

Att-ConvFE-ResNet: A Spatial Attention and Multi-Scale Feature Extraction Framework for Alzheimer's Disease Diagnosis

Upendra Singh*, Vidit Kumar and Bhaskar Pant

Department of Computer Science and Engineering, Graphic Era
(Deemed to be University), Dehradun, Uttarakhand, India.

*Corresponding Author E-mail:aswal.upendra2010@gmail.com

<https://dx.doi.org/10.13005/bpj/3439>

(Received: 16 October 2025; accepted: 12 February 2026)

Alzheimer's disease (AD) presents a neurological challenge, requiring accurate and early diagnosis for effective intervention. In this study an efficient deep learning (DL) framework is proposed, which integrates spatial attention mechanisms with the ResNet50 core network to improve the classification of cognitive conditions, including AD, mild cognitive impairment (MCI), and normal cognitive impairment (CN). The novel aspect of our work lies in the strategic integration of two dedicated spatial attention blocks within the middle and deep layers of the pre-trained ResNet50 backbone. This architecture improves traditional feature extraction by capturing spatial dependencies and focusing the model on relevant regions. We also introduced two dedicated convolutional feature extraction blocks that receive input from the intermediate layers of the ResNet50 and allowing the model to leverage both high-level contextual information and fine-grained positional spatial properties simultaneously. Our efficient and attention-enhancing design significantly improves discriminatory power. The proposed model demonstrates excellent performance, achieving an accuracy rate of 88.46% (AD vs. MCI vs. CN) and 89.23% (AD vs. CN) with the ADNI dataset, which is a significant improvement over previous methods. These results demonstrate the benefit of incorporating spatial attention into deep convolutional neural networks for AD diagnosis.

Keywords: Artificial intelligence(AI); Convolutional Feature Extractor Blocks(CFEB); Spatial Attention Block(SAB); Deep Learning(DL); Alzheimer's Disease Neuroimaging Initiative(ADNI); Positron Emission Tomography(PET); Machine learning(ML); Image-Classification.

AD is a progressive neurological syndrome that affects millions of individuals globally and is known as the principal cause of dementia. AD is primarily related to reduced brain activity and impaired blood flow.

Currently, there is no cure or technique to stop its progression.¹ In later stages of the disease, individuals may experience difficulties with language, mood swings, and significant behavioral

changes. Early diagnosis of AD relies on medical data² advanced computational algorithms, and clinical expertise. Artificial intelligence (AI)-based healthcare solutions are being rapidly researched to support this diagnostic process. Researchers have utilized a variety of tools to effectively study and detect Alzheimer's disease, including magnetic resonance imaging (MRI) and biomarkers such as chemical indicators and blood flow metrics.

Previous studies suggest that integrating AI with medical imaging³ can significantly aid in the early detection of AD. Traditional diagnostic methods and human judgment often fail to accurately identify Alzheimer's disease, leading to inconsistencies in diagnosis and treatment. To overcome these limitations, there has been a growing need for advanced and computationally intensive techniques based on artificial intelligence (AI). By incorporating such techniques into clinical data, they empower healthcare professionals to make evidence-based decisions, assist healthcare professionals in conducting robust assessments, and ultimately contribute to improving patients' lives. Machine learning (ML) and deep learning (DL) methods show significant potential for enhancing diagnostic accuracy and enabling early intervention. The automation of image analysis through AI algorithms not only facilitates real-time monitoring but also reduces inter-observer variability and provides consistent evaluations. ^u

Beyond AD focused research, significant progress has also been made in related healthcare domains using advanced ML techniques. Li *et al.*⁶ proposed a method for brain tumor segmentation study; they employed federated learning combined with differential privacy to train deep neural networks across decentralized MRI data sources. Zhou, Wang & Zhou⁷ applied federated learning with EfficientNet-B0 to enhance diagnostic accuracy while maintaining privacy in brain tumor classification tasks. The ensemble applies a weighted average of predictions and achieves notably lower MAE and RMSE. MRI-based biomarkers have been widely studied to support early diagnosis and differentiation between AD, CN, and MCI cases. ^{8,9} Traditional diagnostic approaches, however, often suffer from subjectivity and limited reproducibility. Some studies have explored the classification of AD and MCI with sMRI, comparing various algorithms and feature extraction techniques. ^{10,11} Current advancements in DL techniques have proved the potential to classify AD, making them more acceptable for clinical use. ¹² Furthermore, decision support systems that integrate imaging data with computational intelligence are being developed to assist clinicians in predicting disease severity and its expansion. ¹³

Moreover, attention-based approaches have shown potential in improving CNN

performance. In this regard, there are some works in computer vision and natural language domains.

¹⁴⁻¹⁷ There are also some efforts in AD classification.

^{18,19} Attention reduces overfitting by encouraging the network to generalize across varied spatial patterns, especially useful in heterogeneous clinical datasets. By weighting spatial or channel-wise features differently, attention modules enable adaptive feature recalibration, which enhances the model's capacity to capture elusive and localized patterns important for medical tasks like AD staging. Therefore, in this paper we proposed and incorporate attention mechanism with base CNN to expand feature representation.

Our contributions to this work are the following:

- A systematic evaluation of six distinct model configurations is carried out by integrating ResNet50 with the proposed spatial attention mechanism and dual convolutional feature extraction pathways. Unlike prior works that typically report a single architecture, this evaluation enables identifying the most effective configuration for enhancing feature discrimination in AD stage classification.

- Two novel modules are designed, namely the Convolutional Feature Extractor Blocks (CFEB) and the Spatial Attention Block (SAB), which together improve feature representation. CFEB block captures multi-scale feature patterns, while SAB block emphasizes spatially important regions, thereby enhancing class separability between AD, MCI, and CN.

- An efficient attention-based model, Att-ConvFE-ResNet, is proposed by combining the designed CFEB and SAB modules with ResNet50. The proposed approach leverages dual convolutional pathways with spatial attention to achieve richer feature learning.

- The proposed approach demonstrates excellent performance compared with recent existing methods when evaluated on the ADNI dataset, highlighting its effectiveness in AD classification.

The remainder of this paper is organized as follows: Section 2 reviews the related work on deep learning approaches for Alzheimer's Disease diagnosis. Section 3 describes the methodology of proposed attention-guided residual network-based framework in detail. Section 4 provides the experimental details with dataset description, pre-processing, and training strategy and ablation study.

In section we have discussion. Finally, Section 6 concludes the study and highlights potential future research directions.

Related Work

Most of the works are based on usage of single CNN models; some are discussed here. For instance, an explainable DL model that incorporates a local data-driven interpretation technique was reported by Ekuma, Hier, & Obafemi-Ajayi²⁰ to predict the severity of AD. Xu et al.,²¹ presented an approach to classify AD using MRI images. Aghaei, Ebrahimi Moghaddam, & Malek²² suggested an approach to address the early-stage diagnosis of AD. They used a transfer learning technique that combines SVM and Softmax classifiers with Inception-V3. An attention-based technique is investigated to enhance the deep CNN's feature representation capacity. Iglesias²³ presented an Easy-Reg model which combines the best aspects of deep learning techniques, traditional registration tools, and their latest work on domain randomization. Early work by Fang et al.²⁴ presented an approach aiming at enhancing early detection of AD using sMRI scans. This approach demonstrated promising classification performance by preserving discriminative information in reduced feature spaces. Building on this, several studies have proposed hybrid DL models for the uncovering and monitoring of AD progression. Further, Abuhmed, El-Sappagh, & Alonso²⁵ developed a robust approach using CNNs to improve accuracy and reduce overfitting in longitudinal datasets.

Recently, machine learning techniques have been extensively applied across diverse medical domains, offering new possibilities for disease forecast and diagnosis. For example, Çalýþkan, Abidin²⁶ proposed an approach for diagnosis of malaria disease by integrating chi-square feature selection algorithm with convolutional neural networks and auto encoder network. Further, C. Aslým, A. Çalýþkan²⁷ developed a model for improving brain tumor detection using VGG-19, ResNet-101 and DenseNet-121 DS pre-trained CNN with MRI. The growing attention in multimodal research has also led to the development of more comprehensive frameworks. A DL architecture presented by Venugopalan, Tong, Hassanzadeh, & Wang²⁸ that integrated neuroimaging and clinical data to improve early

detection of AD stages and highlighting the advantages of combining sMRI with additional modalities. Similarly, Singh, Kumar, & Pant²⁹ enhanced CNN-based diagnosis by combining spatial attention mechanisms which enable the network to emphasize discriminative sections in MRI scans. Further, Ajagbe, Amuda, Oladipupo, AFE, & Okesola³⁰ explored multiclass classification using deep CNNs, which outperformed traditional methods. Moreover, Haq, Huang, Kang, Haq, & Zhan³¹ proposed a novel DL-based classification approach for multi-class calcification using MRI scans. Another approach is also given by Zhang, Liu, An, Gao, & Shen³² for AD diagnosis. Moreover, Yagis et al.³³ advised about potential pitfalls such as data leakage, emphasizing the importance of proper cross-validation and data partitioning in 2D CNN applications. Latest research has also shifted toward the development of interpretable and explainable DL systems. Further, Atnafu & Diciotti³⁴ presented an interpretable framework that improves transparency in prediction. An attention based mechanisms is presented by Hoang, Lee, & Kim³⁵ to highlight clinically relevant sections in MRI scans for improving early detection of AD.

