Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Alzheimer's disease Recognition Classification Study Using MRI Images Based on Deep Learning and Dual Multilayer Attention Mechanisms

Peng Xiao[1*], Yan Chen[1], Meiqing Wu[1], Jiacui Tang[1], Wei Ma[1]

1. Chengdu University of Information Technology, Chengdu, Sichuan, China

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Apr 30, 2025
Accepted: Oct 06, 2025

Keywords:
Alzheimer's Disease
Deep Learning
Artificial Intelligence
Magnetic Resonance Imaging
Classification

ABSTRACT

Introduction: Current deep learning-based computer-aided diagnosis (CAD) techniques face challenges in hierarchical feature extraction and computational efficiency. Traditional convolutional neural networks (CNN) often focus on local or single-scale information, neglecting global correlations of brain atrophy and multiscale pathological features. Additionally, the parameter explosion problem in deep networks limits model's generalization ability on small and medium-sized datasets. While the introduction of attention mechanisms has significantly improved feature extraction and enhanced CNN recognition capabilities, existing attention mechanisms are mostly single-scale, focusing on feature maps at specific hierarchical levels and ignoring the correlations between features of different layers.

Material and Methods: To address these issues, this study proposes a lightweight model combining a shallow feature pyramid CNN with a Dual Multi-level Attention (DMA) mechanism. Experiments using the public OASIS-1 dataset, which contains 86,437 MRI images across 4 categories, employ a focal loss function to handle class imbalance.

Results: The results show that the model including DMA outperforms both the baseline CNN and the single-scale attention mechanism in terms of accuracy (ACC), sensitivity (SEN), and specificity (SPE). Specifically, compared to CNN and CNN+CBAM: ACC improved by 3.33% and 1.26%, SEN improved by 13.2% and 0.9%, and SPE improved by 1%.

Conclusion: The model demonstrates significant advantages in distinguishing small-sample classes and differentiating between very mild dementia and normal controls, highlighting its superiority in fine-grained pathological discrimination.

►*Please cite this article as:*
Xiao P, Chen Y, Wu M, Tang J, Ma W. Alzheimer's disease Recognition Classification Study Using MRI Images Based on Deep Learning and Dual Multilayer Attention Mechanisms. Iran J Med Phys 2025; 22: 267-275. 10.22038/ijmp.2025.87893.2542.

Introduction

Alzheimer's Disease (AD), the most common neurodegenerative disease, is characterized by progressive brain atrophy, β-Amyloid deposition, and neurofibrillary tangles, leading to irreversible cognitive decline. It ranks fifth in the global mortality list, causing a large number of deaths annually [1-2]. As a persistent neurodegenerative brain disorder, AD causes progressive destruction of brain cells after onset, leading to memory loss, cognitive impairment, and an accelerated decline in the ability to perform daily living activities [3-4].

AD is fatal and incurable at present. Patients not only endure lifelong physical suffering, but their families also bear a heavy burden, encompassing economic pressures and enormous mental and physical tolls [4]. Additionally, it imposes a significant burden on global healthcare systems [5]. Accurate early diagnosis is crucial for intervening in disease progression [6], and magnetic resonance imaging (MRI), with its advantages of non-invasiveness and high resolution, has become a core tool for capturing subtle structural changes in the AD brain, such as hippocampal atrophy and cortical thinning [7-8].

In recent years, deep learning-based computer-aided diagnosis (CAD) technologies have shown great potential in MRI analysis [9-13]. However, existing methods still face two key challenges: one-sidedness in feature extraction and limitations in computational efficiency. Convolutional neural networks extract features through multiple convolutional layers, but they often focus on local regions (such as the hippocampus and entorhinal cortex) or single-scale information, neglecting the global correlations of brain atrophy and multi-scale pathological features [14]. Additionally, AD MRI data often exhibit class imbalance (significant differences between mild cognitive impairment (MCI) and normal control (CN) samples), leading to insufficient learning ability for early lesion features in models. On the other hand, the parameter explosion problem in deep networks exacerbates computational resource consumption and

*Corresponding Author: Tel: + 17828439189; Email: 1844494673@qq.com

limits the generalization ability of models on small and medium-sized datasets.

The introduction of attention mechanisms has significantly enhanced the feature representation capabilities of medical image analysis models. The classic Convolutional Block Attention Module [15] (CBAM) enhances the model's focus on key regions by jointly optimizing channel [16] and spatial attention [17]. However, existing methods typically rely on single-scale feature design, neglecting the correlations between features at different hierarchical levels. Existing methods fail to fully leverage the advantages of multi-scale feature fusion and cannot utilize features at different scales, leading to limited generalization ability in complex scenarios. The size of the dataset also imposes constraints on model performance—for smaller datasets, more complex models may suffer from overfitting, resulting in decreased generalization ability in feature extraction. Additionally, traditional attention mechanisms often generate redundancy or cause information loss when processing multi-scale features, making it difficult to adapt to the heterogeneous manifestations of AD lesions at different resolutions.

