assessment and cognitive impairment tests, which are excessively subjective. So, the traditional methods are lack of accuracy and reliability. In order to overcome the difficulties posed by traditional medical diagnosis, there is a pressing need to develop more objective and accurate early diagnostic protocols, thereby creating greater possibilities for diagnosing Alzheimer's disease.

In recent years, with the development of medical imaging technology, in particular, the common application of magnetic resonance imaging (MRI), which provides hopes, offers new approaches for the accurate diagnosis of Alzheimer's disease. Neural network methods rely on their robust feature learning capabilities and end-to-end pattern recognition advantages. It could capture complex linear relationship when processing high-dimensional data, to overcome the limitation of low accuracy in traditional clinical diagnostics methods, and it is able to find the complex pattern that are difficult for the human brain to design and quantify, yet are crucial for diagnosis. The MRI method provides new technical means for achieving a more precise diagnosis. Therefore, the image-based diagnosis approaches for Alzheimer's disease are increasingly becoming a research priority, this approach demonstrates tremendous potential because of its powerful faeture learning capabilities and pattern recognition capabilities. It is worth noting that in order to get over the limitations of early feature representation methods, some scientists began to detect deep learning technology to automatically learn more discriminative high-level features. This approach provides a reliable basis for enhancing the classification accuracy and robustness of the model [1].

Significant progress has been made in the neural network-based diagnosis of Alzheimer's disease. However, there are still numerous challenges that persist in this field, including image data heterogeneity [2], handling incomplete imaging modalities [2], and model interpretability [3]. This review aims to systematize existing research findings on the classification and diagnosis of Alzheimer's disease patients using neural network methods, the application of multimodal imaging data in particular. This review will also analyse key issues in current research by summarizing data augmentation [4], feature fusion [5], and other mainstream applications of deep learning. Concurrently, the trend in neural network-based diagnostic approaches will be explored. This will provide theoretical guidance and technical support for the precise diagnosis of Alzheimer's disease.

2. Application of Multimodal Imaging in Alzheimer's Disease Diagnosis

2.1. Heterogeneity Issues in Multimodal Imaging and Fusion Strategies

Heterogeneity refers to the inconsistency and variations of multimodal data when they are consolidated for storage [2]. In the field of multimodal medical image analysis, the heterogeneity issue makes it difficult the improve model performance and represents a key challenge in translating deep learning models into clinical applications [2,6]. These discrepancies do not originate from the disease itself but stem from data inconsistencies introduced by external factors such as imaging equipment, acquisition protocols, and subject variability. Such noise severely disrupts model training and generalisation capabilities [2]. Consequently, developing integrated approaches capable of effectively mitigating the adverse effects of heterogeneity issues represents a key priority for advancement within this field.

Among multimodal fusion strategies, the preparatory methods for model-driven fusion strategies are often overlooked. This approach implies the model obtains robust resistance to heterogeneous interference by optimising the model architecture and training strategy, thereby laying a solid foundation for subsequent fusion. Multimodal fusion strategies should encompass not only data-level concatenation and feature-level integration, but also model-level approaches. This implies a need to design more robust deep learning models to extract insights from a single or a limited number of modalities. Because this approach can effectively address the challenge of heterogeneity. The work by EL-Geneedy et al. [5] demonstrates the significant advantages of model fusion strategies. Confronted with significant intra-class variability in MRI data (e.g., patients with varying degrees of dementia exhibit different imaging manifestations), this research pursued an unconventional path distinct from conventional multimodal studies. Researchers have significantly enhanced the ability to extract discriminative features from single-modality MRI data by constructing a shallow CNN architecture [5]. In traditional thinking, high-precision medical diagnosis must rely on multimodal data, yet this trial has taken the opposite approach. Its model achieved an astonishing accuracy rate of 99.68% by using only MRI data [5]. This achievement challenges the conventional notion that high-precision diagnosis necessitates multimodal data, demonstrating that deep exploration of single-modality potential offers an effective approach to addressing heterogeneity challenges. Consequently, when confronted with issues such as the high cost of acquiring multimodal data and missing modalities, the research focus could shift towards improving deep learning models for specific modules. This is going to be one of the most productive alternative solutions. This strategy, centred on optimizing deep learning models, complements fusion techniques that directly incorporate multimodal data, precisely forming a complementary advantage in their respective paradigms. These two approaches offer a more comprehensive research direction for addressing the issue of medical imaging variability.

