DMA-HPCNet: Dual Multi-Level Attention Hybrid Pyramid Convolution Neural Network for Alzheimer's Disease Classification

Shiguan Mu, Shixiao Shan, Lanlan Li, Shuiqing Jing, Ruohan Li, Chunhou Zheng, and Xinchun Cui

***Abstract*— Computer-aided diagnosis (CAD) plays a crucial role in the clinical application of Alzheimer's disease (AD). In particular, convolutional neural network (CNN)-based methods are highly sensitive to subtle changes caused by brain atrophy in medical images (e.g., magnetic resonance imaging, MRI). Due to computational resource constraints, most CAD methods focus on quantitative features in specific regions, neglecting the holistic nature of the images, which poses a challenge for a comprehensive understanding of pathological changes in AD. To address this issue, we propose a lightweight dual multi-level hybrid pyramid convolutional neural network (DMA-HPCNet) to aid clinical diagnosis of AD. Specifically, we introduced ResNet as the backbone network and modularly extended the hybrid pyramid convolution (HPC) block and the dual multi-level attention (DMA) module. Among them, the HPC block is designed to enhance the acquisition of information at different scales, and the DMA module is proposed to sequentially extract different local and global representations from the channel and spatial domains. Our proposed DMA-HPCNet method was evaluated on baseline MRI slices of 443 subjects from the ADNI dataset. Experimental results show that our proposed DMA-HPCNet model performs efficiently in AD-related classification tasks with low computational cost.**

***Index Terms*— Alzheimer's disease, pyramid convolution, computer-aided diagnosis, attention mechanism, MRI.**

Manuscript received 6 January 2024; revised 29 March 2024 and 27 April 2024; accepted 6 May 2024. Date of publication 8 May 2024; date of current version 22 May 2024. This work was supported in part by the Research Star-Up Funds for High-Level Recruited Talents of University of Health and Rehabilitation Sciences under Grant TLH2022030. *(Corresponding authors: Chunhou Zheng; Xinchun Cui.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Alzheimer's Disease Neuroimaging Initiative.

Shiguan Mu, Shixiao Shan, Shuiqing Jing, and Ruohan Li are with the School of Computer Science, Qufu Normal University, Rizhao 276826, China (e-mail: sg_mu543@foxmail.com; ssxshine@126.com; jsq_sd@163.com; 122788772@qq.com).

Lanlan Li is with Shidao People's Hospital of Rongcheng City, Weihai 264308, China (e-mail: lan13455856778@126.com).

Chunhou Zheng is with the School of Artificial Intelligence, Anhui University, Hefei 230601, China (e-mail: zhengch99@126.com).

Xinchun Cui is with the School of Computer Science, Qufu Normal University, Rizhao 276826, China, and also with the School of Foundational Education, University of Health and Rehabilitation Sciences, Qingdao 266113, China (e-mail: cuixinchun@uor.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3398640

I. Introduction

ALZHEIMER'S disease (AD), characterised by progressive brain atrophy, is the most common neurological disorder that ultimately triggers an irreversible decline in cognitive function [1]. Dementia-associated brain atrophy is a key biomarker for aiding the clinical diagnosis of AD and is essential for early diagnosis of AD and intervention in the progression of the disease [2], [3]. Recently, magnetic resonance imaging (MRI) has provided a non-invasive imaging approach that can detect subtle morphological changes in the brain [4], [5]. On this basis, many studies [6], [7], [8], [9], [10], [11], [12], [13] have applied machine learning or deep learning methods to assist in the clinical detection of AD and its prodrome, i.e., mild cognitive impairment (MCI).

Conventional MRI-based computer-aided diagnosis (CAD) [14], [15], [16], [17], [18], [19], [20] processes for AD typically involve three steps, i.e., 1) image processing, 2) feature extraction, and 3) diagnostic model construction. MRI preprocessing can extract specific regional information to enhance the accuracy and efficiency of AD diagnosis. According to the feature extraction methods [13], these CAD methods can be divided into three categories, i.e., 1) region-level, 2) patch-level, and 3) slice-level. Due to the abundance of structural information in the brain, traditional methods [9] extract whole-brain information (e.g., brain volume density and gray/white matter volume) as input, which often suffers from the curse of high-dimensional features [21], resulting in difficult network fitting. To alleviate this problem, region-level methods [12], [14], [17], [22] have been proposed to focus on specific brain regions that tend to be highly associated with the development of dementia (e.g., the hippocampus [7] and entorhinal area [8]). However, this method mostly relies on expert knowledge or pre-segmentation of regions of interest (ROIs) (e.g., automated anatomical labeling, AAL [18]) and only extracts quantitative regional features, ignoring global changes in brain atrophy. Through further refining the feature extraction scale, the patch-level methods [10], [15], [20], [23] can extract more subtle features, and by combining different patch blocks, they can effectively represent global atrophy changes. However, these methods mostly involve 3D CNN to construct classifiers, which is resource-intensive for training numerous patch blocks. In contrast to processing whole images or specific regions, slice-level methods [19], [24], [25] focus

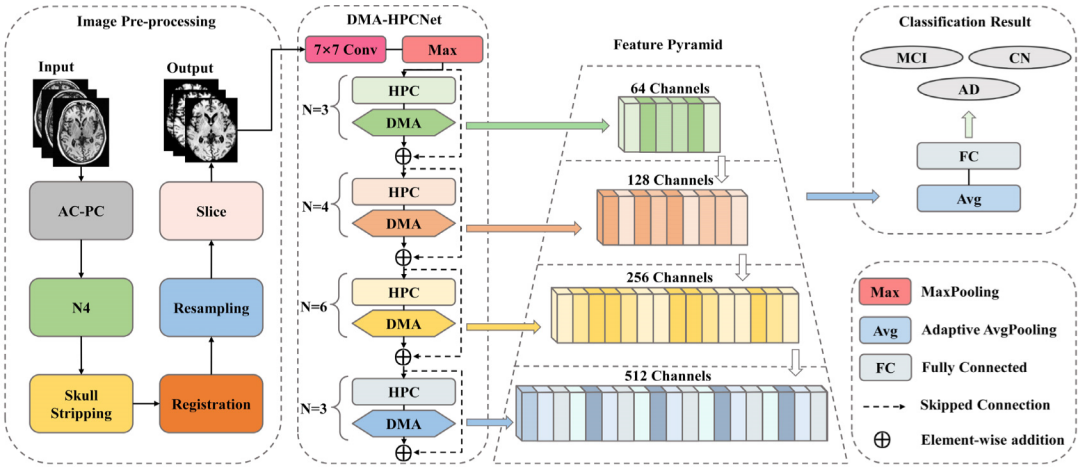

Fig. 1. Illustration of our dual multi-level attention hybrid pyramid convolutional neural network (DMA-HPCNet), which is developed on a slice dataset after image pre-processing. DMA-HPCNet was composed of residual network as the backbone network and extended hybrid pyramid convolution (HPC) block with dual multi-level attention (DMA) module. The classification results of AD-related classification tasks are obtained by extracting features (i.e., Feature Pyramid) to the pooling layer and finally mapping to the fully connected layer.

on 2D slice images, potentially offering greater computational efficiency in processing and analysis. However, over-focusing on local features while neglecting the global nature of images remains a challenge in slice-level methods.