Several recent works have focused on attention-guided and transformer-based architectures. Gao, Shi, Shen, & Liu³⁶ utilized a multimodal transformer to handle incomplete imaging data in AD diagnosis. In the area of explain-ability, Mulyadi et al.³⁷ introduced a clinically guided prototype learning method to produce interpretable AD likelihood maps. Advanced generative approaches have also been employed. Gao, Shi, Shen, & Liu³⁸ used attention and GANs for classification in multimodal datasets, and Hao et al.³⁹ applied hyper-graph convolutional networks for longitudinal analysis. Additionally, Mora-Rubio et al.⁴⁰ performed stage-wise classification of AD using MRI and deep learning, showing significant performance improvements over traditional techniques. Further, J. Zhang et al.⁴¹ suggested a cross-attention-based multimodal network that effectively integrates multiple imaging sources, while Poloni & Ferrari⁴² developed an automated hippocampal landmark classification system, further enhancing the reliability of AD diagnosis.

To this end, an efficient attention-based approach is explored to boost the feature depiction

ability of the deep CNN (ResNet50) for AD diagnosis with sagittal MRI data. The above related work is summarized in table 1.

MATERIALS AND METHODS

Proposed architecture

This work examines the use of spatial attention and convolutional extraction blocks with ResNet50 as base model, for classifying AD. Spatial attention module helps the model to emphasis on important features and convolutional extraction blocks extract the features, which makes it more useful. Due to this, a new architecture i.e. Att-ConvFE-ResNet is proposed with deep learning architecture based on ResNet50, augmented with spatial attention mechanisms and multi-branch convolutional modules for AD. Figure 1 depicts our proposed approach, it consists of three main blocks: 1) base model feature extractor, 2) Spatial Attention module and 3) Convolutional Feature extraction block1 and Convolutional Feature extraction block2. The block takes an input image of size $224 \times 224 \times 3$ and passes it through a backbone for hierarchical feature extraction. The base model is composed of five convolutional stages (conv1 to conv5), with the spatial resolution progressively reduced and the number of channels increased from 64 up to 2048. The proposed approach integrate spatial attention

blocks at two critical stages: after the conv4 output ($14 \times 14 \times 1024$) and the conv5 output ($7 \times 7 \times 2048$) by which it can find the most relevant features of the image. The attention module help the network to emphasize spatial locations that contribute more significantly to the classification task, improving the model's discriminative capability. Moreover, two convolutional feature extraction blocks are presented in parallel. Both utilized the conv5 feature map. Each of these blocks applies convolutional operations to extract different levels of semantic and spatial information, resulting in feature maps of size 512 and 1024 respectively. The outputs from both convolutional feature blocks and the spatial attention block are concatenated along the channel dimension, yielding features size 2560. This feature map is then concatenate with output of the final dense(512) layer in ResnetNet50 network. Finally it maps the features to three output neurons, corresponding to the classes AD, MCI, and CN. Proposed architecture utilize the strength of deep residual learning with attention mechanisms to emphasis on the most informative spatial features. Further, the use of multi-branch convolution and feature fusion allows the network to harmonizing the information from different pathways which results in improved classification output. This makes the proposed model a powerful tool for the automated diagnosis of AD using MRI scans. Algorithm1 provides a step-by-step

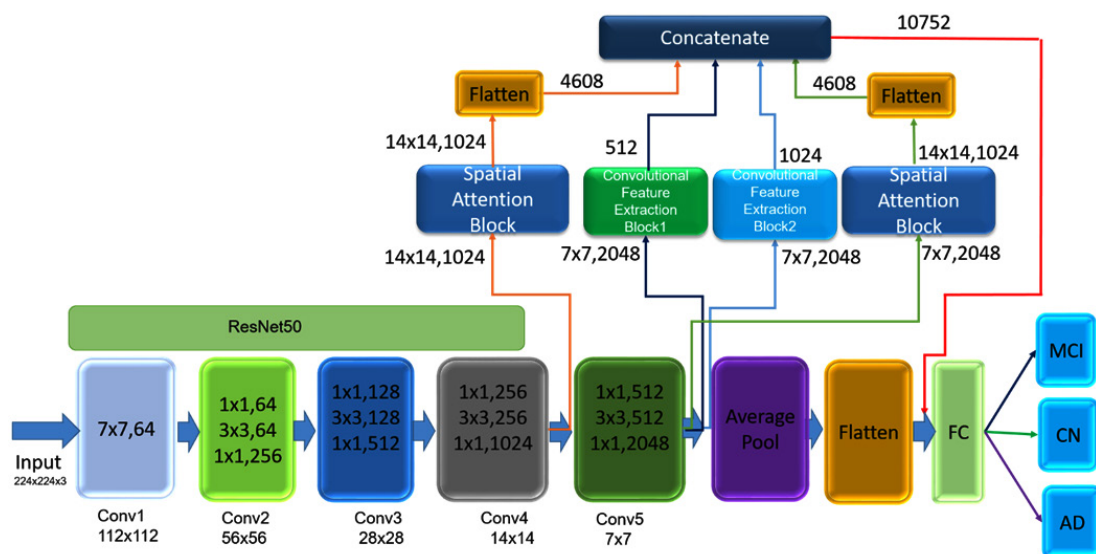


Fig. 1. Proposed architecture

representation of the proposed architecture, capturing the sequential flow of operations from input to final classification.

Algorithm 1:

Step 1. **Input:** Image of size $224 \times 224 \times 3$

Step 2. **Feature Extraction via base backbone:**

Pass input through initial backbone layers:

Conv1: 7×7 convolution with 64 filters $\rightarrow$ output size $112 \times 112 \times 64$

Conv2: $1 \times 1, 64$; $3 \times 3, 64$; $1 \times 1, 256 \rightarrow$ output $56 \times 56 \times 256$

Conv3: $1 \times 1, 128$; $3 \times 3, 128$; $1 \times 1, 512 \rightarrow$ output $28 \times 28 \times 512$

Conv4: $1 \times 1, 256$; $3 \times 3, 256$; $1 \times 1, 1024 \rightarrow$ output $14 \times 14 \times 1024$

Conv5: $1 \times 1, 512$; $3 \times 3, 512$; $1 \times 1, 2048 \rightarrow$ output $7 \times 7 \times 2048$

Step 3. **Attention and Feature Enhancement:**

Apply Spatial Attention Block 1 to Conv4 output $\rightarrow$ output $14 \times 14 \times 1024$

Apply Spatial Attention Block 2 to Conv5 output $\rightarrow$ output $7 \times 7 \times 2048$

Step 4. **Convolutional Feature Extraction:**

From Conv5 output ($7 \times 7 \times 2048$), extract features using Convolutional Feature Extraction Block 1: Conv: 7×7 , 512 filters $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Flatten $\rightarrow$ Dense(512)

Step 5. From attention-refined Conv5 output ($7 \times 7 \times 2048$), extract features using Convolutional Feature Extraction Block 2: Conv: 5×5 , 512 filters $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Flatten $\rightarrow$ Dense (1024).

Step 6. Feature Fusion and Pooling: Concatenate outputs of Spatial Attention Blocks, Convolutional Feature Extraction Block1, and Feature Extraction Block2.

Step 7. Apply dense(512) to concatenated features (Con_Features).

Step 8. Finally concatenate the output of ResNet50 after Dense(512) with Con_Features to produce, Final_Features

Step 9. Output: Pass the *Final_Features* to a Fully Connected (FC) layer. Apply softmax activation to predict class probabilities. Output class with the highest probability among: AD, MCI, and CN.

Mathematical Explanation of the above algorithm for attention-enhanced architecture:

Let the input image be, $X \in R^{224 \times 224 \times 3}$. Then it fed to initial convolution $F_1 = \text{Conv}_{7 \times 7, 64}(X)$,

with stride=2, $\Rightarrow F_1 \in R^{112 \times 112 \times 64}$, followed by *Residual Blocks*:

Conv2 (after 3 residual blocks):

$$F_2 = \text{ResBlock}_2(F_1) \in R^{56 \times 56 \times 256}$$

$$\text{Conv3: } F_3 = \text{ResBlock}_3(F_2) \in R^{28 \times 28 \times 512}$$

$$\text{Conv4: } F_4 = \text{ResBlock}_4(F_3) \in R^{14 \times 14 \times 1024}$$

$$\text{Conv5: } F_5 = \text{ResBlock}_5(F_4) \in R^{7 \times 7 \times 2048}$$

where each ResBlock follows:

$$\text{ResBlock}(F) = F + \text{ReLU}(\text{BN}(\text{Conv}(F)))$$

Spatial Attention Block (SAB)

Let $F \in R^{H \times W \times C}$, $F = \{F_1, F_2, F_3, F_4, F_5\}$. The attention map is computed as:

$$\text{Mavg} = \text{AvgPool}_{\text{channel}}(F \in R^{H \times W \times 1})$$

Concatenation & Convolution:
 $M_s = \sigma(\text{Conv}_{7 \times 7}(\text{Mavg}))$

$$\text{Attention output: } S(F) = M_s \odot F$$

Where σ is the sigmoid activation and $\odot$ dot element-wise multiplication.