Based on the above problems, this paper proposes a new model of shallow feature pyramid CNN combined with a DMA [18] based on the public OASIS-1dataset [19]:

(1) The model adopts a shallow feature pyramid CNN network to reduce the computational redundancy, reduce the computational complexity of the model, extract the image features in different scale directions, and realize the full use of the image feature information.

(2) The DMA module used reduces one channel of the module according to the network structure, reduces the number of weights that the module needs to calculate, and makes the whole network lightweight.

(3) The DMA module realizes the prominence of important features extracted from non-scale features of the convolution block through weighting, and splices and fuses the prominent important features of different scales in the subsequent process, which enhances the ability of the model to extract and identify features.

Experimental results show that the new model significantly improves the discrimination ability of small sample categories and early lesions.

In this section, we will briefly introduce the previous research on CAD methods for AD, the research progress of CNN architecture and attention mechanism in AD classification, analyze the technical bottlenecks of existing methods, and propose the innovative direction of this research.

The core challenge of lightweight CNN in the medical imaging field lies in balancing computational efficiency and feature representation capability, especially when AD pathological features exhibit multi-scale and high-dimensional characteristics. MobileNetV2 [10] reduces the number of parameters to 5.3 million through depth wise separable convolutions, but it fails to achieve high classification accuracy on the ADNI dataset [20]. Its bottleneck stems from the fact that depth wise separable convolutions weaken inter-channel information interaction, leading to insufficient sensitivity to subtle structural changes such as hippocampal atrophy. The study in [11] optimizes computational complexity through channel shuffling and grouped convolutions, but the receptive field of its model in AD classification only covers a 16×16 pixel region, making it difficult to capture the global patterns of whole-brain cortical thickness changes. Tan et al. [12] used the B0 model with compound scaling design to achieve an AUC value of 0.892 in AD classification, but the lack of attention mechanism guidance results in insufficient suppression of non-salient features such as ventricular enlargement.

The introduction of attentional mechanisms improves the accuracy of deep learning in recognizing Alzheimer's disease. Woo et al. [15] embedded CBAM into ResNet-50, constructing channel attention maps via global average pooling and multi-layer perceptrons, and combined spatial attention kernels to improve feature focusing ability. This approach increased the AUC for AD classification from 0.901 to 0.934. However, in lightweight networks, CBAM introduces an additional 1.2 million parameters, leading to a 28% increase in floating-point operations and higher inference latency, which struggles to meet real-time diagnostic requirements. Limitation of single-scale attention [15-17]: Existing studies only apply attention mechanisms to deep feature maps, neglecting the collaborative optimization between anatomical details in shallow features (e.g., entorhinal cortex thickness) and global patterns in deep features (e.g. cortical atrophy distribution). The DMA mechanism [18] consists of multi-layer channel attention (MCA) and multi-layer spatial attention (MSA) modules. By calculating attention in channel and spatial dimensions for feature maps of different scales, it achieves feature screening and weighting, thereby improving model performance.

Materials and Methods

Dataset

The dataset used in this study was sourced from the publicly available OASIS-1 dataset [19], which comprises a cross-sectional sample of 416 participants aged 18 to 96 years. The dataset comprises 86437 brain MRI images, which are classified into four categories according to the progression of Alzheimer's Disease. Based on the provided metadata and clinical dementia rating values, the patient dataset was categorized into four groups: Mild Dementia（Mild）,Moderate Dementia（Moderate）, Non-Demented（Non）and Very Mild Dementia（Very Mild）as shown in Figure

1. This allows for the detection and study of different stages of Alzheimer's Disease progression, providing a valuable resource for analyzing and detecting early signs of the disease. Table 1 describes the distribution of images for each category.

Table 1. OASIS-1 dataset distribution

No.	Class	Images OASIS-1
0	Mild Dementia	5002
1	Moderate Dementia	488
2	Non Demented	67222
3	Very Mild Dementia	13725

Model Structure

The model proposed in this paper takes the shallow feature pyramid CNN network combined with the DMA module (Figure 2) as its core components. The input channels of the four convolution blocks (Figure 2) are set as 16, 32, 64, and 128 in a progressively increasing order. This design of progressively increasing channel numbers across multiple layers allows the network to deeply analyze MRI images at multiple levels from local to global, thereby capturing both subtle features in fine details and macroscopic characteristics from a global perspective. After each convolution block completes feature extraction, the output feature maps are sequentially stored to construct a multi-scale feature stack with channel identifiers (16, 32, 64, 128). This feature stack is then input into the DMA attention module, where key information is effectively highlighted through precise weighting of important features. Concatenation and fusion operations efficiently integrate multi-scale features, making the processed feature information richer and more comprehensive. Finally, the finely processed feature information is mapped to a fully connected layer, which realizes the four-class classification of MRI images through comprehensive processing.

