






Multiresolutional ensemble PartialNet for Alzheimer detection using magnetic resonance imaging data

Imran Razzak¹  | Saeeda Naz²  | Abida Ashraf² |
Fahmi Khalifa³  | Mohamed Reda Bouadjenek¹  |
Shahid Mumtaz⁴ 

¹School of Information Technology,
Deakin University, Geelong, Victoria,
Australia

²Department of Computer Science,
Govt Girls Postgraduate College No. 1,
Abbotabad, HED, KP, Pakistan

³Electronics and Communications
Engineering, Mansoura University,
Mansoura, Egypt

⁴Instituto de Telecomunicações Aveiro,
Aveiro, Portugal

Correspondence

Imran Razzak, Level 5, Building KA,
School of Information Technology,
Deakin University, Waurn Ponds,
Geelong 3217, Victoria, Australia.
Email: imran.razzak@deakin.edu.au and
imran.razzak@ieee.org

Abstract

Alzheimer's disease (AD) is an irreversible and progressive disorder where a large number of brain cells and their connections degenerate and die, eventually destroy the memory and other important mental functions that affect memory, thinking, language, judgment, and behavior. Not a single test can effectively determine AD; however, CT and magnetic resonance imaging (MRI) can be used to observe the decrease in size of different areas (mainly temporal and parietal lobes). This paper proposes an integrative deep ensemble learning framework to obtain better predictive performance for AD diagnosis. Unlike DenseNet, we present a multiresolutional ensemble PartialNet tailored to Alzheimer detection using brain MRIs. PartialNet incorporates the properties of identity mappings, diversified depth as well as deep supervision, thus, considers feature reuse that in turn results in better learning. Additionally, the proposed ensemble PartialNet demonstrates better characteristics in terms of vanishing gradient, diminishing forward flow with better training time, and a low number of parameters compared with DenseNet. Experiments

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *International Journal of Intelligent Systems* published by Wiley Periodicals LLC

performed on benchmark AD neuroimaging initiative data set that showed considerable performance gain (2 + %↑) and (1.2 + %↑) for multiclass and binary class in AD detection in comparison to state-of-the-art methods.

KEYWORDS

Alzheimer, brain disorder, DenseNet, ensemble learning, PartialNet

1 | INTRODUCTION

Alzheimer's disease (AD) is an irreversible and progressive neurological disorder that generally affects elderly people. According to recent statistics, AD is the fourth worldwide leading cause of death following cardiovascular disease, cancer, and stroke.¹ Clinically, AD is characterized by abnormal accumulation of amyloid plaques and neurofibrillary tangles in human brains, which slowly destroys memory and thinking power, and, eventually, the ability to carry out the simplest tasks. An illustration of the biomarkers that show the progress of AD disease is shown in Figure 1.* Although AD affects people with an average age above 65, if diagnosed early it is referred to as early onset of AD. In 2019, about 46 million persons worldwide suffer from AD, 56% of them was at an early stage of onset.² Symptoms of early-onset AD can include mild memory loss and trouble concentrating or finishing daily life tasks. Similar to Parkinson's there is no cure; however, its progression can be delayed through effective management plans that temporarily slow the worsening of dementia symptoms and improve the quality of life. Only definitive approach to diagnose AD is to examine the brain tissue after one's death. However, physicians can use other clinical examinations and tests to assess patients' mental abilities, diagnose dementia, and rule out other conditions. Additionally, specialized brain imaging, such

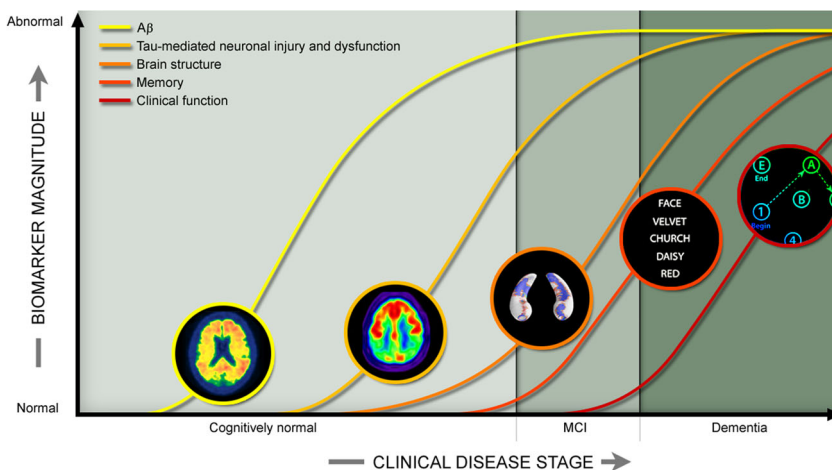


FIGURE 1 Graphical depiction of biomarkers as indicators of Alzheimer's disease. MCI, mild cognitive impairment [Color figure can be viewed at wileyonlinelibrary.com]

as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET), is widely used for AD diagnosis and follow-up of disease progression,^{3,4} see example in Figure 2.

In recent years, extensive research interest has grown exponentially toward computer-aided diagnosis (CAD) tools for brain disorder diagnosis.^{6–8} Particularly, machine learning (ML) approaches have been extensively explored to improve AD diagnosis and its prodromal dementia stage, mild cognitive impairment (MCI), from normal controls (NC). Literature methods can be generally categorized into four major categories: region-of-interest (ROI)-based methods, voxel-based methods, patch-based methods, and methods based on the whole image as input. ROI-based methods are confined to a coarse-scale limit region, thus they may ignore the important fine-scaled information within the region. On the other hand, voxel-based methods are prone to overfitting due to high-dimensional data. While patch-based methods are often desired, they ignore brain representation and focus on fixed-size patches. Finally, methods based on the whole image are unable to identify the subtle disease progression in the brain structures through changes with time. Leveraging the trade-off between local and global representations may, therefore, better help understand the progression of the disease, while not overemphasizing the one aspect only.

