Electronic Letters on Computer Vision and Image Analysis 25(3):81-100, 2026

Investigation of Alzheimer's Disease Biomarkers using Deep Learning on MR Images: A Systematic Review

Muhammed B Ceesay[1], Adhi Harmoko Saputro[1], and Syahril Siregar[1]

[1]*Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia*
Received 20th August 2025; accepted 8th August 2026

Abstract

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder and the leading cause of dementia, particularly among the elderly. Although no cure currently exists, early diagnosis and intervention can slow disease progression and improve quality of life. Magnetic Resonance Imaging (MRI) plays a crucial role in the early detection of AD by capturing characteristic brain changes such as ventricular enlargement, entorhinal cortex degeneration, hippocampal atrophy, and cortical volume loss. In recent years, Deep Learning (DL) has shown exceptional potential in medical image analysis, including disease detection and classification. This systematic review investigates recent studies that combine DL techniques with brain MRI for AD diagnosis and classification. A total of 101 peer-reviewed articles published between 2020 and 2024 were initially identified from five major scientific databases, of which 53 met the inclusion criteria. Several studies also utilized publicly available datasets, particularly from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The reviewed literature reports encouraging results, with some DL models achieving classification accuracies as high as 99.74%. However, issues such as inconsistent data splitting protocols especially the lack of patient-wise separation in some studies raise concerns about potential data leakage, which could lead to overestimated model performance.

Key Words: Alzheimer's Disease, Biomarkers, Deep Learning, Magnetic Resonance Imaging (MRI).

1 Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that leads to cognitive decline and memory loss, particularly in the aging population [1]. It is primarily characterized by the accumulation of beta-amyloid (Aβ) plaques and tau neurofibrillary tangles in the brain, resulting in neuronal loss and degeneration across multiple regions [2]. As AD progresses, it leads to widespread brain dysfunction, affecting memory, language, behavior, and problem-solving abilities. Early-stage AD often involves degeneration in the hippocampus and temporal lobes, regions associated with memory and learning [3]. Although there is currently no cure for AD, early diagnosis and timely intervention have been shown to slow its progression and improve quality of life.

The initial symptoms of AD may begin years even decades before clinical diagnosis. These symptoms often present subtly, making early detection particularly challenging. AD is the leading cause of dementia,

Correspondence to: syahril.siregar@ui.ac.id

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.2330
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

accounting for approximately 60-80% of all dementia cases worldwide [4]. According to the 2018 Alzheimer Report, a new case of dementia is diagnosed every 3 seconds globally [5]. It is estimated that over 44 million individuals are currently affected by AD, with the number expected to rise to 131.5 million by 2050 [6].

The early stage of AD is often diagnosed as Mild Cognitive Impairment (MCI), a condition in which individuals show cognitive decline beyond normal aging but not severe enough to interfere significantly with daily life [5]. Detecting AD at this stage is particularly challenging, as the brain tissue of MCI patients can closely resemble that of cognitively normal (CN) individuals [7]. Consequently, more accurate and robust detection methods are needed. Statistically, individuals with MCI have a 16% chance of progressing to AD within four years, compared to only 1-2% among CN individuals [5]. However, not all MCI patients will progress; some remain stable or even revert to a normal cognitive state. Therefore, MCI is often classified into progressive MCI (pMCI) and stable MCI (sMCI), depending on the likelihood of progression to AD.

While the underlying cause of AD is not yet fully understood, amyloid plaques and tau tangles are widely considered the primary pathological hallmarks [8]. The extracellular deposition of Aβ and the intracellular accumulation of hyperphosphorylated tau protein disrupt synaptic function and accelerate neurodegeneration [9]. Neuroimaging techniques such as fluorodeoxyglucose positron emission tomography (FDG-PET), functional magnetic resonance imaging (fMRI), structural MRI (sMRI), and diffusion tensor imaging (DTI) are used to identify these structural and functional changes in the brain [10].

Recent advancements in artificial intelligence (AI) and machine learning (ML) have significantly enhanced the potential for early AD diagnosis. AI refers to computer systems designed to perform tasks that typically require human intelligence, such as learning, reasoning, and decision-making [11]. ML, a subset of AI, enables systems to learn from data and improve their performance over time. Deep learning (DL), a further subset of ML, employs multi-layered neural networks inspired by the human brain's structure and function [12]. Figure 1 illustrates the hierarchical relationship among AI, ML, and DL.

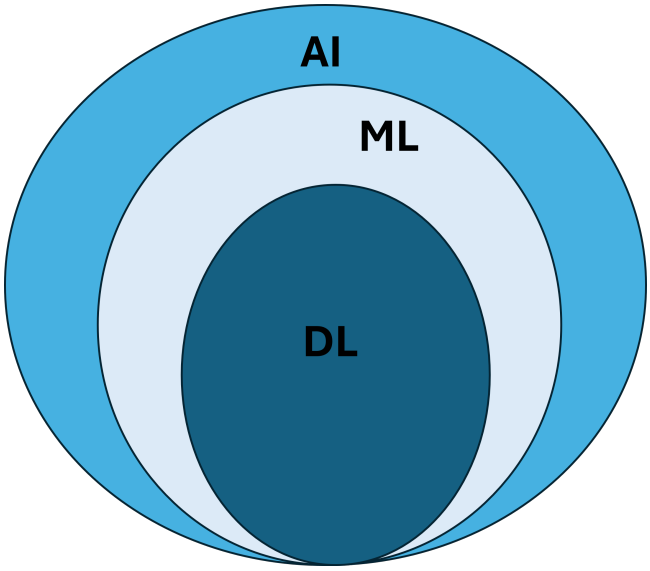


Figure 1: Hierarchical relationship among AI, ML, and DL

DL techniques, particularly convolutional neural networks (CNNs), have been successfully applied to analyze neuroimaging data such as MRI, PET, and DTI scans. These methods have demonstrated strong performance in distinguishing between CN, MCI, and AD, with reported accuracy levels up to 95% for AD vs. CN, 85% for CN vs. MCI, and 75.8% for MCI vs. AD [13]. CNNs are especially well-suited for image analysis due to their ability to automatically extract spatial features without requiring manual input [12, 14]. They consist of convolutional layers, pooling layers, and fully connected layers that together form a powerful architecture for pattern recognition [15]. Other DL architectures used in AD research include recurrent neural networks

(RNNs), long short-term memory networks (LSTMs), autoencoders (AEs), and restricted Boltzmann machines (RBMs) [16].

Given the increasing availability of neuroimaging data and the growing capabilities of DL methods, there is a strong rationale for applying CNN-based models in the early diagnosis of AD.

In this study, we focus on structural MRI data to evaluate its effectiveness in AD classification using CNN models. Our goal is to assess how effectively deep learning, particularly CNNs, can distinguish between CN, MCI, and AD stages using brain imaging data.

2 Evaluation Metrics

To assess the performance of DL models in classifying AD, various evaluation metrics are commonly employed, particularly in binary classification tasks. These include Accuracy, Specificity, Sensitivity (Recall), Precision, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) [17, 18]. These metrics help quantify how well a model distinguishes between AD and non-AD cases.

The following definitions use four fundamental quantities:

- **True Positive (TP)**: AD cases correctly predicted as AD.
- **True Negative (TN)**: Non-AD cases correctly predicted as non-AD.
- **False Positive (FP)**: Non-AD cases incorrectly predicted as AD.
- **False Negative (FN)**: AD cases incorrectly predicted as non-AD.

The commonly used evaluation metrics are defined as follows:

1. **Accuracy**: The overall correctness of the model across all predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2. **Specificity (True Negative Rate)**: The model's ability to correctly identify non-AD cases.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

3. **Sensitivity / Recall (True Positive Rate)**: The model's ability to correctly identify AD cases.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

4. **Precision (Positive Predictive Value)**: The proportion of predicted AD cases that are correctly classified.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

5. **F1-Score**: The harmonic mean of Precision and Recall, particularly useful for imbalanced datasets.

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{5}$$

6. **ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)**: Measures the trade-off between Sensitivity and Specificity. A higher AUC indicates better model performance.

Table 1 provides a summary of these metrics.