2.2. Challenges and Processing Strategies for Incomplete Modal Data

The prevalence of incomplete modalities is widespread in multi-modality Alzheimer's disease diagnosis and is the current core challenge facing contemporary clinical practice. A complete set of symptoms enables a more precise diagnosis of Alzheimer's disease. So, in an ideal research setting, researchers would expect each subject to possess complete data across all modalities (e.g., T1-weighted MRI, DTI, fMRI) [1-3][7]. However, clinical practice frequently presents uncontrollable issues of modality missing and data heterogeneity, severely limiting the direct application of multimodal diagnostic models designed for complete datasets. As regards the current stage of research progress. The challenges posed by incomplete modal images are typically primarily manifested in the following three aspects. Firstly, modal data is prone to loss, resulting in incomplete image data. This implies that the absence of different modalities (Such as T1-MRI, DTI, fMRI) results in incomplete medical imaging data, which in turn compromises diagnostic accuracy and affects the performance of multimodal models. Furthermore, the issue of data heterogeneity persists, as features extracted from different devices are inevitably subject to variation. Even within the same modality, differences can be introduced due to variations in the underlying mechanisms of the equipment. Discrepancies between modal data result in their incompleteness, thereby complicating the diagnosis of Alzheimer's disease. Thirdly, the image dependency issue in classification models. Existing approaches typically synthesize only one missing modality when handling incomplete data, resulting in limited effectiveness for data modality augmentation [2]. The above issues all fall under the category of incomplete modality problems, resulting in the inability to directly apply multimodal models designed for complete data. This severely limits their clinical utility, thereby preventing clinical application plans from being implemented as intended.

To address these challenges, this review primarily summarizes three major approaches from the conclusions drawn by researchers. The first strategy involves compositing the images from both stages and performing category-based learning. Initially, techniques such as Generative Adversarial Networks (GANs) can be used to complete models with missing modalities [2]. The generated image is then fused with the real image, and the final classification is performed based on the feature of the fused image to obtain the final result (e.g., "healthy" or "Alzheimer's disease"). This approach offers a clear workflow, though the image generation task remains relatively independent from classification. Thus, it is difficult to guarantee that the generated images will prioritise serving classification objectives [2]. The second strategy is end-to-end federated learning [1,7,8]. The core of such methods is the simultaneous placement of image generation and classification tasks within a unified framework for joint optimization. By incorporating mechanisms such as transformed features, the model learns during training to generate images that most effectively assist the classifier in making correct judgments, rather than merely pursuing visual realism. For instance, the framework proposed by Chen Zhaodong achieves coordinated optimization between the image generator and classifier. This manner enables the modal to dynamically acquire the situation where the modal is absent, and significantly enhancing diagnostic accuracy and robustness [2]. The third strategy involves information completion and augmentation based on large language models (LLMs). Beyond image-generation approaches, Du et al.'s pioneering work [4] introduced an alternative paradigm—employing LLMs for "information completion" [4]. This approach does not directly generate incomplete imaging modalities. Instead, it employs universal textual sources such as clinical notes as information carriers. Through meticulously designed prompt engineering, it guides LLMs (e.g., GPT-4) to deeply interpret and extract cognitive function-related features from these texts [4]. This approach essentially employs LLMs as "virtual modality generators," transforming unstructured textual information into structured knowledge usable for diagnostic support. It thus effectively addresses the challenge of missing specific imaging modalities, offering a highly flexible new approach for handling incomplete modal data.

2.3. Exploring Model Interpretability

The clinical application of diagnostic models requires gaining doctors trust, meaning they must possess a high degree of interpretability. In research on model explaination, a core contradiction arises that precise diagnosis demands high model performance, yet the higher the model's performance, the more complex its decision-making process becomes, and the greater the challenge of explanation.

The research by Sorour et al. demonstrates the high accuracy of deep learning models, yet their models are overly complex, thereby raising issues concerning model interpretability [9]. Their study systematically compared five deep learning models. It is noteworthy that CNN-LSTM mixture model achieved a remarkable accuracy of 99.92% [9]. The result deeply demonstrates the immense potential of complex deep learning architectures for AD classification tasks. However, this high performance leads the model excessively complex, making it a classic black-box model. Clinicians struggle to comprehend which specific features within MRI images underpin the model's diagnostic decisions. Thus, while achieving high performance, this study profoundly highlights a core obstacle to clinical deployment of such models: the lack of decision transparency [1,3,6,7,9]. This outcome demonstrates conversely that high accuracy in any model is only meaningful when accompanied by guarantees of explainability.