In recent years, deep convolution neural networks (CNNs) have made significant progress in MRI image processing [26], [27]. Unlike traditional machine learning methods, CNNs have the ability to automatically learn advanced features from MRI data, driving the establishment of end-to-end models in brain disease research. Although deeper networks aid in extracting discriminative dementia features from complex brain structures, this exacerbates the problem of computational resource strain [28], [29] especially when using 3D CNNs [30]. Furthermore, most deep learning methods used to aid clinical diagnosis of AD still focus on specific areas within the same template space, neglecting the global nature of the images. Therefore, constructing a lightweight model that comprehensively incorporates both local features and global features is crucial for MRI-based assisted clinical diagnosis.

To tackle aforementioned challenges, we propose a lightweight dual multi-level attention hybrid pyramid convolution neural network (DMA-HPCNet) for aiding in the clinical diagnosis of AD. Specifically, as illustrated in Fig. 1, DMA-HPCNet introduces ResNet-50 as the backbone network and modularly extends the hybrid pyramid convolution (HPC) block and dual multi-level attention (DMA) module. Through the HPC block with hierarchical feature fusion, DMA-HPCNet can learn discriminative features at different scales from MRI images, aiding in obtaining the global features of the images. Then through the DMA module, the attention of channels and spatial domains were successively applied to different scales of dementia features, and the attention weights of different scales were concatenated to obtain the recalibrated whole-level dementia attention. On this basis, a global classifier is constructed to support clinical diagnosis of AD. We evaluated baseline MRI scans of 443 participants from a public dataset (i.e., the Alzheimer's Disease Neuroimaging Initiative, ADNI) and experimental results on multiple AD-related classification tasks show that our proposed DMA-HPCNet method achieves excellent performance at low computational cost in clinical diagnosis using MRI slices. In contrast to existing approaches, our main contributions can be summarized as follows.

1) A lightweight dual multi-level attention hybrid pyramid convolution neural network (DMA-HPCNet) is proposed for enhancing AD diagnosis performance, which can automatically learn various local and global discriminative dementia features from MRI scan slices and perform well in AD-related classification tasks.
2) The hybrid pyramid convolution (HPC) block with hierarchical feature fusion (HFF) is designed to extract discriminative features at different scales and characterise the global features of MRI images, which contributed to a better understanding of local brain atrophy and global abnormal changes induced by AD.
3) The dual multi-level attention (DMA) module is proposed to gradually filter key features of the channel and spatial domains from different local spaces, and to obtain whole-level representation by recalibrating their attention weights, which can further improve the ability to represent local and global brain structures.

II. Related Works

In this section, we will briefly introduce previous work on CAD methods for AD with MRI data. Then, we will separately review relevant studies on pyramid network and attention mechanism in the domain of medical analysis.

A. MRI-Based Computer Aided Diagnosis for AD

According to the feature extraction method adopted by the network, most of the existing MRI-based computer aided diagnosis methods for AD could be categorized into region-, patch-, and slice-level.

Generally, region-level methods extract quantitative features from pre-segmented brain ROIs for building classification models. For example, Turkson et al. [12] pre-segmented whole brain MRI images into 138 unique anatomical structures and extracted AD-related features from the entorhinal area, amygdala, hippocampus and posterior cingulate cortex.

Khan et al. [16] used FreeSurfer [31] to process sMRI data and extract cortical features of brain regions to improve the classification performance of the three classifiers through feature selection. Ayyar et al. [18] pre-aligned sMRI data with the AAL atlas and used 3D CNN to extract ROIs features from the hippocampus for binary classification of AD. Although emphasizing specific changes in brain regions is critical for determining disease progression, subjective biases and incomplete understanding of pathogenic mechanisms, region-level approaches often require integration across levels and approaches.

Compared with region-level methods, patch-level methods divide the whole brain space into many smaller patches, which enables the comprehensive capture of the correlation and influence between different brain regions. For instance, Ashtari-Majlan et al. [10] utilized anatomical markers in MRI images to extract patches and train the model to classify MRI images. Instead of a single imaging method, Suk et al. [20] proposed a constrained Boltzmann machine (BM) to construct patch-level deep networks by extracting pairs of 3D patches from MRI and PET images to obtain potential feature representations of AD. Liu et al. [32] proposed a patch-based multimodal deep learning framework that utilizes discriminative location discovery strategies and multi-modal image feature integration to effectively capture multi-perspective representations of brain diseases. In order to capture the diverse changes of dementia, Lian et al. [23] proposed to partition patches from a unified linear alignment space and obtain deeper AD representations through combination and pruning strategies, thereby building a hierarchical AD diagnostic model. These custom patch methods could capture the diversity and global information of brain images, which can better reflect the distribution and influence of AD in the whole brain. However, the training of numerous patch blocks may require more computing resources.

To alleviate the computational resource strain, slice-level methods were proposed to construct a slice-level classifier by selecting 2D slices from 3D MRI scans. For instance, Sarraf and Tofighi [19] extracted the displacement and scale-invariant features of slice images from different layers of CNN and LeNet-5 architectures to discriminate the structural MRI data of AD. For many slices, Kumar and Nandhini [25] proposed entropy-based feature extraction combined with transfer learning to improve the efficiency of AD diagnosis by selecting MRI slices with the most information in the training stage. Kim et al. [33] applied slice-selective learning and generative adversarial networks to build general models of AD and NC features in different PET image acquisition environments. The slice-level approach provides rapid and targeted information for clinical diagnosis of AD. However, how to improve the ability to obtain the global features of the image is still a challenge for this method.

B. Pyramid Network

In medical imaging, pyramid networks [34], [35] extend parallel branches or layers to capture diverse features, accurately capturing pathological characteristics specific to the disease under different perceptions. As a distinct neural network architecture, pyramid networks often connect and merge different layers using different methods (e.g., pyramid pooling operation, feature pyramid fusion, and multi-scale convolution [36]). For example, Chen et al. [37] achieved feature extraction and integration at different scales by introducing expansion ratios in convolution kernels without increasing network parameters. Building on this, Mohamed et al. [38] introduced a spatial pyramid pooling module to improve the automated screening of various eye diseases. Ge et al. [39] employed multi-scale convolution network in parallel over multiple tissue regions to extract dementia features of various sizes. While these pyramid networks can acquire local information at different scales, an efficient method is still required for deeper structural information.