In this study, we develop an integrative deep ensemble learning framework to obtain better predictive performance. The proposed framework is based on ensemble PartialNet learning that incorporates a deep multiresolutional ensemble PartialNet, which possesses the properties of identity mappings and diversified depth. Additionally, the proposed pipeline integrates deep supervision and transition blocks to provided better feature representation. To further improve the gradient propagation and information flow between layers, we utilize a partial connectivity pattern by connecting some of its subsequent layers but not all. The following summarizes our *major contributions*:

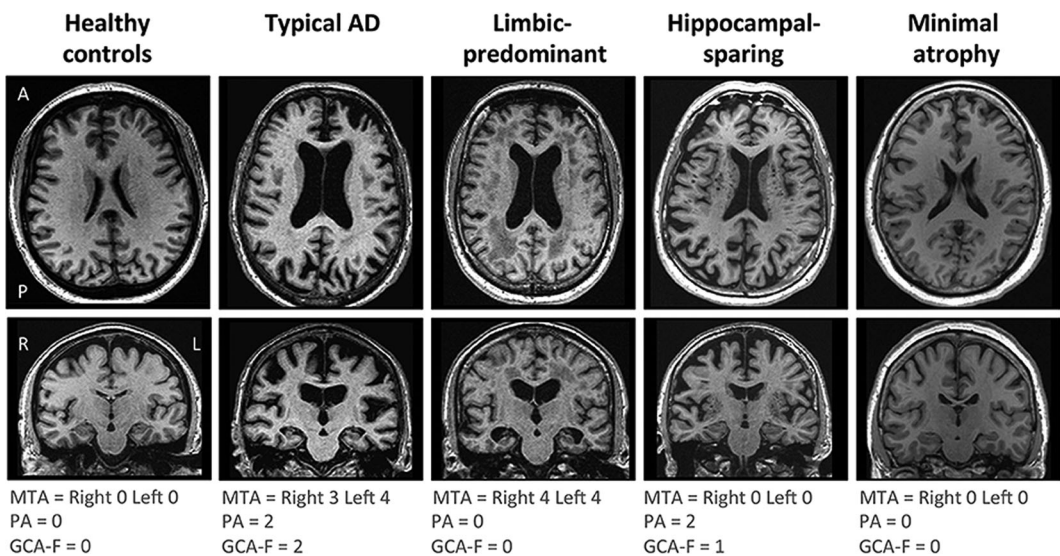


FIGURE 2 Visual examples of the brain atrophy patterns in the different AD subtypes.⁵ Note that GCA-F, MTA, and PA stand for medial temporal atrophy, posterior atrophy, and global cortical atrophy-frontal, respectively. AD, Alzheimer's disease

- Unlike DenseNet that utilizes each layer in convolution neural network to all preceding layers, PartialNet considers the feature reuse efficiently by limiting the connection to its preceding layers, which in turn results in better learning.
- We utilize a block concept that limits the skip connection within a block and utilizes a multiresolution ensemble; hence, every block behaves like a unique structure and contributes more to gradient magnitude than a deeper network.
- Supervision and transition blocks are exploited in each PartialNet block to help in intermediate features learning.
- Besides, we bound the path in each block as well as limit the number of dense connections; hence, gradient flow is enhanced when compared with a partially dense path.
- Both two-level or binary (i.e., AD–MCI, MCI–CN, and AD–CN) and multiclass (AD, MCI, and CN) classifications have been studied and evaluated using Alzheimer's disease neuroimaging initiative (ADNI) data set that showed a considerable gain in performance in comparison to benchmark methods.

The remainder of the paper is organized into five sections. In Section 2, an overview of the literature's related work is presented. This is followed by the details and description of the proposed multiresolutional ensemble PartialNet framework in Section 3.2. Experimental settings and results are fully discussed in Section 4. Results discussions, observations, and limitations are provided in Section 5. Finally, Section 6 provides conclusions and future directions.

2 | RELATED WORK

AD is a neurodegenerative disorder and its diagnosis at an early stage is of immense importance. A great deal of research work for early-stage AD diagnosis or prognosis has been developed in recent years. This has been supported by recent advances in ML approaches, especially the deep learning (DL) techniques.^{9–11} This section provides a review of different ML- and DL-based classification methods employed in this direction.

Particularly, Suk et al.¹² developed a model for AD detection using hierarchical feature representation and multimodel (MM) fusion. Their study utilized both MRI and PET scans of 398 subjects selected from ADNI data set: 93 AD, 104 MCI, and 101 NC. The MR images were preprocessed to remove gradient nonlinearity and b1 field in homogeneity. The PET images were preprocessed to intensity normalization, spatially aligned, and smoothed. Authors have utilized feature learning at patch level and presented MM Deep Boltzmann Machine (DBM) and utilize the densities tissues in MRI image patch as well as voxel intensities of patch from PET image. Latent and hierarchical features from the trained MM DBM are extracted for the paired patches and then supplied to the multilevel restricted Boltzmann machine (RBM) classifier. The achieved accuracy, sensitivity, and specificity of AD versus cognitive normal (CN) were 95.35%, 94.65%, and 95.22%, respectively, and 85.67%, 95.37%, and 65.87%, respectively, for MCI versus CN.

Sarraf et al.¹³ introduced a DL framework for the Alzheimer diagnosis from healthy and Alzheimer patients. The study is conducted on 28 Alzheimer and 15 healthy subjects collected from ADNI data set. Extensive preprocessing methods such as spatial smoothing, motion correction, skull stripping, noise removal, and registration are applied to improve the quality of input data. After preprocessing, data were passed to the DL model called LeNet, which achieved an accuracy of 96.85%. Similarly, Mathew et al.¹⁴ considered a subset of ADNI database (151 MRI images from patients, including 71 NC and 87 AD) and applied several preprocessing methods, such as image