Regarding the issue of explainability in diagnostic models, current research is not without its achievements. A collaborative, end-to-end learning framework centered on the introduction of transformative features was proposed in Chen Zhaodong's research. This model enables the synergistic integration of image generation and muiltmodal classification [2]. This model generates pseudo-images based on real images, partially compensating for missing features. It incorporates both real and generated image features into decision-making. This approach enables the model to activate a generative pathway when modal data is missing, thereby supplementing information through the generation of pseudo-images and engancing diagnostic robustness. This approach partially transforms the model's hollow decision-making process into the intuitive task of data completion process. This methodology enhances the transparency of decision-making, and provides crucial support for the model's practical clinical application. By ingeniously converting portions of the opaque decision-making process into an understandable and verifiable data generation task, this method markedly enhances the model's clinical credibility.

3. Deep Learning-Based Classification Algorithms for Alzheimer's Disease

3.1. Data Augmentation Algorithms

In the intelligent diagnosis of Alzheimer's disease MRI images, data augmentation serves as a pivotal technique for enhancing model generalisation capabilities and preventing overfitting. Whilst conventional data augmentation techniques (such as rotation, flipping, and filtering) effectively expand datasets, the diversity of generated samples remains limited, and they struggle to simulate complex image variations associated with Alzheimer's pathology.Traditional methods form a fundamental and necessary foundation. Among these, augmentation based on spatial filtering enhances model robustness in complex clinical settings by simulating image degradation (e.g., blurring, noise). However, when confronted with large volumes of complex samples, traditional approaches often exhibit limitations.

Current research has expanded towards deeper augmentation strategies: firstly, towards more comprehensive data preprocessing. The work by Battineni et al. provides an exemplary model for this. Although this work does not directly generate new images [3], it achieves effective "enhancement" of classification models based on structurally optimised features (such as clinical scores and brain volume metrics) through a systematic data preprocessing workflow. Its core lies in improving data quality: imputed missing values using the mode, and selecting highly discriminative features via correlation analysis. This approach constructs high-quality training datasets with superior signal-to-noise ratios at the source [3].

To overcome the limitations of traditional approaches, enhancing feature-level capture has become one of the most effective methods today. This type of method has significant characteristics. By employing deep learning models, it learns from raw data to derive more discriminative high-level features, which results in the enhancement of data information at the feature level. Suk and Shen's research has achieved data augmentation through algorithmic models in this direction [1]. They innovatively proposed employing Stacked Auto-Encoders (SAE) to learn high-level latent feature representations from multimodal neuroimaging data (such as grey matter volume from MRI and average intensity from PET). Through its deep encoding-decoding architecture, the SAE captures hidden nonlinear relationships and analysis its complex patterns within raw features. The pre-training and fine-tuning strategy adopted by Suk et al. [1] significantly enhanced the quality of feature learning. Ultimately, they concatenated the high-level features learned by the SAE with the original features to form an augmented feature vector, serving as input for downstream classifiers. This approach does not directly generate new images but achieves enhancement by elevating the representational power of features, providing models with more informative inputs. It demonstrates particular promise in detecting early, subtle pathological changes.

Furthermore, current research encompasses enhancements in model algorithms. Chen Zhaodong proposed gradient boosting algorithms, employing techniques such as CatBoost and XGBoost [2]. The fundamental principle involves combining multiple weak classifiers (particularly decision trees) to form a strong classifier. Each weak classifier builds upon its predecessor, progressively enhancing the model's predictive capability by minimising the loss function gradient [2].

Enhancing classification performance, particularly excelling when handling high-dimensional data. In summary, to optimize traditional data augmentation techniques that struggle to address complex pathological changes, current research has provided a systematic approach to addressing the issue. Firstly, data refinement is achieved through pre-processing to ensure data quality from the outset. Subsequently, deep learning models such as stacked auto-encoders are employed to mine and enhance features at the characteristic level, capturing pathological. Ultimately, through the application of powerful algorithms such as gradient boosting, the final decision-making process is rendered optimized. This signifies that the role of data augmentation technology has evolved from simple data enhancement to a core technology which comprehensively improves the robustness and accuracy of the diagnostic process.

