Sādhanā (2023) 48:173
https://doi.org/10.1007/s12046-023-02219-8

Early diagnosis of Alzhiemer's disease using wavelet-pooling based deep convolutional neural network

MANU RAJU[1], VARUN P. GOPI[2,*], V S ANITHA[3] and ABISHEK SHERAWAT[4]

[1]Department of Electronics and Communication Engineering, Government Engineering College Wayanad, APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India
[2]Department of Electronics and Communication Engineering, National Institute of Technology Tiruchirappalli, Tiruchirappalli, India
[3]Department of Computer Science and Engineering, Government Engineering College Wayanad, APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India
[4]Department of Radiodiagnosis and Imaging, All India Institute of Medical Sciences Bhopal, Bhopal, Madhya Pradesh, India
e-mail: manugecw108@gmail.com; varun@nitt.edu; anithavs@gecwyd.ac.in; abhishek.radiodiagnosis@aiimsbhopal.edu.in

MS received 25 February 2023; revised 5 April 2023; accepted 14 May 2023

Abstract. Coronal anatomic slices of structural MRI images clearly show the topographical structures of the Hippocampus and Amygdala, which are essential for early diagnosis of Alzheimer's disease (AD). MR coronal sections are best appreciated for studying the complex topographical relationships of the amygdala and the topographical structures of the hippocampus, which helps in the early detection of disease. Early diagnosis helps prevent the disease's progression to its final stage. It allows the patient to be aware of the severity of the disease and can take the necessary therapeutic medications to prevent its progression. A coronal view study of MR images is proposed in this paper for early diagnosis of disease using a wavelet-pooling-based multi-path and multi-scale convolutional neural network. This work aims to perform a three-way classification of 2D coronal slices of MRI images to diagnose Mild Cognitive Impairment, AD, and Normal Control in a single algorithm and learn the brain-affected regions through Gradcam visualization. wavelet-pooling is utilized to extract the texture details of the image and thus provide spatial attention to the texture features of the image, which is impossible using Max-pooling or Average-pooling. Multi-scale feature learning is incorporated using parallel multiple low-rank convolutional kernels to capture varying scales of atrophy regions. Multi-path mode compensates for the early loss of features and avoids vanishing gradient problems. The proposed model is trained and tested on the ADNI dataset comprising 900 subjects to give an accuracy of 96.5% with ten-fold cross-validation. The multi-scale and multi-path methods significantly reduce the number of learnable parameters.

Keywords. Alzheimer's disease; mild cognitive impairment; wavelet-pooling; multi-scale CNN; multi-path CNN.

1. Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease mainly occurring in the elderly. As per WHO statistics, around 55 million people will be affected by dementia worldwide in 2025. It is anticipated to double every 20 years, reaching 78 million by 2030 and 139 million by 2050. AD is the leading cause of dementia and contributes to 60–70% of cases. It deteriorates cognitive function and causes memory impairment. It is a progressive disease in which, at its late stage, they lose the complete cognitive ability and memory and depend on others for daily activities. AD has no cure but can be delayed by taking necessary medications if known at earlier stages. Changes in the brain can begin years before the first symptoms appear. Mild Cognitive Impairment (MCI) is the precursor of AD, and it is crucial to know the disease at this stage to prevent further progression. Structural MRI is an excellent biomarker to predict AD, especially in the MCI stage, since it gives the anatomic structure of atrophy regions. Even though SMRI gives the structural details, identifying the MCI stage is challenging since only subtle changes exist between AD, MCI, and NC. Ternary classification can classify three classes since only a single

*For correspondence
Published online: 19 August 2023

algorithm is required to classify them. It also has the added advantage of training with both MCI and AD, which helps to learn the atrophy features common to both classes. Often, Alzheimer's disease symptoms are dismissed as part of the normal aging process [1]. We can see the difference between regular aging changes and Alzheimer's patients in the MRI slices by including Normal Control (NC). Coronal studies of MRI images can reveal the role of the amygdala and hippocampus in early disease detection. Also, it can bring forth other regions of the disease.

Machine learning techniques have become attractive for computer-assisted diagnosis as they can find the correlation between regions and automate the classification process. Thus, it is widely used to analyze neuropsychiatric disorders and diagnose them. Although several models are used in the literature, Support Vector Machine (SVM) and deep learning-based models are prominent among those that give good results for Alzheimer's detection. SVM needs hand-crafted features extracted from the images that may be included in three categories: voxel-based, vertex-based, and ROI-based methods [2]. In the voxel-based method, the features are defined at the level of an MRI voxel, and probabilities are assigned to each tissue class (gray matter, white matter, and CSF) [2]. The vertex-based method measures the cortical thickness of the regions used as features. In [3], texture features of cortical thickness in the various areas are trained using an SVM classifier to give good results. But the disadvantage is the need for domain expertise to find the region of interest area to extract the cortical thickness. The ROI-based method uses segmentation of the hippocampus, Amygdala, or any other regions relevant to the disease to measure volume, shape, and texture and uses any machine learning classifier to classify the features. In [4] SVM classifier predicts dementia based on the volume, shape, and texture features. Combined volume and texture gave the maximum Area Under the Curve value, and texture can be used to predict the earlier stage of AD, reflecting pathological changes of dementia such as neurofibrillary tangles and amyloid-beta plaques. Even though the SVM classifier performs well, it is criticized for performing poorly on raw data and needs good feature engineering to extract informative features.