Apply SAB to $F_4 \Rightarrow A_1 = S(F_4) \in R^{14 \times 14 \times 1024}$ and $F_5 \Rightarrow A_2 = S(F_5) \in R^{7 \times 7 \times 2048}$

Then we applied Convolutional Feature Extraction Blocks (Block1 and Block2) to F_5 as:

$$z_1 = \text{ReLU}(\text{BN}(\text{Conv}_{7 \times 7, 512}(F_5))) \in R^{1 \times 1 \times 512}$$

$$z_1 = \text{Flatten}(z_1) \in R^{512}$$

$$z_2 = \text{ReLU}(\text{BN}(\text{Conv}_{5 \times 5, 512}(F_5))) \in R^{1 \times 1 \times 1024}$$

$$z_2 = \text{Flatten}(z_2) \in R^{512}$$

Then computed features are concatenated:

$$F_{\text{Concat}} = \text{Concat}(A_1, A_2, z_1, z_2) \in R^{2560}$$

$$\text{Apply } F_{f\text{concat}} = \text{Dense}_{512}(F_{\text{Concat}})$$

Concatenate the output of ResNet50 after Dense_{512} with $F_{f\text{concat}}$

$$F_{\text{features}} = \text{Concatenate}(A_{\text{Dense}_{512}}, F_{f\text{concat}}) \in R^{512}$$

Final Classification: Apply a fully connected layer followed by softmax:

$$y = \text{Softmax}(W_f^T F_{\text{features}} + b_f) \in R^3,$$

where $W_f \in R^{512}$, $b_f \in R^3$ and y contains probabilities for each class AD, MCI, and CN

Convolutional Feature extraction blocks

The figure 2 integrates two parallel Convolutional Feature Extraction Blocks (CFEBs) to improve the variety of features captured

from the high-dimensional output of ResNet50 ($7 \times 7 \times 2048$). Each block is designed with a distinct convolutional kernel and output dimension to extract complementary representations. The block1 begins by applying a 2D convolution operation with a 7×7 kernel and 512 filters to the input tensor. The output of the convolution is then normalized using Batch Normalization (BN) and passed to ReLU (Rectified Linear Unit). These features are then flattened into a one-dimensional vector, which is subsequently passed through a dense (fully connected) layer with 512 units to produce the final feature representation $F_1 \in R^{512}$ for this block. Block2 follows the same sequence of operations as Block1 but instead applies a convolution with a smaller 5×5 kernel and 512 filters. This produces feature vector $F_2 \in R^{1024}$. Both blocks transform the input feature map into vectorised representations of different dimensions (512 and 1024), which are later concatenated and fed into the classification pipeline. The distinct configurations of Block1 and Block2 ensure that both global and local spatial dependencies are captured. Let $I_1 \in R^{7,7,2048}$ be the input to both the blocks.

$$F_1 = Dense_{512} \left(Flatten \left(ReLU \left(BN \left(Conv_{7 \times 7, 512} (I_1) \right) \right) \right) \right) \\ F_1 \in R^{512} \quad \dots(1)$$

$$F_2 = Dense_{512} \left(Flatten \left(ReLU \left(BN \left(Conv_{5 \times 5, 512} (I_1) \right) \right) \right) \right) \\ F_2 \in R^{1024} \quad \dots(2)$$

Spatial-Attention Block

To extract extra representative features for improving disease area in MRI images, a spatial attention module is integrated into the backbone CNN. The spatial attention module is tailored to suitably extract local structural features from patches of 2D feature maps. Figure 3 displays the spatial attention block's construction. Working of the block is given below.

The attention module combines four different branches, processes them through convolutions and fusions, and generates a spatial attention map using a series of operations. Mathematical formulation of the attention-module: Let $I \in R^{X \times Y \times Z}$ be the features extracted, in branch1 following operations are applied on the features $X1 = Upsampling(Resaphae(GAP(I))) \in R^{X \times Y \times Z}$, where $GAP(X) = \frac{1}{XY} \sum_{i=1}^X \sum_{j=1}^Y I_{ij} \in R^Z$

In branch 2 following operations are performed on the feature extracted.

$$X2 = ReLU(W * I + b) \in R^{X \times Y \times 1} \quad \dots(4)$$

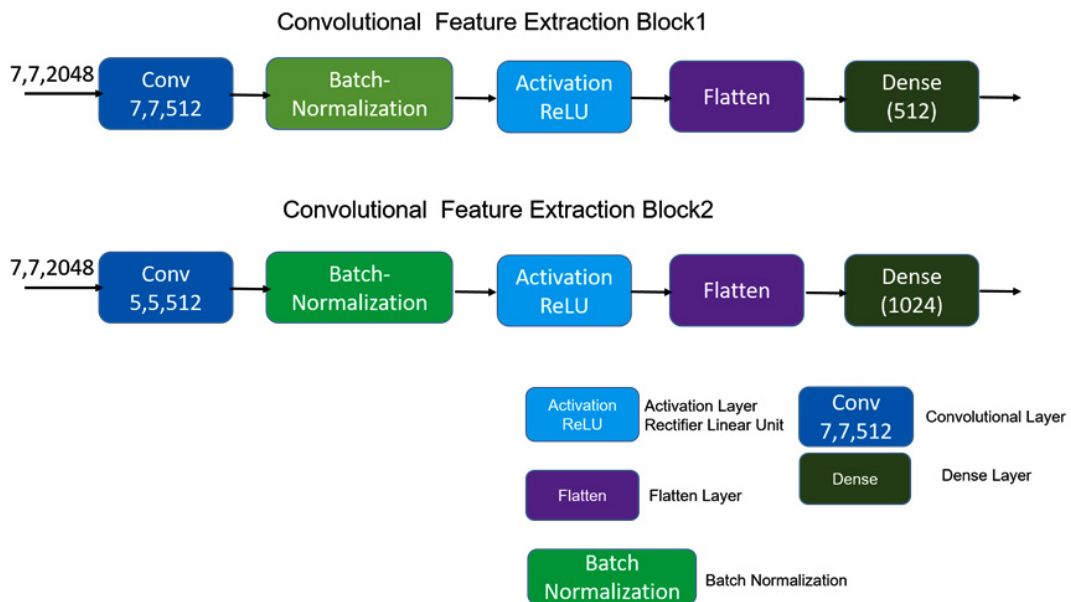


Fig. 2. Covolutional Fteature extraction blocks (CFEBs)

Table 1. Key related works

Ref.	Model / Approach	Hyper-parameters	Accuracy (if reported)	Research Gap
Ekuma, Hier, & Obafemi-Ajayi ²⁰	Explainable DL with local data-driven interpretation	Learning rate(lr) = 0.001, batch size(bs) = 32, optimizer = Adam and epochs = 50	84.6% (AD vs CN)	Focused on interpretability but lacks multi-class classification (AD vs MCI vs CN).
Xu et al. ²¹	CNN-based MRI classification	lr = 0.001, bs = 32, optimizer = Adam and epochs = 50, kernel size = 3×3	82%	Limited to binary classification, no attention mechanisms applied.
Aghaei, Ebrahimi Moghaddam, & Malek ²²	Transfer learning (Inception-V3 + SVM/Softmax)	Pre-trained ImageNet weights, fine-tuned last 2 layers, dropout = 0.5, batch size = 32, optimizer = Adam	85%	Limited feature representation; struggles with early-stage AD.
Fang et al. ²³	Gaussian discriminative component analysis (sMRI)	Dimensionality reduction applied, standard ML classifier settings	80%	Feature reduction may discard useful spatial info.
Abuhmed, El-Sappagh, & Alonso ²⁴	Hybrid CNN for longitudinal datasets	CNN layers tuned via cross-validation, lr = 0.001, batch size = 32, epochs = 100	86%	Improved robustness but lacks attention-based refinement.
Venugopalan, Tong, Hassanzadeh, & Wang ²⁶	Multimodal DL (sMRI + clinical data)	Late fusion strategy, CNN backbone, lr = 0.0005, bs = 16, epochs = 100	87%	Requires multimodal data, limiting clinical usability.
Singh, Kumar, & Pant ²⁷	CNN + Spatial Attention	ResNet backbone, spatial attention module, lr = 0.0001, bs = 16, epochs = 100	88%	Strong discriminative focus, but computationally intensive.
Ajagbe, Amuda, Oladipupo, AFE, & Okesola ²⁸	Multi-class CNN (AD/MCI/CN)	Conv layers = 3–5, kernel = 3×3, bs = 32, epochs = 100, optimizer = Adam	83%	No attention or multi-branch features; limited feature richness.
Yagis et al. ³¹	2D CNNs on MRI	Standard 2D CNN training, batch size = 32, epochs = 50	79%	Highlighted risk of data leakage; weak generalization.
Atnafu & Diciotti ³²	Interpretable DL framework	Prototype learning, learning rate = 0.001, batch size = 32, epochs = 50	84%	Improves transparency, but limited benchmark validation.
Hoang, Lee, & Kim ³³	Guided attention CNN	CNN backbone + guided attention, lr = 0.0001, batch size = 16, epochs = 50	89%	Requires high-quality MRI, not validated on multimodal data.

Then X1 and X2 are added such that

$$\text{Part1} = \text{ReLU}(W_{12} * (X_1 + X_2) + b_{12}) \in R^{X \times Y \times 1} \quad \dots(5)$$

Next in branch 3 separable convolution layer applied to again same features.

$$X3 = \text{ReLU}(\text{SepConv}(I)) \in R^{X \times Y \times Z} \quad \dots(6)$$

In branch 4 two convolutional layers are used.

$$X4 = \text{ReLU}(W_2 * (\text{ReLU}(W_{3*} I + b_{11})) + b_2) \in I^{X \times Y \times Z} \quad \dots(7)$$

After this branch 3 and branch4 are added as part2 in the attention module.

$$\text{Part2} = \text{ReLU}(W_{34} * (X_3 + X_4) + b_{34}) \in R^{X \times Y \times 1} \quad \dots(8)$$

Element wise multiplication:

$$M = I \odot \text{Part1} \odot \text{Part2} \in R^{X \times Y \times 1} \quad \dots(9)$$

After this channel wise average pooling is applied.

$$M_{avg} = \frac{1}{C} \sum_{c=1}^C M; ; c \in R^{X \times Y \times Z} \quad \dots(10)$$

Final spatial attention map.

$$A = \sigma(W_f * M_{avg} + b_f) \in R^{X \times Y \times Z} \quad \dots(11)$$

σ is the sigmoid activation function.

$$\text{Output: } Y = X \odot A \quad \dots(12)$$

$$Y = I \odot \sigma(W_f * (\frac{1}{C} \sum_{c=1}^C I \odot \text{ReLU}(W_{12} * (X_1 + X_2)) \odot \text{ReLU}(W_{34} * (X_3 + X_4)))) \quad \dots(13)$$

Multiclass Cross Entropy loss: It is used in our experimental setup to reduce the loss. For a data set with N instances, Multiclass Cross Entropy loss is calculated as

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (y_{i,j} \cdot \log(P_{i,j})) \quad \dots(14)$$

Where C= number of classes, $y_{i,j}$ = the true labels for class j for instance i and $P_{i,j}$ = predicted probability for class j for instance.