Each convolutional block consists of two 3×3 convolutional layers, batch normalization layers, a max pooling layer, two RELU activation functions, and a Dropout layer. The first convolutional layer extracts initial features, followed by the first batch normalization and first RELU activation function, which normalize the initial features (such as edges and corners) and apply sufficient non-linear activation to form basic features. The second convolutional layer extracts combined features, and the subsequent second batch normalization and second RELU activation function normalize the textures, shapes, and other attributes of these combined features while applying non-linear activation. This process stabilizes the feature distribution after the two convolutional layers, preventing drastic changes in the input distribution of the second layer caused by the sparsity of the first RELU activation.

Non-linear activation using RELU is defined as:

$$A^{(n)}{}_{i,j,k} = ReLU(F^{(n)}{}_{i,j,k}) = max(0, F^{(n)}{}_{i,j,k}) \quad (1)$$

For the output feature map of the convolutional layer $F^{(n)}{}_{i,j,k}$, where n denotes the n-th convolutional block, (i, j) represents the spatial location, and k is the channel index.

Using max pooling to reduce the spatial dimensions of the feature map, retain local maxima (salient features), and enhance translation invariance, it is defined as:

$$P^{(n)}{}_{i',j',k} = max\{A^{(n)}{}_{i,j,k} | (i, j) \in Pooling\ Window\} \quad (2)$$

The pooling window size is 2×2. For the input feature map $A^{(n)}{}_{i,j,k}$, the output after pooling is $P^{(n)}{}_{i',j',k}$, where (i, j) represents the spatial location after pooling.

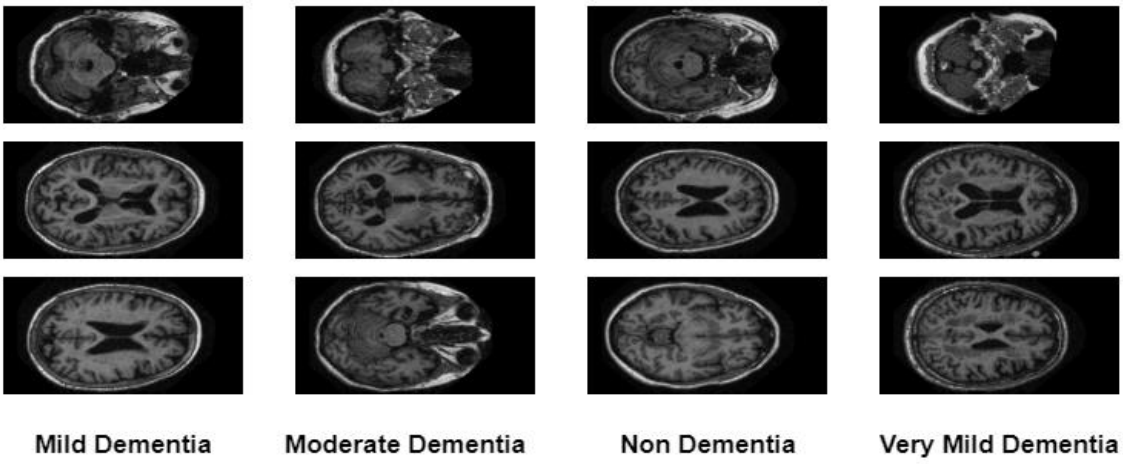

Figure 1. Example of image Samples 4 classes from the OASIS-1 dataset.