Table 1: Summary of Common Evaluation Metrics for AD Classification

Metric	Definition and Interpretation
Accuracy	Overall proportion of correctly classified instances.
Specificity	Proportion of correctly identified non-AD cases (true negative rate).
Sensitivity / Recall	Proportion of correctly identified AD cases (true positive rate).
Precision	Proportion of predicted AD cases that are truly AD (positive predictive value).
F1-Score	Harmonic mean of Precision and Recall, balancing false positives and false negatives.
ROC-AUC	Area under the ROC curve, indicating the model's ability to discriminate between classes.

3 Magnetic Resonance Imaging

3.1 MRI and Its Physics

The human body is composed predominantly of water, accounting for approximately 80% of its content, and hence contains a large number of hydrogen atoms [16]. Each hydrogen nucleus consists of a single proton and behaves like a small bar magnet, spinning around its own axis. In their natural state, these magnetic moments are randomly oriented.

When the body is placed in a strong external magnetic field (B_0), the hydrogen nuclei align either parallel or anti-parallel to the field. This alignment causes the protons to precess around the axis of the magnetic field at a characteristic frequency known as the **Larmor frequency**, given by:

$$\omega_0 = \gamma B_0 \tag{6}$$

where ω_0 is the Larmor frequency, B_0 is the external magnetic field, and γ is the gyromagnetic ratio (42.5 MHz/T for hydrogen) [19].

To generate a measurable signal, a radio-frequency (RF) pulse is applied perpendicular to B_0, which perturbs the aligned spins, causing them to move from the longitudinal (Z) plane to the transverse (XY) plane [20, 21]. After the RF pulse is switched off, the hydrogen nuclei re-emit energy as they relax back to equilibrium. The emitted signals, which vary based on tissue properties, are detected and used to form tomographic images via Fourier transformation [16, 22].

Relaxation Mechanisms Relaxation is the process by which the net magnetization returns to its equilibrium state after the RF pulse is removed [19]. The signal emitted during this relaxation is called **Free Induction Decay (FID)** [23]. Two primary types of relaxation occur:

- **T_1 relaxation** (spin-lattice relaxation): the recovery of longitudinal magnetization.
- **T_2 relaxation** (spin-spin relaxation): the decay of transverse magnetization due to dephasing.

The effectiveness of contrast agents in shortening these relaxation times is referred to as **relaxivity**. Higher relaxivity results in stronger contrast in the MRI images [23].

T_1 Relaxation T_1 is defined as the time it takes for the longitudinal magnetization to recover to 63% of its equilibrium value following a 90° RF pulse [23]. As the protons realign with B_0, the transverse magnetization

decays due to dephasing, while longitudinal magnetization gradually restores [24]. T_1-weighted MRI is optimized using paramagnetic contrast agents such as gadolinium (Gd^{3+}), manganese (Mn^{2+}), and iron (Fe^{3+}) [16]. These images are commonly used for anatomical analysis, where fat appears bright and fluids appear dark.

T_2 Relaxation T_2 is the time required for transverse magnetization to decay to 37% of its initial value, primarily due to spin-spin interactions. Stronger interactions lead to faster dephasing and darker image contrast. A related phenomenon, denoted T_2^*, accounts for additional magnetic field inhomogeneities. Superparamagnetic nanoparticles such as iron oxide are commonly used to enhance T_2-weighted imaging [25]. T_2-weighted MRI is particularly useful for detecting pathological changes, where fluid appears bright and fat appears dark [16].

Figure 2 illustrates T_1 and T_2-weighted MRI brain scans obtained from ADNI database.

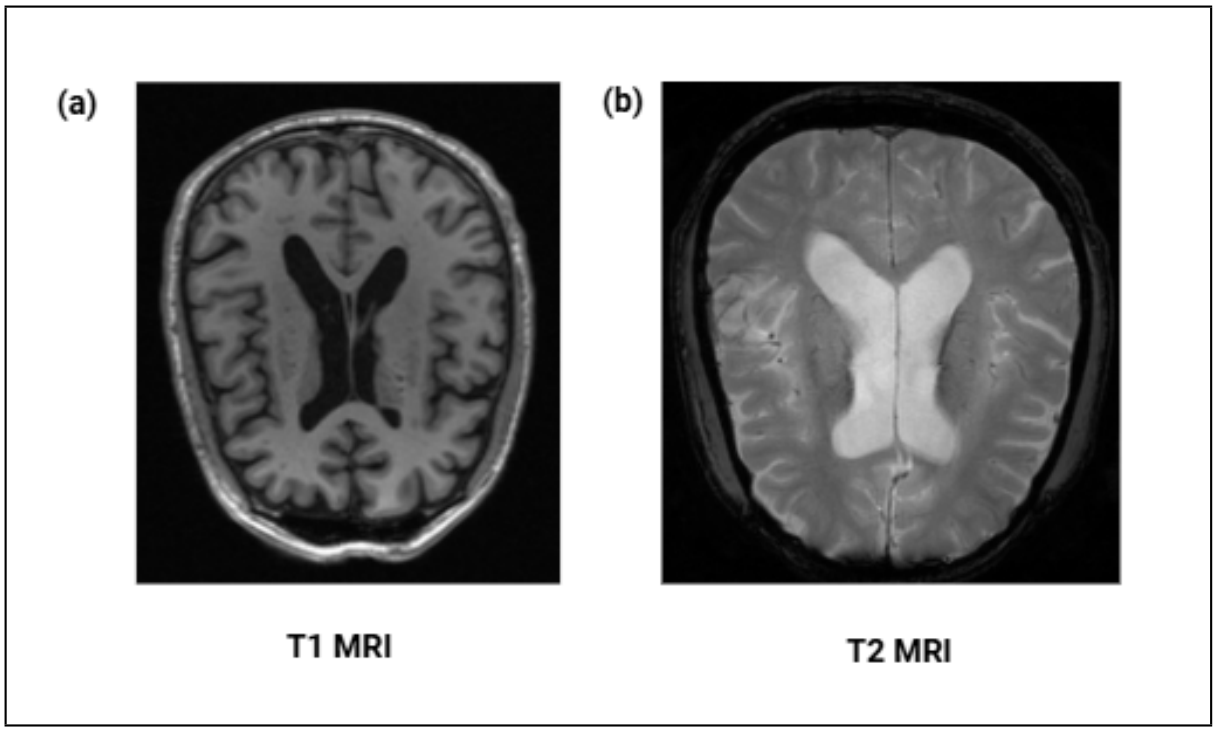


Figure 2: T1 and T2-weighted MRI brain Scans

3.2 MRI in Alzheimer's Disease Diagnosis

MRI provides high-resolution, non-invasive imaging in multiple planes, making it a critical tool in AD research and diagnosis [26]. It offers detailed structural and functional information about the brain, enabling the detection of anatomical changes associated with AD [27].

Many studies have utilized AI techniques to analyze MRI data for features such as cortical atrophy, hippocampal volume reduction, and degeneration of the entorhinal cortex and other subcortical structures [28]. MRI can effectively distinguish between gray matter (GM) and white matter (WM), allowing for robust tissue segmentation [27].

The availability of open-source datasets such as ADNI and OASIS has accelerated research in this domain. Numerous studies have assessed the performance of MRI-based models in diagnosing AD, though early-stage detection remains challenging and not yet 100% accurate [29]. One of the most reliable biomarkers for AD is the reduction in hippocampal volume [30], which correlates strongly with neuronal loss. MRI scans have consistently shown that in early AD, the hippocampus, amygdala, and entorhinal cortex are affected. As the disease progresses, brain structures such as the ventricles enlarge while the hippocampus and cerebral cortex

shrink [31]. Volumetric MRI measurements of these regions are considered valid and correlate well with pathological findings [32].

4 Open Access MRI Data Sources

The most common and and widely used database for the collection of AD related data is Alzheimer's Disease Neuroimaging Initiative (ADNI). The database consist of series of image data, such as MRI, fMRI, PET DTI, and as well as genetic information, demographic and clinical data, cognitive scores and CSF biomarkers obtained from more than 1,800 participants. The database was initiated in 2010 and still active with several phases of study, such as ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. ADNI data collection was through multiple follow-up sessions with the objective of developing biomarkers to monitor disease progression and understanding the underlying neurobiological changes associated with AD[33].