cropping, resize, normalization, and reorientation. Principal component analysis (PCA) and discrete wavelet transform (DWT) are used to extract features followed by classification using support vector machine (SVM) and achieved an accuracy of 91% and 84% for MCI versus CN, and AD versus CN, respectively. In another work, Iftikhar and Idris proposed an ensemble classifier to differentiate MCI and AD patients.¹⁵ Volumetric cortex and cortical thickness-based features are extracted and forwarded to the ensemble classifier. The study is conducted on 180 subjects (60 MCI, 60 AD, and 60 NC) and achieved a specificity of 89% and a sensitivity of 92% with an accuracy of 91.66% to differentiate AD versus MCI. A deep three-dimensional (3D) convolutional neural network (3D-CNN) by Hosseini-Asl et al.¹⁶ was proposed for AD diagnosis. The framework is based on the extraction of local features from 3D skull-stripped and spatially normalized input image using Convolutional AutoEncoder (CAE). Features dementia case study is utilized as a biomarker, and fine-tuned transfer learning is utilized to identify Alzheimer's patients on ANDI. The model training and experiments were carried out on CAD dementia data set which contains 70 CN, 70 MCI, and 70 AD. The fine-tune approach showed significantly better performance by achieving 97.6% accuracy for CN versus AD task. Recently, Ashraf et al.¹⁷ employed multiple convolution neural networks for the classification of AD, MCI, and CN. Thirteen deep transfer learning-based networks were evaluated and compared using augmented data of ADNI data set and reported the highest accuracy up to 99.05 with fivefold cross-validation using the Dens-Net model. Furthermore, the authors investigate the freeze features of multiple CNN architectures in Reference [18]. More recently, Ju et al.¹⁹ utilized multimodal data by integrating textual data (gender, age, and genetic information) along with MRI images for the diagnose of Alzheimer's patient. The study is conducted on 91 MCI and 79 NC MRI images extracted from ADNI-2 data set. Besides, MRI images, the age, gender, and genetic information are also extracted and used to find prevalence between MCI and gender, age, and ApoE. Their analysis pipeline used Data Processing and Analysis of Brain Imaging (DPABI) for preprocessing. Similarly, the study is conducted on functional MRI (fMRI) time-series data as well as correlation coefficient using SVM, logistic regression (LR), linear discriminant analysis (LDA), and autoencoder. The study showed gain in performance (accuracy/sensitivity/specificity) 67.72%/65%/66%, 71.38%/77%/62%, 78.91%/79%/64%, and 86.47%/92%/81% using LDA, LR, SVM, and autoencoders, respectively. The results showed that correlation coefficient could be used to improve the diagnosis performance. In another work, Farooq et al.²⁰ examined multiple DL techniques for multiclass classification of AD. Namely, they applied ResNet-152, GoogLeNet, and ResNet-18 for diagnosis of Alzheimer patients on ADNI data set (33, 22, 449, and 45 cases of AD, LMCI, MCI, and CN, respectively). The experiment showed that GoogLeNet showed better performance with an accuracy of 98.8%, while ResNet-152 and ResNet-18 (98.14% and 98.01%, respectively) also achieved competing performance.

Bäckström et al.²¹ described a simple, yet effective method to detect AD called 3D CNN architecture (3D ConvNet) using brain MRIs. Their framework started with cortical reconstruction, edge trimming, image resizing, and intensity normalization as preprocessing steps. Then automated features were extracted from the preprocessed images using the proposed DL technique. Experimental data were gathered from ADNI data set, namely, 340 subjects were used which include 1190 MRI scans of 199 AD patients (103 male and 96 female) and 141 NC (75 male and 66 female) and achieved 98.78% Alzheimer diagnosis accuracy.

Kazemi and Houghten²² considered fMRI images to different AD patients at different stages. Several preprocessing methods such as extraction of brain, spatial smoothing, slice timing correction, spatial normalization, high pass filtering, and image conversion were applied to improve the quality of input data. Finally, AlexNet is used for classification on 197 subjects (90 male and 107 female) to differentiate patients among five classes: CN, AD, Late Mild

Cognitive Impairment (LMCI), Early Mild Cognitive Impairment (EMCI), and subjective memory complaints (SMCs). Data split for each experiment was conducted as in Reference [21]. Overall accuracy was 97.63%, and per-class accuracies were 94.97%, 95.64%, 95.89%, 98.34%, and 94.55% for AD, EMCI, LMCI, CN, and SMC, respectively. Another approach based on transfer learning was proposed by Ebrahimi-Ghahnavieh et al.²³ to detect AD using MRIs from ADNI data set. Recurrent neural network along with CNN to better observe the association between sequences of input slices. CNN is used to extract features, and a recurrent neural network is applied to consider the relationship between slices that result in improvement of diagnostic performance. A summary of the literature related work is summarized in Table 1.

3 | METHODOLOGY

The proposed integrative deep ensemble learning framework for AD detection using MRI is schematized in Figure 3. As demonstrated in the system block diagram, our developed framework incorporates a deep multiresolutional ensemble PartialNet to obtain better predictive performance. The input to the proposed pipeline is MRI data obtained from the ADNI data sets and our system provides both binary and multiclass classification. To analyze the data, data preparation, that is, preprocessing, is required to help in targeting the ROI. Next, we describe the preprocessing steps employed in our analysis pipeline followed by full description of the deep multiresolutional ensemble approach.

3.1 | Preprocessing

In this study, we have applied several preprocessing methods to improve the performance of the proposed PartialNet. At first, we have converted the raw MR images into one-channel images of different sizes. We further performed resizing and cropping to remove the white spaces and enhance the quality of the images. We have extracted ROI to determine the extreme points in contours along with the x - and y -coordinates. Besides, we have applied several data augmentations, such as rotation (90° , 180° , and 270°), illumination, zoom in and zoom out, vertical flipping, and horizontal flipping. With the application of different data augmentations, the number of database images is increased from 3925 images to 37,590.

3.2 | Proposed PartialNet ensemble framework

Deeper the DenseNet means an exponential increase in computational and space complexity due to the increased dense block depth. In addition, it results in a much larger number of parameters. Unlike DenseNet that utilizes each layer in convolution neural network to all preceding layers, we present a deep multiresolutional ensemble PartialNet that incorporates the properties of identity mappings, diversified depth, and deep supervision, thus, considers the feature reuse, which in turn results in better learning. In addition, we have limited the structure to blocks aided with supervision and transition block in each that forces the network to consider intermediate features as well as low-level features.

Deeper layers do not contribute to gradient propagation and behave like ensembles of the same network. Besides, the ResNet and DenseNet, paths are different lengths (i.e., skip connection from

TABLE 1 Related work summary for Alzheimer's disease (AD) detection using ADNI data set

| Study, year | Preprocessing | Features | Data set detail | Classifier | Accuracy (%) |
|-----------------------------------------|-------------------------------------------------------------------------------------------|----------------------------------------------|----------------------------------------------------|---------------------------------|--------------------------------------------------------------------|
| Suk et al., ¹² 2014 | Anterior commissure, posterior commissure correction, skull stripping, cerebellum removal | Model based | MR and PET images of 93 AD, 204 MCI, and 101 CN | Deep Boltzmann machine | 95.35 (AD vs. CN), 85.67 (MCI vs. CN) |
| Sarraf et al., ¹³ 2016 | Motion correction, skull stripping, spatial smoothing, noise removal | Model based | MRI of 28 AD and 15 CN | LeNet | 96.86 (AD vs. CN) |
| Mathew et al., ¹⁴ 2016 | Cropping, skull stripping, normalization | DWT, PCA, FDR | MRI of 71 CN and 87 AD | SVM | 84 (AD vs. CN), 91.7 (MCI vs. CN) |
| Ifthikhar and Idris, ¹⁵ 2016 | Spatial correction, normalization, registration, cortical thickness, volumetric cortex | Hybrid features | MRI of 60 AD, 60 MCI, and 60 CN | Ensemble SVM | 91.66 (AD vs. MCI), 98.83 (AD vs. CN), 90.83 (MCI vs. CN) |
| Hosseini-Asl et al., ¹⁶ 2016 | Skull stripping, spatially normalization | Model based | T1 weighted MRI scans of 70 AD, 70 CN, and 70 MCI | 3D convolutional neural network | 97.6 (CN vs. AD) |
| Ju et al., ¹⁹ 2019 | Motion correction, slice timing correction, skull stripping, normalization, smoothing | AAL Atlas identifies the region-of-interests | MRI of 91 MCI and 79 CN | LDA, LR, SVM, and SAE | 67.72 (LDA), 71.38 (LR), 78.91 (SVM), and 86.47 (SAE) (CN vs. MCI) |
| Bäckström et al., ²¹ 2018 | Cortical reconstruction, trim edges, resize, normalization | Model based | 3D MRI scans of 199 AD, 141 CN, total scans = 1198 | 3D ConvNet | 98.78 (CN vs. AD) |
| Kazemi and Houghten, ²² 2018 | Slice timing correction, spatial smoothing, high pass filtering, spatial normalization | Model based | 3D MRI of 199 AD and 141 CN, total 1198 | AlexNet | 98.74 (CN vs. AD) |