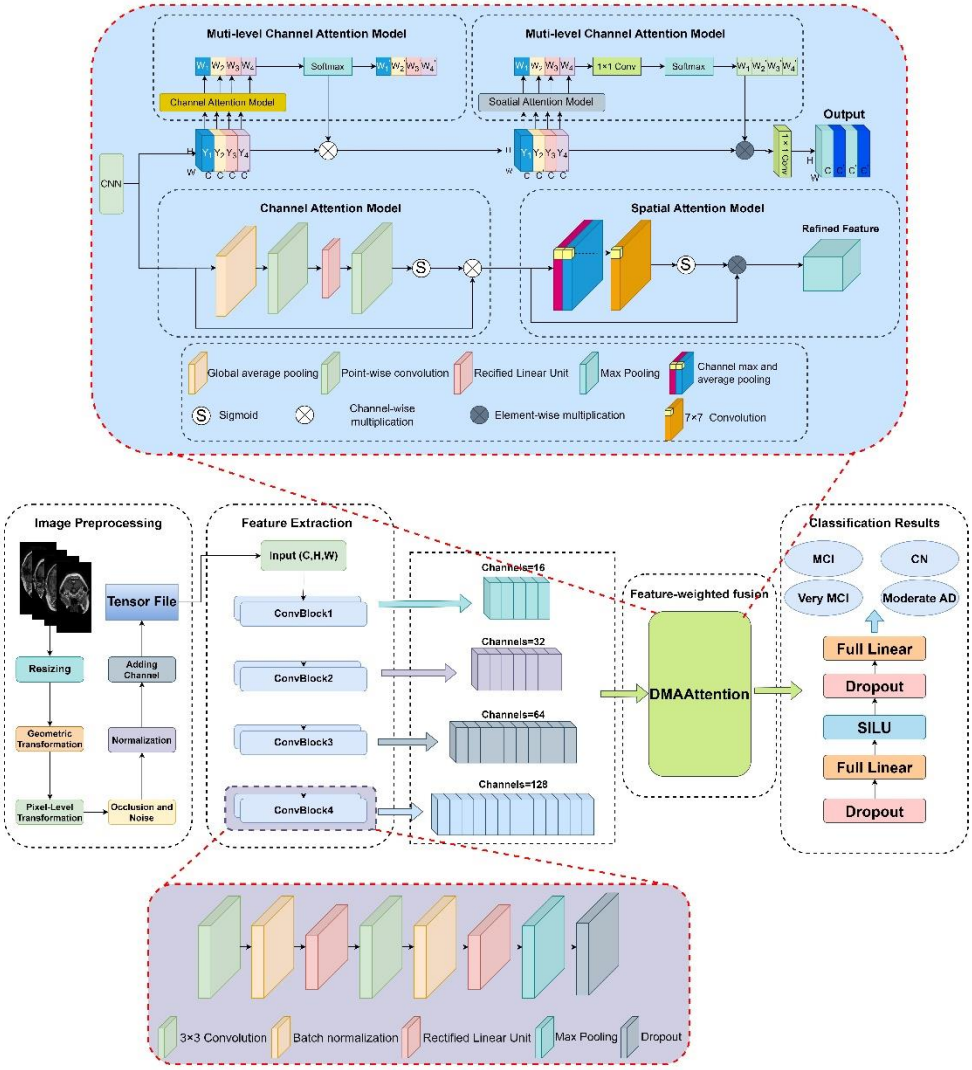

Figure 2. The shallow feature pyramid CNN +Dual Multi-level Attention [16] modules.

The first 3×3 convolutional layer extracts initial features, capturing local information such as simple edges and lines. Subsequently, the first batch normalization layer normalizes these initial features, adjusting their mean and variance to stabilize the feature distribution and accelerate model training convergence. Immediately after, the first RELU activation function applies a nonlinear transformation to the features, highlighting positive values and suppressing negative values to fully activate information such as edges and corners, forming basic features. The second 3×3 convolutional layer extracts combined features on this basis, capturing more advanced textures, shapes, etc. Subsequently, the second batch normalization layer normalizes the features again, stabilizing the feature distribution after the two convolutional layers and alleviating the problem of drastic changes in the input distribution of the second layer caused by the sparsity of the first RELU (where some neurons output zero), thus avoiding instability in model training. Then, the second RELU activation function further applies nonlinear activation to the textures, shapes, and other attributes of the combined features, enhancing the feature representation ability and enabling the model to better capture complex patterns. Finally, the max pooling layer reduces the spatial dimensions of the feature map through down sampling, lowering computational complexity while retaining important features; the Dropout layer randomly discards some neurons during training to prevent overfitting and enhance the model's generalization ability.

Dual Multi-Level Attention (DMA)

The DMA [16] mechanism enhances feature representation capabilities through cross-scale interactions in channel and spatial dimensions. By employing feature attention methods that integrate the channel and spatial domains, it deeply delves into the mechanism of attention from local to global scales, improving the model's diagnostic performance. DMA is constructed by cascading the Multi-Level Channel Attention and Multi-Level Spatial Attention modules

(Figure 2), sequentially optimizing attention in the channel and spatial domains for multi-scale features. In the MCA module, for the input feature list of different scales (feature maps with 16, 32, 64, 128 channels), global adaptive average pooling is first applied independently to each scale of features to compress the spatial dimensions into 1×1 while preserving channel dimension information. Subsequently, a bottleneck structure composed of two 1×1 convolutions (with a reduction factor) is used for channel importance modeling: the first convolution reduces the number of channels to max(in_channels /reduction, 1) and activates it with ReLU, and the second convolution restores the original number of channels to generate channel-wise attention weights. After concatenating the attention weights of each scale along the channel dimension, global SoftMax is used for cross-scale normalization to make the channel weights of different scales globally comparable. Finally, the weights are split according to the original number of channels, and element-wise multiplied back to the feature maps of the corresponding scales to complete the cross-scale fusion of channel attention. The MSA module focuses on cross-scale interactions of spatial attention. For each scale of features, average pooling and max pooling features are first calculated separately, concatenated along the channel dimension, and then input into a 7×7 convolution to generate a single-channel spatial attention map. To achieve cross-scale interaction, the attention maps of each scale are cropped to the spatial size (H, W) of the first scale and concatenated, then activated by a Sigmoid function to generate a global spatial attention map. Subsequently, the attention map is split into single channels, cropped back to the original scale's spatial size, and element-wise multiplied back to the corresponding feature maps, realizing multi-scale alignment and weighting of spatial attention. The DMA mechanism establishes cross-domain dependency relationships among features of different scales through channel-first then spatial attention hierarchical processing, effectively enhancing key feature representations and providing more discriminative feature inputs for subsequent classification tasks.