The AD Research unit at Washington University developed the Open Access Series of Imaging Study (OASIS) database, containing significant imaging data from both CN and AD subjects[34]. The database provides three (3) distinct datasets, which includes OASIS-1, OASIS-2 and OASIS-3[35]. The OASIS-1 dataset, was released in 2007 with a cross sectional dataset containing 416 participants, 100 of whom are aged over 60 diagnose with very mild or moderate AD based on Clinical Dementia Rating (CDR). OASIS-2 was released in 2010 with longitudinal studies containing 150 participants with 373 imaging sessions. OASIS-3 is the most comprehensive series, released in 2019 containing longitudinal data from 1,090 participants, with 493 of distinct AD stages and 650 CN subjects[33].

The Kaggle dataset consist of MRI data of 5,121 axial images, labeled based on the corresponding severity level of dementia. The dataset dose not includes patient age, clinical or demographic information. For class distribution, it contains 2,560 CN and 2,561 of other AD related stages. All images has a fixed resolution of 176×208 pixels[36].

5 Related Work

Recent years have seen a surge in research applying DL methods for AD classification using MRI data. This section presents a critical review of existing literature, comparing approaches in terms of model architectures, classification strategies (binary vs. multi-class), datasets, evaluation metrics, and experimental design. Table 2 and 3 provides a concise summary of binary and multi-class classification approaches.

5.1 Binary Classification Studies

Binary classification studies primarily aim to differentiate AD patients from CN individuals. A variety of DL architectures have been employed for this task. Mehmood et al[3] proposed a VGG16 model from Transfer Learning (TL) to predict and classify AD stages. Their approach acquired an accuracy of 98.73% in classifying AD vs CN subjects and 83.72% in classifying sMCI vs pMCI patients. Meng et al. [17] implemented a lightweight 2D CNN using the ADNI dataset, achieving 94.2% accuracy in distinguishing AD from CN. Similarly, Hicks et al. [18] applied transfer learning with VGG16 and ResNet50, obtaining over 91% accuracy. Loddo et al[36] employed AlexNet, ResNet50, ResNet101, GoogleNet and inception-ResNet-V2 from Transfer Learning (TL) to classify AD stages. Final classification was perform by ensembling the predictions of the top 3 DL models. The dataset used in their study was obtained from OASIS, Kaggle and ADNI database, which were utilize to trained and evaluated the models independently. For binary classification, their approach yields an accuracy of 98.51% using the OASIS dataset, 96.57% using the Kaggle dataset and 99.74% using the ADNI dataset to classify AD and CN subjects. Salehi et al. [37] used an LSTM-based approach on a Kaggle dataset

and achieved an impressive 98.62% accuracy for AD-CN classification. Similarly, Yagis E et al[38] build a 3D-CNN based model inspired by VGG-16 on OASIS and ADNI dataset and achieved an accuracy of 73.4% on ADNI dataset and 69.9% on OASIS dataset to classify AD and CN subjects. However, many of these studies do not clearly report whether data splitting was conducted on a patient-wise basis, raising concerns regarding data leakage.

Additional binary classification efforts include those by Ahmad et al. [39], who utilized a 2D CNN-based computer-aided diagnosis (CAD) system after segmenting MRI into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), achieving up to 96.4% accuracy in AD-MCI classification. Similarly, Kim et al. [40] evaluated a commercial decision-support system, VUNO Med-DeepBrain AD (DBAD) 2D CNN-based, against expert diagnosis and reported 87.1% classification accuracy. Lin et al[41] proposed a pre-trained ResNet50 model as a backbone, integrated with post-fusion algorithm and attention mechanism for 3D MRI data. They acquired an accuracy of 91.0% in classifying AD and CN subjects. Fathi et al. [42] employed DenseNet201, DensNet169, DenseNet121, ResNet50, Inception-ResNetV2, and VGG16 as ensemble prediction to classify AD stages. The data used in their study was obtained from ADNI and Firoozgar Hospital (FH) in Tehran. For binary classification, their approach acquired an accuracy of 98.57% (AD vs CN), 96.37% (CN vs sMCI), 94.22% (sMCI vs pMCI) and 99.83% (pMCI vs AD).

5.2 Multi-Class Classification Approaches

Multi-class classification introduces greater complexity, typically involving differentiation among AD, CN, and mild cognitive impairment (MCI). For instance, the study of Loddo et al, as described in section 5.1, also performs a multi-class classification. They acquired an accuracy of 98.24% using the OASIS dataset and 97.71% using Kaggle dataset to classify, CN,sMCI, pMCI and AD subjects. Similarly, an accuracy of 99.22% was acquired using ADNI dataset to classify CN, MCI and AD subjects. Fathi et al as described in section 5.1, further extended their study to include multi-class classification, obtaining 93.88% accuracy in classifying sMCI, pMCI, CN, and AD.

Srividhya L et al[43] employed 26 Keras pre-trained models to classify AD stages and identify the best performing algorithm. They obtained their MRI data from ADNI database consisting of CN, EMCI, MCI, LMCI and AD subjects. They utilized SMOTE to mitigate class imbalance and employed Grad-Cam and saliency maps to address interpret-ability. ResNet50-V2 outperforms all the other models with an accuracy of 96.07%.

Raju M et al[44] proposed a novel approach for multi-class AD classification. They employed an embedded 3D-CNN with SVM model to predict and classify CN, MCI and AD subjects. They obtained their data from ADNI database consisting of 152 CN, 181 MCI and 132 AD subjects. The features extracted by the 3D-CNN are feed to the SVM for final classification which yielded and accuracy of 97.77%. To address interpret-ability, they employed saliency maps generated by the 3D-CNN obtained using Grad-Cam highlighting specific brain regions such as, temporal lobe (hippocampus, amygdala, entorhinal cortex, para-hippocampus) and the parietal lobe as the most discriminative areas to differentiate CN, MCI and AD subjects. Faisal et al. [45] demonstrated a 2D-CNN model capable of achieving 96.41% accuracy across AD, CN, and MCI classes.

More recent work by Jin et al. [46] combined 3D DenseNet for segmentation with MobileNet V3 for classification, reporting a multi-class accuracy of 92.63%. Aaraji and Abbas [47] designed an ensemble of 3D-CNNs that independently extract local (hippocampus) and global (whole brain) features before classification using an SVM, achieving 85% in multi-class settings.

5.3 Architectures and Preprocessing Strategies

Across studies, commonly used CNN backbones include VGG, ResNet, DenseNet, GoogleNet, and LeNet. Some works utilize shallow or lightweight networks to reduce computational overhead, while others apply attention mechanisms [41] or Grad-CAM [44] for better interpretability. Preprocessing strategies vary widely—

Ref	Model / Approach	Dataset	Accuracy	Remarks
[3]	VGG16	ADNI	83.72% (sMCI-pMCI)	Transfer learning applied
[17]	CNN (2D slices)	ADNI	94.2% (AD vs CN)	Lightweight model; unclear patient-wise split
[18]	VGG16, ResNet50	ADNI	91.5% (binary)	High performance; no external validation
[36]	Ensemble-CNNs	OASIS Kaggle ADNI	98.51% 96.57% 99.74%	Top 3 transfer-learning CNN ensemble for final classification on each dataset
[37]	LSTM	Kaggle	98.62% (AD vs CN)	Large dataset; high binary performance
[38]	3D-CNN	OASIS ADNI	69.9% 73.4%	Relatively poor classification performance
[39]	2D CNN + CAD	ADNI	96.4% (AD-MCI)	Includes segmentation preprocessing
[40]	DBAD (2D CNN-based)	AMC/KVHS	87.1% (AD-CN)	Compared against medical experts
[41]	ResNet + Attention	ADNI	91% (AD-CN)	Post-fusion with attention
[42]	Ensemble CNNs	ADNI + FH	98.57% 96.37% 94.22% 99.83%	Best performance for binary classification

Table 2: Summary of Deep Learning Approaches for Binary AD Classification Using MRI

from using representative 2D slices to full 3D volumes. For example, Chen et al. [48] evaluated the effect of MRI field strength (1.5T vs 3T) on model performance, noting slight variations in classification accuracy across CNN architectures.