3.2. Feature Fusion and Classification Algorithm Models

Integrating information from different modalities is the core method for overcoming the issue of incomplete information across modalities. The key process involves utilising the distinct characteristics between different modalities to achieve complementary information, resulting in relatively complete modal data. Two technical pathways have been primarily identified by me from the present research.

3.2.1 Deep Fusion Based on Complex Network Structures

The first approach involves deep integration of models or features. At the feature fusion level. This study can find in Li Yaozu's research, the author proposed a federated network based on multimodal fusion, enhancing the accuracy of AI diagnosis and improving the processing capability for multi-source data. Its principle involves achieving more precise data integration across multiple [10]. Another scholar, Zang Xuefeng, has proposed an Alzheimer's disease aid-diagnosis method based on a Transformer network with variable convolutional kernels. First, through pretretment, a grey matter-weighted image combination is obtained from the original weighted image combination [11]. Then, by comstructing a similarity feature matrix, different modalities are merged. Finally, a Transformer network with variable convolutional kernels was employed to extract deep features, successfully achieving feature fusion within images. Moreover, the research by Sathiyamoorthi et al. [6] presents an innovative method. This study does not directly perform end-to-end deep learning. Instead, it utilizes the two-dimensional grey-level co-occurrence matrix (2D-GLCM) to extract texture features with clear physical significance from MRI images. These features are then fused with clear physical significance from MRI images. These features are then combined with abstract features extracted by a deep convolutional neural network (DCNN). At the model fusion level, it is noteworthy that the research by Chen et al. holds exceptional reference value in the realm of deep model integration [8]. They devised an end-to-end ensemble learning framework employing Faster R-CNN for precise localisation of abnormal brain regions, alongside an enhanced ResNet50 for extracting deep spatio-temporal features. This was further augmented by bidirectional gated recurrent units (Bi-GRU) to model contextual dependencies within feature sequences. Finally, the above methods were seamlessly integrated, with the Soft-NMS algorithm introduced into Faster R-CNN to optimize multi-object detection. This three-tiered sequential model—spanning anomaly region localisation, feature extraction, and model integration—achieved cross-level collaborative optimization from pixel-level to feature-level to sequence-level. Experimental results demonstrated an exceptional classification accuracy of 98.91%. The approaches outlined above demonstrate the substantial potential of system-level deep integration. This methodology merges the strong interpretability of traditional radiological features with deep learning models, providing another crucial foundation for addressing the challenge of deep integration based on complex network architectures.

3.2.2 Feature-Splicing-Based Classical Hybrid Models

In feature fusion strategies, the classical hybrid model based on feature splicing also demonstrates robust performance. For instance, the research by Houria et al. [7] can be referenced. The researchers employed a customisation strategy using convolutional neural networks (CNNs) to extract high-level features from different modalities. These extracted high-level features were subsequently spliced at the feature level before being input into a support vector machine (SVM) for classification.This CNN-SVM hybrid model fully leverages the advantages of deep learning feature extractors while achieving the high efficiency of traditional classifiers under small sample conditions. The approach has demonstrated outstanding performance across numerous classification tasks, conclusively proving the strong complementarity achievable between multimodal information.

Therefore, for the present study, the quality of modal data can be enhanced through model fusion and feature fusion approaches. Despite the technical limitations of current research, future work will pursue the fusion of these two methodologies to advance the interpretability of medical models. It is believed that deep learning models will undoubtedly play a role in clinical diagnosis in the future.

4. Challenges and Development

In Alzheimer's disease classification tasks, deep learning models have demonstrated remarkable potential. However, the further advancement of deep learning models faces significant challenges, primarily reflected in the following three aspects. First and foremost, challenges at the data level must be confronted. The quality of model data primarily depends on the quality of annotated data. Nevertheless, acquiring large and high-quality datasets necessitates addressing challenges such as the high cost of medical data, category imbalance, and multi-center data heterogeneity. Concurrently, issues like the subjectivity of annotation severely constrain [1,6,7,9]. Secondly, challenges exist at the model level: The most advanced medical diagnostic models currently in use are predominantly typical black-box models. The decision-making process of the model suffers from a lack of transparency. Clinicians struggle to comprehend and trust their diagnostic rationale, presenting a key barrier to widespread adoption. Furthermore, the generalisability of these models on independent prospective datasets remains to be validated, and their clinical utility—rather than merely laboratory accuracy—requires further demonstration. Finally, the challenge of clinical integration needs to be addressed. The majority of existing research constitutes retrospective analyses of models, with no effective solutions yet identified for achieving clinical deployment. Seamlessly integrating models into clinical workflows to enable real-time, efficient diagnostic support and ultimately improve patient outcomes remains an unresolved systems engineering problem.