C. Attention Mechanism

Recently, attention mechanism [40], [41] has shown remarkable performance in addressing the suboptimal ability of models to detect pathological changes in medical images caused by irrelevant or noisy information (e.g., motion artifacts and radiographic artifacts). For instance, Aetesam and Maji [42] proposed symmetric group attention blocks to enhance the practicality of MRI image processing within noise level ranges. Sinha et al. [43] used attention-guided deep networks to harmonize images from multiple MRI datasets, addressing noise effects caused by different scanning protocols and devices. Additionally, Yan et al. [44] proposed various image filtering methods and a pyramid squeeze attention mechanism for AD image classification, considering diverse pathological variations of the disease in different regions. Illakiya et al. [45] integrated enhanced non-local attention (ENLA) and coordinate attention modules to extract both global and local features from MRI images. Although attention mechanisms have excelled in computer-aided diagnosis for AD, research focusing on applying attention at different scales of pathology while considering the whole nature of images remains limited. It's worth noting that combining pyramid network with attention mechanism might be a promising direction to explore.

III. Materials and Method

In this section, we will introduce the materials used in the experiments, followed by a detailed analysis of the framework of the method, and then, an explanation of the loss function employed.

A. Materials

The dataset used in this study is sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database [46], which has collected a large amount of clinical, imaging, and biomarker data through long-term follow-up of participants. Based on clinical assessments (e.g., Mini-Mental State Examination, MMSE [47]), as shown in Table I, we downloaded the baseline T1-weighted MRI scans of 443 subjects from ADNI and categorized them into three groups: 154 AD subjects, 150 MCI subjects, and 139 CN subjects.

All original T1-weighted MRI scan data were uniformly pre-processed using Medical Image Processing Analysis and

TABLE I
DEMOGRAPHIC INFORMATION FOR SUBJECTS INCLUDED IN DIFFERENT GROUP TYPES (I.E. AD, MCI, AND CN). THE GENDER IS EXPRESSED AS MALE AND FEMALE. THE AGE, YEARS OF EDUCATION, AND MINI-MENTAL STATE EXAMINATION (MMSE) SCORES ARE PRESENTED AS MEAN ± STANDARD DEVIATION (STD)

Group Type	Gender	Age	Education	MMSE
AD	91/63	75.3 ± 7.5	14.8 ± 3.1	23.2 ± 2.1
MCI	97/53	74.8 ± 6.9	15.6 ± 2.7	26.4 ± 1.8
CN	77/62	73.9 ± 5.7	16.1 ± 2.8	29.1 ± 1.1

Visualization (MIPAV) software[1] [48] for subsequent analysis and research purposes. As shown in Fig. 1, image preprocessing workflow in MIPAV follows standard steps, which can be summarized as follows. 1) The anterior commissure (AC) and posterior commissure (PC) corrections were applied to all 3D MRI anatomical structures, aligning them to a standardized coordinate system. 2) Implementation of the N4 algorithm to eliminate intensity variations caused by magnetic field inhomogeneity [49]. 3) Application of skull stripping to eliminate non-diagnostic structures [50]. 4) Linear alignment with the Colin27 template [51] to eliminate global linear differences, including translation, scaling, and rotational disparities. 5) Data resampling to ensure uniform spatial resolution (i.e., $1 \times 1 \times 1\text{mm}^3$) across all datasets.

Based on the pre-processed 3D MRI data, we first used the Nifti tool to select images by volumetric center positioning [25], then selected the clearest brain structural segments as samples (i.e., the parts of the MRI images that do not contain motion artifacts or distortions), and finally performed data augmentation operations, including random rotation, contrast adjustment, and left and right movement, to generate 2D slice dataset. Our slice dataset consists of 4390 slices, including 1510 AD slices, 1500 MCI slices, and 1380 CN slices. It is worth noting that our data segmentation (i.e., into the training set, test set, and validation set according to 7:2:1) is performed before the slice selection step to prevent images of the same subject from appearing in both the training and test sets. The implementation of these steps provides a reliable basis for subsequent analysis and research, and the experimental results confirm the feasibility of the method.

B. Network Architecture

Our proposed dual multi-level attention hybrid pyramid convolution neural network (DMA-HPCNet) is developed in a slice dataset after linear registration. As shown in Fig. 1, we employ the ResNet-50 as the backbone network, comprising five main convolution blocks. Firstly, the initial convolution block (kernel size: 7×7, padding: 3, and stride: 2) increases the input channels from 3 to 64. Subsequently, the remaining convolution blocks consist of hybrid pyramid convolutions with dual multi-level attention, designed to extract features from MRI images instead of the bottleneck block of ResNet-50. Finally, these convolution blocks are stacked in a modular manner with a configuration of 3, 4, 6, and 3, each module has corresponding channel numbers of 64, 128, 256, and 512.

[1] http://mipav.cit.nih.gov/index.php

1) Deep Residual Learning Network: ResNet introduces the concept of residual learning (i.e., residual connections or skip connections) [30], which allows the network to automatically learn the difference between the original input and the desired output, solving the problem of network degradation caused by traditional networks as they continue to deepen. Briefly, this cross-layer connection operation improves the efficiency of information propagation, contributes to the continuous deepening of the network, and avoids the performance degradation caused by the disappearance of gradient or network degradation. Compared to the traditional network, ResNet has demonstrated excellent performance in large-scale image classification tasks such as ImageNet [52]. However, in medical image diagnostics, when dealing with varying sizes of pathological features, networks need the capability to capture more nuanced and comprehensive features to support accurate diagnosis. ResNet shows low sensitivity to various pathological features, necessitating improvement in learning crucial feature information. Therefore, to address these issues, we have improved the network structure.

2) Hybrid Pyramid Convolution Block: Multi-scale convolution (MSC) block adopts the form of pyramid convolution layer, enhancing the model's perception and detection abilities, especially in medical image diagnostics. Due to the complex brain structure information present in MRI brain images, MSC is sensitive to the local and global changes in brain regions caused by AD, but this is accompanied by the increase of computational complexity and parameter amount. Therefore, it is necessary to balance the fusion methods and structures of different scale information.

In this paper, a simple but effective MSC block (i.e., HPC) is proposed to extract different scale features from MRI images, as shown in Fig. 2, where common MSC blocks either combine convolution kernels of different sizes (e.g., 3×3, 5×5, 7×7, and 9×9) or introduce dilation rates of different sizes (i.e., dilated convolution). Different from them, we have adopted a cascading method to combine dilated convolution and grouped convolution [53]. Then the hierarchical feature fusion (HFF) [54] is introduced to progressively superimpose different perceptual spatial information, alleviating the gridding artifacts of dilated convolution at low computational cost.