5.4 Critical Gaps and Challenges

Despite high reported accuracies, several challenges remain unaddressed in the literature. A common issue is the lack of clarity around data partitioning strategies. Studies that fail to use patient-wise splits risk data leakage, potentially inflating accuracy due to overlapping subject images in training and testing sets. External validation using independent datasets is also rare, limiting generalization. Furthermore, while explainability methods such as Grad-CAM have been introduced, few studies integrate explainable AI (XAI) into their workflows to clarify *why* certain predictions are made—a key requirement for clinical deployment. Finally, most studies focus exclusively on MRI, though future work could benefit from multi-modal fusion involving PET, DTI, or clinical biomarkers.

In summary, deep learning has shown considerable promise in the automated classification of AD from MRI

Ref	Model / Approach	Dataset	Accuracy	Remarks
[36]	Ensemble-CNNs	OASIS Kaggle ADNI	98.24% 97.71% 99.22%	Top 3 transfer-learning CNN ensemble across multiple datasets
[42]	Ensemble CNNs	ADNI + FH	93.88%	Best performance among multi-dataset ensemble methods
[43]	ResNet50-V2	ADNI	96.01%	Compared multiple Keras pre-trained models; Grad-CAM used for interpretability
[44]	3D-CNN + SVM	ADNI	97.77%	Patch-based 3D-CNN combined with SVM; includes Grad-CAM analysis
[45]	2D-CNN	ADNI	96.41%	Strong performance for 3-class classification
[46]	DenseNet + MobileNetV3	ADNI	92.63%	Combined 3D segmentation with slice-based analysis
[47]	3D-CNN + SVM	ADNI	85%	Uses both hippocampal and global structural features

Table 3: Summary of Deep Learning Approaches for Multi-Class AD Classification Using MRI

data. However, for reliable clinical translation, future studies must emphasize rigorous experimental design, transparent methodology, explainability, and broader validation using diverse datasets.

6 Methods and Data

This section outlines the methodology employed in this review, including the formulation of research questions, data collection procedures, inclusion and exclusion criteria, and how relevant information was extracted from the selected literature. Additionally, the MRI data source used for analysis is also described.

6.1 Research Questions

Formulating clear and relevant research questions is essential for guiding a systematic review. The following five questions were designed to direct the literature search and analysis process, focusing on the intersection of DL, evaluation metrics, AD biomarkers, and MRI:

1. How accurately can CNNs detect and classify Alzheimer's disease?
2. What evaluation metrics are used to assess the performance of CNN models in AD classification?
3. What is the basic physical principle behind MRI operation?
4. How does MRI contribute to the diagnosis of Alzheimer's disease?
5. What neuroanatomical features distinguish healthy brains from those affected by AD?

6.2 Literature Data Collection

Following the formulation of research questions, a literature search was conducted using five major databases: Google Scholar, Association for Computing Machinery (ACM) Digital Library, IEEE Xplore, CORE, and Nature. Articles published between 2020 and 2024 were considered. Any articles published outside this range were excluded.

Initially, 101 articles were retrieved. After applying inclusion and exclusion criteria, 53 articles were selected for detailed analysis. The search employed keywords such as "Alzheimer's disease," "deep learning," and "MRI."

Additionally, snowballing techniques were used by checking the reference lists of selected papers to identify further relevant literature. While this approach broadened the scope, it also introduced older publications. These were used sparingly and only if critically relevant.

6.3 MRI Dataset

To support our review of imaging-based approaches, we used sample MRI scans obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The dataset was accessed on December 17, 2024, and includes T1- and T2-weighted MRI brain images. ADNI is widely recognized as one of the most comprehensive and commonly used datasets in Alzheimer's research [49, 50, 51]. It is a longitudinal, multi-center study designed to improve the development and validation of clinical, genetic, imaging, and biochemical biomarkers for early diagnosis and progression monitoring of AD.

6.4 Inclusion and Exclusion Criteria

After downloading the articles, a filtering process was performed. The abstracts and methodology sections were reviewed to assess relevance and scientific quality. Selected articles were organized into a dedicated folder for ease of access.

Inclusion criteria:

- Articles published between 2020-2024.
- Studies focusing on CNN or DL applied to AD classification using MRI.
- Clear methodological description and presentation of results.

Exclusion criteria:

- Articles without clearly defined methodology or insufficient explanation of results.
- Duplicate studies or review papers that did not offer new empirical data.
- Papers outside the defined publication year range.

To manage and annotate the selected literature, Microsoft Edge was used for highlighting relevant content, and Mendeley was used for reference management. The manuscript was written and compiled using Overleaf.

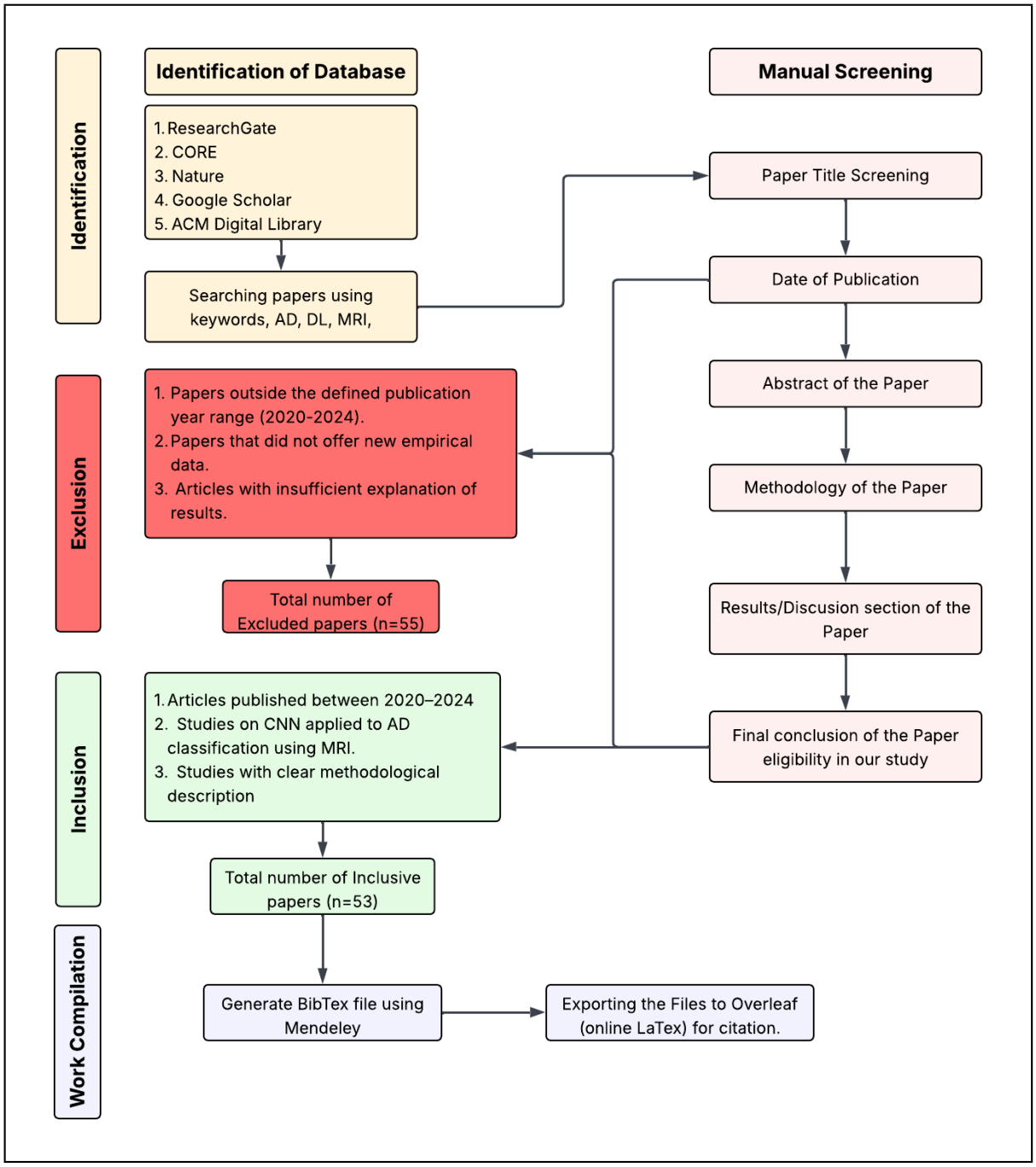


Figure 3: The Flowchart highlights the literature selection processes and stages involved in compiling the results in this study

7 Results and Discussion

The choice of input representation plays a significant role in the performance and computational efficiency of classification models, specifically CNN algorithms. However, the most common input approaches includes, slice-based, voxel-based (3D Volumetric) patch-based and Region of Interest (ROIs)-based. Several studies employed 3D volumetric MRI data or solely focused on Cerebrospinal Fluid (CSF), GM, WM or the hippocampus to classify AD stages. However, these approaches retain significant anatomical information and required sub-

stantial higher computational resources and larger training samples to enhance model performance and good generalization. For more computational efficient alternative approach, 2D slice-based reduce model complexity while still capturing meaningful and spatial information for disease classification [52].