Global Adaptive Average Pooling is defined as:

$$y_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{n,c,i,j} \qquad \forall_n \in [1, N], c \in [1, C] \tag{3}$$

For an input feature map $X \in R^{N \times C \times H \times W}$ (where N is the batch size, C is the number of channels, and H×W are the spatial dimensions), the output result is $Y \in R^{N \times C \times 1 \times 1}$, which represents the average value of pixel values at all spatial positions for each channel.

Classification Module

As shown in Figure 2, in the classification module, the first fully connected layer performs a linear transformation on the output of weighted feature fusion to preprocess the features. The second fully connected layer, after being processed by the SILU activation function and Dropout, maps the features to the classification dimension and outputs class scores, thereby providing the basis for the final classification decision.

The SILU (Sigmoid Linear Unit) activation function is a self-gated activation function. Compared to ReLU, it enables the model to learn more complex patterns and has desirable properties such as smoothness and non-monotonicity, which help improve model performance. It is defined as:

$$SILU(x) = x \cdot \frac{1}{1 + e^{-x}} \tag{4}$$

x is the output of the first fully connected layer. By using x as the input to SILU, this design helps better enhance the model's generalization capability.

Loss Function

To enhance the discriminative ability of multi-level dementia classification, the Focal Loss function is employed to optimize the model's training process. Addressing the class imbalance issue in the dataset, this loss function introduces a focus parameter γ and class balance factor α based on cross-entropy, dynamically adjusting the training weights of different samples. The final definition of Focal Loss is:

$$F_{loss} = \alpha \cdot (1 - p_t)^{\gamma} \cdot -log(\frac{e^{xy}}{\sum_j e^x j}) \tag{5}$$

p_t is the prediction confidence, which reflects the difficulty level of samples, and is defined as:

$$p_t = e^{-log(\frac{e^{xy}}{\sum_j e^x j})} \tag{6}$$

The Focal Loss function enables the model to allocate more training attention to confusable early-stage lesions (such as Mild Dementia and Moderate Dementia) and small-sample classes in the dementia grading task, significantly improving the clinical accuracy of classification.

Results

Experimental Environment

Table 2 shows the experimental equipment experimental environment and model parameters in this paper, and the Adam optimizer is used for network training. The experimental dataset consisted of 86,437 images that were divided into subsets at a scale of 0.85:0.2:0.15.

Table 2. Experimental Setting

Parameter	Setting
GPU	NVIDIA GeForce RTX 4060 Ti 8 GB
PyTorch	2.5.0
Python	3.11
Learning Rate	1×10^{-4}
Epochs	40
Batch Size	32

Evaluation Metrics

This paper uses common evaluation metrics to comprehensively assess the effectiveness of the proposed model. These metrics include Specificity (SPE), Sensitivity (SEN), Accuracy (ACC), Recall, F1-Score, and the Area Under the ROC Curve (AUC).

Definitions for some of the metrics in the experiment are provided below: $SPE = \frac{TN}{TN+FP}$, $REcall = SEN = \frac{TP}{TP+FN}$, and $ACC = \frac{TP+TN}{TP+TN+FN+FP}$. In the formula, TP represents true positives, FN represents false negatives, TN represents true negatives, and FP represents false positives. AUC (Area Under the ROC Curve) is calculated based on the False Positive Rate (FPR =1 − SPE) and True Positive Rate (TPR = SEN).

Experiment

In order to verify that multi-scale DMA attention can better improve the recognition ability of the model than single-scale attention, we conducted a comparative experiment based on the CNN model used in the experiment and the single-scale attention CBAM. The experimental results are presented in Table 3.

As shown in Table 3, shallow CNN+DMA is superior to shallow CNN and CBAM single-scale attention. The model improved the recognition rate ACC for the entire dataset by 3.3% and 1.3%, respectively, compared to the two baselines. This improvement reflects the new model's ability to effectively identify and classify data features, thereby improving the overall classification accuracy. Recognition rates were also significantly improved in four categories compared to CNN and CNN+CBAM: 5% and 2% for the mild category, 18% and 2% for the moderate category, 4% for the non-category, and 5% and 6% for the very mild category. Other indicators also improved by 2% to 12% in each of the four types of identification categories.

Table 3. The evaluation metrics of the three models under different categories of data.