Abbreviations: AAL, automated anatomical labeling; ADNI, Alzheimer's disease neuroimaging initiative; CN, cognitive normals; DWT, discrete wavelet transform; FDR, Fisher discriminant ratio; LDA, linear discriminative analysis; LR, logistic regression; MCI, mild cognitive impairment; MR, magnetic resonance; MRI, magnetic resonance imaging; PCA, Principal component analysis; PET, positron emission tomography; SAE, stacked autoencoder; SVM, support vector machine.

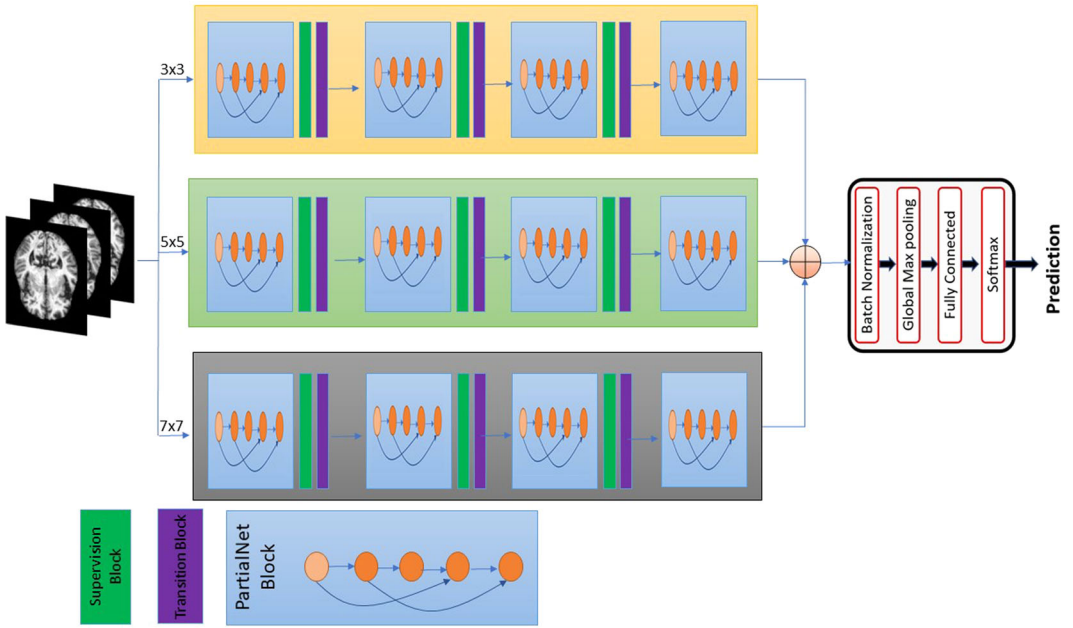


FIGURE 3 Schematic illustration of the proposed framework-multiple partialNet fusion [Color figure can be viewed at wileyonlinelibrary.com]

input to output); hence, shallow network contributes more to gradient magnitude. To overcome the aforementioned challenge, we utilize the block concept that limits the skip connection within a block and adopt a multiresolution policy. Therefore, every block behaves like a unique structure and contributes more to gradient magnitude than a deeper network. The path length plays a major role in gradient magnitude; thus, gradient flow is better over a partially dense path. The proposed framework is based on a partially dense network; therefore it has the benefits of feature reuse, which results in better learning. Figure 3 illustrates the proposed framework. The partially connected layers from different levels help improve the information flow between layers. Besides, it also alleviates the vanishing gradient problem due to direct connectivity between and later layers. In addition, we have introduced supervisor block and transition layer.

The proposed ensemble framework consists of three multiresolutional networks as shown in Figure 3 and each of which consists of four blocks. Each block consists of a partially connected dense layer, supervision layer followed by transition layer. Unlike the densely connected layer, we have considered partial connectivity which has threefold benefits: reduce the parameter and better feature learning, and avoid overfitting. Unlike DenseNet, PartialNet improves the flow of information by directly connecting some of its subsequent layers and concatenating the feature map, that is, the feature map received at the l th layer simply receive the information from the preceding layers $l - 1/2$ (Figure 4).

$$x_l = H_l([\mathbf{x}_0, \mathbf{x}_2, \dots, \mathbf{x}_{l-2}]), \tag{1}$$

here “++” is the concatenation operation, H_l denotes the composite nonlinear transformation consists of batch normalization (BN), leaky rectified linear unit (ReLU), and $3 \times 3 \times 3$

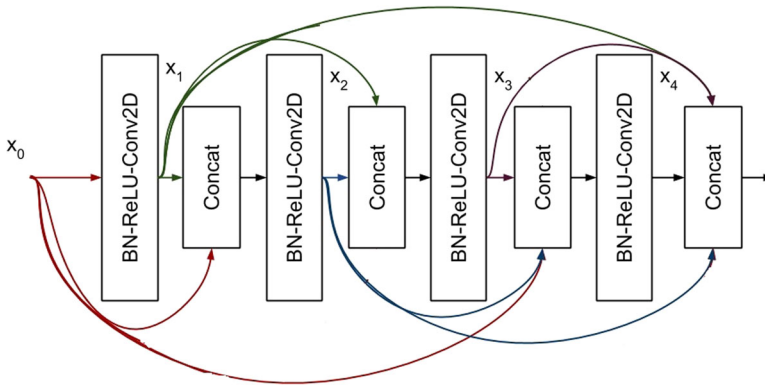


FIGURE 4 Basic PartialNet structure. As can be readily observed there is less number of connections to its preceding layers. BN, batch normalization; ReLU, rectified linear unit [Color figure can be viewed at wileyonlinelibrary.com]

convolution. In each partially connected block, each layer l_i receives a features map from l_n preceding layers. $[x_0, x_1, \dots, x_{l-1}]$ is the connectivity of earlier $l - 1$ layers into a single tensor.