Across various studies reviewed, CNN architectures demonstrated a strong discriminative capabilities in classifying AD stages for both binary and multi-class classification. However, depending on the modeling strategy and methodological rigor, the magnitude of the reported accuracies varies significantly across different studies. In contrast, 2D-CNN provides high classification performance, often exceeding 90% due to their computational efficiency and increasing number of training samples through data augmentation. Despite the classification performance of 2D-CNN, they are more susceptible to data leakage if proper care is not taken during model training and evaluation. For instance, Ahmed et al, Meng et al etc achieve a high classification accuracy in their study, which might be due to data leakage.

In contrast, 3D-CNN is computational expensive compared to 2D-CNNs. They offer more physiologically coherent feature representations due to the preservation of spatial information. The model performance across studies that utilized 3D-CNN varies significantly due to dataset size, pre-processing steps and methodological approaches. For example, in Raju et al proposed approach achieved a significant competitive accuracy of 97.77%. Similarly, other studies reported substantial lower results, such as Yagis et al due to limited data size and rigorous methodological framework.

Transfer Learning approaches, such as VGG16, ResNet50, DensNet or MobileNet, achieved a significant classification accuracies, usually above 90%. For instance, Mehmood et al.[3] utilized VGG16 to classify CN, AD, pMCI, and sMCI subjects. Their model achieved an impressive accuracy of 98.73% for distinguishing AD from CN subjects, and 83.72% for sMCI and pMCI. Similarly, attention-based approaches, such as the one described in Lin et al study improve model performance. Notably, ensemble CNNs consistently outperform individual model across distinct datasets, achieving accuracies above 98%.

Moreover, the choice of dataset significant affects model performance. ADNI dataset is the most widely use database in most studies to accurately predict and classify AD. Studies that utilized ADNI dataset provides strong classification performance compared to other database, such as OASIS and Kaggle. For example, both Loddo et al and Yagis et al study has proven that DL models provides substantial higher classification accuracy on ADNI dataset compared to OASIS and Kaggle, highlighting the benefits of using quality dataset. However, models trained and evaluated exclusively on a single dataset tends to report biased and optimistic results, especially when tested on on homogeneous samples. In contrast, models trained and evaluated on different dataset or real world cohort provides more conservative and unbiased classification results, reflecting the vulnerability and diagnostic complexity of medical images. In Fathi et al study, they combined dataset from ADNI and database and FH dataset in Tehran to classify AD stages, illustrating how multi-dataset integration can improve model generalization and provides a more realistic assessment of model performance.

Other author's approach such as Srividhya et al, Raju et al and others, has shown the effectiveness of CNN algorithms in distinguishing AD stages in both multi-class and binary classification. These accuracies raised concern in most studies were patient-wise data splitting is not explicitly mention, potentially contributing to inflated accuracies in the 95-99%, often caused by data leakage during model training and validation.

To assess the performance of deep learning algorithms, several evaluation metrics are commonly employed, including accuracy, precision, recall (sensitivity), specificity, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). These metrics provide a more comprehensive evaluation beyond simple accuracy, especially when dealing with imbalanced classes (e.g., CN vs. AD or sMCI vs. pMCI).

MRI Physics and its Diagnostic Value in AD: MRI plays a vital role in AD diagnosis. It is a non-invasive imaging modality that uses magnetic fields and radiofrequency pulses to produce high-resolution two- and three-dimensional images of the human body. By manipulating relaxation times after the radiofrequency pulse is turned off, two primary contrasts are generated: T1-weighted and T2-weighted MRI images.

MRI has been instrumental in detecting early AD-related brain changes. It enables visualization of brain regions such as the entorhinal cortex, amygdala, and hippocampus areas known to show atrophy during the early stages of AD. Moreover, MRI provides volumetric measurements that reveal changes in gray matter (GM) and white matter (WM), as well as the enlargement of ventricles and cortical atrophy associated with AD progression.

Discriminative Brain Features for AD Classification: Specific brain features extracted from MRI scans aid in distinguishing between AD, CN, and MCI subjects. Figure 4 illustrates the clear structural differences between the brains of AD and CN subjects. AD patients typically exhibit enlarged ventricles, reduced hippocampal volume, and significant cortical thinning. These morphological changes serve as reliable biomarkers for differentiating AD from CN.

However, the classification of MCI particularly distinguishing between stable and progressive MCI remains challenging. Brain tissue structures in CN and MCI individuals can be remarkably similar, making manual diagnosis difficult. Here, DL models offer a promising solution by detecting subtle patterns that may not be visible through conventional radiological examination. The ability of CNNs to extract discriminative features from raw MRI images enables accurate differentiation between CN, MCI, and AD stages, thereby aiding early and precise diagnosis.

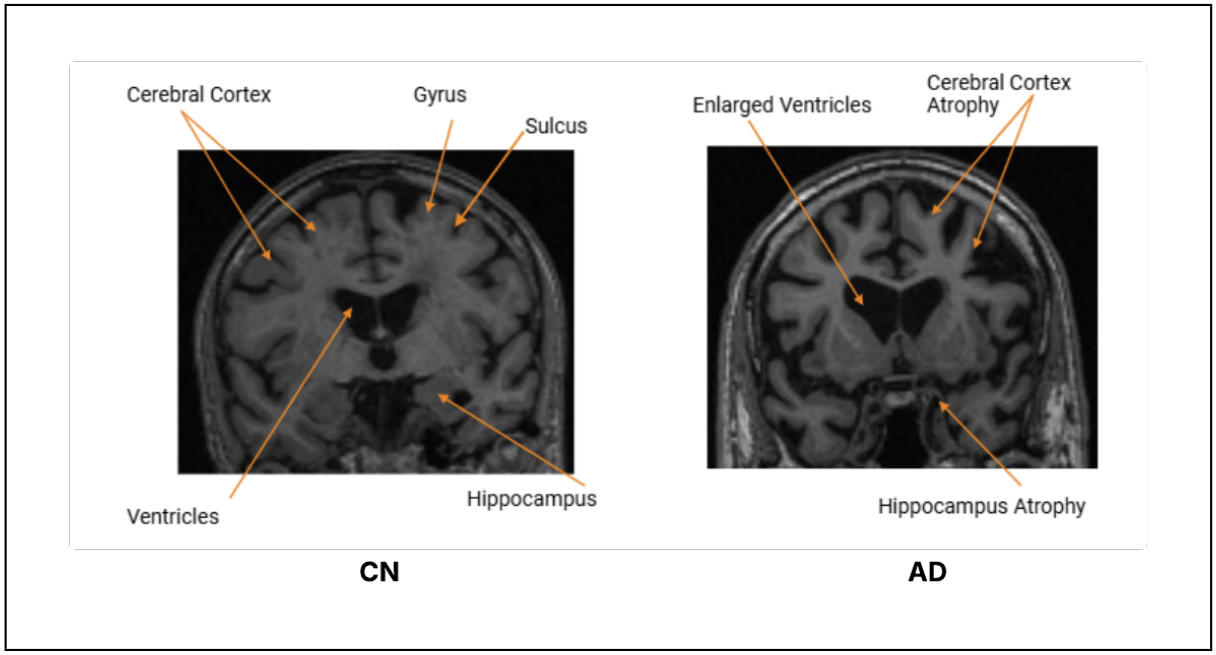


Figure 4: Structural brain differences between CN and AD subjects. The image shows noticeable enlargement of ventricles and hippocampal atrophy in AD compared to CN.