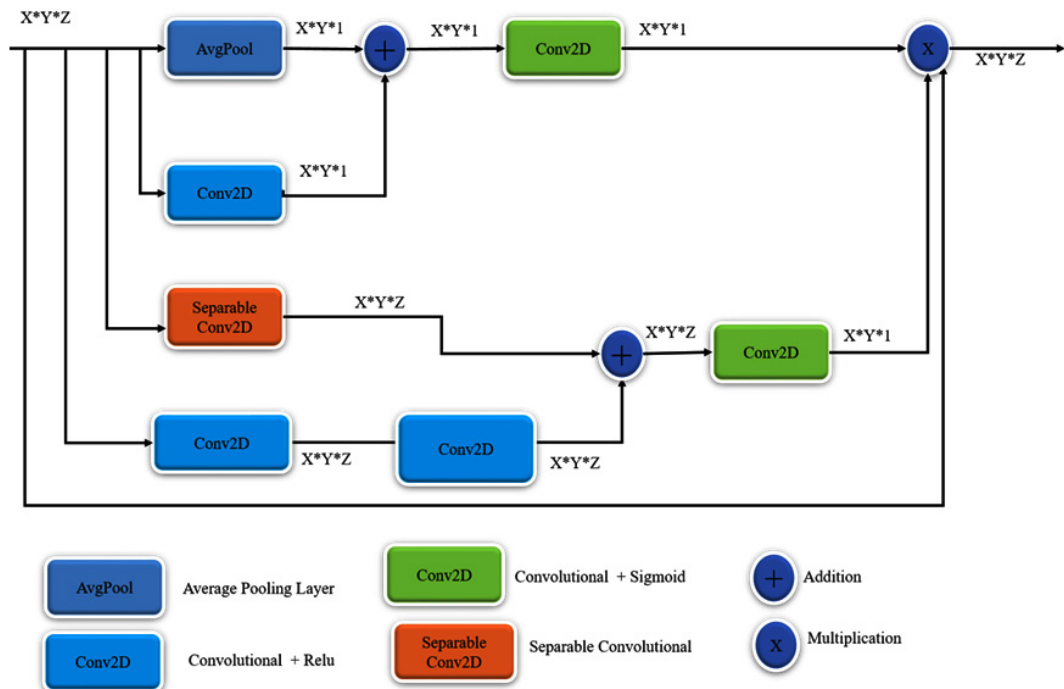


Fig. 3. Spatial-Attention Block (SAB)

RESULTS

We used ADNI⁴³ dataset in all the experiments, 5255 Sagittal MRI images (CN: 3294, MCI: 658, AD: 1303) have been taken into consideration. For training purpose, we used 70% images and for testing 30% images are used. The 1576 images are used for testing, and 3679 images are used for training as depicted in Table 2.

Result Analysis

In this study, the proposed approach is designed and evaluated using six distinct models denoted as M1 through M6. To investigate the impact of integrating spatial attention and convolutional feature extraction at different stages of the ResNet50 architecture. Each model has three components: The base backbone (CNN), a spatial attention block, and two dedicated convolutional feature extraction blocks. Models M1 to M5 utilize the specific convolutional stage with the attention mechanism and feature extraction modules (Applied in ResNet50). Model M1 applies attention to conv3_block and conv4_block, model M2 to conv2_block and conv3_block, model M3 to conv2_block and conv4_block, model M4 to conv3_block and conv5_block, model M5 to conv1_block and conv5_block, and model M6 to conv1_block and conv4_block. Based on the performance analysis of these configurations, it is observed that applying attention to deeper blocks yields more significant performance gains compared to earlier layers. Consequently, a final model is proposed, which integrates spatial attention and convolutional feature extraction at conv4_block and conv5_block, leveraging the most semantically rich features for optimal classification performance. This proposed architecture aims to leverage the most informative spatial and hierarchical features for improved classification performance. Results are analyzed over ADNI

dataset. Table 3 presents performance evaluation of various models.

The comparative analysis of various models: The model M1 attains an accuracy of 0.87 and 0.91 F1-score. The model struggles with MCI and shows recall of 0.58 where as M2 and M3 show an accuracy of 0.86 and show an improvement in MCI class. Further, M3 offers balanced performance for all classes, with F1 scores of 0.78 for AD and MCI, and 0.90 for CN. Model M4 matches M1 in accuracy (0.87) and shows the highest precision for MCI (0.92). However, its recall for MCI is 0.63. Next the model M5 shows lower accuracy (0.85) with 0.61 recall for MCI which suggests that MCI cases remain a challenge. Next, the model M6 reaches an accuracy of (0.84), recall 0.88 for AD, and low precision (0.69) which indicates normal misclassification of non-AD cases as AD. In contrast, the proposed model proves the most reliable and balanced performance for all classes. It achieves the highest accuracy of 0.88. Further, it attains a precision of 0.86 for AD class and a recall of 0.78 which leads to F1-score of 0.82 and achieved balance between detecting AD cases and minimizing false positives. Next, for CN class it shows a recall of 0.95 and an F1-score of 0.91 which suggests that the model accurately classifies CN class. Further, the proposed model handles MCI better than the other models by achieving both high precision (0.89), improved recall (0.67), and F1-score of 0.77. Hence the use of spatial attention and advanced convolutional feature extraction in proposed approach, helps in capturing delicate structural and textural differences in brain imaging data.

Further, in figures (4 to 10), the performance evaluation plots for all models are presented (M1 to proposed model). Figures include Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves which carry a thorough understanding of each model's behavior for the AD, MCI, and CN classes. Figure 4 illustrates the performance of model M1. The left panel displays the ROC curves for each class. Model M1 proves its discriminative ability with area under the ROC curve (AUC) values of 0.96 for AD, 0.93 for CN, and 0.95 for MCI. The right panel presents the PR curves, which are especially useful for evaluating performance in the presence of class imbalance. The average precision (AP)

Table 2. Dataset Distribution

	No. of Sagittal- MRI images	
	# Training images	# Testing images
AD	931	372
CN	2288	1006
MCI	460	198

values are 0.91 for AD, 0.95 for CN, and 0.80 for MCI. These results suggest that the model maintains a high precision and recall for AD and CN, while performance for MCI is relatively lower. The drop in precision at higher recall levels for MCI shows that the model struggles more with correctly identifying MCI cases, likely due to the subtler and

overlapping characteristics of this class compared to AD and CN.

In figure 5, the left panel shows the ROC curves. Model M2 achieves high AUC scores: 0.95 for AD, 0.93 for CN, and an impressive 0.97 for MCI. The right panel presents the PR curves, which are especially informative for imbalanced

Table 3. Performance evaluation of several models including proposed method.

Model	Accuracy		Precision	Recall	F1-Score
M1	0.87	AD	0.84	0.82	0.83
		CN	0.88	0.95	0.91
		MCI	0.91	0.58	0.71
M2	0.86	AD	0.79	0.76	0.78
		CN	0.89	0.91	0.90
		MCI	0.82	0.78	0.80
M3	0.86	AD	0.82	0.74	0.78
		CN	0.87	0.92	0.90
		MCI	0.84	0.72	0.78
M4	0.87	AD	0.83	0.82	0.83
		CN	0.88	0.94	0.91
		MCI	0.92	0.63	0.75
M5	0.85	AD	0.83	0.77	0.80
		CN	0.86	0.93	0.90
		MCI	0.81	0.61	0.69
M6	0.84	AD	0.69	0.88	0.77
		CN	0.90	0.86	0.88
		MCI	0.94	0.68	0.79
Proposed Method	0.88	AD	0.86	0.78	0.82
		CN	0.88	0.95	0.91
		MCI	0.89	0.67	0.77

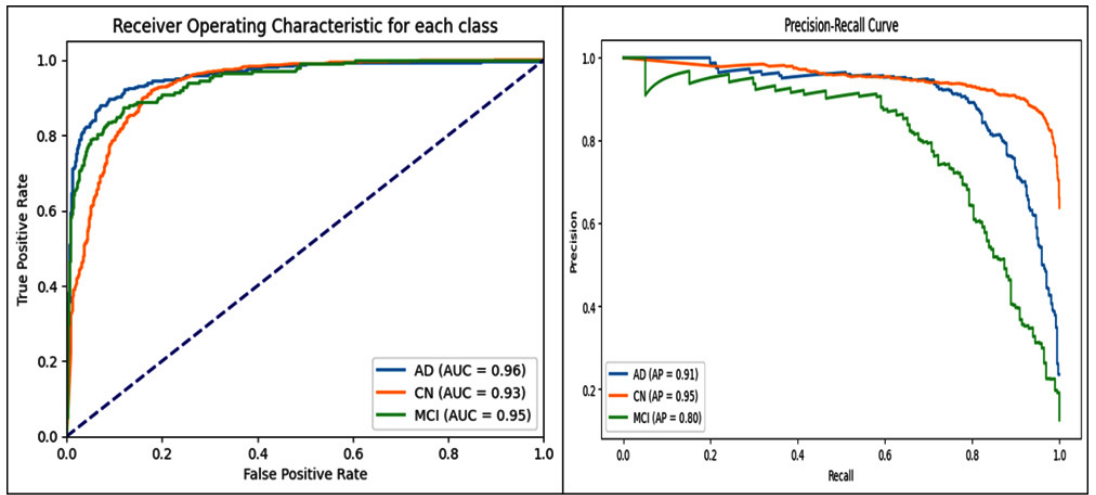


Fig. 4. M1 ROC , Precision – Recall curves

datasets. The AP values are 0.86 for AD, 0.96 for CN, and 0.85 for MCI. CN shows the strongest performance, maintaining high precision even at higher recall values. MCI, although slightly lower in AP than CN, outperforms AD in terms of recall-precision. The relatively lower AP for AD suggests that while the model can distinguish it well based on the ROC, it may produce more false positives or lower precision when trying to recall all AD instances.