Deep learning is an end-to-end learning model where the feature extraction and classification are automated by the network itself [5]. The primary advantage of end-to-end learning is optimizing all steps in the processing pipeline, leading to optimal performance. Early deep learning models [6] needed feature extraction before feeding the one-dimensional feature vector to the network. Also, it may flatten 2 Dimensional (2D) and 3 Dimensional (3D) images into one single vector to learn and classify. Handcrafted feature extraction leads to data scarcity and high dimensionality, and the need for feature selection throws off many features [7]. Convolutional Neural Networks (CNN) is an end-to-end, powerful deep learning method where spatial relationships of the image are utilized, and features are learned automatically from the image. With a successful implementation of AlexNet [8] for natural image classification, CNN has expanded its application to diverse fields. CNN is famous for 2D images application, and now it has extended to 3D images like MRI. In AD diagnosis, the paper [9] uses 2D image slices of MRI to feed into the Inception v4-based architecture, which consists of 487 layers to classify AD and NC. The ensemble learning method combines CNN in [10] to learn features using 2D CNN from sagittal, coronal, and axial planes. 2D images are readily available in clinical settings. The 3D images can provide more information than 2D images, but the computational load and high dimensionality lead to less use of such images in clinical settings [9]. A 3D CNN pretrained with a 3D convolutional autoencoder is used to perform binary and three-way classification [11]. To reduce the complexity of 3D CNN architecture and increase the feature learning capability, a long-range dependency mechanism that uses ResNet as a backbone is used in [12] to detect the MCI stage. The complexity is reduced using a P3D block, which decouples a 3D convolution into two 2D convolution operations. Some work [13, 14] use patch-based methods to reduce the complexity, but the image size is still big. Therefore, the 2D image-based methods are preferable over the 3D image for reducing the complexity and learning features at higher levels of depth. Texture-based features can enhance efficiency, especially for early AD diagnosis [4, 15]. Both works have used Hippocampus texture features to classify AD and MCI. But both works used handcrafted features using a time-consuming segmentation method, which needs domain expertise. CNN with wavelet-pooling is proposed in [16, 17], which can perform texture-based image classification where texture content is mainly preserved and able to provide a reduction in size without any translational invariance. wavelet-pooling also increases computational efficiency and regulates the overfitting of the network [18]. AD involves the degeneration of different brain parts. The structures affected are of varying scales, and multi-scale networks can extract ROI at different scales, which are local and global [19]. Further, to reduce the computational complexity and trainable parameters and increase the feature learning ability, a multi-scale and multi-path ensemble method is proposed in [20]. By taking advantage of the texture-based feature and multi-scale and multi-path methods, we propose a multi-scale multi-path wavelet-pooling-based method to perform a three-way classification of AD, MCI, and NC. In this paper, we propose

- A coronal study of AD to diagnose the disease at an early stage and markings of biomarkers involved in the disease.
- Texture-based feature learning incorporated using wavelet-pooling.
- A comparison study of different pooling methods that showcases the capability of wavelet-pooling in AD detection.

- A multi-scale and multi-path method to enhance the early detection of the disease.

2. Methodology

2.1 *Dataset*

Data used to prepare this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI)1 database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The dataset is taken from ADNI 1 and ADNI 2 randomly. 200 NC from ADNI 1 and 100 from ADNI 2. 300 MCI from ADNI 1. 170 AD from ADNI 1, and 130 from ADNI 2. A total of 900 subjects with T1-weighted MRI image scans were downloaded, of which 300 are AD, 300 are MCI, and 300 are NC. The demographic details are shown in table 1. The preprocessing steps were conducted using freesurfer software: Motion correction, Non-uniform intensity normalization, Tailairach Transformation, Intensity Normalization, and Skull Stripping. Further, we found that all images don't lie in the same orientation. So, non-linear registration was performed using FSL software to orient with the MNI152 template and obtain an image size of 182×218×182. 30 coronal slices are extracted from the mid-temporal region of the brain MRI, which covers the entire hippocampus area. Coronal slices of NC, MCI, and AD are shown in figure 1.

2.2 *Proposed method*

This work proposes a multi-path and multi-scale wavelet-pooling-based CNN architecture to classify AD, MCI, and NC using T1-weighted MRI 2D coronal slices. There are several advantages to using 2D images over 3D images, even though 3D images have more information. As 3D images consist of a huge dimensionality, it is required to build deep layers to learn good representation for efficient classification. Still, it is infeasible due to the computational complexity, limited availability of computational resources, and GPU memory. The processing time requirement and computational resources for 2D images are lower and can be used in most clinical settings. Another advantage is that

Table 1. Demographic details of subjects.

	AD	MCI	CN
Age	75± 7.81	76.15± 6.69	75.3 ± 7.8
Gender	169 M and 131 F	206 M and 96 F	164 M and 136 F

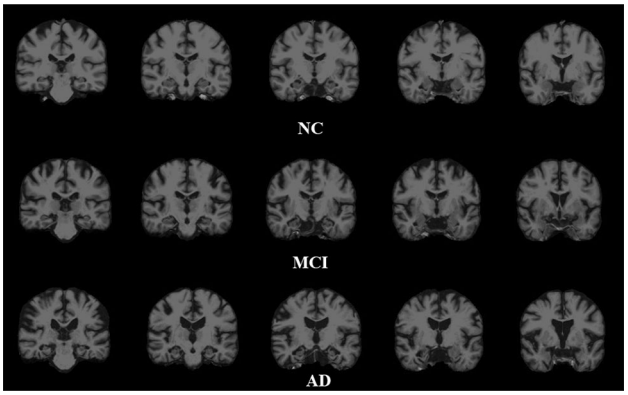


Figure 1. Samples of coronal slices after preprocessing.

such images are widely applicable in clinical settings rather than 3D images. Also, the availability of numerous standard public datasets like ImageNet, CIFAR, and classic architectures like GoogleNet, ResNet, etc., aids 2D image-based methods. We use 2D coronal slices of the Medial Temporal Lobe (MTL) area. The algorithm will learn the MTL Atrophy scale, a standard clinical measure for AD-related neurodegenerative diseases. As per the National Institute on Aging and Alzheimer's Association research guidelines [21], this scale is used as evidence for AD-related neurodegenerative disease. MTL-based detection is commonly used and more specific to AD diagnosis, even though other areas provide AD information. The other atrophy areas indicate inter-subject variability [22] and are inconsistent with AD. The algorithm learned with such patterns may confuse and lead to a decrease in accuracy. We extracted 30 slices of coronal views from the MTL starting from the hippocampus corpus area (from the level of anterior pons). We can cover the entire hippocampus area, which is the most useful information. The slices are converted to 224×224×3 standard image size by stacking the same slice three times and using zero padding. Most CNNs require high parameters, more computational load, and huge deep layers to achieve the desired output, limiting further applications. Here we use the concatenation of shallow and deep layers in a multi-path, which helps avoid the vanishing gradient problem and fuses multi-layer features. The network consists of low-rank kernels used at multi-scale levels to increase the feature extraction capability. Also, low-rank kernels compress the network with fewer parameters and space [20]. Finally, wavelet pooling enables extracting the texture features of disease-related regions.