8 Discussion and Limitations

This systematic review highlights the rapid advancements in deep learning (DL) techniques applied to MRI-based diagnosis of Alzheimer's Disease (AD). While the majority of studies report high classification accuracies, several limitations must be considered when interpreting these results.

Model Interpretability and Clinical Adoption

One of the key concerns in deploying DL models in clinical settings is the lack of interpretability. Although many models achieve high performance metrics, understanding how and why a model makes certain decisions remains a challenge. Clinicians are more likely to adopt models that can provide visual explanations or highlight relevant brain regions contributing to a diagnosis. Some studies address this by integrating visualization tools such as Grad-CAM or saliency maps, but more consistent efforts are needed to enhance trust in DL predictions.

Data Leakage and Validation Strategy

A major issue identified in the reviewed articles is the potential for data leakage, particularly in studies that split MRI images randomly instead of by patient. Since multiple slices from the same individual are highly correlated, slice-level random splitting can result in overly optimistic performance. Patient-wise data separation is essential to ensure true generalization and to avoid inflated accuracy metrics. For instance, Young et al[53] conduct a review study on data leakage in AD classification using DL algorithms. They reported that, studies that correctly implement patient-wise data splitting often achieve classification accuracies between 60-90%, whereas those with higher risk of data leakage reported accuracies between 95-99%. The authors argued that, models with extreme high accuracy performance due data leakage consistently fail to generalize when applied to real world clinical trials. However, methodological rigorous approaches, such as patient-wise data splitting and external validation provides more robust yet realistic performance estimates. Additionally, data augmentation plays a significant role in model generalization and should strictly be applied to the training set to prevent data leakage. Studies such as Mehmood et al applied data augmentation before splitting the data into training, validation and testing set, which results into higher accuracy due to data leakage. Figure 5 illustrated the right approach for data augmentation and splitting, which helps to prevent the risk of data leakage during training, validation and testing phase. Unfortunately, several studies failed to clearly state whether their validation was conducted on a per-patient basis.

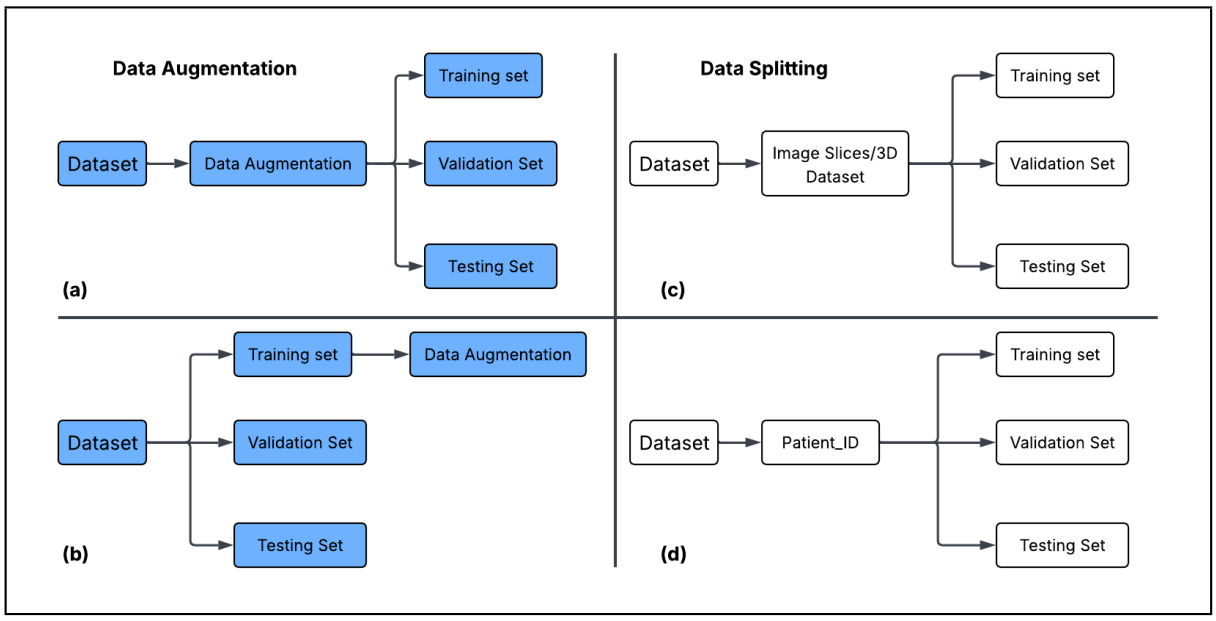


Figure 5: A summary of Data Augmentation and Splitting procedures.The figure illustrate (a) an improper Data Augmentation, (b) correct Data Augmentation, (c) an improper data splitting and (d) Correct data splitting approach.

Class Imbalance and Label Ambiguity

Another limitation is the challenge posed by imbalanced datasets, especially when distinguishing between stable MCI (sMCI) and progressive MCI (pMCI). The relatively small number of pMCI cases may lead models to favor the majority class, resulting in biased performance. Additionally, diagnostic boundaries between CN, sMCI, and pMCI can be ambiguous, even among expert clinicians. This ambiguity affects the quality of labels used during training, potentially reducing model accuracy.

Generalizability Across Datasets

Most studies rely heavily on the ADNI dataset, which, while comprehensive, may not fully represent the diversity of MRI acquisition protocols, scanners, or populations across different regions. The generalizability of trained models to other datasets or clinical environments remains uncertain. Future work should emphasize cross-dataset validation and model robustness in real-world clinical settings.

Computational Resources and Model Complexity

Many high-performing models require substantial computational resources for training and inference, which may not be readily available in all healthcare environments, particularly in resource-limited settings. This underscores the need for lightweight, efficient architectures that retain diagnostic accuracy while enabling deployment on standard hospital equipment.

Ethical and Privacy Considerations

The use of large-scale neuroimaging data raises important ethical considerations regarding patient privacy and data sharing. Although de-identified datasets such as ADNI offer a valuable resource for research, broader efforts are needed to ensure privacy compliance and data governance when developing clinical AI systems.

Overall, while DL-based AD classification using MRI is a promising field, robust validation, explainability, and clinical integration remain open challenges that require ongoing attention.

9 Conclusion

AD is a progressive neurodegenerative disorder primarily caused by the deposition of amyloid-beta (Aβ) plaques and neurofibrillary tangles (NFTs) in the brain, leading to the death of neurons responsible for memory, thinking, and learning. As the global population ages, the prevalence of AD continues to rise. The prodromal stage of AD, known as MCI, lies between cognitively normal individuals and those with full-blown AD. However, due to the high similarity in brain tissue between CN and MCI subjects, early diagnosis of AD remains a significant challenge.

MRI serves as a non-invasive imaging modality that enables the diagnosis of AD at the preclinical stage. When combined DL techniques, MRI-based diagnosis can be further enhanced for better prediction and classification. MRI offers detailed anatomical insights, such as hippocampal atrophy, ventricular enlargement, volumetric reduction in white matter (WM) and gray matter (GM), and cortical thinning key features associated with AD progression.

DL algorithms have demonstrated remarkable performance in the classification and prediction of AD, particularly when differentiating between CN, MCI, and AD subjects. Despite these advances, several challenges remain, particularly in explainability and model generalization.

Future studies should focus on improving explainable artificial intelligence (XAI) frameworks to provide better insight into the decision-making processes of DL models. Enhancing interpretability will foster greater trust among clinicians and support the clinical adoption of DL tools in early AD diagnosis and monitoring.

Acknowledgments

This research was supported by Hibah Publikasi Pascasarjana FMIPA under contract number PKS-38/UN2.F3D /PPM.00.02/2025. M. B. C. acknowledges UI-Great scholarship for financing his study in Univeritas Indonesia.

References

[1] S. K. Kim, Q. A. Duong, and J. K. Gahm, "Multimodal 3d deep learning for early diagnosis of alzheimer's disease," *IEEE Access*, vol. 12, pp. 46 278–46 289, 2024. doi: 10.1109/ACCESS.2024.3381862 .