In figure 6, the left panel illustrates the ROC curves. Model M3 shows strong

discriminative performance with AUC values of 0.95 for AD, 0.93 for CN, and 0.96 for MCI. The right panel displays the PR curves, providing insights into how well the model balances precision (positive predictive value) and recall, especially under class imbalance. Model M3 achieves an AP of 0.89 for AD, 0.96 for CN, and 0.85 for MCI. AD also performs well with improved precision compared to previous models. MCI, while showing strong ROC performance, still has a relatively lower AP, indicating that despite being well-separated in ROC space, it may be prone to more

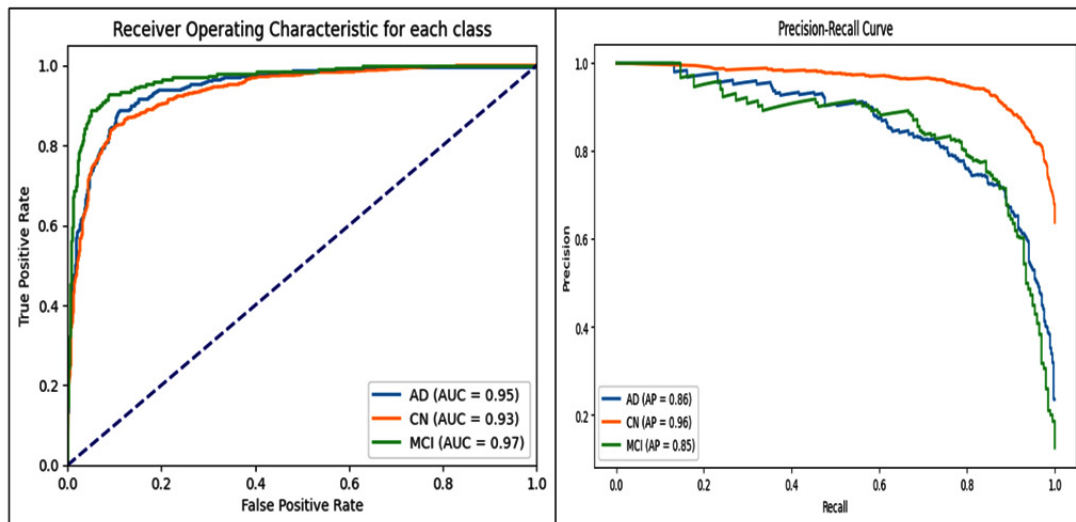


Fig. 5. M2 ROC , Precision – Recall curves

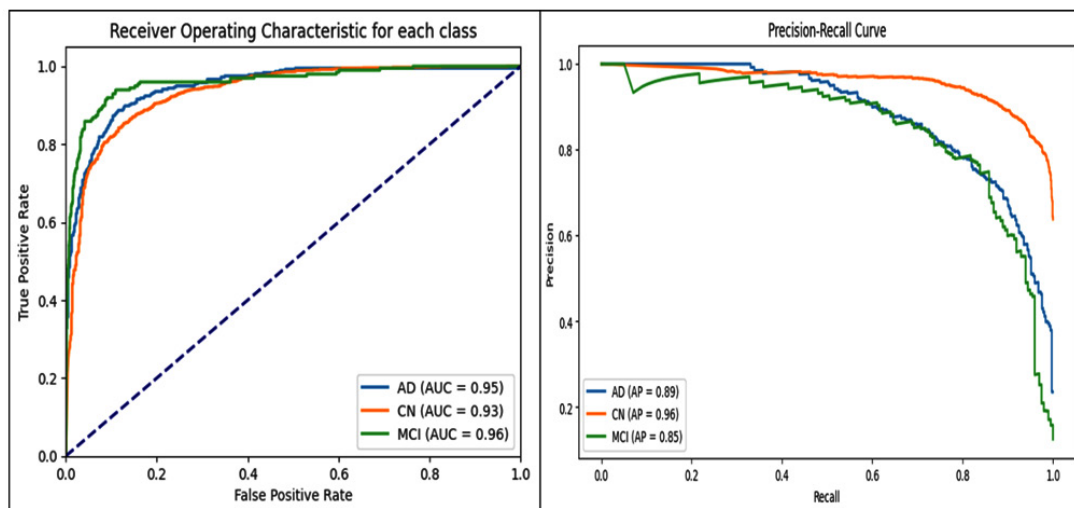


Fig. 6. M3 ROC , Precision – Recall curves

false positives or less precision when identifying MCI cases.

In figure 7 in the left panel, the ROC curves are given. Model M4 achieves AUC values of 0.95 for AD, 0.93 for CN, and 0.96 for MCI. The right panel presents the PR curves, which assess the model's balance between precision and recall, especially valuable for datasets with potential class imbalance. The AP scores are 0.89 for AD, 0.96 for CN, and 0.85 for MCI. CN again demonstrates the highest reliability with consistently high precision across all recall levels. AD shows improved precision compared to previous models, while MCI continues to exhibit slightly lower precision, suggesting that although M4 can separate MCI

well in ROC space, it may still generate more false positives or less consistent predictions when identifying MCI cases.

In figure 8, the left panel represents the ROC curves. Model M5 achieves excellent AUC scores: 0.96 for AD, 0.93 for CN, and 0.95 for MCI. These values suggest the model has a strong capacity to correctly differentiate between all three classes, with especially high sensitivity and specificity for AD and MCI. The right panel shows the PR curves, which are particularly informative in cases of class imbalance. The AP values are 0.89 for AD, 0.96 for CN, and 0.76 for MCI. CN continues to demonstrate exceptional performance with high precision across most recall values. AD

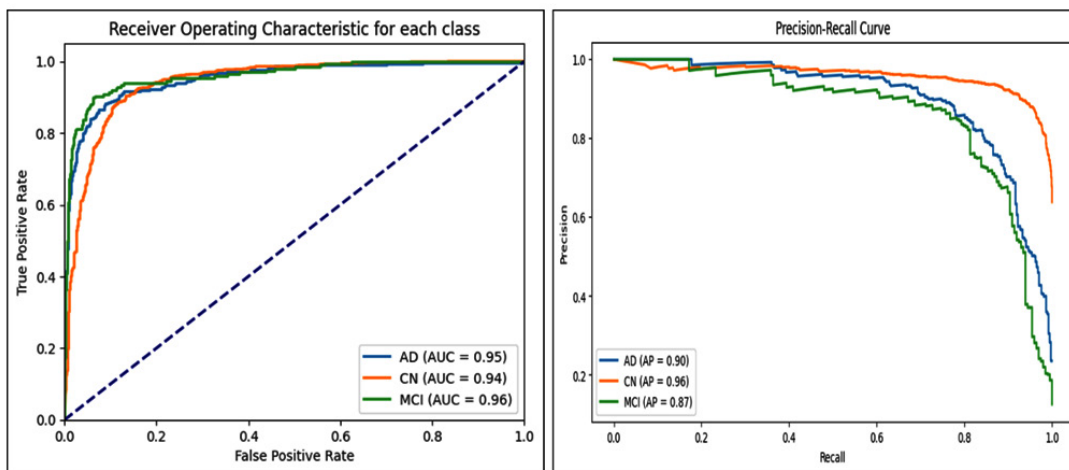


Fig. 7. M4 ROC , Precision – Recall curves

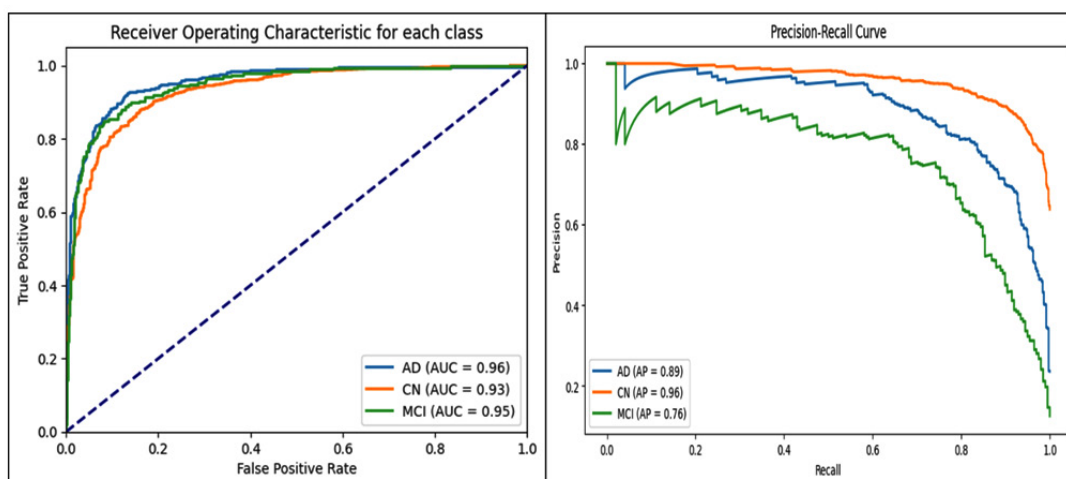


Fig. 8. M5 ROC , Precision – Recall curves

also maintains strong performance. However, MCI presents a significant drop in precision, particularly at higher recall levels, which indicates a higher rate of false positives and reduced confidence in the model's forecasts for this class.