2.3 *Proposed architecture*

As shown in Figure 2, the proposed architecture uses a 224×224×3 size image as input. It has two paths to incorporate multi-scale and multi-path computation; the primary path includes two Single-Scale Convolution Layers

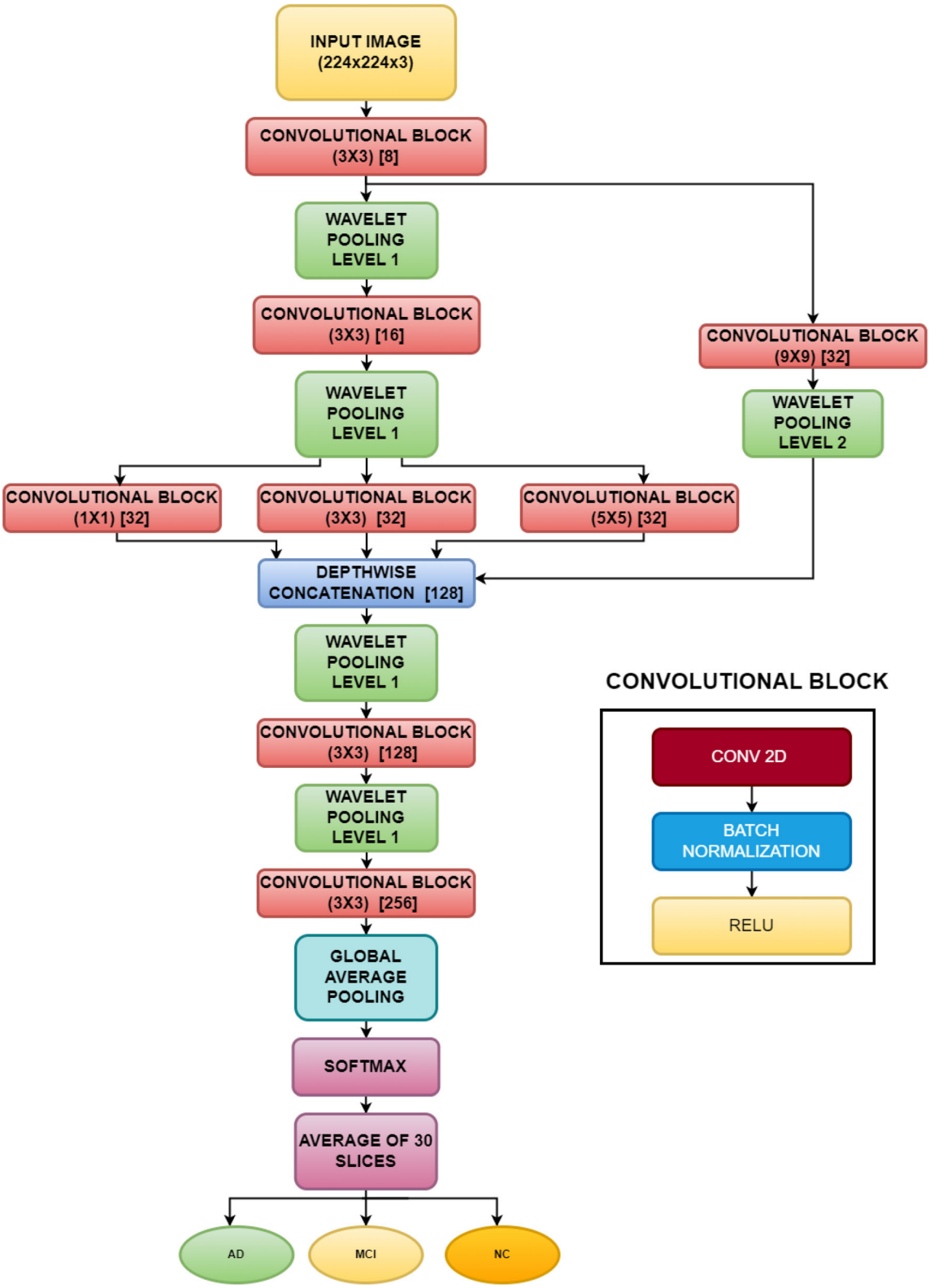


Figure 2. Schematic of proposed multi-scale and multi-path wavelet-pooling based Architecture.

(SSCL) followed by one multi-scale convolution layer and two SSCL. Finally, it is converted to a one-dimensional vector by averaging the feature maps for ternary classification. The other path includes a single-scale convolution whose features are concatenated with the multi-scale features on the primary path, which helps to learn

local and global features [23]. The multi-path also helps to avoid the vanishing gradient problem that may occur in the primary path while going through many layers.

First, an SSCL is used in the main path using 3×3 size with eight filters to give eight feature maps. The second SSCL provides 16 feature maps using 3×3 size with 16 filters. Subsequent low-rank kernels extract features at multiple scales using 1×1, 3×3, and 5×5 with 32 filters each to give 96 feature maps at the output. A parallel path is formed from the output of the first SSCL layer in the primary feedforward network using a 9×9 size of 32 filters, which generates 32 feature maps. The multi-path output is merged depth-wise with the multi-scale output in the primary path to produce 128 feature maps. The shortcut connection combines early and late features, extracting local and global structures from the data. Next, the combined features pass through two SSCL layers to give 128 and 256 feature maps, respectively. wavelet-pooling is included after each convolutional layer to cut short the spatial dimension by retaining the essential texture features. The primary path has one wavelet decomposition level, whereas second-level decomposition is utilized in the multi-path. Global average pooling is applied to the average of all feature maps from the layers to provide a single feature vector of size 256. Global average pooling is preferred over a fully connected layer as the former will reduce overfitting by reducing the total number of parameters in the model. It also reduces the computational complexity of the architecture. Finally, AD, MCI, and NC classifications are done using a softmax. The average of 30 slices for each subject is considered while taking the output. ReLu is the activation function used after each convolution process. It suppresses all the negative values to zero and passes the positive values. ReLu increases the speed and accuracy of the computation compared to other activation functions and exhibits more gradient shifts. Batch normalization helps to maintain the stability of the network.

