Automated Detection of Alzheimer's Disease and Mild Cognitive Impairment Using Whole Brain MRI

FAZAL UR REHMAN FAISAL AND GOO-RAK KWON[ID], (Senior Member, IEEE)

Department of Information and Communication Engineering, Chosun University, Gwangju 61452, South Korea

Corresponding author: Goo-Rak Kwon (grkwon@chosun.ac.kr)

ABSTRACT Early diagnosis is critical for the development and success of interventions, and neuroimaging is one of the most promising areas for early detection of Alzheimer's disease (AD). This study is aimed to develop a deep learning method to extract valuable AD biomarkers from structural magnetic resonance imaging (sMRI) and classify brain images into AD, mild cognitive impairment (MCI), and cognitively normal (CN) groups. In this work, we adapted and trained convolutional neural networks (CNNs) on sMRI images of the brain from ADNI datasets available in online databases. Our proposed mechanism was used to combine features from different layers to hierarchically transform the images from magnetic resonance imaging into more compact high-level features. The proposed method has a reduced number of parameters which reduces the computation complexity. The method is compared with the existing state-of-the-art works for AD classification, which show superior results for the widely used evaluation metrics, including accuracy, an area under the ROC curve, etc., suggesting that our proposed convolution operation is suitable for the AD diagnosis

INDEX TERMS Alzheimer's disease (AD), magnetic resonance imaging (MRI), convolutional neural network (CNN), image classification.

I. INTRODUCTION

Alzheimer's disease (AD), a neurological degenerative disease, is a prominent challenge of the 21st century. According to the World Alzheimer's report statistics, more than 55 million people have been diagnosed with Alzheimer's disease, and this number is increasing day by day, with 78 million expected by 2030 [1]. Clinically, the disease manifests as memory loss, disorientation, and visuospatial disturbances [2], to circumvent which many efforts have been made for early detection and diagnosis of AD. Partial AD treatment can be done by observing the symptoms; however, a method to identify specific biomarkers in the cerebrospinal fluid (CSF) is being developed for a more accurate diagnosis [3]. This is an intrusive investigation that may cause harm

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng[ID].

to the patient [4]. In addition, advanced imaging techniques such as magnetic resonance imaging (MRI) and position emission tomography (PET) can also be used to diagnose AD-related structural and molecular biomarkers [5]. In particular, magnetic resonance imaging (MRI) is a non-invasive and powerful tool for understanding and evaluating AD's anatomical and functional brain changes. They are recognized as indispensable in clinical practice and play an important role in AD disease progression [6]–[8].

Integration of large-scale, high dimensional, and multi-modal data from rapidly advancing neuroimaging techniques imposes difficulty for contemporary methods to identify the disease. As a result, interest in computational machine learning methods for integrative analysis has exploded [9]–[12]. These methods have been used to generate the desired output from training data, such as voxel intensity, tissue density, and shape descriptor features. However, to use these

machine learning algorithms for the classification of disease, pre-processing techniques must be adopted, which are primarily time-consuming and computationally intensive. Classification studies using machine learning algorithms usually require four steps: feature extraction, feature selection, dimension reduction, and feature-based classification algorithm selection [13], [14]. These steps require specialized knowledge and multiple optimization steps, which are time-consuming and laborious. In contrast, researchers have been exploring alternative techniques to traditional machine learning, such as Deep Learning (DL) algorithms. Deep learning has emerged as one of the widely adopted machine learning algorithms, which has shown optimal results in various domains, including speech recognition, computer vision, and natural language understanding [15]–[18]. In general, deep learning methods are a subset of representation-learning methods in a way that they can automatically identify the best representation from raw data without the need for prior feature selection [19]. This is accomplished by employing a hierarchical structure with varying levels of complexity, as well as applying consecutive nonlinear transformations to the raw data. These transformations produce increasingly higher levels of abstraction, with higher-level features being more insensitive to noise in the input data than lower-level features [20]. Consequently, researchers are focusing their efforts on developing a deep learning model in medical imaging that can accurately diagnose the disease. Recently, deep learning models have shown significant success in various medical image analysis problems, including CT scans, MRIs, X-rays, ultrasounds, and sentiment analysis [21]. It has produced notable results in detecting and classifying specific diseases in the domains of the lungs, abdomen, brain, cardiovascular, retina, and others [22].