Figure 9, the left panel shows the ROC curves. Model M6 achieves AUC values of 0.95 for AD, 0.92 for CN, and 0.96 for MCI. The right panel presents the PR curves, which are especially important for evaluating performance on imbalanced datasets. The AP values are 0.83 for AD, 0.95 for CN, and 0.88 for MCI. CN maintains the highest AP score, reflecting consistently high precision across all levels of recall. MCI also

performs very well in terms of precision-recall balance, which aligns with its strong ROC AUC. AD shows slightly lower AP compared to its AUC, indicating that although the model can distinguish AD cases well, it may experience more false positives when identifying them

Diagram 10 for the proposed model. The left panel represents the ROC curves. The proposed model achieves AUC scores of 0.95 for AD, 0.94 for CN, and 0.96 for MCI. These high AUC values indicate excellent discriminative capability across all three classes, particularly strong for MCI and AD. In the right panel, the PR curves are given. The model attains AP scores of 0.90 for AD, 0.96

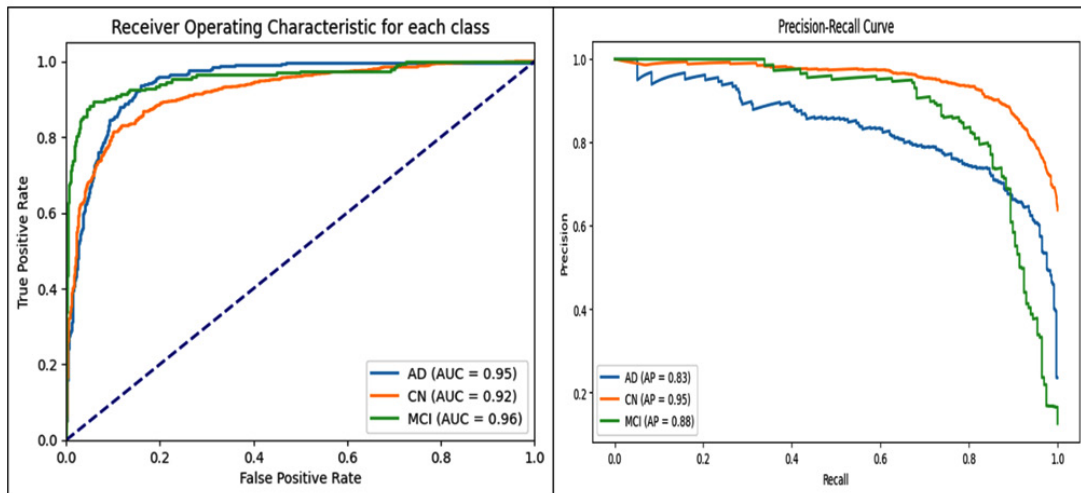


Fig. 9. M6 ROC , Precision – Recall curves

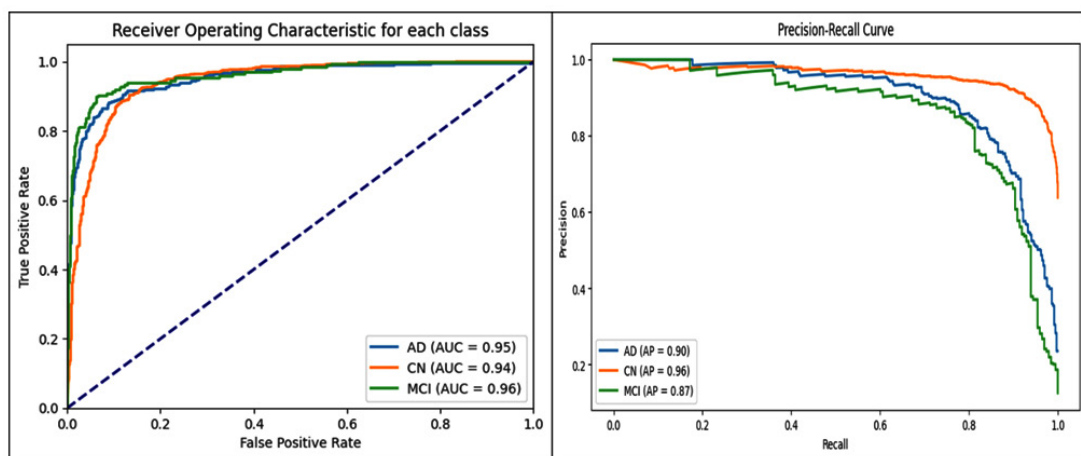


Fig. 10. Proposed Method , ROC , Precision – Recall curves

for CN, and 0.87 for MCI. CN achieves the highest AP, reflecting consistent and confident predictions with minimal false positives.

In addition to ROC analysis, confusion matrices are presented in figure 11, for all model configurations, ranging from M1 to the final proposed model, to provide a more granular view of classification performance. The confusion matrix

diagram compares the classification performance of six baseline models (M1 to M6) and the proposed model across three classes: AD, CN, and MCI. The ideal matrix has high values along with the diagonal (correct classifications) and minimal values elsewhere (misclassifications). M1 shows strong performance for AD (306 correct) and CN (952 correct), but misclassifies a notable portion of MCI

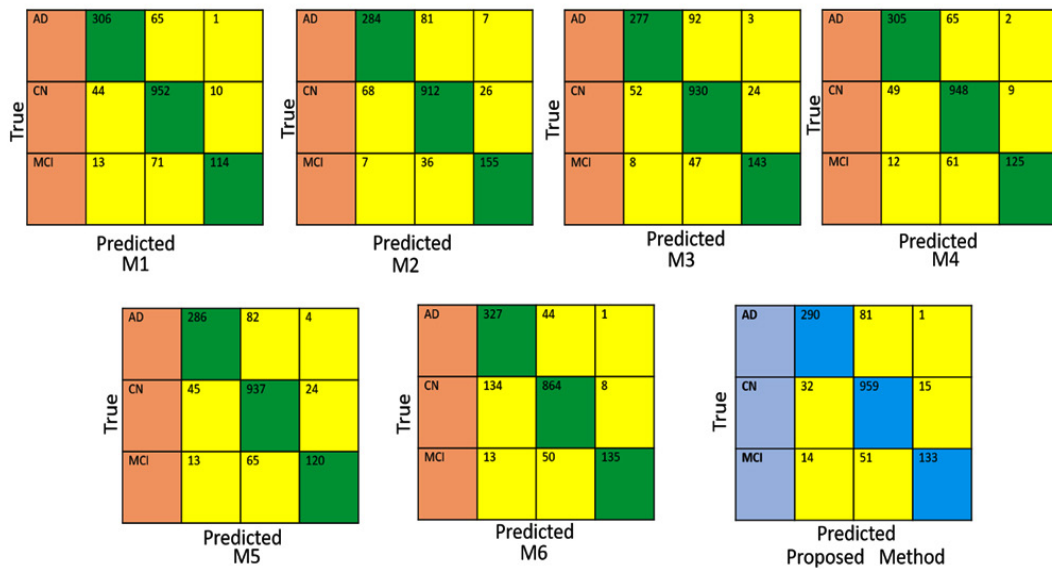


Fig. 11. Confusion Matrices

Table 4. Comparative Analysis in two classes (AD vs CN)

Model	Accuracy (AD vs CN)	Sensitivity	Specificity	AUC
Bae et al., ⁴⁴	83.33%	85.25%	87.35%	88.45%
Beheshti et al., ⁴⁵	84.07%	86.35%	88.45%	-
Gao et al., ⁴⁶	75.30%	77.30%	74.10%	78.60%
Murugan et al., ⁴⁷	84.83%	84.96%	83.67%	88.86%
Xu et al., ⁴⁸	84.38%	--	--	--
Proposed approach	89.23 ±0.87%	86.24±0.65%	90.45±0.78%	92.67±0.54%

Table 5. Comparative Analysis in three classes (AD vs MCI vs CN)

Model	Accuracy (AD vs MCI vs CN)	Sensitivity	Specificity	AUC
Angkoso et al., ⁴⁹	85.25%	83.30%	85.65%	-
Ban et al., ⁵⁰	82.58%	84.39%	90.98%	98.42%
Ekuma et al., ⁵¹	75.12%	76.48%	---	77.47%
Stamate et al., ⁵²	82.30%	96.45%	94.50%	90.25%
Proposed approach	88.46±0.91%	80.00±0.82%	88.78±0.74%	91.56±0.63%

Table 6. Ablation study results

Images	Accuracy (AD vs CN)	Sensitivity	Specificity	AUC	Accuracy (AD vs MCI vs CN)	Sensitivity	Specificity	AUC
DTI	90.25%	91.54%	92.35%	93.45%	89.88%	88.37%	87.78%	90.45%
PET	89.57%	90.56%	91.56%	92.46%	89.24%	89.23%	91.34%	89.56%

Table 7. Comparison of dataset size and hyper-parameter settings

Study (Year)	Dataset size/Split	Learning Rate	Batch Size	Dropout	Optimizer	Accuracy (%)
Beheshti <i>et al.</i> , ⁴⁵	NR	NR	NR	NR	NR	84.07
Gao <i>et al.</i> , ⁴⁶	NR			NR	NR	75.30
Murugan <i>et al.</i> , ⁴⁷	NR	0.001 (initial LR reported)	16		SGD	84.83
Angkoso <i>et al.</i> , ⁴⁹	1,500 (500 each class; 80/20 split)	NR	NR	0.5	NR	85.25
Ekuma <i>et al.</i> , ⁵¹	stratified 5-fold CV	1e-5	32	NR	NR	75.12
Ban <i>et al.</i> , ⁵⁰	Development/test split (ADNI/SNUBH): development 80% / test 20% (per dataset) — see Methods	base LR 0.001 (decayed by 0.1 three times).	64 (mini batch)	NR	SGD	83.33
Proposed approach	ADNI dataset, 5255 Sagittal MRI images in total (CN: 3294, MCI: 658, AD: 1303)With 50% training data With 70% training data	Lr=.0003 (For first 70 epochs,) Lr=.00003 (For next 30 epochs) Lr=.0003 (For first 70 epochs,)Lr=.00003 (For next 30 epochs) Lr=.0003 (For first 70 epochs,)Lr=.00003 (For next 30 epochs)	32	0.5	Adam	86.16±0.62
	With 100% training data		32	0.5	Adam	88.46±0.91

cases as CN (71).M2 performs well for CN (912 correct), but with slightly lower AD (284 correct) and MCI (36 correct) performance, indicating some confusion, particularly between MCI and CN.M3 demonstrates high accuracy for CN (930 correct) and improved MCI recognition (143 correct), though a few AD instances are misclassified.M4 maintains strong CN classification (948 correct), but misclassifies some MCI and AD instances more than M3.M5 shows balanced performance with AD (286 correct), CN (937 correct), and moderate MCI classification (120 correct).M6 improves AD (327 correct) and CN (864 correct) predictions, while MCI classification (135 correct) remains solid. The proposed model (highlighted in green) outperforms the others with superior classification across all classes: 290 correct for AD, 959 for CN, and 133 for MCI. Notably, it minimizes misclassifications significantly, indicating an accurate model. This demonstrates that incorporating spatial attention and convolutional feature extraction improves the model's capacity to differentiate refined features among the three clinical categories.