Model	Category	ACC	Precision	Recall	F1-Score	Support
CNN	Mild	0.915	0.94	0.65	0.78	1238
	Moderate		0.82	0.71	0.82	121
	Non		0.91	0.98	0.95	16638
	Very Mild		0.89	0.66	0.76	3397
CNN+CBAM	Mild	0.935	0.97	0.76	0.84	1238
	Moderate		0.98	0.77	0.79	121
	Non		0.95	0.98	0.96	16638
	Very Mild		0.88	0.79	0.83	3397
CNN+DMA (Ours)	Mild	0.948	0.99	0.77	0.86	1238
	Moderate		1.00	0.76	0.86	121
	Non		0.95	0.99	0.97	16638
	Very Mild		0.94	0.80	0.87	3397

Special note: Since the data of the Mild class and the Moderate class is much smaller than that of the other two classes, the recognition accuracy of these two categories is very high on the model.
The ACC parameter in the table is the accuracy of the model over the entire dataset.

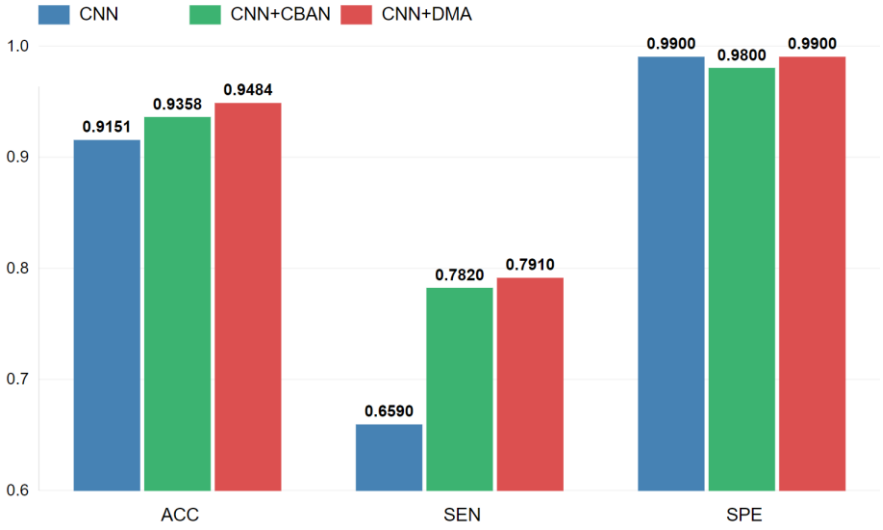

Figure 3. Evaluation results of ACC, SPE, and SEN for the three models on the entire dataset.

As shown in Figure 3, the model that introduced the attention mechanism stood out among the ACC, SEN, and SPE evaluation results of the three models (CNN, CNN+CBAM, CNN+DMA) across the entire dataset. Specifically, in the ACC metrics, the CNN is 0.9151, the CNN+CBAM is 0.9358, and the CNN+DMA reaches 0.9484, representing a 3.33% improvement compared to the CNN and a 1.26% increase over the CNN+CBAM. In terms of SEN metrics, CNN+DMA improved by 13.2% over CNN, while CNN+CBAM improved by 0.9%. In terms of SPE metrics, CNN+DMA increased by 1% compared to CNN+CBAM. These data clearly show that the accuracy, sensitivity, and specificity of the model in multi-classification tasks are significantly improved after the introduction of attention mechanisms in the underlying shallow feature pyramid CNN, especially the CNN+DMA model, which outperforms shallow CNN and CBAM on all metrics.

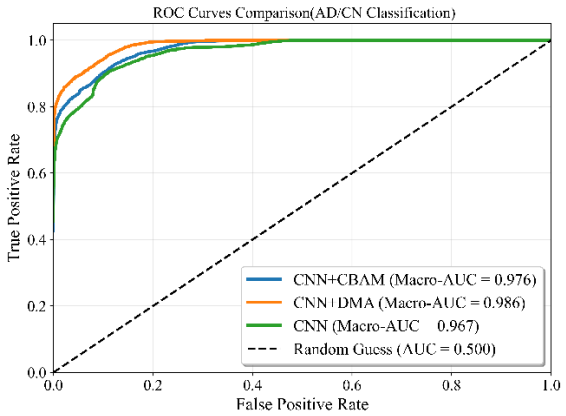

Figure 4. ROC classification curves of the three models for the entire dataset.

Figure 4 shows the ROC classification curve and area under the curve (AUC) for the entire dataset for the three models (CNN (AUC = 0.959), CNN+CBAM (AUC = 0.974), and CNN+DMA (AUC = 0.986)), as well as the dashed line representing random guessing. Among them, the AUC of the CNN+DMA model was 0.986, which was significantly greater than that of the other two models (CNN and CNN+CBAM). Experiments have confirmed that the multi-scale attention mechanism can effectively enhance the feature representation and enhance the discrimination ability of complex lesions, especially in scenes of sample imbalance and small sample size. Therefore, this mechanism is an effective solution to optimize medical imaging diagnostic models, and provides a powerful strategy to solve the inherent challenges in medical image analysis.