[2] R. Rasi and A. Guvenis, "Predicting amyloid positivity from fdg-pet images using radiomics: A parsimonious model," *Computer Methods and Programs in Biomedicine*, vol. 247, p. 108098, 4 2024. doi: 10.1016/j.cmpb.2024.108098 .

[3] A. Mehmood, S. Yang, Z. Feng, M. Wang, A. S. Ahmad, R. Khan, M. Maqsood, and M. Yaqub, "A transfer learning approach for early diagnosis of alzheimer's disease on mri images," *Neuroscience*, vol. 460, pp. 43–52, 4 2021. doi: 10.1016/j.neuroscience.2021.01.002 .

[4] S. L, S. V, V. Ravi, G. E.A, and S. K.P, "Deep learning-based approach for multi-stage diagnosis of alzheimer's disease," *Multimedia Tools and Applications*, vol. 83, pp. 16 799–16 822, 7 2023. doi: 10.1007/s11042-023-16026-0 .

[5] K. Chen, Y. Weng, A. A. Hosseini, T. Dening, G. Zuo, and Y. Zhang, "A comparative study of gnn and mlp based machine learning for the diagnosis of alzheimer's disease involving data synthesis," *Neural Networks*, vol. 169, pp. 442–452, 1 2024. doi: 10.1016/j.neunet.2023.10.040 .

[6] Y. AbdulAzeem, W. M. Bahgat, and M. Badawy, "A cnn based framework for classification of alzheimer's disease," *Neural Computing and Applications*, vol. 33, pp. 10 415–10 428, 8 2021. doi: 10.1007/s00521-021-05799-w .

[7] M. Khojaste-Sarakhsi, S. S. Haghighi, S. F. Ghomi, and E. Marchiori, "A 3d multi-scale cyclegan framework for generating synthetic pets from mris for alzheimer's disease diagnosis," *Image and Vision Computing*, vol. 146, p. 105017, 6 2024. doi: 10.1016/j.imavis.2024.105017 .

[8] H. Ashrafian, E. H. Zadeh, and R. H. Khan, "Review on alzheimer's disease: Inhibition of amyloid beta and tau tangle formation," *International Journal of Biological Macromolecules*, vol. 167, pp. 382–394, 1 2021. doi: 10.1016/j.ijbiomac.2020.11.192 .

[9] L. V. Hoecke, J. Castelein, J. Xie, L. V. Acker, G. V. Imschoot, E. V. Wonterghem, I. Vlaeminck, D. Verhaege, W. Claeys, K. Wierda, Z. Callaerts-Vegh, and R. E. Vandenbroucke, "An immunological puzzle: The adaptive immune system fuels alzheimer's disease pathology," *Brain, Behavior, and Immunity*, vol. 117, pp. 122–134, 3 2024. doi: 10.1016/j.bbi.2023.12.026 .

[10] M. Odusami, R. Maskeliūnas, R. Damaševičius, and S. Misra, "Machine learning with multimodal neuroimaging data to classify stages of alzheimer's disease: a systematic review and meta-analysis," *Cognitive Neurodynamics*, vol. 18, pp. 775–794, 6 2024. doi: 10.1007/s11571-023-09993-5 .

[11] L. C. Adams, K. K. Bressem, K. Ziegeler, J. L. Vahldiek, and D. Poddubnyy, "Artificial intelligence to analyze magnetic resonance imaging in rheumatology," *Joint Bone Spine*, vol. 91, p. 105651, 5 2024. doi: 10.1016/j.jbspin.2023.105651 .

[12] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, cnn architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, p. 53, 3 2021. doi: 10.1186/s40537-021-00444-8 .

[13] N. Pradhan, S. Sagar, and A. S. Singh, "Analysis of mri image data for alzheimer disease detection using deep learning techniques," *Multimedia Tools and Applications*, vol. 83, pp. 17 729–17 752, 7 2023. doi: 10.1007/s11042-023-16256-2 .

[14] A. M. H. Taha, Q. M. M. Zarandah, B. S. Abunasser, and A. M. B. S. S. A.-N. Z. K. D. Alkayyali, "Advancements in ai for medical imaging: Transforming diagnosis and treatment," *International Journal of Academic Engineering Research(Ijaer)*, vol. 8, pp. 8–15, 2024.

[15] M. A. Bravo-Ortiz, S. A. Holguin-Garcia, S. Quiñones-Arredondo, A. Mora-Rubio, E. Guevara-Navarro, H. B. Arteaga-Arteaga, G. A. Ruz, and R. Tabares-Soto, "A systematic review of vision transformers and convolutional neural networks for alzheimer's disease classification using 3d mri images," *Neural Computing and Applications*, vol. 36, pp. 21 985–22 012, 12 2024. doi: 10.1007/s00521-024-10420-x .

[16] R. Sharma, T. Goel, M. Tanveer, C. T. Lin, and R. Murugan, "Deep-learning-based diagnosis and prognosis of alzheimer's disease: A comprehensive review," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, pp. 1123–1138, 9 2023. doi: 10.1109/TCDS.2023.3254209 .

[17] X. Meng, J. Liu, X. Fan, C. Bian, Q. Wei, Z. Wang, W. Liu, and Z. Jiao, "Multi-modal neuroimaging neural network-based feature detection for diagnosis of alzheimer's disease," *Frontiers in Aging Neuroscience*, vol. 14, 5 2022. doi: 10.3389/fnagi.2022.911220 .

[18] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, 12 2022. doi: 10.1038/s41598-022-09954-8 .

[19] S. D. Serai, "Basics of magnetic resonance imaging and quantitative parameters t1, t2, t2*, t1rho and diffusion-weighted imaging," *Pediatric Radiology*, vol. 52, pp. 217–227, 2 2022. doi: 10.1007/s00247-021-05042-7 .

[20] J. Kim, M. Jeong, W. R. Stiles, and H. S. Choi, "Neuroimaging modalities in alzheimer's disease: Diagnosis and clinical features," *International Journal of Molecular Sciences*, vol. 23, p. 6079, 5 2022. doi: 10.3390/ijms23116079 .

[21] P. N. Furtado, "Influence of mri modality on accuracy of deep learning segmentation of abdominal organs," in *2020 3rd International Conference on Digital Medicine and Image Processing*. ACM, 11 2020, pp. 56–62. doi: 10.1145/3441369.3441379 .

[22] D. Kumar, B. Pratap, N. Boora, R. Kumar, and N. K. Sah, "A comparative study of medical imaging modalities," *International Journal of Radiology Sciences*, vol. 3, pp. 9–16, 1 2021. doi: 10.33545/26649810.2021.v3.i1a.11 .

[23] G. M. Currie, P. Kamvosoulis, and S. Bushong, "Pet/mri, part 2: Technologic principles," *Journal of Nuclear Medicine Technology*, vol. 49, pp. 217–225, 9 2021. doi: 10.2967/JNMT.120.261862 .

[24] S. Caspani, R. Magalhães, J. P. Araújo, and C. T. Sousa, "Magnetic nanomaterials as contrast agents for mri," 6 2020. doi: 10.3390/ma13112586 .

[25] B. Brito, T. W. Price, J. Gallo, M. Bañobre-López, and G. J. Stasiuk, "Smart magnetic resonance imaging-based theranostics for cancer," pp. 8706–8737, 2021. doi: 10.7150/thno.57004 .

[26] H. Hampel, J. Hardy, K. Blennow, C. Chen, G. Perry, S. H. Kim, V. L. Villemagne, P. Aisen, M. Vendruscolo, T. Iwatsubo, C. L. Masters, M. Cho, L. Lannfelt, J. L. Cummings, and A. Vergallo, "The amyloid- pathway in alzheimer's disease," *Molecular Psychiatry*, vol. 26, pp. 5481–5503, 10 2021. doi: 10.1038/s41380-021-01249-0 .

[27] S. Miao, Q. Xu, W. Li, C. Yang, B. Sheng, F. Liu, T. T. Bezabih, and X. Yu, "¡scp¿mmtfn¡/scp¿ : Multimodal multi-scale transformer fusion network for alzheimer's disease diagnosis," *International Journal of Imaging Systems and Technology*, vol. 34, 1 2024. doi: 10.1002/ima.22970 .