Firstly, assuming the input feature map is X, where $X \in \mathbb{R}^{C \times H \times W}$, C, H, and W denote the number of channels, height, and width of the feature map, respectively. Then X is split into S parts of parallel along the C (i.e., channel direction), denoted by $[x_0, x_1, \ldots, x_{s-1}]$, the channel number of each segmented part is $C' = C/S$.

For each part of X, we employ dilated convolutions with dilation rates of d to extract features at different scales, where $d \in \{1, 2, \ldots, n\}$. Inspired by the EPSANet [55], we combine different kernels with varying numbers of groups applied in dilated convolutions to reduce the network's parameter count. This dilated grouped convolutions can be defined as follows:

$$D_i = Dconv_{3\times3}(d_i, G_i)(X_i) \tag{1}$$

where $i \in \{1, 2, \ldots, n\}$, $G_i = 2^{d_i+1}$ is the number of groups based on dilation rate variation and $Dconv_{3\times3}$ is the dilated convolution with 3×3 kernel size.

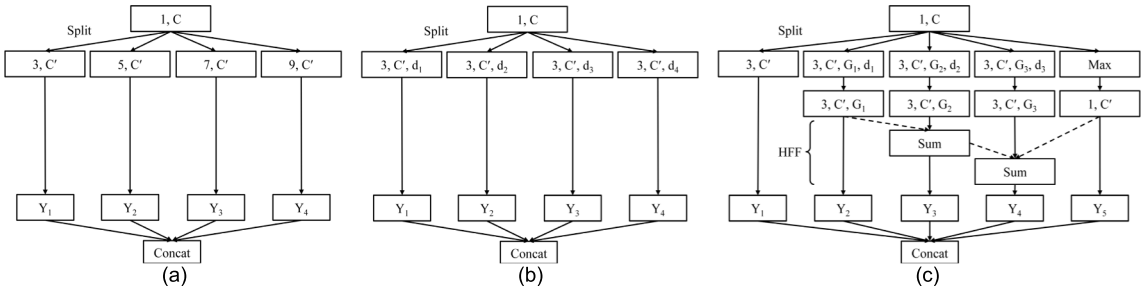

Fig. 2. The specific structure of MSC Block. (a) is split multi-scale convolution (SPC) block, including different size convolution kernels. (b) is atrous split multi-scale convolution (ASPC) block with different atrous rates (or dilated rates). (c) is our proposed HPC block. We denote each module as (# kernel size, # number of kernels, # number of groups, # dilated rate). Here, # Sum represents the addition operation of corresponding spatial positions, HFF represents hierarchical feature fusion, and Y_i represents the corresponding output results of different convolutions.

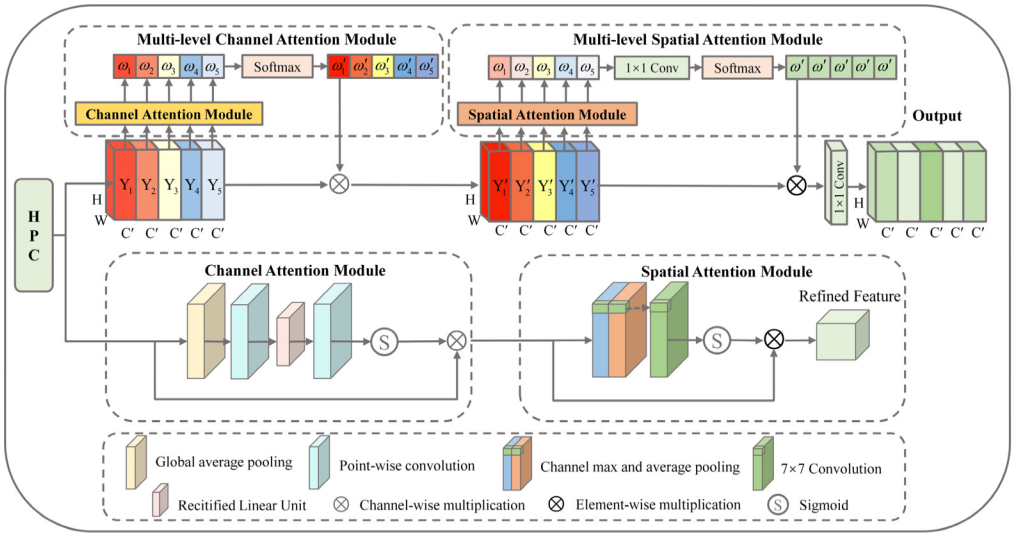

Fig. 3. Illustration of our dual multi-level attention (DMA) module, which consists of a channel attention module and a spatial attention module, is used in different input parts. Then the attention weights of different scales in the channel and spatial domain are recalibrated by the Softmax function to obtain the whole-level representation.

Then the original feature is mapped into a grouped convolutions $Gconv_{3\times3}$ (kernel size: 3×3, stride: 1, and padding: 1), whose output is denoted as F_0. Besides, after each dilated grouped convolution, a grouped convolution is cascaded to ensure equal receptive fields with other MSC blocks in Fig. 2. This can be expressed as:

$$F_i = Gconv_{3\times3}(D_i, G_i) \tag{2}$$

Finally, prior to connecting features at each scale, HFF is introduced to hierarchically add output features and integrate supplementary information $F_{\max}$, where $F_{\max}$ generated by max pooling and point-wise convolution. The specific process of HFF is as follows:

$$Y = [F_0; F_1'; F_2'; F_3'; F_{max}] \tag{3}$$

where $F_i' = \sum_{k=1}^{i} F_k$, [;] is concatenation, and Y is the obtained feature map after multi-scale feature fusion.

This HPC block improves the representational capability of the model by cascading different convolution layers and integrating different spatial information. Subsequent ablation studies verified the effectiveness of this block.

3) Dual Multi-Level Attention Mechanism: Inspired by the CBAM [41], we have adopted the integrated feature attention method of channel and spatial domain, and analyze the function of attention from the part to the whole, and then design dual multi-level attention (DMA) mechanism to further improve the diagnostic performance. As shown in Fig. 3, the DMA mainly consists of multi-level channel attention (MCA) and multi-level spatial attention (MSA) modules by successively use channel attention and spatial attention for different parts of different domain spaces of the HPC block.