Unlike DenseNet, partial dense blocks consist of partial connectivity of earlier layer. The layer in partial block consists of $1 \times 1 \times 1$ and $3 \times 3 \times 3$ convolution layers. The denser block is followed by supervision and transition layers. The supervision block consists of 1×1 and 3×3 convolution, and the transition block consists of BN, 1×1 convolution, and 2×2 pooling. The aim of the supervision block is to filter the information and force the network to learn intermediate features.

Every ensemble network of multipath networks processes the information of different scales and depth levels. To learn the intermediate features, PartialNet is aided with supervision and transition block to have supervised feature transformation. The supervision block consists of 1×1 convolution and 3×3 convolution whereas the transition block consists of 1×1 , 3×3 , and 1×1 convolution filters. The supervision block filters the information and learns intermediate features. It is worth mentioning that we have not utilized the supervision and transition block after the last PartialNet block. As the last PartialNet block does not have any further block, thus there is no need to bound the skip connection. Besides, it also degrades the performance. The feature maps from ensemble PartialNets are concatenated. Table 2 lists the output size and the parameters of each network layer in the DenseNet model.

$$X_i = x_1 + x_2 + x_3, \tag{2}$$

$$Z = X_0 + X_1 + X_2 + \dots + X_r. \tag{3}$$

Where Z is the final concatenated features, X_i represents the output from PartialNet, and r is the total number of ensemble networks.

In traditional DenseNet, a layer has input from all its preceding layers, thus multiplicity of the network is $2nk$; however, multiresidual network has much less number of connections. In comparison to DenseNet and ResNet, PartialNet has a moderate number of skip connections ($ResNet < PartialNet < DenseNet$). One of the key aspects of PartialNet is the growth rate that describes the rate at which the size of each layer within each block of PartialNet grows. The growth rate in each block individually acts as a regulator to control the flow of information from a layer to its following layers. For example, the growth rate $k = 16$ shows that a filter size of 16 is

TABLE 2 PartialNet configurations

| Layers | Configurations |
|---------------------|------------------------------|
| Input | 224 × 224 × 3 |
| Convolution | 7 × 7 stride 2 |
| Max pooling | 3 × 3 stride 2 |
| Partial block 1 | (1 × 1 conv, 3 × 3 conv) × 2 |
| Supervision block 1 | (1 × 1 conv, 3 × 3 conv) |
| Transition block 1 | (1 × 1 conv, 3 × 3 max pool) |
| Partial block 2 | (1 × 1 conv, 3 × 3 conv) × 2 |
| Supervision block 2 | (1 × 1 conv, 3 × 3 conv) |
| Transition block 2 | (1 × 1 conv, 3 × 3 max pool) |
| Partial block 3 | (1 × 1 conv, 3 × 3 conv) × 2 |
| Supervision block 3 | (1 × 1 conv, 3 × 3 conv) |
| Transition block 3 | (1 × 1 conv, 3 × 3 max pool) |
| Partial block 4 | (1 × 1 conv, 3 × 3 conv) × 2 |
| Global average pool | |
| Fully connected | |
| Softmax | |

used at each layer in each block. We have noticed that a smaller growth rate showed better performance and transfer information efficiently between the layers. Similar to Lodhi and Kang,²⁴ we have further introduced the bottleneck layer (1 × 1 convolution) before each 3 × 3 convolution in each block that helps reduce the input feature maps. Unlike DenseNet, the shortcut connections were used to cross two or three convolutional layers. Two 3 × 3 convolution layers with a bottleneck of 1 × 1 were used which also concatenate multiple convolutional features. Then, these feature maps are fed to the transition layer. We have ensembled multiresolutional PartialNet based on probability by integrating the probabilities of the softmax layer as shown in Equation (4).

$$P^i = (\alpha_1^i, \alpha_2^i, \dots, \alpha_n^i), \quad (4)$$

where α_j^i indicates the probabilities of the class j .

P^i in Equation (4) can be normalize as

$$P^i = \frac{P^i}{\max[\alpha_1^i, \alpha_2^i, \dots, \alpha_n^i]}. \quad (5)$$

The prediction can be determined based on output of multiresolution PartialNet based on probability as

$$Y = \arg \max \left(\sum_{i=1}^m \log(\alpha_1^i), \sum_{i=1}^m \log(\alpha_2^i), \dots, \sum_{i=1}^m \log(\alpha_n^i) \right). \quad (6)$$

4 | EXPERIMENTAL RESULTS

This section details the experimental design, data set, and the classification results of the proposed framework compared with other state-of-the-art (SOTA) methods. This study is conducted on benchmark ADNI data set. We implemented PartialNet using MATLAB and tested using various block sizes, multiresolution networks, and growth rates. We have used the stochastic gradient descent (SGD) method for training and performed 10-fold cross-validation by partitioning the data set randomly. Evaluation is conducted using several metrics, including accuracy (for both test and validation), sensitivity, and specificity. Additionally, the results of ablation studies for the proposed model conducted using multi-DenseNet fusion with fully and partially connected dense blocks are also discussed. All the experiments were performed on a system with NVIDIA RTX5000 GPU.

Evaluation is performed using the neuroimaging data that are obtained from the publicly available ADNI database.²⁵ DR. Michael W. Weiner launched ADNI is multisite, longitudinal back in 2004, which was financially supported by both private and public partnership (27 million by 20 companies and 40 million from National Institute on Aging). ADNI develops clinical, imaging, genetic, and biospecimen biomarkers for the early diagnosis of AD.²⁵ The primary goal of ADNI has been to assess the capabilities of the integration between imagine-derived biomarkers (e.g., MRI and PET), clinical and other neurological assessments to detect AD at the early stage of MCI. The data sets (ADNI, ADNI 2, ADNI 3, and ADNI GO) include 1800 female and male subjects. In our study, we have considered 350 subjects and collected T1 weighted structural MRI images (95 CN, 95 AD, and 146 MCI) from ADNI data set. The data set consists of multiple scans of each user performed at different times. In this study, we have used a minimum and maximum scan number of 3 and 15, respectively. Table 3 describes the statistics of data set used in this study.