The confusion matrix in Figure 5 shows the prediction results for the four classifications (Mild Dementia, Moderate Dementia, No Dementia, and Very Mild Dementia). 21,393 images from 15% of the dataset were used as a test set to validate the predictive categorization of the best model. The test set consisted of a total of 21,393 images, of which 1,167 images were from the Mild class, 125 images were from the Moderate class, and Non class had 16716 images, and 3,385 images were from the Very Mild class. As can be seen from the figure, since the dataset samples in the moderate dementia category are very few compared to the other categories, it achieves a completely correct prediction, and the other three categories are also predicted correctly at more than 95%, with only a small number of samples with misclassification, it can be obtained that the model performs well in predicting the diagnosis of dementia.

Figure 5. Four types of prediction outcome confusion matrices.

Discussion

This study successfully addresses two critical challenges in deep learning-based Alzheimer's disease classification: one-sided feature extraction and computational efficiency limitations. Our proposed model, combining a shallow feature pyramid CNN with a dual multi-level attention (DMA) mechanism, achieved an overall accuracy of 94.84% on the OASIS-1 dataset, representing a 3.33% improvement over the baseline CNN and a 1.26% improvement over CNN+CBAM. More importantly, the model demonstrated substantial gains in sensitivity (13.2% over CNN, 0.9% over CNN+CBAM) and maintained high specificity (1% improvement over CNN+CBAM), with an AUC of 0.986.

The experimental results validate our hypothesis that multi-scale attention mechanisms can more effectively capture the heterogeneous manifestations of AD pathology across different spatial resolutions. Traditional CNNs and even single-scale attention mechanisms like CBAM focus predominantly on features at specific hierarchical levels, missing the correlations between shallow anatomical details (such as entorhinal cortex thickness) and deep global patterns (such as cortical atrophy distribution). The DMA mechanism's ability to establish cross-scale dependencies through sequential channel and spatial attention processing enables the model to integrate diagnostic information across multiple resolution levels,

resulting in more robust and discriminative feature representations.

The shallow feature pyramid architecture with progressively increasing channels (16, 32, 64, 128) provides a hierarchical representation of brain MRI images, capturing both fine-grained local features and coarse-grained global patterns. This design philosophy differs fundamentally from previous lightweight architectures such as MobileNetV2, ShuffleNetV2, and EfficientNet, which prioritize parameter reduction through techniques like depthwise separable convolutions or compound scaling but often sacrifice feature richness in the process.

The DMA mechanism introduces two key innovations that distinguish it from existing attention approaches. First, the multi-level channel attention (MCA) module employs cross-scale normalization using global SoftMax, enabling the model to compare and weight channel importance across different feature scales globally rather than independently. This cross-scale interaction is crucial for AD diagnosis because pathological changes manifest at multiple anatomical scales—from microscopic hippocampal atrophy to macroscopic ventricular enlargement—and their relative importance varies across disease stages.

Second, the multi-level spatial attention (MSA) module performs spatial attention alignment by cropping attention maps to a common spatial size before concatenation and sigmoid activation. This design ensures that spatial attention weights from different scales are directly comparable and can be effectively fused, avoiding the information loss or redundancy that occurs when processing multi-scale features with traditional single-scale attention mechanisms. The 7×7 convolutional kernel in MSA provides an appropriate receptive field size to capture local spatial context while maintaining computational efficiency.

The OASIS-1 dataset exhibits severe class imbalance, with the Non-Demented class comprising 77.8% of all images, Very Mild Dementia 15.9%, Mild Dementia 5.8%, and Moderate Dementia only 0.6%. This distribution reflects real-world clinical scenarios where advanced dementia cases are relatively rare in cross-sectional studies, but it poses significant challenges for model training.

The focal loss function addresses this challenge through two mechanisms. The focus parameter down-weights the contribution of easily classified samples, allowing the model to concentrate learning capacity on hard-to-classify examples at decision boundaries. The class balance factor provides explicit weighting to compensate for class frequency imbalances. Together, these mechanisms prevent the model from becoming biased toward the majority class while ensuring adequate learning for minority classes.The effectiveness of focal loss is evident in the classification results for moderate dementia. Despite this category comprising only 0.6% of the training data, the model achieved high recall and precision, indicating it successfully learned the discriminative features of this rare category without generating false positives. This balance is crucial for clinical deployment—requiring both the sensitivity to detect rare severe cases and the specificity to avoid false positives.

Despite these promising results, several limitations warrant consideration. First, the cross-sectional nature of the current study limits our ability to track disease progression over time, which is clinically crucial for monitoring treatment responses and predicting outcomes in individual patients. Longitudinal studies would provide valuable insights into how the model performs across different disease trajectories and progression rates. Second, the model relies solely on structural MRI data, potentially missing complementary diagnostic information from functional imaging (fMRI, PET) or biochemical markers (cerebrospinal fluid biomarkers, blood-based markers). Third, while the DMA mechanism demonstrates superior performance compared to single-scale attention, the fixed multi-scale architecture may not optimally adapt to the varying complexity of different brain regions or disease stages, suggesting room for more adaptive approaches.