Firstly, during the HFF stage, the MCA module adopts global average pooling (GAP) to obtain importance scores for each channel of Y_i. The specific calculation is as follows:

$$Z = GAP(Y_i) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} Y_{ic}(i, j) \tag{4}$$

where Z is the result of averaging each channel over the entire space. The attention weight of cth channel in the feature map of different scales can be written as:

$$\omega_i = \sigma(\tau(\delta(\tau(Z_i)))) \tag{5}$$

where $\omega_i \in \mathbb{R}^{C'\times1\times1}$ is the attention weight of feature at different scales, δ is the Rectified Linear Unit (ReLU)

operation, σ is the sigmoid activation, and τ is a point-wise convolution (kernel size: 1×1, stride: 1, and padding: 0). By employing two point-wise convolutions to combine linear information between channels, yielding fewer parameters compared to fully connected layers.

Secondly, to facilitate information interaction among various scales channel attentions and capture overall channel correlations, we concatenate individual channel attention vectors and denote as:

$$\omega = [\omega_0 \ ; \ \omega_1; \ldots; \omega_{s-1}] \tag{6}$$

where ω is the whole channel attention vector. And we represent Softmax function as ρ to recalibrate the whole attention vector. Then the recalibrated attentions are connected to achieve the interaction of part and whole attention.

$$Catt = [Catt_0 \ ; \ Catt_2; \ldots; Catt_{s-1}] \tag{7}$$

where $Catt_i = \rho(\omega_i)$ is the attention weight of different parts and $Catt$ is the attention weight of whole. Then the recalibrated attention weight is multiplied with the corresponding scale feature map and expressed as:

$$Y' = Y_i \odot Catt_i \tag{8}$$

where $\odot$ is the channel-wise multiplication and Y' is the channel attention weighted feature map of the MCA module.

In addition, MSA module is introduced to extract spatial attention weights of channel attention-weighted feature maps at different scales. Among them, each part of the attention-weighted feature map performs two different pooling operations (i.e., channel average pooling and channel max pooling) along the channel axis, respectively, whose output are denoted as avg_i and max_i. And then these outputs are mapped to a convolution layer $Conv_{7\times7}$ (kernel size: 7×7, stride: 1, and padding: 3), and the activation function is adopted to generate spatial attention weight.

$$Aatt_i = \sigma(Conv_{7\times7}([avg_i; max_i])) \tag{9}$$

where $Aatt_i$ is the spatial attention weights of different scales.

Spatial attention vectors of different scales are concatenated and then synthesized using a point-wise convolution and the activation function to represent the attention weights of the whole space. The specific calculation is as follows:

$$Aatt' = \sigma(\tau([Aatt_0 \ ; \ Aatt_1; \ldots; Aatt_{s-1}])) \tag{10}$$

Finally, the recalibrated spatial attention vector is multiplied with the channel attention-weighted feature map to obtain the channel and spatial information enhanced feature map.

$$Out_i = Y_i' \otimes Aatt_i' \tag{11}$$

where $\otimes$ is element-wise multiplication. Then the refined output obtained after dual multi-level attention is expressed as:

$$Out_i' = [Out_0 \ ; \ Out_1; \ldots; Out_{s-1}] \tag{12}$$

This DMA module improves network performance by adopting channel and spatial attention on each local and global feature map. We demonstrate this module and its components (i.e., MCA and MSA) in the ablation study in Section IV.

C. Loss Function

To precisely diagnose whether an individual has dementia and its extent, the cross-entropy loss function is utilized to optimize our HPC method. During the training process, measuring the difference between model predictions and true labels, along with minimizing the cross-entropy loss function, enables to obtain model for each MRI image classification results. The cross-entropy loss function can be denoted as:

$$L(y, p) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right] \tag{13}$$

where N represents the number of images, y_i is the true label of image i, and p_i is the prediction probability of image i.

IV. Experiments and Discussion

In this section, we initially provide a brief overview of the experimental setting and assessment criteria, then gradually validate the effects of each component of the proposed method on the performance of the model through, and finally the proposed model is compared with popular models and advanced studies to assess its performance.

A. Experimental Setting

In this study, the experiments were run on a computer equipped with an NVIDIA GeForce RTX 2080 Ti 11GB GPU. We have used MIPAV for the same preprocessing of all medical data, and all experiments were implemented based on PyTorch 1.7.1 and Python 3.9[2]. The model parameters are set as follows: the batch size is 32, the optimizer for the training network is Adam, the number of epochs is 80, and the learning rate is 1e-5. The experimental dataset consisted of 4390 images, which were segmented at a ratio of 7:2:1.

B. Assessment Criteria

We have adopted several common assessment metrics to comprehensively evaluate the effectiveness of the proposed model. These metrics encompass specificity (SPE), sensitivity (SEN), accuracy (ACC), and area under the ROC curve (AUC). They are defined as follows, $SPE = \frac{TN}{TN+FP}$, $SEN = \frac{TP}{TP+FN}$, and $ACC = \frac{TP+TN}{TP+TN+FP+FN}$, where TP is true positives, FN is false negatives, TN is true negatives, FP is false positives. The AUC is calculated based on false positive rate ($FPR = 1 - SPE$) and true positive rate ($TPR = SEN$), which are obtained from the threshold (default is 0.5) performed on the classification score generated by the network.

C. Ablation Study

This section conducted ablation experiments to verify the effectiveness of the HPC block and attention mechanism in our proposed method. All methods were fairly compared using the same test subjects and assessment criteria.

1) Effectiveness of Hybrid Pyramid Convolution Block: The effectiveness of HPC block is verified by its presence or absence in multiple classification tasks, as shown in Table II.

[2]https://github.com/SG-Python3/PyTorch

TABLE II
ACC, SEN, SPE, AND AUC OF VARIOUS MSC MODELS FOR AD, MCI, AND NC IN SLICE DATASET

	Method	ACC	SEN	SPE	AUC
AD vs CN	ResNet-50	0.923	0.917	0.933	0.978
	ResNet-50+SPC	0.944	**0.970**	0.905	0.985
	ResNet-50+ASPC	0.931	0.947	0.906	0.977
	HPCNet	**0.952**	0.960	**0.939**	**0.987**
AD vs MCI	ResNet-50	0.887	0.892	0.880	0.951
	ResNet-50+SPC	0.909	**0.910**	0.907	0.973
	ResNet-50+ASPC	0.902	0.878	**0.937**	0.964
	HPCNet	**0.917**	**0.910**	0.928	**0.974**
MCI vs CN	ResNet-50	0.798	0.833	0.763	0.874
	ResNet-50+SPC	0.821	0.914	0.740	0.884
	ResNet-50+ASPC	0.811	0.832	**0.792**	0.878
	HPCNet	**0.825**	**0.923**	0.740	**0.891**

TABLE III
THE PARAMETERS, FLOPS, MEMORY, AND RUNNING TIME OF VARIOUS MSC BLOCKS IN THE MCI/CN OF SLICE DATASET

	ResNet-50+SPC	ResNet-50+ASPC	HPCNet
Params	63.7×10^6	23.5×10^6	18.1×10^6
FLOPs	10.7×10^9	4.1×10^9	3.2×10^9
Memory	244.93M	91.25M	70.80M
Time (ms)	72.4	50.3	53.1