2.3.1 *Advantages of multi-path structure* The multi-path connected from the first SSCL layer to the multi-scale layer output helps compensate for the loss of global features incurred while moving through a single-path convolutional layer. There is a possibility of focusing more on the hippocampus region alone in a brain MRI image since the anatomical boundaries of the hippocampus are distinct in T1-weighted MRI images [24]. It may lead to the loss of information from other regions of the MRI, which is critical for the detection of AD that is MCI. To elevate the problem of learning local patterns alone and shrinking to a particular region while going deep in a single path convolution layer, multi-path enables us to provide the features learned at the early part of layers [25–27]. Hence, increasing the learning capability of the architecture. A big kernel size of 9×9 is chosen for the multi-path to extract global structures on early AD detection. Another advantage of the parallel path is to reduce vanishing gradient [28], which is always a problem for deep learning architectures. No gradient changes towards the deeper layers will result in poor architectural training. Parallel path connections go through fewer non-linear activations, reducing the squashing of the derivatives and improving the derivative of the overall architecture.

2.3.2 *Advantages of multi-scale structure* Learning features at different scales enables the architecture to explore local patterns at different dimensions [19]. The Alzheimer's brain MRI image has different atrophy regions of various sizes and structures throughout the image. So it is challenging for a single-scale convolution kernel alone to extract complex features for early AD detection. So, in the proposed architecture, we have included a multi-scale convolutional layer to focus on sparse local atrophy regions pertaining to the disease conditions. We use three different filter sizes for this purpose. First, the 1×1 filter size enables channel-wise pooling to increase the feature maps from 16 to 32 without losing dimensions. Next 3×3 size filter learns larger local patterns related to the disease. Finally, the expanded 5×5 size filter can learn even larger local structures. The feature maps obtained in all three filters are concatenated depth-wise. Zero padding is applied to maintain the same output dimension for all the feature maps. One hundred twenty-eight feature maps of different scales are obtained through this operation, which leads to learning various local structures and is very useful for MCI detection.

2.3.3 *Wavelet-pooling* Earlier identification of people at risk of AD can be found using texture details of the hippocampus, precuneus, and Posterior Cingulate Cortex (PCC) with more accuracy than the hippocampal volume measure [29]. In [4], it is shown that not only the volume and shape of the hippocampus but also texture can be used for early prediction of the disease. So, wavelet-pooling is used in our architecture to extract texture features and improve early detection of the disease [29]. The selection of wavelet functions depends on the textural properties of the image. Daubechies wavelet can provide fine spatial frequency localization with narrow high and wide low frequencies simultaneously [30]. Besides, they are more efficient in recognizing fine structure details, and usage of overlapping windows results in capturing all changes in the pixel intensities. These characteristics of the wavelet enable us to choose an orthogonal four-tap Daubechies filter for obtaining the very subtle atrophy regions of the disease. We have experimented with the Haar wavelet, but it resulted in low performance. Haar wavelet has only two filter coefficients and does not have overlapping windows, capturing only adjacent pixel variations, so it is unsuitable for obtaining texture properties. The Daubechies-4 transform has four wavelet and scaling coefficients. Wavelet-based pooling can replace max-

pooling and average-pooling, as it can improve network efficiency by incorporating texture features of atrophy regions and aiding early detection at the MCI stage [16, 31]. wavelet-pooling reduces the feature map dimension with fewer discontinuities and artifacts, thereby enhancing the classification accuracy and improving the regularization of the network [18, 32]. This work uses 2D Discrete Wavelet Transform (DWT) to implement wavelet-pooling. The mathematical expression for DWT is given as

$$W_\psi(c,d) = \frac{1}{\sqrt{c}} \sum_n \psi^* \left(\frac{n-d}{c} \right) \tag{1}$$

where c and d represent the scaling factor and translational parameter, respectively. $\psi(.)$ is the basic wavelet function, and $\psi^*(.)$ is the conjugate function of $\psi(.)$. The 2D DWT is computed using the row-column method, in which first, the 1D DWT coefficients are used to transform the row data of the image and then perform column transform to get the transformed image. From paper [33] 2D DWT for an image $x \in R^{m \times m}$ can be defined as

$$X = W^T x W \tag{2}$$

where $\mathrm{W} = \begin{bmatrix} L \\ H \end{bmatrix}$ we get

$$X = \begin{bmatrix} LxL^T & LxH^T \\ HxL^T & HxH^T \end{bmatrix} = \begin{bmatrix} X_\phi^j & X_H^j \\ X_V^j & X_D^j \end{bmatrix} \tag{3}$$

Where X_ϕ^j is the approximation component, X_H^j, X_V^j and X_D^j are Horizontal, Vertical and Diagonal detail components of the image.

2.3.4 *Loss function* We have three classes, AD, MCI, and NC, to be classified. So we have taken categorical cross-entropy as our loss function. The categorical cross-entropy loss is shown in the equation.

$$L(w) = -\frac{1}{M} \sum_{k=1}^{M} \sum_{i=1}^{3} [\mathrm{y}_{ij}^k log(f(\mathrm{x}_{ij}^k; w))] \tag{4}$$

Where x_{ij} is the jth coronal slice of ith class of brain MRI and y_{ij} is the true label of the slice. M corresponds to the total number of samples of each batch, and w is the weight parameter. During the testing and validation process, the average probability of all slices of each subject (x_{i1}, x_{i2}, x_{i3},....,x_{i30}) is taken into consideration for diagnosing AD, MCI, and NC.

3. Results

A multi-path and multi-scale architecture with wavelet-pooling is proposed to classify AD, MCI, and NC. The algorithm's performance is evaluated using the ADNI dataset by a 10-fold cross-validation method. The dataset is split as per the patient level, and the output is taken by averaging the thirty slices which belong to each subject. A total of 900 subject MRI scans are used for training and testing (300 AD, 300 MCI, and 300 NC). 30 coronal slices from the mid-temporal lobe of each subject are extracted and resized into 224×224×3. So there are, in total, 27000 images. Each slice is labeled AD, MCI, or NC; it is averaged for each subject while obtaining results. We validated the algorithm's performance using average weighted accuracy, obtained after ten-fold cross-validation. The accuracy is computed by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Further, a detailed performance evaluation is validated using the metrics precision, recall, and F1 score, as given below.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{8}$$

The proposed method has an average accuracy of 96.52% with only 0.484 million learnable parameters, which outperforms most existing works. The validation accuracy for the ternary classification of the proposed model with different pooling techniques is demonstrated in table 2. The Daubechies-4 wavelet achieved a maximum accuracy of 96.52%. The wavelet-pooling techniques can perform better than the max-pooling and average-pooling methods. The detailed performance study is shown in table 3 with other pooling techniques to show wavelet-pooling efficiency using the Daubechies-4 wavelet in classifying AD, MCI, and NC. The precision, recall, and F1-score obtained for the proposed method with Daubechies-4 wavelet are 0.967, 0.966, and 0.966, respectively. The high precision and recall values determine the model's ability to discriminate among the three classes. The high recall value of MCI, i.e., 0.957, determines the potential of the model to perform early prediction of the disease without much error.