Experiments were performed on the ADNI data set, described above, using the proposed multiresolutional ensemble PartialNet. We compare the performance of our pipeline against its counterpart networks. All networks were trained using the SGD method. Learning rate, weight decay, and Nesterov moment were set to 0.1, 10^{-4} , and 0.9, respectively. We reduce the learning rate by 10% at 50% and 75% of training epochs. We have randomly portioned the available data samples into 60%, 20%, and 20% for training, testing, and validation, respectively. The number of ensemble networks is set to 3 and the number of PartialNet/DenseNet/ResNet blocks is set to 4.

In this study and by using the available data set, we performed both binary classification (i.e., MC vs. AD, AD vs. CN, and MCI vs. CN) and multiclass classification. In our first experiment, we have considered a binary class problem to differentiate patients among different classes. The results are summarized in Table 4. We further have performed different ensemble methods (voting, averaging, and probability). Table 5 describes the ablation study. As readily seen in Tables 5 and 4, PartialNet achieved significantly better performance than DenseNet.

TABLE 3 Demographic details of data set used in our study

| Class | # Subjects | Male/female | Total # scans |
|-------|------------|-------------|---------------|
| AD | 95 | 44/51 | 1425 |
| MCI | 138 | 66/72 | 1021 |
| CN | 146 | 87/59 | 1479 |

Abbreviations: "AD," "MCI," and "CN" stand for Alzheimer's disease, mild cognitive impairment, and cognitive normal, respectively.

TABLE 4 Comparative analysis of each class and overall recognition results using different ensembling approaches

| Modalities | Multiresolution architecture | |
|--------------------------------|------------------------------|------------|
| | DenseNet | PartialNet |
| <i>AD versus CN versus MCI</i> | | |
| AD | 96.77 | 97.87 |
| MCI | 97.43 | 98.74 |
| CN | 94.43 | 97.31 |
| Overall | 96.72 | 98.23 |
| <i>AD versus CN</i> | | |
| AD | 76.43 | 85.11 |
| CN | 83.22 | 88.65 |
| Overall | 80.11 | 87.11 |
| <i>AD versus MCI</i> | | |
| AD | 95.98 | 98.43 |
| MCI | 97.23 | 99.87 |
| Overall | 96.65 | 99.26 |
| <i>CN versus MCI</i> | | |
| CN | 95.23 | 100 |
| MCI | 97.43 | 100 |
| Overall | 96.77 | 100 |

Abbreviations: AD, MCI, and CN stand for Alzheimer's disease, mild cognitive impairment, and cognitive normal, respectively.

TABLE 5 Overall recognition rate for both binary and multiclass classification comparing the proposed ensemble multiresolution PartialNet architecture with DenseNet

| Multiclass classification | Binary classification | | | | |
|---------------------------|------------------------------|-------------------|---------------|---------------|-------------|
| | Multiresolution architecture | AD vs. CN vs. MCI | AD vs. CN | AD vs. MCI | MCI vs. CN |
| Max voting | Ensemble DenseNet | 93.42% | 78.35 | 94.60 | 94.87 |
| | Ensemble PartialNet | 96.41% | 81.22% | 97.89% | 99.43% |
| Model averaging | Ensemble DenseNet | 94.88% | 79.43 | 9.65 | 93.98 |
| | Ensemble PartialNet | 97.10% | 83.74% | 98.43% | 99.13% |
| Probabilistic | Ensemble DenseNet | 96.72% | 80.11 | 96.65 | 97.76 |
| | Ensemble PartialNet | 98.23% | 83.74% | 99.26% | 100% |

Note: Bold values indicate highest performance achieved by proposed framework.

Abbreviations: AD, MCI, and CN stand for Alzheimer's disease, mild cognitive impairment, and cognitive normal, respectively.

Besides the accuracy, we can notice that PartialNet has significantly less computational and space complexity due to less number of connections in comparison to DenseNet.

To achieve the best network and parameter, we have performed several experiments. Tables 5 and 6 describe the ablation study. Notice that performance degraded by adding the supervision and transition block after the last PartialNet block consistent with earlier blocks. As the last PartialNet block does not have any further block, thus there is no need to bound the skip connection. Besides, gradient flow is also better. Similarly, we can notice that PartialNet should better perform in comparison to the same multiresolution ensemble of DenseNet and ResNet. It might be due to the fact that both DenseNet and ResNet have paths of different lengths. For example, skip connection from input to output, we bound the path in each block as well as limit the number of dense connections; hence, gradient flow is better over the partially dense path.

To highlight the advantage of the proposed pipeline, we compared its performance against several SOTA methods. The comparison of our work with other literature work is summarized in Table 7. The table also summarizes the modalities, techniques, and accuracy of all compared methods. Table 7 lists the methods only that are evaluated on ADNI data set.

5 | DISCUSSION AND OBSERVATIONS

In this paper, we sought an integrative deep pipeline for the detection of AD. The proposed analysis pipeline is based on an ensemble learning technique that incorporates a deep multi-resolutional ensemble PartialNet. Compared with DenseNet, the proposed approach demonstrated high performance due to efficient feature reuse. That can be explained in part by limiting the connection in each block and utilizing the block concept, which limits the skip connection within a block and utilized a multiresolution ensemble, as every block behaves like unique structure and contribute more to gradient magnitude than deeper network. The proposed pipeline has demonstrated high accuracy for both binary and multiclass classifications when evaluated on ADNI data set. Namely, our experiments showed a considerable gain in detection performance in comparison to SOTA methods. PartialNet achieved a maximum classification accuracy of 100% for MCI versus CN, 99.26% for MCI versus AD, 88.71% for NC versus AD, and 98.23% for NC versus MCI versus AD with sensitivity and specificity 98 plus.

Additionally, our pipeline showed a considerable gain in performance in comparison to benchmark methods. As evident from data presented in Tables 6–8, PartialNet showed considerably outperformed all the other approaches. We have compared the performance of the proposed PartialNet with its variant DenseNet with the same set of parameters and ensemble frameworks. Table 6 summarizes the evaluation results of PartialNet and DenseNet. We can notice that PartialNet showed significantly better performance 98.23%/87.11%/99.26%/100% for the task of AD versus CN versus

TABLE 6 Overall recognition rate for both binary and multiclass classifications comparing the proposed ensemble multiresolution PartialNet architecture

| Multiresolution architecture | AD vs. CN vs. MCI | AD vs. CN | AD vs. MCI | MCI vs. CN |
|------------------------------|-------------------|---------------|---------------|-------------|
| Ensemble DenseNet | 96.72% | 80.11 | 96.65 | 97.76 |
| Ensemble PartialNet | 98.23% | 87.11% | 99.26% | 100% |

Note: Bold values indicate highest performance achieved by proposed framework.

Abbreviations: AD, MCI, and CN stand for Alzheimer's disease, mild cognitive impairment, and cognitive normal, respectively.