Comparative analysis

Now the proposed method is compared with recent, existing approaches of similar types. For comparative analysis in 2 classes (AD vs. CN), some recent methods i.e. Bae *et al.*,⁴⁴ Beheshti *et al.*,⁴⁵ Gao *et al.*,⁴⁶ Murugan *et al.*,⁴⁷ and Xu *et al.*,⁴⁸ are used and revealed in table 4. Table 4 presents a comparison of various models used for AD Vs CN classification. Bae *et al.*⁴⁴ reported a slightly lower accuracy of 83.33% but maintained a strong balance between sensitivity (85.25%) and specificity (87.35%), with an AUC of 88.45%. Beheshti *et al.*⁴⁵ attained 84.07% accuracy with slightly higher sensitivity (86.35%) and specificity (88.45%), though AUC was not provided. Gao *et al.*⁴⁶ presented lower overall performance, with 75.30% accuracy, 77.30% sensitivity, 74.10% specificity, and a relatively modest AUC of 78.60%. Murugan *et al.*⁴⁷ achieved 84.83% accuracy, with similar sensitivity (84.96%) and specificity (83.67%), and an AUC of 88.86%. Xu *et al.*⁴⁸ reported an accuracy of 84.38%, but did not include the other metrics. In contrast, the proposed approach outperforms all listed methods with the highest accuracy of 89.23%, along with superior sensitivity (86.24%), specificity (90.45%), and AUC (92.67%).

For three classes (AD vs. CN Vs MCI) the following methods⁴⁹⁻⁵² are used and shown in table 5. Angkoso *et al.*,⁴⁹ achieved an accuracy of 85.25%, with a sensitivity of 83.30% and specificity of 85.65%, though the AUC was not reported. Ban *et al.*,⁵⁰ performed well, with an accuracy of 82.58%, sensitivity of 84.39%, high specificity of 90.98%, and an impressive AUC of 98.42%. Ekuma *et al.*,⁵¹ reported a lower accuracy of 75.12% and sensitivity of 76.48% with an AUC of 77.47%, but no specificity value was provided. Stamate *et al.*,⁵² reported 82.30% accuracy, 96.45% sensitivity, 94.50% specificity, and an AUC of 90.25%. Among all, the proposed approach achieved the highest overall performance, with an accuracy of 88.46%, sensitivity of 80.00%, specificity of 88.78%, and AUC of 91.56%.

Ablation Study

In this the proposed approach is tested with DTI and PET images of (ADNI), DTI images are used in first row and PET are used in second row. PET data consists of 100k samples and DTI data consists of 13k images. The results are reported in table 6.

The impact of dataset size, learning rate schedules, and training configurations

The comparative analysis of prior studies and the proposed approach highlight the impact of dataset size, learning rate(lr) schedules, and training configurations is given in table 7. For instance, Ban *et al.*,⁵⁰ employed a multi-dataset setup (ADNI/SNUBH) with an 80/20 test split, using an SGD optimizer with an initial lr of 0.001 and a batch size of 64, reaching 83.33% accuracy. Murugan *et al.*⁴⁷ reported an initial lr of 0.001 with a smaller batch size of 16 under SGD, achieve 84.83%, while Beheshti *et al.*,⁴⁵ and Gao *et al.*,⁴⁶ provided limited training details, reporting 84.07% and 75.30% accuracies, respectively. Ekuma *et al.*,⁵¹ used a stratified 5-fold cross-validation strategy with a lr. of 1e-5 and batch size 32, achieving 75.12% accuracy, whereas Angkoso *et al.*⁴⁹ applied an 80/20 split across 1,500 samples with dropout 0.5, achieved 85.25% accuracy. The proposed approach employed a considerably larger dataset of 5,255 sagittal MRI images (CN: 3,294; MCI: 658; AD: 1,303) and incorporated a staged lr. schedule: 0.0003 for the first 70 epochs and 0.00003 for the final 30 epochs with a batch size of 32, dropout of 0.5, and Adam optimizer. Proposed

approach achieved the following performance scaled accuracy of 86.16% with 50% data, 87.62% with 70% data, and 88.46% with the full dataset.

DISCUSSION

The present study proposes an attention-oriented CNN framework (Att-ConvFE-ResNet) that integrates dual convolutional feature extraction (CFEB) blocks with a spatial attention block (SAB) to improve feature representation from MRI scans. Our findings demonstrate that this approach improves classification performance across both binary and multi-class and its superior performance is validated through extensive benchmarking against recent conventional and DL techniques. The results indicate that the proposed model achieves superior performance in terms of accuracy, sensitivity, specificity, and AUC compared to previous studies. For example, in the binary classification task, the proposed approach outperformed models by Ban *et al.*,⁵⁰ Beheshti *et al.*,⁴⁵ Gao *et al.*,⁴⁶ Murugan *et al.*,⁴⁷ and Xu *et al.*,⁴⁸ highlighting the robustness of the spatial attention mechanism in focusing on discriminative regions of the brain. Similarly, in the three-class scenario, the model demonstrated higher overall performance than methods reported by Angkoso *et al.*,⁴⁹ Ekuma *et al.*,⁵¹ and Stamate *et al.*,⁵² supporting the generalizability of the attention-enhanced feature extraction across heterogeneous classes. The dual convolutional feature extraction blocks capture multi-scale representations of MRI data, allowing the model to detect subtle structural changes associated with AD and MCI. Further, the spatial attention mechanism effectively recalibrates feature maps. Third, integrating CFEB and SAB with a residual backbone allows efficient gradient propagation, mitigating the vanishing gradient problem and facilitating the training of deeper networks.

Despite the promising results, several limitations should be considered. First, although our performance has been validated against results from several cohorts reported in the literature, this study relies entirely on the ADNI dataset for training and testing, which may not fully represent the diversity of the population. Future work will include external validation on independent cohorts such as OASIS or AIBL. Second, the

model currently only utilizes structural MRI data; the exclusion of multimodal biomarkers such as PET scans, cerebrospinal fluid measurements, and clinical assessments may limit the model's diagnostic sensitivity in some cases. Third, while the spatial attention module improves feature attention, its interpretability at the individual patient level remains limited; clinicians may require visual annotation tools such as Grad-CAM to gain practical insights. Fourth, the small sample size for some classes (especially MCI) may lead to class imbalance and affect the model's robustness. Finally, the computational complexity of the attention-based CNN construct may pose a challenge for deployment in clinical settings with limited hardware resources. The proposed attention-guided CNN approach demonstrates robust performance in differentiating AD, MCI, and CN classes. Therefore, combining multi-scale feature extraction with spatial attention enhances the model's discriminative power and provides a promising approach for automated MRI-based diagnosis.

CONCLUSION

In this research, a novel deep learning approach is presented that enhances ResNet50 with spatial attention mechanisms and multi-path feature extraction to improve the classification of AD, MCI, and CN classes. By integrating spatial attention blocks at key stages of the network, the model effectively emphasizes informative regions in the input, leading to more robust and discriminative feature representations. The use of dedicated convolutional feature extraction blocks followed by feature fusion further enriches the learned representations by capturing complementary spatial and contextual information. Experimental evaluation demonstrates that our proposed approach achieves higher accuracy 88.46% in three class classification and 89.23% accuracy for two class classification. The results emphasize the potential of attention-guided deep models in neuroimaging-based Alzheimer's Disease diagnosis and lay the foundation for more interpretable and accurate computer-aided diagnostic tools. Moreover, studies may concentrate on creating a hybrid model that combines PET, DTI, and fMRI data to further improve AD diagnosis accuracy. Future

work should focus on external validation across diverse populations, integration of multimodal biomarkers, enhancing interpretability, and optimizing computational efficiency to facilitate real-world clinical adoption.

ACKNOWLEDGEMENT

The author would like to express deepest gratitude to Graphic Era Deemed to be University for providing the opportunity, resources, and academic environment necessary to carry out this research. The Department of Computer Science and Engineering, Graphic Era Deemed to be University of Dehradun is highly appreciated for allowing the computer laboratory work.

The author sincerely thankful to the research supervisor, Dr. Kamlesh Purohit, and Dr. Manoj Diwakar for their invaluable guidance, continuous support, and insightful feedback throughout the course of this study. Their expertise and encouragement have been instrumental in shaping the direction and quality of this work.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article

Ethics Statements

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials.

Permission to reproduce material from other sources

Not Applicable

Author Contributions

Upendra Singh: Writing – Original Draft, Conceptualization, Methodology, Data Collection;

Vidit Kumar: Supervision, Writing – Review, Analysis; Bhaskar Pant: Resources, Supervision.