The high precision and recall values for MCI claim that this architecture is good for early diagnosis of the disease. Performance comparison with other pooling techniques is

Table 2. Validation accuracy for ternary classification.

Pooling method	Mean accuracy(%)(SD) (AD, MCI, NC)
db4 wavelet-pooling	96.52 (0.8)
Haar wavelet-pooling	94.6 (1.1)
Average-pooling	94.5 (2.2)
Max-pooling	91.8 (1.8)

also shown in Table 3, which confirms the feature extraction capability of wavelet-pooling using the Daubechies-4 (db4) wavelet. Compared to the Haar wavelet, db4 gives more results because it uses overlapping windows to average more pixel intensities than the Haar wavelet to reflect all changes between pixels. The db4 wavelet has four wavelet coefficients, whereas Haar has only two. Max-pooling and average-pooling are simple to implement and take less computational load but depend on neighborhood subsampling. This leads to loss of information and introduces discontinuities and artifacts, resulting in less classification accuracy [18], as visible in table 2.

4. Discussions

The proposed work utilizes multi-scale, multi-path, and texture features to efficiently classify AD, MCI, and NC. The feature extraction ability of multi-scale and fusing of multilayer features using multi-path mode helps to learn local and global features relevant to disease [25]. The vanishing gradient problem is also solved using multi-path mode. Low-rank 1×1, 3×3, and 5×5 kernels replace large-rank kernel computations with lesser parameters. A comparison is made in table 4 with existing standard architectures. Each model pretrained with the ImageNet dataset is used to train with the MRI dataset. The table shows that the proposed model performs better than any other standard architecture because of multi-scale and multi-path techniques with wavelet-pooling. Training big architectures and getting efficient output with a limited dataset is challenging. The multi-scale feature extraction and multi-path fusing of early features enable the model to perform well with fewer parameters and without going into much deeper layers. It is fast and easy to train the architecture from scratch and takes only 25 epochs to converge to the optimum global value.

Another Comparison is made with existing works in table 5 which shows that our method with 2D images and 2D CNN architecture can achieve higher accuracy for ternary classification than with 3D images and 3D CNN-based method. This is because of the architecture efficiency and selection of slices from maximum atrophy regions. Selection of coronal slices from the medial temporal lobe area contributes to a significant portion of AD which covers most of the hippocampus area. So the best way to assess the

Table 3. Performance comparison of different pooling methods.

Pooling method	Parameters	AD mean (SD)	MCI mean (SD)	NC mean (SD)	Weighted average (SD)
db4 wavelet-pooling	Precision	0.968 (0.01)	0.968 (0.01)	0.964 (0.01)	0.967 (0.008)
	Recall	0.969 (0.01)	0.957 (0.02)	0.971 (0.01)	0.966 (0.009)
	F1-score	0.969 (0.009)	0.963 (0.008)	0.967 (0.008)	0.966 (0.009)
Haar wavelet-pooling	Precision	0.945 (0.02)	0.945 (0.02)	0.954 (0.03)	0.948 (0.01)
	Recall	0.957 (0.02)	0.936 (0.03)	0.942 (0.02)	0.95 (0.01)
	F1-score	0.946 (0.02)	0.941 (0.01)	0.946 (0.015)	0.95 (0.013)
Average-pooling	Precision	0.95 (0.04)	0.942 (0.05)	0.96 (0.03)	0.952 (0.016)
	Recall	0.96 (0.02)	0.932 (0.07)	0.944 (0.05)	0.944 (0.02)
	F1-score	0.953 (0.01)	0.933 (0.03)	0.95 (0.02)	0.946 (0.02)
Max-pooling	Precision	0.938 (0.03)	0.9 (0.04)	0.927 (0.03)	0.925 (0.01)
	Recall	0.911 (0.04)	0.928 (0.03)	0.916 (0.04)	0.919 (0.01)
	F1-score	0.925 (0.02)	0.916 (0.02)	0.919 (0.02)	0.92 (0.01)

Table 4. Comparison of the proposed model with existing Standard networks on ADNI Dataset.

Methods	Precision (Mean)	Recall (Mean)	F1-score (Mean)	Accuracy(Mean)	Parameters (Millions)
Proposed model	0.967	0.966	0.966	0.9652	0.484
VGG16 [35]	0.884	0.874	0.894	0.877	15.04
Xception [36]	0.815	0.815	0.815	0.813	21.9
ResNet50 [37]	0.88	0.87	0.876	0.873	24.7
NASNetMobile [38]	0.756	0.742	0.751	0.746	4.87
EfficientNetB0 [39]	0.933	0.926	0.926	0.926	4.7
DensNet121 [40]	0.83	0.82	0.82	0.821	7.03

Table 5. Comparison with existing works.

Methods	Accuracy	Modality	Techniques used
Janani *et al* [5]	79	MRI, Genetic and clinical	3D CCN and Autoencoder
Juan Song et.al [41]	74.54	MRI+PET	Fusion using 3D CNN multi-scale
Karasawa *et al* [42]	87	MRI	3D CNN with 39 layers
Hosseini-Asl E [11]	89.1	MRI	3D CNN with 3DCAE
Adrien Payan [43]	85.53	MRI	2DCNN
H.Lei *et al* [44]	85.30	MRI	Multitask Sparse Low rank learning
Billone *et al* [45]	91.85	MRI	DemNet 2DCNN
Proposed method	96.52	MRI	Multi-scale and multi-path 2D CNN

Table 6. Comparison of the computational cost with pre-trained networks.