TABLE 7 Comparative accuracy of various SOTA DNN-based AD Identification systems on ADNI data set

| Study | Technique | Modality | Classification | Accuracy (%) |
|-------------------------------------|--------------------------|-----------------|----------------------------------|--------------|
| Bäckström et al. ²¹ | 3D ConvNet | MRI | 2 ways | 98.41 |
| DenseNet | Multiresolution Ensemble | MRI | 2 ways | 97.45 |
| Gorji and Haddadnia ²⁶ | Pseudo-Zernike moments | MRI | 2 ways | 95.69 |
| Hon and Khan ²⁷ | Transfer learning | MRI | 2 ways | 96 |
| Proposed PartialNet | Multiresolution ensemble | MRI | 2 ways | 98.82 |
| Lin et al. ²⁸ | NGF | MRI + PET + CSF | 3 ways (AD, MCI, NC) | 60.2 |
| Suk et al. ²⁹ | DW-S MTL | MRI + PET + CSF | 3 ways (AD, MCI, NC) | 62.93 |
| Gupta et al. ³⁰ | SAE | MRI | 3 ways (AD, MCI, NC) | 85 |
| Hosseini-Asl et al. ^{2,16} | DSA-3D CNN | MRI | 3 ways (AD, MCI, NC) | 94.8 |
| Lui et al. ³¹ | SAE-Zeromask | MRI + PET | 4 ways (AD, cMCI, neMCI, NC) | 53.8 |
| Farooq et al. ²⁰ | GoogleNet | MRI | 4 ways (AD, MCI, LMCI, NC) | 98.88 |
| Farooq et al. ²⁰ | ResNet-152 | MRI | 4 ways (AD, MCI, LMCI, NC) | 98.14 |
| Kazemi et al. ²² | AlexNet | MRI | 4 ways (AD, SMC, EMCI, LMCI, CN) | 97.63 |
| Proposed PartialNet | Multiresolution ensemble | MRI | 3 ways (AD, MCI, CN) | 98.23 |

Note: Bold values indicate highest performance achieved by proposed framework.

Abbreviations: 3D CNN, three-dimensional convolutional neural network; AD, Alzheimer's disease; ADNI, Alzheimer's disease neuroimaging initiative; CSF, cerebrospinal fluid; DNN, deep neural network; DSA, deeply supervised adaptable; MRI, magnetic resonance imaging; NGF, nerve growth factor; PET, positron emission tomography; SAE, stacked autoencoder; SOTA, state-of-the-art.

TABLE 8 Comparative accuracy of various SOTAs on ADNI (baseline) with 813 (228 NC, 187 AD, and 398 MCI) subjects

| Study | AD vs. CN | MCI vs. AD |
|---------------------------------|--------------|--------------|
| Hon and Khan ²⁷ | 84.14 | 82.26 |
| Wang et al. ³³ | 79.93 | 76.86 |
| Snapshot ensemble ³⁴ | 54.93 | 68.02 |
| Ensemble transfer ³⁵ | 85.27 | 83.11 |
| Proposed PartialNet | 87.11 | 99.26 |

Note: Bold values indicate highest performance achieved by proposed framework.

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's disease neuroimaging initiative; CN, cognitive normal; MCI, mild cognitive impairment; NC, normal control; SOTA, state-of-the-art.

MCI, AD versus CN, AD versus MCI, and MCI versus CN in comparison to 96.72%/80.11%/96.65%/97.76% using DenseNet. With 3-way classification, we can notice that PartialNet so far has produced better results with 98.23% accuracy in comparison to 96.72 using the DenseNet ensemble. The classwise comparison of SOTA models is described in Table 7. In competing with SOTA methods, PartialNet showed better performance for both 2-way and 3-way Alzheimer diagnoses. We can notice a gain in performance in terms of all parameters (sensitivity, specificity, and accuracy). The proposed PartialNet framework achieved 98.67/99.27/98.22 sensitivity in comparison to Farooq et al.,²⁰ Gupta et al.,³² and Hosseini-Asl et al.² that achieved 98/97/97, 96/74/88, and 100/80/47 sensitivity, for AD/MC/NC, respectively. Experiment results showed that our proposed PartialNet has produced more promising and consistent results for all the three classes the sensitivity, specificity, and accuracy remained above 98%, 97.7%, and 98.23%, respectively. PartialNet performs well because it follows a simple connectivity rule in comparison to DenseNet and forces the network to learn representation at each block that incorporates identity mappings naturally as well as it diversified the depth and considers deep supervision.

We can summarize that an ensemble of PartialNet showed considerably better generalization than its parent network. PartialNet incorporates the properties of identity mappings, diversified depth as well as deep supervision, thus, considers the feature reuse which in turn results in better learning, thus better in terms of vanishing gradient, diminishing forward flow with better training time, and a low number of parameters in comparison to DenseNet. An experiment is performed on benchmark ADNI data set that shows considerable gain ($2 + \% \uparrow$) and ($1.2 + \% \uparrow$) for multiclass and binary class in Alzheimer detection performance in comparison to SOTA methods. In summary, we have the following *key observations*:

- The block concept forces the network to extract a unique structure and contribute more to gradient magnitude than a deeper network.
- The number of blocks depends upon the problem, and larger data sets with complex nature require a more number of blocks.
- Gradient flow is enhanced when compared with its variants DenseNet that helped to improve the performance, especially for CN class.

Despite the demonstrated effectiveness of the proposed ensemble PartialNet and significant improvement in performance for AD diagnosis, the proposed pipeline has some limitations.

First, the number of blocks depends upon the data set complexity and size, which will increase the number of parameters. Second, even though, the model showed a notable gain in performance in comparison to SOTA, it showed poor performance of AD versus CN in comparison to other classes, which may be due to the complexity of the images.

6 | CONCLUSION

In this paper, we presented multiresolutional ensemble PartialNet tailored to Alzheimer detection using brain MR imaging data. PartialNet incorporates the properties of identity mappings, diversified depth, and deep supervision, thus, considers feature reuse that in turn results in better learning. Compared with DenseNet, the proposed multiresolutional ensemble PartialNet has demonstrated better performance in terms of vanishing gradient, diminishing forward flow with better training time, and a low number of parameters. An experiment is performed on benchmark ADNI data set that showed a considerable gain in detection performance in comparison to SOTA methods. PartialNet achieved a maximum classification accuracy of 100% for MCI versus CN, 99.26% for MCI versus AD, 88.71% for NC versus AD, and 98.23% for NC versus MCI versus AD with sensitivity and specificity 98 plus.