Notably, the most widely used deep learning design, Convolutional Neural Network (CNN), has received a lot of attention in medical image analysis due to its success in image analysis and classification [23], [24]. These accomplishments have piqued the interest of researchers to improve CNN-based systems for AD detection. Despite the fact that the existing methods have provided good diagnostic results, little has been done to optimize the architecture of CNN for practical AD detection. Inspired by the success of deep learning methods and patch-based mechanisms in medical imaging, in this work, we propose an improved convolutional neural network for AD diagnosis and prediction using magnetic resonance images. The proposed architecture significantly reduces the number of parameters and computational costs compared to conventional neural networks. Our research contributions are fourfold:

1) This work explores the convolutional layer's computational cost and concentrate on alternative methods to perform convolution operation in order to reduce the number of parameters and calculation costs while maintaining classification accuracy.
2) Each layer is composed of a block containing a series of operations on the input. The resolution and size of the input to each operation in the block are kept the same. The improvement demonstrated its usefulness in classification performance of the proposed method.
3) The proposed model learns valuable features from the input data without pre-processing and shows superior performance.
4) The performance of the proposed method has been compared with two state-of-the-art classification models: ResNet50 and VGG, which show better accuracy results using less number of parameters.

The rest of the paper is organized as follows: Section II introduces the relevant literature review, and Section III summarizes the materials and methodology used in this article. Sections IV and V presents the results of the experiments and their interpretation. Section VI concludes with an outlook on future projects.

II. RELATED WORK

Several approaches for neuroimaging classification have been proposed in the recent years to improve classification performance. We examine some machine learning (ML) classification structures used in neuroimaging, as well as methods based on a convolutional neural network.

A. MACHINE LEARNING FOR NEUROIMAGING

In recent years, many machine learning-based techniques have been used for multi-class classification and binary classification for the early detection of Alzheimer's disease. Kim *et al.* [25] proposed a fully automated classification method in which cortical thickness features for diagnosis have been used. Long *et al.* [26] investigated the regional morphological differences in the brain. They observed the hippocampus and amygdala deformation to identify the progressive MCI. Where, diffusive morphological changes in the whole-brain gray matter (GM) were responsible for determining mild or moderate AD. Subsequently, these subjects were classified using a linear support vector machine (SVM). Guo *et al.* [27] proposed extracting and combining the brain region and subgraph features of functional magnetic resonance imaging, which were then utilized for training the multi-kernel SVM for classification. This approach retains the global topological information and the sensitivity to change in the brain region. In contrast to Guo *et al.*, Khedher *et al.* [28] proposed a method for classifying AD using independent component analysis to extract features from regions of white matter (WM), GM, and cerebrospinal fluid (CSF) for training SVM classifier. Moreover, Tong *et al.* [29] proposed multiple learning instances for dementia classification, where features were formed by extracting pockets of MRI voxel patches and mapping them to graphs. An SVM classifier was then used to distinguish between AD patients and NC (Normal Control) subjects. In another work, Gupta *et al.* [30] used machine learning approaches (SVM, k-nearest neighbor (KNN), and Random Forest (RF)) to classify atrophic states (AD, NC/healthy

control (HC), asymptotic Alzheimer's disease (aAD), mild Alzheimer's disease (mAD)) using combined voxel-based morphometry (VBM) features, cortical and subcortical volumetric features (CSC) and hippocampal volumetric features.