In the future, researchers should mainly break through the following aspects. First, future research should focus on the matching of medicine and models, exploring ways to incorporate medical-related knowledge into the deep learning model, so as to create an explainable AI model for diagnosis. Concurrently, researchers should pay attention to the interpretability and generalization ability of the model, and can give priority to the development of explainable AI(XAI) technology. Its decision-making process must be transparent to doctors. At the same time, appropriate clinical trials should be adopted to strictly evaluate the diagnostic model. It is essential to ensure that the generalization ability and clinic value of the model are sufficient to support its clinical application. In addition, interdisciplinary research combining AI research with clinical diagnosis should be promoted. AI model researchers must have close contact with clinicians in order to achieve consensus. Breaking down discipline barriers and allowing AI researchers to work closely with doctors to design effective medical diagnosis models is the key to the application of deep learning models to clinical practice today.

5. Conclusion

This paper provides a systematic review of research advances in Alzheimer's disease (AD) MRI image classification and diagnosis using neural network-based methods. Core findings indicate that the research paradigm in this field is undergoing a profound shift: from "single-modality analysis" to "multimodal fusion", and from "pursuing performance" to "balancing interpretability".

Simultaneously, this study highlights persistent challenges within the field. Key obstacles include reliance on large-scale, high-quality annotated datasets; insufficient interpretability stemming from complex models' "black box" nature, which hinders clinical adoption; and the predominance of retrospective data in existing research, who’s clinical generalisability requires validation through prospective trials.

Looking ahead, the next frontier in AD diagnostic support research will be "knowledge-data dual-driven" intelligent systems. Future endeavours should concentrate on developing deep learning models that integrate medical a priori knowledge (such as pathological mechanisms) to enhance the rationality and credibility of decision-making. Efforts should also focus on deploying lightweight, explainable AI technologies in clinical endpoints and conducting large-scale, multicentre clinical validation trials. This will ultimately achieve the critical transition of AI from "laboratory precision" to "clinical utility," providing robust technological support for the early prevention and control of AD.

References

[1] Suk H I, Shen D. Deep learning-based feature representation for AD/MCI classification. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Springer, Berlin, Heidelberg, 2013: 583-590.

[2] Chen Z. Joint optimization of cross-modal image synthesis and classification for Alzheimer’s disease. Yunnan University, 2024.

[3] Battineni G, Hussain M A, Chintalapudi N, et al. Improving Alzheimer’s Disease Detection by MRI Using Multimodal Machine Learning Algorithms. IEEE Transactions on Medical Imaging, 2021, 40(11): 2103-2115.

[4] Du X, Novara-Laurentiev J, Prasack J M, et al. Enhancing early detection of cognitive decline in elderly a comparative study utilizing large language models in clinical notes. arXiv preprint arXiv:2403.12345, 2024.

[5] El-Geneedy M, Moustafa H E, Khalifa F, et al. An MRI-based deep learning approach for accurate detection of Alzheimer’s disease. Technical Report TR2023-001, Brigham and Women’s Hospital, Harvard Medical School, 2023.

[6] Sathiyamoorthi V, Ilavarasi A K, Murugeswari K, et al. A deep convolutional neural network-based computer-aided diagnosis system for predicting Alzheimer’s disease from MRI images. Measurement, 2021, 171: 108838.

[7] Houria L, Belkhamsa N, Cherfa A, et al. Multi-modalty MRI for Alzheimer's disease detection using deep learning. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2022: 1043-1053.

[8] Chen Y, Wang L, Ding F, et al. Automatic classification of Alzheimer's disease integrating soft NMS and improved ResNet50 with deep learning model. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2024: 100782-100792.

[9] Sorour S E, El-Mageed A A A, Albarrrak K M, et al. Classification of Alzheimer's disease using MRI data based on Deep learning Techniques. In Proceedings of the 2024 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2024: 101940.

[10] Li Y. Research on classification algorithm for Alzheimer's disease based on deep learning joint network and multi-modal feature fusion. Qufu Normal University, 2025.

[11] Zang X. Research on auxiliary diagnosis method for Alzheimer's disease based on structural magnetic resonance imaging and deep learning. Shandong Normal University, 2025.