TABLE IV
THE ACC AND STANDARD DEVIATION OF VARIOUS MSC MODELS IN THE MCI/CN OF SLICE DATASET

	ResNet-50+SPC	ResNet-50+ASPC	HPCNet
ACC	0.821	0.811	0.825
Std	0.036	0.020	0.011

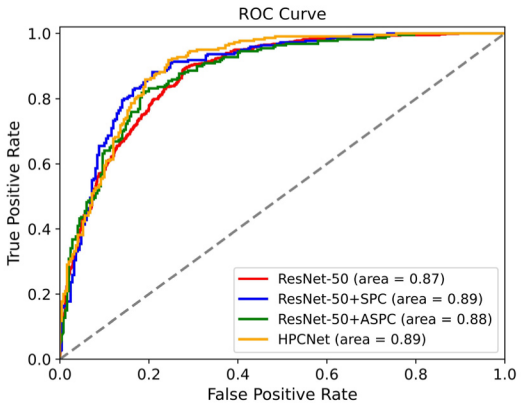

Fig. 4. Introduction of different multi-scale structures (i.e., SPC, ASPC, and HPC) in ResNet-50 for the classification of ROC curves in MCI/CN.

Introducing different MSC blocks to the baseline ResNet-50 effectively improved the classification performance of the model. Among them, introducing HPC blocks to construct HPCNet model in several classification tasks, such as AD/CN, AD/MCI, and MCI/CN, ACC improved by 2.9%, 3%, and 2.7%, respectively. In comparison to the SPC and ASPC blocks, our proposed HPC blocks exhibited superior ACC and AUC values in AD-related tasks. This improvement is attributed to the HPC block's fusion of various perceptual information, compensating for the limited capability of a single convolution layer to extract global features.

To further validate the effectiveness of the HPC block, we conducted a comprehensive comparison between the HPC block and two types of MSC blocks in the MCI/CN group, as depicted in Table III and Table IV. Compared to the traditional SPC structure, which consumes excessive parameters, the HPC block tends to enhance the ASPC structure, achieving more efficient classification performance with fewer parameters and FLOPs with only a small increase in running time. Meanwhile, the ACC of the HPC block improved by 0.4% and 1.4% compared to SPC block and ASPC block, respectively, and the standard deviation was lower. As shown in Fig. 4, the area under ROC curve corresponding to the HPCNet built with HPC block is competitive with ResNet-50+SPC. These results demonstrate that our HPC block is not only more efficient, but also lighter.

2) Effectiveness of Dual Multi-Level Attention: The effectiveness of the DMA block depends on its presence within the HPCNet model. Firstly, Table II and Table V show that the introduction of the DMA block results in significant improvements in several classification tasks. In the AD/CN and AD/MCI classification experiments, HPCNet exhibited increases of 2.8% and 2.9% in ACC, 2.1% and 3.6% in SEN, 4.0% and 2.0% in SPE, and 1.0% and 1.1% in AUC, respectively. In the MCI/CN identification experiments, HPCNet's ACC improved by 2.2%, achieving SEN, SPE, and AUC of 0.859, 0.836, and 0.921, respectively. From the results, DMA module realized the screening and weighting of features, and improved the classification performance of the model. Secondly, in the AD/CN, AD/MCI, and MCI/CN classification tasks, compared with single channel attention (SE) and multi-level channel attention (MCA), introduced the DMA ACC increased by 2.0% and 1.5%, 2.0% and 1.7%, and 1.5% and 1.1%, respectively. Compared with CBAM, the ACC improved by 1.0%, 1.1%, and 0.7%, correspondingly. Finally, compared to the two components of the DMA (i.e., MCA and MSA), as shown in Fig. 5, DMA respectively improved ACC by 1.1% and 1.3%, achieving ACC, SEN, SPE, and AUC values of 0.847, 0.859, 0.836, and 0.921. These results demonstrate that integrating MCA and MSA aids in further improving the classification performance.

TABLE V
ACC, SEN, SPE, AND AUC OF VARIOUS ATTENTION MECHANISM METHODS FOR AD, MCI, AND NC IN SLICE DATASET

	Method	ACC	SEN	SPE	AUC
AD vs CN	HPCNet+SE	0.960	0.951	0.973	0.992
	HPCNet+MCA	0.965	0.970	0.958	0.993
	HPCNet+CBAM	0.970	0.964	**0.979**	0.996
	DMA-HPCNet	**0.980**	**0.981**	**0.979**	**0.997**
AD vs MCI	HPCNet+SE	0.926	0.925	0.926	0.978
	HPCNet+MCA	0.929	0.932	0.924	0.982
	HPCNet+CBAM	0.935	**0.964**	0.892	0.982
	DMA-HPCNet	**0.946**	0.946	**0.947**	**0.985**
MCI vs CN	HPCNet+SE	0.832	0.827	**0.836**	0.908
	HPCNet+MCA	0.836	**0.900**	0.780	0.905
	HPCNet+CBAM	0.840	0.877	0.808	0.911
	DMA-HPCNet	**0.847**	0.859	**0.836**	**0.921**

TABLE VI
COMPARISON WITH POPULAR MODELS IN THE MCI/CN

Model	Dataset	ACC	SEN	SPE	AUC
ResNet-50 [30]	Slice dataset	0.798	0.833	0.763	0.874
ResNeXt-50 [53]	Slice dataset	0.803	0.849	0.759	0.882
ShuffleNet V2 [28]	Slice dataset	0.755	0.741	0.768	0.815
EfficientNet [29]	Slice dataset	0.845	**0.882**	0.812	0.915
EPSANet [55]	Slice dataset	0.832	0.850	0.816	0.901
Proposed method	Slice dataset	**0.847**	0.859	**0.836**	**0.921**

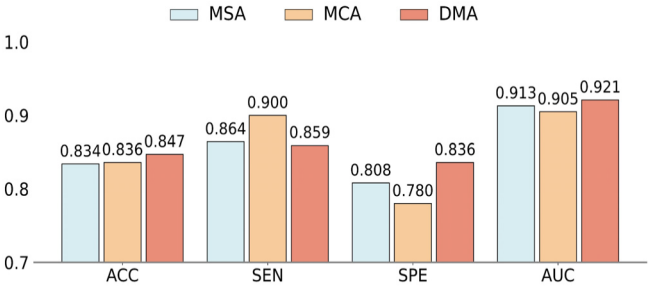

Fig. 5. Comparison of the effects of DMA and its different parts (i.e., MSA and MCA) in the MCI/CN.