[28] N. B. Balakrishnan, S. P. S, and J. J. Panackal, "Alzheimer's disease diagnosis using machine learning: A review," *International Journal of Engineering Trends and Technology*, vol. 71, pp. 120–129, 3 2022. doi: 10.14445/22315381/IJETT-V71I3P213 .

[29] S. Afzal, M. Maqsood, U. Khan, I. Mehmood, H. Nawaz, F. Aadil, O.-Y. Song, and Y. Nam, "Alzheimer disease detection techniques and methods: A review," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, p. 26, 2021. doi: 10.9781/ijimai.2021.04.005 .

[30] C. Villa, M. Lavitrano, E. Salvatore, and R. Combi, "Molecular and imaging biomarkers in alzheimer's disease: A focus on recent insights," *Journal of Personalized Medicine*, vol. 10, p. 61, 7 2020. doi: 10.3390/jpm10030061 .

[31] T. J. Saleem, S. R. Zahra, F. Wu, A. Alwakeel, M. Alwakeel, F. Jeribi, and M. Hijji, "Deep learning-based diagnosis of alzheimer's disease," *Journal of Personalized Medicine*, vol. 12, p. 815, 5 2022. doi: 10.3390/jpm12050815 .

[32] J. Gobom, A. Brinkmalm, G. Brinkmalm, K. Blennow, and H. Zetterberg, "Alzheimer's disease biomarker analysis using targeted mass spectrometry," *Molecular & Cellular Proteomics*, vol. 23, p. 100721, 2 2024. doi: 10.1016/j.mcpro.2024.100721 .

[33] C. L. Saratxaga, I. Moya, A. PicÃ³n, M. Acosta, A. M.-F. de leceta, E. Garrote, and A. Bereciartua-Perez, "Mri deep learning-based solution for alzheimerâ€™s disease prediction," *Journal of Personalized Medicine*, vol. 11, 9 2021. doi: 10.3390/jpm11090902 .

[34] B. S. Rao and M. Aparna, "A review on alzheimer's disease through analysis of mri images using deep learning techniques," *IEEE Access*, vol. 11, pp. 71 542–71 556, 2023. doi: 10.1109/ACCESS.2023.3294981 .

[35] A. D. Arya, S. S. Verma, P. Chakarabarti, T. Chakrabarti, A. A. Elngar, A. M. Kamali, and M. Nami, "A systematic review on machine learning and deep learning techniques in the effective diagnosis of alzheimerâ€™s disease," *Brain Informatics*, vol. 10, 12 2023. doi: 10.1186/s40708-023-00195-7 .

[36] A. Loddo, S. Buttau, and C. D. Ruberto, "Deep learning based pipelines for alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method," *Computers in Biology and Medicine*, vol. 141, 2 2022. doi: 10.1016/j.compbiomed.2021.105032 .

[37] W. Salehi, P. Baglat, G. Gupta, S. B. Khan, A. Almusharraf, A. Alqahtani, and A. Kumar, "An approach to binary classification of alzheimer's disease using lstm," *Bioengineering*, vol. 10, 8 2023. doi: 10.3390/bioengineering10080950 .

[38] E. Yagis, L. Citi, S. Diciotti, C. Marzi, S. W. Atnafu, and A. G. D. Herrera, "3d convolutional neural networks for diagnosis of alzheimer's disease via structural mri," in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, vol. 2020-July. Institute of Electrical and Electronics Engineers Inc., 7 2020, pp. 65–70. doi: 10.1109/CBMS49503.2020.00020 .

[39] M. F. Ahmad, S. Akbar, S. A. E. Hassan, A. Rehman, and N. Ayesha, "Deep learning approach to diagnose alzheimer's disease through magnetic resonance images," in *2021 International Conference on Innovative Computing (ICIC)*. IEEE, 11 2021, pp. 1–6. doi: 10.1109/ICIC53490.2021.9693041 .

[40] J. S. Kim, J. W. Han, J. B. Bae, D. G. Moon, J. Shin, J. E. Kong, H. Lee, H. W. Yang, E. Lim, J. Y. Kim, L. Sunwoo, S. J. Cho, D. Lee, I. Kim, S. W. Ha, M. J. Kang, C. H. Suh, W. H. Shim, S. J. Kim, and K. W. Kim, "Deep learning-based diagnosis of alzheimer's disease using brain magnetic resonance images: an empirical study," *Scientific Reports*, vol. 12, 12 2022. doi: 10.1038/s41598-022-22917-3 .

[41] Y. Lin, X. Li, Y. Zhang, and J. Tang, "Attention-based efficient classification for 3d mri image of alzheimer's disease," in *Proceedings of the 2023 6th International Conference on Sensors, Signal and Image Processing*. ACM, 10 2023, pp. 34–39. doi: 10.1145/3653863.3653865 .

[42] S. Fathi, A. Ahmadi, A. Dehnad, M. Almasi-Dooghaee, and M. Sadegh, "A deep learning-based ensemble method for early diagnosis of alzheimer's disease using mri images," *Neuroinformatics*, vol. 22, pp. 89–105, 1 2024. doi: 10.1007/s12021-023-09646-2 .

[43] S. L, S. V, V. Ravi, G. E.A, and S. K.P, "Deep learning-based approach for multi-stage diagnosis of alzheimerâ€™s disease," *Multimedia Tools and Applications*, vol. 83, pp. 16 799–16 822, 2 2024. doi: 10.1007/s11042-023-16026-0 .

[44] M. Raju, V. P. Gopi, V. S. Anitha, and K. A. Wahid, "Multi-class diagnosis of alzheimerâ€™s disease using cascaded three dimensional-convolutional neural network," *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 1219–1228, 12 2020. doi: 10.1007/s13246-020-00924-w .

[45] F. U. R. Faisal and G.-R. Kwon, "Automated detection of alzheimer's disease and mild cognitive impairment using whole brain mri," *IEEE Access*, vol. 10, pp. 65 055–65 066, 2022. doi: 10.1109/ACCESS.2022.3180073 .

[46] Z. Jin, J. Gong, M. Deng, P. Zheng, and G. Li, "Deep learning-based diagnosis algorithm for alzheimer's disease," *Journal of Imaging*, vol. 10, 12 2024. doi: 10.3390/jimaging10120333 .

[47] Z. S. Aaraji and H. H. Abbas, "Automatic diagnosis of alzheimer's disease using deep learning techniques," *NeuroQuantology*, vol. 19, pp. 126–140, 2021. doi: 10.14704/nq.2021.19.11.NQ21183 .

[48] S. Chen, J. Zhang, X. Wei, and Q. Zhang, "Alzheimer's disease classification using structural mri based on convolutional neural networks," in *Proceedings of the 2020 2nd International Conference on Big-data Service and Intelligent Computation*. ACM, 12 2020, pp. 7–13. doi: 10.1145/3440054.3440056 .

[49] A. Ebrahimi, S. Luo, and R. Chiong, "Deep sequence modelling for alzheimer's disease detection using mri," *Computers in Biology and Medicine*, vol. 134, p. 104537, 7 2021. doi: 10.1016/j.compbiomed.2021.104537 .

[50] R. Janghel and Y. Rathore, "Deep convolution neural network based system for early diagnosis of alzheimer's disease," *IRBM*, vol. 42, pp. 258–267, 8 2021. doi: 10.1016/j.irbm.2020.06.006 .

[51] J. Li, Y. Wei, C. Wang, Q. Hu, Y. Liu, and L. Xu, "3-d cnn-based multichannel contrastive learning for alzheimer's disease automatic diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022. doi: 10.1109/TIM.2022.3162265 .

[52] M. Khojaste-Sarakhsi, S. S. Haghighi, S. M. Ghomi, and E. Marchiori, "Deep learning for alzheimer's disease diagnosis: A survey," 8 2022. doi: 10.1016/j.artmed.2022.102332 .

[53] V. M. Young, S. Gates, L. Y. Garcia, and A. Salardini, "Data leakage in deep learning for alzheimerâ€™s disease diagnosis: A scoping review of methodological rigor and performance inflation," 9 2025. doi: 10.3390/diagnostics15182348 .