Model	Trainable parameters	No. of FLOPs (BFLOPs)	GPU Memory Requirement(GB)	Memory Required by Model Weights (MB)
VGG16 [35]	138,357,544	15.50	1.6186	57.38
Xception [36]	23,851,784	11.00	1.6034	87.42
ResNet 50 [37]	25,636,712	3.8	9.2258	226.92
NASNetMobile [38]	1,115,139	0.749	8.263	83.83
EfficientNetB0 [39]	7,21,923	0.3066	5.9924	18.2018
DenseNet121 [40]	5,90851	0.0567	11.910	29.099
Proposed method	**4,84,595**	**0.0233**	**1.5075**	**1.8486**

Highlighted the results of proposed method

atrophy of the medial temporal lobe is by taking coronal slices of T1-weighted MRI [34]. Also, 2D images are more applicable in clinical settings than 3D images due to their availability. This justifies using 2D images of the brain for doing this work.

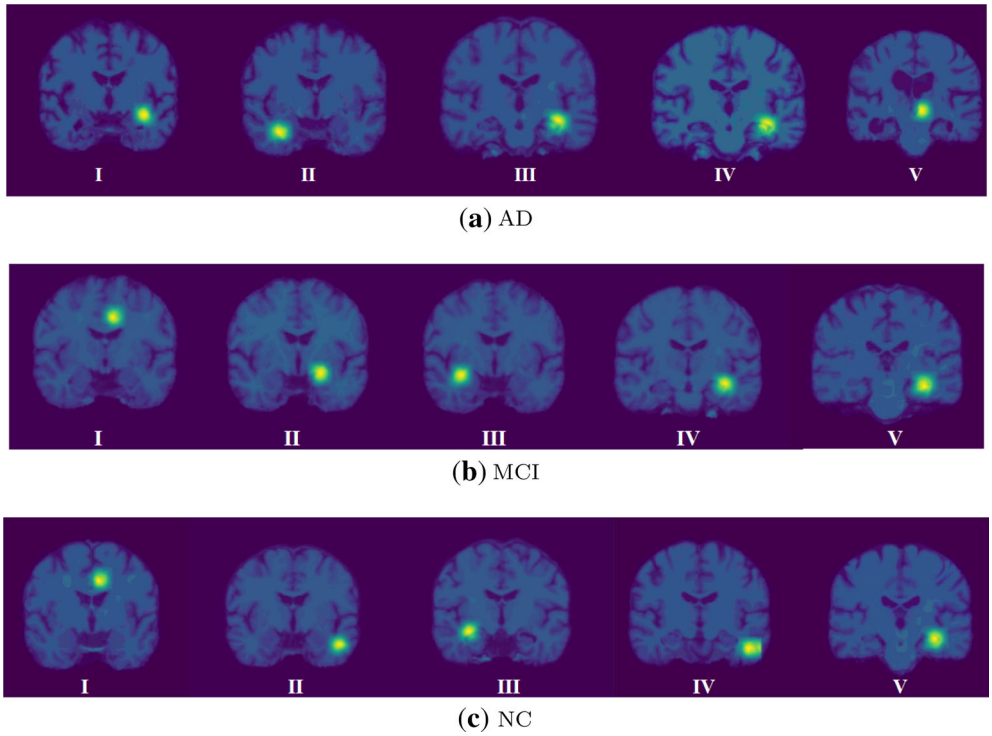
4.1 *Computational cost of the proposed network with existing pre-trained networks*

The computational cost of the proposed method is compared with pre-trained standard architectures in table 6. Our approach has low computational costs regarding parameters, computational time, and memory utilization. To evaluate the computational load with pretrained networks, we have considered total learnable parameters, Floating Point Operations Per Second (FLOPs), GPU memory, and Model weight memory. The complexity of the network architecture is determined by its trainable parameters; the more parameters, the more complexity. As the table shows, our model requires significantly fewer parameters when compared to VGG16, Xception, ResNet50, NASNetMobile, EfficientNetB0, and DenseNet models. The proposed network comprises only 4,84,595 learnable parameters, decreasing the model complexity. The CNN execution time is measured using FLOPS, which tells the floating point operations required by the system. The proposed method has floating point operations of 0.0233 BFLOPs, which is lower than all standard networks. The computational resource utilization is measured using GPU memory and memory for model weights. The proposed method utilizes only 1.50 GB of GPU memory, much less than all others. The model weight memory is substantially less, with 1.8486 MB.

4.2 *Grad-CAM for qualitative analysis*

Gradient-weighted Class Activation Mapping (Grad-CAM) describes the internal architecture of the proposed architectural model by attention-weighted visualization to localize relevant image regions in the feature maps [13, 46–48].

Figure 3 describes the major brain-affected AD, MCI, and NC regions in coronal slices. Thirty coronal slices of the mid-temporal region were selected for AD, MCI, and NC classification. The highlighted areas of heat maps show that the hippocampus and amygdala areas are the most affected part of the brain, and it helps discriminate between the classes. In figure 3a, Slice No. I is cut through the



(**a**) AD

(**b**) MCI

(**c**) NC

Figure 3. Saliency map.

anterior columns of the fornix, and the highlighted part shows the affected right amygdala area. The highlighted part also shows the enlargement of the temporal horn region of the lateral ventricles. It confirms with the existing references of AD diagnosis that the enlargement of the lateral ventricular region increases more while progressing to AD than NC. In Slice No. II, the left hippocampus is seen highlighted. Slices III and IV indicate right hippocampus has been affected. Finally, slice V reflects the thalamus area as an atrophy region. In AD, the hippocampus region is the most discriminated region. The atrophy caused due to the hippocampus volume reduction by 15–30% than an NC [34] will result in a significant change in tissue characteristics which help in texture classification of AD. Since there is a drastic change in the volume and thickness of the hippocampus region, the central affected region is highlighted in the AD coronal slices. The hippocampus structure is easier for automated algorithms to identify than the amygdala, entorhinal cortex, or parahippocampal gyrus, as the anatomical boundaries of the hippocampus are distinct on T1 weighted MRI.