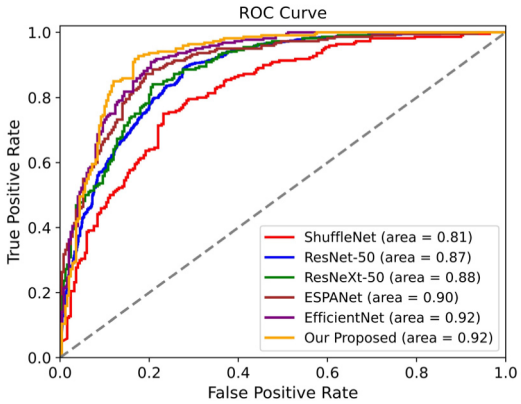

Fig. 6. The ROC curves of the proposed model and several popular models for the classification in the MCI/NC.

D. Result Analysis

To evaluate the performance of the proposed DMA-HPCNet model, several popular models [28], [29], [30], [53], [55] were introduced into our slice dataset for fair comparison. As depicted in Table VI and Fig. 6, in the MCI/CN identification experiment, the following conclusions can be drawn as follows. Firstly, compared with several popular models, our DMA-HPCNet model has achieved significant improvement in various assessment metrics such as ACC, SPE, and AUC, achieving 0.847, 0.836, and 0.921, respectively. Secondly, our DMA-HPCNet model demonstrates competitive performance compared to the advanced models like EfficientNet [29] (0.845, 0.882, 0.812, and 0.915 for ACC, SEN, SPE, and AUC, respectively). Finally, the ROC curve corresponding to our DMA-HPCNet model in this paper approached closest to the upper boundary (1,0) and left boundary (0,1), indicating superior classification performance.

To further evaluate the performance of our DMA-HPCNet model, we conducted a comprehensive comparison with various advanced studies, as shown in Table VII, which includes network feature extraction forms, dataset sizes, and experimental results. Based on the network feature extraction forms, we summarised recent research on AD classification tasks, including two region-level methods [12], [14], two patch-level methods [10], [20], and two slice-level methods [19], [24]. From Table VII, several conclusions were drawn as follows. Firstly, compared to various advanced methods, our dataset showed relative balance and performed well in various classification tasks. For example, in the MCI/CN discrimination experiment, although the accuracy is generally not high, our DMA-HPCNet model achieved an accuracy of 0.847 in extracting slice-level features, which is an improvement of 0.8% and 0.5% compared to the region-based and patch-level approaches adopted by Turkson et al. [12] and Suk et al. [20], respectively. Additionally, our DMA-HPCNet model achieved an accuracy of 0.946 in the AD/MCI classification task,

TABLE VII
THE RESULTS OF EACH METHOD IN THE AD/NC, AD/MCI, AND MCI/NC

Reference	CNN Model	Subjects	Accuracy		
			AD vs CN	AD vs MCI	MCI vs CN
Turk et al. [12]	ROI-level	150AD+150MCI+150CN	0.901	0.873	0.839
Liu et al. [14]	ROI-level	180AD+374MCI+204CN	0.914	-	0.821
Ashtari-Majlan et al. [10]	Patch-level	200AD+267MCI+231CN	0.977	-	-
Suk et al. [20]	Patch-level	93AD+204MCI+101CN	0.924	-	0.842
Sarraf et al. [19]	Slice-level	211AD+91CN	**0.988**	-	-
Chen et al. [24]	Slice-level	43AD+97MCI+60CN	0.946	0.921	-
Proposed method	Slice-level	154AD+150MCI+139CN	0.980	**0.946**	**0.847**

outperforming other methods listed in the table. Secondly, as a crucial basis for the clinical diagnosis of AD, in the AD/CN group, our DMA-HPCNet model achieved a classification accuracy of 0.980, which is competitive with the patch-based method adopted by Ashtari-Majlan et al. [10]. Most of these methods involve extracting 3D information to model, which is difficult to achieve in a lightweight manner. It is worth noting that our DMA-HPCNet model selects cross sections from 3D MRI images after registration to construct a lightweight CNN model, which is superior to other slice-level methods [19], [24] in AD/MCI and MCI/CN classification tasks. Finally, based on these results, we conclude that our proposed DMA-HPCNet model, which uses hybrid pyramid convolution block and dual multi-level attention mechanism, proves its validity in AD diagnosis with excellent performance.

V. CONCLUSION AND FUTURE WORK

In this study, we propose a lightweight dual multi-level hybrid pyramid convolutional neural network (DMA-HPCNet) to assist clinical diagnosis of AD. Our model extends hybrid pyramid convolution (HPC) block and double multi-level attention (DMA) module on the basis of ResNet. Among them, the HPC block is designed to improve the multi-scale information capture capability of the model. Based on this, the DMA module is proposed to sequentially extract different local and global representations from the channel domain and spatial domain. Our proposed DMA-HPCNet model was evaluated on baseline MRI slice dataset of 443 subjects. The experimental results show that our model performs well in AD-related classification tasks than popular models and advanced research methods. While our proposed DMA-HPCNet method has shown good performance in AD-related diagnosis, there are still limitations worth exploring and solving in future work. To that end, we have summarized the following limitations and potential solutions. Firstly, the convolution kernel size and number in the feature-extracting HPC block are fixed. However, there are individual differences in brain structures among different participants. While HPC block is highly sensitive to pathological changes, employing adaptive multi-scale kernels for processing inputs seems more reasonable. Moreover, in complex classification tasks (e.g., MCI recognition task), imaging features show no significant differences, and using only single MRI imaging data is insufficient for obtaining precise diagnostic performance. In future work, integrating other modal information (e.g., PET imaging) to identify consistent pathological regions and potential pathogenic features for assisting clinical diagnosis is a worthwhile direction to explore.