REFERENCES

1. Alzheimer's Association, alzheimer's disease facts and figures. *Alzheimers Dement.* 2018;14(3):367–429.
2. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, et al. Mild cognitive impairment. *Lancet.* 2006;367(9518):1262–1270.
3. Mirzaei G, Adeli A, Adeli H. Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Rev Neurosci.* 2016;27(8):857–870.
4. Yiðit A, İplik Z. Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease. *Turk J Electr Eng Comput Sci.* 2020;28(1):196–210.
5. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MAA, Dwivedi YK. A systematic literature review of artificial intelligence in the healthcare sector. *J Innov Knowl.* 2023;8(1):100333.
6. Li W, Milletari F, Xu D, Rieke N, Hancox J, Zhu W, et al. Privacy-preserving federated brain tumour segmentation. *Lect Notes Comput Sci (Mach Learn Med Imaging).* 2019;133–141.
7. Zhou L, Wang M, Zhou N. Distributed federated learning-based deep learning model for privacy MRI brain tumor detection. *arXiv.* 2024;arXiv:2404.10026.
8. Eskildsen SF, Coupé P, García-Lorenzo D, Fonov V, Pruessner JC, Collins DL, et al. Prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neuroimage.* 2013;65:511–521.
9. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I, et al. MRI biomarkers for early diagnosis of Alzheimer's disease. *Front Neurosci.* 2015;9:307.
10. Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, et al. Automatic classification of Alzheimer's disease from MRI. *Neuroimage.* 2011;56(2):766–781.
11. Cao P, Liu X, Yang J, Zhao D, Huang M, Zhang J, et al. Nonlinearity-aware dimensionality reduction and oversampling for AD/MCI classification. *Comput Biol Med.* 2017;91:21–37.
12. Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, et al. Interpretable deep learning framework for Alzheimer's disease classification. *Brain.* 2020;143(6):1920–1933.
13. Bucholtz M, Ding X, Wang H, Glass DH, Wang H, Prasad G, et al. Decision support system for predicting Alzheimer's disease severity. *Expert*

- Syst Appl.* 2019;130:157–171.
14. Buterez D, Janet JP, Oglic D, Liò P. End-to-end attention-based learning on graphs. *Nat Commun.* 2025;16(1):1–16.
 15. Kakuba S, Poulouse A, Han DS. Attention-based multi-learning approach for speech emotion recognition. *IEEE Access.* 2022;10:122302–122313.
 16. Luong MT, Pham H, Manning CD. Effective attention-based neural machine translation. *arXiv.* 2015;arXiv:1508.04025.
 17. Ma J, Zhang H, Yi P, Wang Z. SCSCN: separated channel–spatial convolution net for reconstruction. *IEEE Trans Ind Electron.* 2019;67(10):8649–8658.
 18. Lu B, Li HX, Chang ZK, Li L, Chen NX, Zhu ZC, et al. Alzheimer's disease classifier using deep learning. *J Big Data.* 2022;9(1):101.
 19. Lu P, Hu L, Mitelpunkt A, Bhatnagar S, Lu L, Liang H. Hierarchical attention-based multimodal fusion for AD progression prediction. *Biomed Signal Process Control.* 2024;88:105669.
 20. Ekuma G, Hier DB, Obafemi-Ajayi T. An explainable deep learning model for prediction of severity of Alzheimer's disease. *IEEE Conf Comput Intell Bioinform Comput Biol (CIBCB).* 2023; 1–8.
 21. Xu S. CNN-based MRI analysis of Alzheimer's disease. *Highlights Sci Eng Technol.* 2022;14:1–7.
 22. Aghaei A, Ebrahimi Moghaddam M, Malek H. Interpretable ensemble deep learning model for early detection of Alzheimer's disease using LIME. *Int J Imaging Syst Technol.* 2022;32(6):1889–1902.
 23. Iglesias JE. A ready-to-use machine learning tool for symmetric multi-modality registration of brain MRI. *Sci Rep.* 2023;13(1):6657.
 24. Fang C, Li C, Forouzannezhad P, Cabrerizo M, Curiel RE, Loewenstein D, et al. Gaussian discriminative component analysis for early detection of Alzheimer's disease. *J Neurosci Methods.* 2020;344:108856.
 25. Abuhmed T, El-Sappagh S, Alonso JM. Robust hybrid deep learning models for Alzheimer's progression detection. *Knowl Based Syst.* 2021;213:106688.
 26. Çalýpkın A. Diagnosis of malaria disease by integrating chi-square feature selection algorithm with convolutional neural networks and autoencoder network. *Transactions of the Institute of Measurement and Control.* 2023;45(5):975–985.
 27. 52. C. Aslým and A. Çalýpkın, "Improving Brain Tumor Detection with Deep Learning Models: A Comparative Analysis Using MRI and RO-SVM Classification", *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 14, no. 2, pp. 647–663, 2025.
 28. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep.* 2021;11(1):3254.
 29. Singh U, Kumar V, Pant B. Spatial attention enhanced deep learning model for Alzheimer's disease diagnosis. *Proc Int Conf Comput Commun Netw Technol (ICCCNT).* 2024; 1–6.
 30. Ajagbe SA, Amuda KA, Oladipupo MA, Afe OF, Okesola KI. Multi-classification of Alzheimer's disease using DCNN. *Int J Adv Comput Res.* 2021;11(53):51–60.
 31. Haq EU, Huang J, Kang L, Haq HU, Zhan T. Image-based techniques for identification and classification of brain diseases: A review. *Med Biol Eng Comput.* 2020;58(11):2603–2620.
 32. Zhang J, Liu M, An L, Gao Y, Shen D. Alzheimer's disease diagnosis using landmark-based features. *IEEE J Biomed Health Inform.* 2017;21(6):1607–1616.
 33. Yagis E, Atnafu SW, García Seco de Herrera A, Marzi C, Scheda R, Giannelli M, et al. Effect of data leakage in brain MRI classification using 2D CNNs. *Sci Rep.* 2021;11(1):22544.
 34. Atnafu SW, Diciotti S. Development of an interpretable deep learning system for Alzheimer's disease. *Adv Sci Technol Sustain Manuf Process Eng.* 2023; 27–37.
 35. Hoang GM, Lee Y, Kim JG. Deep learning with guided attention for early diagnosis of Alzheimer's disease. *Phys Scr.* 2025;100(6):065020.
 36. Gao X, Shi F, Shen D, Liu M. Multimodal transformer network for incomplete image generation and AD diagnosis. *Comput Med Imaging Graph.* 2023;110:102303.
 37. Mulyadi AW, Jung W, Oh K, Yoon JS, Lee KH, Suk H-I. Explainable Alzheimer's disease likelihood map via prototype learning. *Neuroimage.* 2023;273:120073.
 38. Gao X, Shi F, Shen D, Liu M. Task-induced pyramid and attention GAN for multimodal brain image imputation. *IEEE J Biomed Health Inform.* 2021;26(1):36–43.
 39. Hao X, Li J, Ma M, Qin J, Zhang D, Liu F, et al. Hypergraph convolutional network for longitudinal analysis in Alzheimer's disease. *Comput Biol Med.* 2024;168:107765.
 40. Mora-Rubio A, Bravo-Ortiz MA, Arredondo SQ, Torres JMS, Ruz GA, Tabares-Soto R. Classification of Alzheimer's disease stages using deep learning. *PeerJ Comput Sci.* 2023;9:e1490.
 41. Zhang J, He X, Liu Y, Cai Q, Chen H, Qing L. Multimodal cross-attention network for Alzheimer's

- disease diagnosis with multi-modality data. *Comput Biol Med.* 2023;162:107050.
42. Poloni KM, Ferrari RJ. Automated detection, selection and classification of hippocampal landmark points for Alzheimer's disease diagnosis. *Comput Methods Programs Biomed.* 2022;214:106581.
 43. Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., ... & Weiner, M. W. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 2008;27(4): 685-691.
 44. Bae JB, Lee S, Jung W, Park S, Kim W, Oh H, et al. Identification of Alzheimer's disease using a convolutional neural network based on T1-weighted MRI. *Sci Rep.* 2020;10(1):22252.
 45. Beheshti I, Maikusa N, Matsuda H, Demirel H, Anbarjafari G, Initiative J-ADNI. Histogram-based feature extraction from gray matter similarity matrix for Alzheimer's disease classification. *J Alzheimers Dis.* 2016;55(4):1571-1582.
 46. Gao X, Shi F, Shen D, Liu M. Multimodal transformer network for incomplete image generation and AD diagnosis. *Comput Med Imaging Graph.* 2023;110:102303.
 47. Murugan S, Venkatesan C, Sumithra M, Gao XZ, Elakkiya B, Akila M, et al. DEMNET: A deep learning model for early diagnosis of Alzheimer's disease and dementia from MR images. *IEEE Access.* 2021;9:90319-90329.
 48. Xu S. CNN-based MRI analysis of Alzheimer's disease. *Highlights Sci Eng Technol.* 2022;14:1-7.
 49. Angkoso CV, Agustin Tjahyaningtjas HP, Purnama I, Purnomo MH. Multiplane convolutional neural network for Alzheimer's disease classification. *Int J Intell Eng Syst.* 2022;15(1):1-8.
 50. Ban Y, Zhang X, Lao H. Diagnosis of Alzheimer's disease using structure-highlighting key slice stacking and transfer learning. *Med Phys.* 2022;49(9):5855-5869.
 51. Ekuma G, Hier DB, Obafemi-Ajayi T. An explainable deep learning model for prediction of severity of Alzheimer's disease. *IEEE Conf Comput Intell Bioinform Comput Biol (CIBCB).* 2023;pp.1-8.
 52. Stamate D, Smith R, Tsygancov R, Vorobev R, Langham J, Stahl D, et al. Applying deep learning to predicting dementia and mild cognitive impairment. *Artificial Intelligence Applications and Innovations (AIAI).* 2020;16-25.