Figure 3b represents the visualization of Grad-CAM for MCI individuals. In the case of MCI, we can see that the left and right amygdala is highlighted in slice no. II and III show that the amygdala texture properties can be used for early disease prediction. Amygdala is affected in the early stage, due to which neuropsychiatric problems are prevalent in the mild stage of AD [5]. Slice No. I give the cingulate gyrus the highlighted region. The anterior cingulate gyrus is part of the motivation, attention, and behaviour network and is associated with the amygdala. The posterior cingulate gyrus is part of the posterior network of learning and memory and is related to the Hippocampus. Slice No. IV and V indicate the hippocampus region as the discriminating region. So in MCI, both amygdala and Hippocampus play a vital role in determining the stage, and the cingulate gyrus also aids in prediction. Figure 3c shows the heat map of NC, which reveals that the Hippocampus and amygdala are the main atrophy regions that help discriminate the three classes.

5. Conclusion

This study proposed a novel architecture for the ternary classification of AD, MCI, and NC using 2D MRI images. A multi-scale and multi-path architecture with wavelet-pooling is proposed to increase the feature extraction ability, prevent vanishing gradient problems, and reduce the computation load. The main path has multi-scale feature extraction, and the parallel path fuses the early features with multi-scale features to learn local and global features. The wavelet pooling ensures the extraction of the texture features of the MCI and AD atrophy regions. The Daubechies-4 wavelet gave the maximum efficiency for the architecture due to its orthogonal and compact nature.

wavelet-pooling can replace Max-pooling and Average-pooling, especially for medical image applications, as it increases the efficiency of the network regardless of computational load. The heat map confirms the hippocampus and amygdala areas as the most affected regions, which helps to discriminate the three classes and aids in the early prediction of the disease. The high accuracy, precision, and recall values ensure that the 2D coronal slices can be used for early AD detection without depending on 3D images. In the future, we aim to use a multi-modal approach to learn the early prediction of AD.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health www.fnih.org. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Declarations

Conflict of interest As authors hereby declare that we have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Porsteinsson A P, Isaacson R S, Knox S, Sabbagh M N and Rubino I 2021 Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021. *J. Prev. Alzheimer's Dis.* 8(3):371–386. https://doi.org/10.14283/jpad.2021.23

[2] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert M O, Chupin M, Benali H and Colliot O 2011 Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *Neuroimage* 56(2): 766–781

[3] Wang S L, Cai Z C and Xu C L 2013 Classification for Alzheimer's disease based on SVM using a spatial texture feature of cortical thickness. In: *2013 10th International Computer Conference on Wavelet Active Media Technology and Information Processing ICCWAMTIP* pp. 158–161. https://doi.org/10.1109/ICCWAMTIP.2013.6716622

[4] Achterberg H C, Sørensen L, Wolters F J, Niessen W J, Vernooij M W, Ikram M A, Nielsen M and de Bruijne M 2019 The value of hippocampal volume, shape, and texture for 11-year prediction of dementia: a population-based study. *Neurobiol Aging* 81: 58–66

[5] Oh K, Chung Y C, Kim K W, Kim W S and Oh IS 2019 Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. *Sci. Reports 2019 91* 9(1): 1. https://doi.org/10.1038/s41598-019-54548-6

[6] Raju M, Sudila T V, Gopi V P, Anitha V S 2020 Classification of Mild Cognitive Impairment and Alzheimer's Disease from Magnetic Resonance Images using Deep Learning. In: *2020 IEEE International Conference on Recent Trends in Electronics, Information &. Communication Technology RTEICT-2020* p. 52. https://doi.org/10.1109/RTEICT49044.2020.9315695

[7] Liu S, Liu S, Cai W, Pujol S, Kikinis R and Feng D 2014 Early diagnosis of Alzheimer's disease with deep learning. In: *2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014*, Institute of Electrical and Electronics Engineers Inc., pp. 1015–1018. https://doi.org/10.1109/isbi.2014.6868045

[8] Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges C, Bottou L, Weinberger K (eds) *Advances in Neural Information Processing Systems*, Curran Associates, Inc., vol 25. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[9] 2020 Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. *Sci. Rep.* 10(1): 1–10. https://doi.org/10.1038/s41598-020-79243-9

[10] Pan D, Zeng A, Jia L, Huang Y, Frizzell T and Song X 2020 Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. *Front. Neurosci.* 14(May): 1–19. https://doi.org/10.3389/fnins.2020.00259

[11] Hosseini-asl E, Keynton R, El-baz A, Drzezga A, Lautenschlager N, Siebner H, Riemenschneider M, Willoch F, Minoshima S, Schwaiger M and Kurz A 2016 Alzheimer's Disease Diagnostics by Adaptation of 3D Convolutional

Network Electrical and Computer Engineering Department, University of Louisville, Louisville, KY, USA. *Eur. J. Nucl. Med. Mol. Imaging* 30(8): 1104. https://doi.org/10.1007/s00259-003-1194-1

[12] Pei Z, Gou Y, Ma M, Guo M, Leng C, Chen Y and Li J 2021 Alzheimer's disease diagnosis based on long-range dependency mechanism using convolutional neural network. *Multimed. Tools Appl.*. https://doi.org/10.1007/s11042-021-11279-z

[13] Raju M, Gopi V P, Anitha V S and Wahid K A 2020 Multi-class diagnosis of Alzheimer's disease using cascaded three-dimensional-convolutional neural network. *Phys. Eng. Sci. Med.* 43(4): 1219. https://doi.org/10.1007/s13246-020-00924-w

[14] Liu M, Cheng D, Wang K and Wang Y 2018 Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Neuroinformatics* 16(3–4): 295–308. https://doi.org/10.1007/s12021-018-9370-4

[15] Dolph C V, Alam M, Shboul Z, Samad M D and Iftekharuddin K M 2017 Deep learning of texture and structural features for multiclass Alzheimer's disease classification. *Proceedings of the International Joint Conference on Neural Networks* 2017-May(1310353):2259–2266. https://doi.org/10.1109/IJCNN.2017.7966129

[16] Fortuna-Cervantes J M, Ramírez-Torres M T, Mejía-Carlos M, Murguía J S, Martinez-Carranza J, Soubervielle-Montalvo C and Guerra-García C A 2022 Texture and Materials Image Classification Based on Wavelet Pooling Layer in CNN. *Appl. Sci.* 12(7): 1–26. https://doi.org/10.3390/app12073592

[17] Fujieda S, Takayama K and Hachisuka T 2017 Wavelet Convolutional Neural Networks for Texture Classification. http://arxiv.org/abs/1707.07394, 1707.07394