By observing the main results from PartialNet and DenseNet, we can summarize that an ensemble of PartialNet showed considerably better generalization than its parent network. This may be due to limiting the connection in each block and utilizing the block concept which limits the skip connection within a block and utilized a multiresolution ensemble, as every block behaves like a unique structure and contributes more to gradient magnitude than a deeper network. In future, we plan to explore group equivariant PartialNet to aid translations, reflections, and rotations capability. Besides, we also plan to explore diver prediction to understand biological relevance that will help observe the disease progression.

ACKNOWLEDGMENT

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID

Imran Razzak  <https://orcid.org/0000-0002-3930-6600>

Saeeda Naz  <https://orcid.org/0000-0002-5665-4615>

Fahmi Khalifa  <https://orcid.org/0000-0003-3318-2851>

Mohamed Reda Bouadjenek  <https://orcid.org/0000-0003-1807-430X>

Shahid Mumtaz  <https://orcid.org/0000-0001-6364-6149>

ENDNOTE

* <http://adni.loni.usc.edu/study-design/>.

REFERENCES

1. Zhang F, Li Z, Zhang B, Du H, Wang B, Zhang X. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing*. 2019;361:185-195.
2. Hosseini-Asl E, Gimel'farb G, El-Baz A. Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. 2016.

3. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: overview, challenges and the future. *Classification in BioApps*. 2018:323-350.
4. Rehman A, Naz S, Razzak I. Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. *Multimedia Syst*. 2021:1-33.
5. Ferreira D, Pereira JB, Volpe G, Westman E. Subtypes of Alzheimer's disease display distinct network abnormalities extending beyond their pattern of brain atrophy. *Front Neurol*. 2019;10:524.
6. Naseer A, Rani M, Naz S, Razzak MI, Imran M, Xu G. Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Comput Appl*. 2020;32(3):839-854.
7. Razzak MI, Imran M, Xu G. Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks. *IEEE J Biomed Health Inf*. 2018;23(5):1911-1919.
8. Rehman A, Naz S, Razzak MI, Akram F, Imran M. A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits Syst Signal Process*. 2020;39(2):757-775.
9. Khawaja A, Khan TM, Naveed K, Naqvi SS, Rehman NU, Nawaz SJ. An improved retinal vessel segmentation framework using Frangi filter coupled with the probabilistic patch based denoiser. *IEEE Access*. 2019;7:164344-164361.
10. Khan MA, Khan TM, Soomro TA, Mir N, Gao J. Boosting sensitivity of a retinal vessel segmentation algorithm. *Pattern Anal Appl*. 2019;22(2):583-599.
11. Tabassum M, Khan TM, Arsalan M, et al. CDED-Net: joint segmentation of optic disc and optic cup for glaucoma screening. *IEEE Access*. 2020;8:102733-102747.
12. Suk HI, Lee SW, Shen D, The Alzheimers Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*. 2014;101:569-582.
13. Sarraf S, DeSouza DD, Anderson JAE, Tofighi G. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv*. 2016:070441.
14. Mathew J, Mekkayil L, Ramasangu H, Karthikeyan BR, Manjunath AG. Robust algorithm for early detection of Alzheimer's disease using multiple feature extractions. In: 2016 IEEE Annual India Conference (INDICON). IEEE; 2016:1-6.
15. Iftikhar MA, Idris A. An ensemble classification approach for automated diagnosis of Alzheimer's disease and mild cognitive impairment. In: 2016 IEEE Annual India Conference (INDICON). IEEE; 2016:78-83.
16. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE; 2016:126-130.
17. Ashraf A, Naz S, Shirazi SH, Razzak I, Parsad M. Deep transfer learning for Alzheimer neurological disorder detection. *Multimedia Tools Appl*. 2021;80:30117-30142.
18. Naz S, Ashraf A, Zaib A. Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset. *Multimedia Syst*. 2021;28:85-94.
19. Ju R, Hu C, Zhou P, Li Q. Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans Comput Biol Bioinf (TCBB)*. 2019;16(1):244-257.
20. Farooq A, Anwar S, Awais M, Rehman S. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: 2017 IEEE International Conference on Imaging systems and techniques (IST). IEEE; 2017:1-6.
21. Bäckström K, Nazari M, Gu IYH, Jakola AS. An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; 2018:149-153.
22. Kazemi Y, Houghten S. A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data. In: 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE; 2018:1-8.
23. Ebrahimi-Ghahnavieh A, Luo S, Chiong R. Transfer learning for Alzheimer's disease detection on MRI images. In: IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). IEEE; 2019:133-138.
24. Lodhi B, Kang J. Multipath-DenseNet: a supervised ensemble architecture of densely connected convolutional networks. *Inf Sci*. 2019;482:63-72.
25. Sandeep C, Kumar AS, Susanth M. The online datasets used to classify the different stages for the early diagnosis of Alzheimer's disease (AD). *Int J Eng Adv Technol*. 2017;6(4):38-45.

26. Gorji H, Haddadnia J. A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. *Neuroscience*. 2015;305:361-371.
27. Hon M, Khan NM. Towards Alzheimer's disease classification through transfer learning. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2017:1166-1169.
28. Lin W, Tong T, Gao Q, et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front Neurosci*. 2018;12:777.
29. Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2013:583-590.
30. Gupta Y, Lee KH, Choi KY, et al. Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images. *PLOS ONE*. 2019;14(10):e0222446.
31. Liu M, Li F, Yan H, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *NeuroImage*. 2020;208(5):116459. doi:10.1016/j.neuroimage.2019.116459
32. Gupta A, Ayhan MS, Maida AS. Natural image bases to represent neuroimaging data. In: *Proceedings of the 30th International Conference on International Conference on Machine Learning (ICML'13)*. Vol 28, 2013:III-987-III-994. JMLR.org. <http://dl.acm.org/citation.cfm?id=3042817.3043047>
33. Wang SH, Phillips P, Sui Y, Liu B, Yang M, Cheng H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J Med Syst*. 2018;42(5):1-11.
34. Zheng C, Xia Y, Pan Y, Chen J. Automated identification of dementia using medical imaging: a survey from a pattern classification perspective. *Brain Inf*. 2016;3(1):17-27.
35. Tanveer M, Rashid AH, Ganaie M, Reza M, Razzak I, Hua KL. Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. *IEEE J Biomed Health Inf*. 2021.

How to cite this article: Razzak I, Naz S, Ashraf A, Khalifa F, Bouadjenek MR, Mumtaz S. Multiresolutional ensemble PartialNet for Alzheimer detection using magnetic resonance imaging data. *Int J Intell Syst*. 2022;1-18. doi:10.1002/int.22856