Conclusion

To address the challenges of one-sided feature extraction and computational efficiency in AD MRI classification, this study proposes a lightweight model based on a shallow feature pyramid CNN and a dual multi-level attention mechanism (DMA). Experimental results show that the DMA mechanism establishes a hierarchical feature selection mechanism through cross-scale interactions across channel and spatial dimensions, effectively enhancing the model's discriminative ability for early-stage lesions (very Mild dementia and normal controls) and small-sample classes (moderate dementia). Specifically, compared to CNN and CNN+CBAM: ACC improved by 3.33% and 1.26%, SEN improved by 13.2% and 0.9%, and SPE improved by 1%. This provides an efficient solution for the early and accurate diagnosis of AD.

Compared to the CNN method mentioned in [21] the classification accuracy of 97.45% for moderate patients our method has improved by 2%. Compared to the classification accuracies of 97.7%, 92.4%, and 95.8% for mild cognitive impairment (Mild), normal (Non), and moderate Alzheimer's disease (Moderate) in [22], our method correspondingly improves the classification accuracies by 1.3%, 2.6%, and 4.2%.

The current study is based on single-modal MRI data from OASIS-1. In the future, we plan to incorporate larger-scale datasets such as ADNI and integrate multimodal information including PET imaging and cerebrospinal fluid biomarkers to further enhance the model's comprehensive representation capability for AD pathological mechanisms. Currently, the DMA mechanism uses fixed multi-scale feature input. Future research could explore dynamic attention weight allocation.

Acknowledgment

OASIS researchers provided experimental data but were not involved in the analysis or writing of this study. For a full list of researchers, please visit: https://direct.mit.edu/jocn/article-abstract/19/9/1498/4427/Open-Access-Series-of-Imaging-Studies-OASIS-Cross?redirectedFrom=fulltext.

References

1. Przedborski S, Vila M, Jackson-Lewis V. Neurodegeneration: what is it and where are we? J Clin Invest. 2003;111(1):3-10.
2. Tufail AB, Ma YK, Zhang QN. Binary classification of Alzheimer ′ s disease using sMRI imaging modality and deep learning. J Digit Imaging. 2020;33(5):1073-90.
3. Tiwari S, Atluri V, Kaushik A, et al. Alzheimer′s disease: pathogenesis, diagnostics, and therapeutics. Int J Nanomedicine. 2019;14:5541-54.
4. De la Torre JC. Alzheimer′s disease is incurable but preventable. J Alzheimers Dis. 2010;20(3):861-70.
5. Alzheimer′s Association. 2020 Alzheimer′s disease facts and figures. Alzheimers Dement. 2020;16(3):391-460.
6. Bateman RJ, Xiong C, Benzinger TLS, Fagan AM, Cruchaga C, Goate AM, et al. Clinical and biomarker changes in dominantly inherited Alzheimer ′ s disease. N Engl J Med. 2012;367(9):795-804.
7. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer′s disease from healthy aging: evidence from functional MRI. Proc Natl Acad Sci U S A. 2004;101(13):4637-42.
8. Guan H, Liu M. Domain adaptation for medical image analysis: a survey. IEEE Trans Biomed Eng. 2022;69(3):1173-85.
9. Hwang S, et al. Multi-scale feature fusion with hierarchical attention for Alzheimer ′ s disease classification. Med Image Anal. 2021;72:102107.
10. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018.
11. Ma N, Zhang X, Zheng HT, Sun J. ShuffleNet V2: practical guidelines for efficient CNN architecture design. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018.
12. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946. 2019.
13. Abedinzadeh Torghabeh F, Hosseini SA. Deep learning-based brain tumor segmentation in MRI images: a MobileNetV2-DeepLabV3+ approach. Iran J Med Phys. 2024;21(6):343-54.
14. Li H, Wei Y, Li L, Tang X. Hierarchical feature extraction with local neural response for image recognition. IEEE Trans Cybern. 2013;43(2):412-24.
15. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018.
16. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018.
17. Zhu X, Cheng D, Zhang Z, Lin S, Dai J. An empirical study of spatial attention mechanisms in deep networks. In: Proceedings of the IEEE International Conference on Computer Vision; 2019.
18. Mu S, Shan S, Li L, Yang Z, Fang Z, Chen Z, et al. DMA-HPCNet: dual multi-level attention hybrid pyramid convolution neural network for Alzheimer′s disease classification. IEEE Trans Neural Syst Rehabil Eng. 2024.
19. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci. 2007;19(9):1498-507.
20. Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer ′ s Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008;27(4):685-91.
21. Roncero-Parra C, Parreo-Torres A, Sánchez-Reolid R, Morales-Sánchez J, de la Torre-Díez I, López-Coronado M. Inter-hospital moderate and advanced Alzheimer′s disease detection through convolutional neural networks. Heliyon. 2024;10(4):e26298.
22. Lim JS. Zoom-in neural network deep-learning model for Alzheimer′s disease assessments. Sensors (Basel). 2022;22(22):8887.