[18] Williams T and Li R 2018 Wavelet pooling for convolutional neural networks. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings* pp. 1–12

[19] Liu Z, Lu H, Pan X, Xu M, Lan R and Luo X 2022 Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network. *Knowledge-Based Syst.* 238: 107942

[20] Wang X, Bao A, Lv E and Cheng Y 2020 Multiscale Multipath Ensemble Convolutional Neural Network. *IEEE Trans. Syst. Man. Cybern. Syst.* p. 1. https://doi.org/10.1109/tsmc.2020.2972695

[21] 2011 The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement* 7(3): 270. https://doi.org/10.1016/J.JALZ.2011.03.008

[22] Pini L, Pievani M, Bocchetta M, Altomare D, Bosco P, Cavedo E, Galluzzi S, Marizzoni M and Frisoni G B 2016 Brain atrophy in Alzheimer's Disease and aging. *Ageing Res. Rev.* 30: 25. https://doi.org/10.1016/J.ARR.2016.01.002

[23] Thomas A, Harikrishnan P, Ramachandran R, Ramachandran S, Manoj R, Palanisamy P and Gopi V P 2021 A novel multiscale and multipath convolutional neural network based age-related macular degeneration detection using oct images. *Comput. Methods Progr. Biomed.* 209: 106294

[24] van der Lijn F, Den Heijer T, Breteler M M and Niessen W J 2008 Hippocampus segmentation in mr images using atlas registration, voxel classification, and graph cuts. *Neuroimage* 43(4): 708–720

[25] Thomas A, Harikrishnan P, Krishna A K, Palanisamy P and Gopi V P 2021 A novel multiscale convolutional neural network based age-related macular degeneration detection using oct images. *Biomed. Signal Process Control* 67: 102538

[26] Nisha J, Gopi V P and Palanisamy P 2022 Automated colorectal polyp detection based on image enhancement and dual-path cnn architecture. *Biomed. Signal Process Control* 73: 103465

[27] Thomas A, Harikrishnan P, Krishna A K, Ponnusamy P and Gopi V P 2021 Automated detection of age-related macular degeneration from oct images using multipath cnn. *J. Comput. Sci. Eng.* 15(1): 34–46

[28] Alzubaidi L, Zhang J, Humaidi A J, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel M A, Al-Amidie M and Farhan L 2021 Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *J. Big Data* 8(1): 1–74

[29] Lee S, Lee H and Kim K W 2020 Magnetic resonance imaging texture predicts progression to dementia due to Alzheimer disease earlier than hippocampal volume. *J. Psychiatry Neurosci.* 45(1): 7–14

[30] Al-Kadi O S, Chung D Y, Carlisle R C, Coussios C C and Noble J A 2015 Quantification of ultrasonic texture intra-heterogeneity via volumetric stochastic modeling for tissue characterization. *Med. Image Anal.* 21(1): 59–71

[31] Tao Z, Wei T and Li J 2021 Wavelet multi-level attention capsule network for texture classification. *IEEE Signal Process. Lett.* 28: 1215–1219. https://doi.org/10.1109/LSP.2021.3088052

[32] Rossetto A M and Zhou W 2019 Improving classification with CNNs using wavelet pooling with nesterov-accelerated adam. *Proceedings of 11th International Conference on Bioinformatics and Computational Biology, BiCOB 2019* 60: 84–93, https://doi.org/10.29007/9c5j

[33] Li Q, Shen L, Guo S and Lai Z 2020 Wavelet integrated cnns for noise-robust image classification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*

[34] Frisoni G B, Fox N C, Jack C R, Scheltens P and Thompson P M 2010 The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* 6(2): 67. https://doi.org/10.1038/nrneurol.2009.215

[35] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition http://www.robots.ox.ac.uk/, 1409.1556v6

[36] Chollet F 2017 Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258

[37] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778

[38] Zoph B, Vasudevan V, Shlens J and Le Q V 2018 Learning transferable architectures for scalable image recognition. In:

Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697–8710

[39] Tan M and Le Q 2019 Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*, PMLR, pp. 6105–6114

[40] Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T and Keutzer K 2014 Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869

[41] Song J, Zheng J, Li P, Lu X, Zhu G and Shen P 2021 An Effective Multimodal Image Fusion Method Using MRI and PET for Alzheimer's Disease Diagnosis. *Front. Digit. Heal.*19. https://doi.org/10.3389/FDGTH.2021.637386

[42] Karasawa H, Liu C L, Ohwada H 2018 Deep 3D Convolutional Neural Network Architectures for Alzheimer's Disease Diagnosis. In: *Lecture Notes in Computer Science (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol **10751**, p. 287, https://doi.org/10.1007/978-3-319-75417-8-27

[43] Payan A and Montana G 2015 Predicting Alzheimer's disease a neuroimaging study with 3D convolutional neural networks. In: *ICPRAM 2015 - 4th The International Conference on Pattern Recognition Applications and Methods Proceedings* 2: 355, arXiv:1502.02506v1

[44] Lei H, Zhao Y and Lei B 2019 Predicting early stages of neurodegenerative diseases via multi-task low-rank feature learning. *Lecture Notes in Computer Science (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 11767: 131. https://doi.org/10.1007/978-3-030-32251-9-15

[45] Billones C D, Demetria O J L D, Hostallero D E D and Naval P C 2016 Demnet: A convolutional neural network for the detection of alzheimer's disease and mild cognitive impairment. In: *2016 IEEE Region 10 Conference (TENCON)*, pp. 3724–3727, https://doi.org/10.1109/TENCON.2016.7848755

[46] Raju M, Gopi V P and Anitha V S 2021 Multi-class Classification of Alzheimer's Disease using 3DCNN Features and Multilayer Perceptron. In: *2021 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* 2021: 368. https://doi.org/10.1109/WISPNET51692.2021.9419393

[47] Raju M, Sudila T, Gopi V P, Anitha V 2020 Classification of mild cognitive impairment and alzheimer's disease from magnetic resonance images using deep learning. In: *2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, IEEE, pp. 52–57

[48] Sunija A, Gopi V P and Palanisamy P 2022 Redundancy reduced depthwise separable convolution for glaucoma classification using oct images. *Biomed. Signal Process Control* 71: 103192