ACKNOWLEDGMENT

The investigators within the ADNI provided experimental data but were not involved in the analysis or writing of this study. A complete list of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

REFERENCES

[1] A. Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 15, no. 3, pp. 321–387, Mar. 2019.
[2] R. J. Bateman et al., "Clinical and biomarker changes in dominantly inherited Alzheimer's disease," *New England J. Med.*, vol. 367, no. 9, pp. 795–804, 2012.
[3] Z. Xia, T. Zhou, S. Mamoon, and J. Lu, "Recognition of dementia biomarkers with deep finer-DBN," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1926–1935, 2021.
[4] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 13, pp. 4637–4642, Mar. 2004.
[5] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.
[6] J. Escudero, E. Ifeachor, J. P. Zajicek, C. Green, J. Shearer, and S. Pearson, "Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 164–168, Jan. 2013.
[7] R. Cui and M. Liu, "Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 5, pp. 2099–2107, Sep. 2019.
[8] E. Feczko, J. C. Augustinack, B. Fischl, and B. C. Dickerson, "An MRI-based method for measuring volume, thickness and surface area of entorhinal, perirhinal, and posterior parahippocampal cortex," *Neurobiol. Aging*, vol. 30, no. 3, pp. 420–431, Mar. 2009.
[9] T.-D. Vu, N.-H. Ho, H.-J. Yang, J. Kim, and H.-C. Song, "Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection," *Soft Comput.*, vol. 22, no. 20, pp. 6825–6833, Oct. 2018.
[10] M. Ashtari-Majlan, A. Seifi, and M. M. Dehshibi, "A multi-stream convolutional neural network for classification of progressive MCI in Alzheimer's disease using structural MRI images," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 3918–3926, Aug. 2022.
[11] Y. Wu, Y. Zhou, W. Zeng, Q. Qian, and M. Song, "An attention-based 3D CNN with multi-scale integration block for Alzheimer's disease classification," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 11, pp. 5665–5673, Nov. 2022.
[12] R. E. Turkson, H. Qu, C. B. Mawuli, and M. J. Eghan, "Classification of Alzheimer's disease using deep convolutional spiking neural network," *Neural Process. Lett.*, vol. 53, no. 4, pp. 2649–2663, Aug. 2021.
[13] M. Rashid, H. Singh, and V. Goyal, "The use of machine learning and deep learning algorithms in functional magnetic resonance imaging—A systematic review," *Expert Syst.*, vol. 37, no. 6, Dec. 2020, Art. no. e12644.

[14] S. Liu et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.

[15] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain MRI," *Med. Image Anal.*, vol. 18, no. 5, pp. 808–818, Jul. 2014.

[16] R. U. Khan, M. Tanveer, and R. B. Pachori, "A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging," *Expert Syst.*, vol. 38, no. 1, Jan. 2021, Art. no. e12566.

[17] C. Wee, P. Yap, and D. Shen, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns," *Hum. Brain Mapping*, vol. 34, no. 12, pp. 3411–3425, Dec. 2013.

[18] M. P. Ayyar, J. Benois-Pineau, A. Zemmari, and G. Catheline, "Explaining 3D CNNs for Alzheimer's disease classification on sMRI images with multiple Rois," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 284–288.

[19] S. Sarraf and G. Tofighi, "Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks," 2016, *arXiv:1607.06583*.

[20] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014.

[21] Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables," *NeuroImage*, vol. 50, no. 4, pp. 1519–1535, May 2010.

[22] M. Fabietti et al., "Early detection of Alzheimer's disease from cortical and hippocampal local field potentials using an ensembled machine learning model," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2839–2848, 2023.

[23] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[24] X. Chen, M. Tang, A. Liu, and X. Wei, "Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI," *Ann. Transl. Med.*, vol. 10, no. 14, pp. 765–776, 2022.

[25] S. S. Kumar and M. Nandhini, "Entropy slicing extraction and transfer learning classification for early diagnosis of Alzheimer diseases with sMRI," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 2, pp. 1–22, May 2021.

[26] F. Duan et al., "Topological network analysis of early Alzheimer's disease based on resting-state EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2164–2172, Oct. 2020.

[27] J. Feng, S.-W. Zhang, and L. Chen, "Extracting ROI-based contourlet subband energy feature from the sMRI image for Alzheimer's disease classification," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 3, pp. 1627–1639, May 2022.

[28] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[29] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[31] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[32] F. Liu, S. Yuan, W. Li, Q. Xu, and B. Sheng, "Patch-based deep multi-modal learning framework for Alzheimer's disease diagnosis using multi-view neuroimaging," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104400.

[33] H. W. Kim, H. E. Lee, S. Lee, K. T. Oh, M. Yun, and S. K. Yoo, "Slice-selective learning for Alzheimer's disease classification using a generative adversarial network: A feasibility study of external validation," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 47, no. 9, pp. 2197–2206, Aug. 2020.

[34] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-Resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4484.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[36] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 46, pp. 26–34, May 2018.

[37] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[38] N. A. Mohamed, M. A. Zulkifley, and S. R. Abdani, "Spatial pyramid pooling with atrous convolutional for MobileNet," in *Proc. IEEE Student Conf. Res. Develop. (SCOReD)*, Sep. 2020, pp. 333–336.

[39] C. Ge, Q. Qu, I. Y.-H. Gu, and A. S. Jakola, "Multi-stream multi-scale deep convolutional networks for Alzheimer's disease detection using MR images," *Neurocomputing*, vol. 350, pp. 60–69, Jul. 2019.

[40] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[41] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2018, pp. 3–19.

[42] H. Aetesam and S. K. Maji, "Attention-based noise prior network for magnetic resonance image denoising," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–4.

[43] S. Sinha, S. I. Thomopoulos, P. Lam, A. Muir, and P. M. Thompson, "Alzheimer's disease classification accuracy is improved by MRI harmonization based on attention-guided generative adversarial networks," in *Proc. 17th Int. Symp. Med. Inf. Process. Anal.*, 2021, pp. 180–189.

[44] B. Yan et al., "Quantifying the impact of pyramid squeeze attention mechanism and filtering approaches on Alzheimer's disease classification," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105944.

[45] T. Illakiya, K. Ramamurthy, M. V. Siddharth, R. Mishra, and A. Udainiya, "AHANet: Adaptive hybrid attention network for Alzheimer's disease classification using brain magnetic resonance imaging," *Bioengineering*, vol. 10, no. 6, p. 714, Jun. 2023.

[46] C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.

[47] P. Skorga and C. F. Young, "Mini-mental state examination for the detection of Alzheimer disease and other dementias in people with mild cognitive impairment," *Clin. Nurse Spec.*, vol. 29, no. 5, pp. 265–267, 2015.

[48] M. J. McAuliffe, F. M. Lalonde, D. McGarry, W. Gandler, K. Csaky, and B. L. Trus, "Medical image processing, analysis and visualization in clinical research," in *Proc. 14th IEEE Symp. Comput.-Based Med. Syst. (CBMS)*, Sep. 2001, pp. 381–386.

[49] N. J. Tustison et al., "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[50] Y. Wang, J. Nie, P.-T. Yap, F. Shi, L. Guo, and D. Shen, "Robust deformable-surface-based skull-stripping for large-scale studies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2011, pp. 635–642.

[51] C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *J. Comput. Assist. Tomogr.*, vol. 22, no. 2, pp. 324–333, Mar. 1998.

[52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[53] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.

[54] S. Mehta, M. Rastegari, A. Caspi, L. Shapiro, and H. Hajishirzi, "ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 552–568.

[55] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," in *Proc. Asian Conf. Comput. Vis.*, 2022, pp. 1161–1177.