B. CNN METHODS FOR NEUROIMAGING

The enormous potential of deep learning has been observed in the medical image diagnosis [31]–[33], where initially, it was used for region segmentation or feature extraction, followed by traditional machine learning algorithms such as SVM and boosting. For instance, Silva *et al.* [34] proposed a convolutional neural network for feature extraction from MRI scan, followed by SVM, KNN, and Random Forest algorithms for Alzheimer's disease classification. Similarly, Liu *et al.* [35] proposed a deep convolutional feature learning method for classification of AD and MCI using both unsupervised and supervised learning. Owing to the recent success of deep learning, especially convolutional neural networks in the field of computer vision for extracting image features, their potential has been explored for AD diagnosis [36]. For instance, Wang *et al.* [37] used brain extraction tools to select the hippocampus-containing slices and fed them to a convolutional neural network for diagnosis. In addition to that a patch-based ensemble classifier was created to predict the AD and NC classes [38]. Furthermore, for the diagnosis of Alzheimer's disease, 2-D and 3-D deep learning models were developed in [39]. Unlike others, Basaia *et al.* [40] performed the extra operation of augmentation of training data by deforming, cropping, rotating, flipping, and scaling MRI data at different angles before feeding it into a convolutional neural network for diagnosis.

In more recent work, Korolev *et al.* [41] proposed two 3D CNN architectures based on VGGNet and ResNet, proving that the step of manual feature extraction is unnecessary for brain MRI image classification. Their 3D models, 3D-VGG and 3D-ResNet, are widely used in research to classify 3D medical images. Also, the usage of 3D CNN was focused in Ehsan Hosseini-Asl *et al.* [7] and [42], in which Alzheimer's Disease Neuroimaging Initiative (ADNI) data was classified. The 3D convolutional neural network was utilized for feature extraction from MRIs and to identify the biomarkers for multiple AD classes. Similarly, Abrol *et al.* [43] also created 3D CNNs based on the ResNet architecture and tested them on various binary and multiclass tasks. They created a training set for cross-validation and a small test set using ADNI data. The results were promising; however, no comparisons were made with other standard evaluation frameworks, which leads to an ambiguity that the model is overfitted on the training examples. In [44], a deep learning-based classifier for AD versus NC classification has been proposed where the discrete volume estimation model with convolutional neural networks has been used to extract deep features of the discrete volume of the left and right hippocampal model (RHM and LHM). Recently, [2] proposed the diagnostic model AD based on a densely connected 3D CNN and an attention-driven mechanism to combine high-level features and spatial information

TABLE 1. Demographic representation of all subjects.

Diagnostic type	Number	Age	Gender(M/F)
AD	163	78.04±6.40	82/81
MCI	163	77.14±7.02	99/64
CN	163	78.02±5.33	78/85

AD: Alzheimer's disease; MCI: Mild Cognitive Impairment; CN: Normal control.

extracted from MRI. J Liu *et al.* [45] proposed the depthwise separable convolution, replacing the conventional convolution with depthwise separable. AlexNet and GoogLeNet transfer learning models were used to train their idea, which significantly reduced the computational cost and parameters.

In this paper, a CNN network structure has been proposed where the standard convolutional layer along-with depthwise-pointwise convolutional layer was embedded to extract more features that are sensitive to brain activities or structural differences in large regions. To the best of our knowledge, such architecture has not been used for the brain image classification. This greatly reduces the number of parameters while maintaining reasonable accuracy compared to other benchmark models.

III. MATERIAL AND METHODS

A. MRI ACQUISITION PROTOCOL

The ADNI was established in 2003 as a public-private partnership led by Michael W. Weiner, MD. The primary goal of ADNI was to determine whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers and clinical and neuropsychological assessment could be used together to track the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). T1-weighted MR images were acquired sagittally in ADNI using a volumetric 3D MPRAGE with a spatial resolution of $1.25 \times 1.25 \text{mm}^2$ in-plane and 1.2mm thick sagittal slices. The majority of these images were captured using 1.5 T scanners. More details about the MR acquisition can be found at the ADNI's website (http://adni.loni.usc.edu/).

B. PARTICIPANTS

We evaluated T1-weighted MRI data from ADN1/ GO of participants, including 163 AD patients, 163 subjects diagnosed with MCI, and 163 normal controls (NC) within 24 months in this study. Table 1 shows the demographic details for each group. The structural MRI scans used in the study had already been reviewed for quality and underwent gradient inhomogeneity correction (gradwarp), B1 non-uniformity correction, and N3 processing (to reduce the residual intensity non-uniformity).

C. IMAGE PATCH GENERATION

Due to the volumetric nature of MRI images, the natural choice of deep learning model is a 3D Convolutional Neural Network (3D CNN) [46]. Compared to 2D CNN
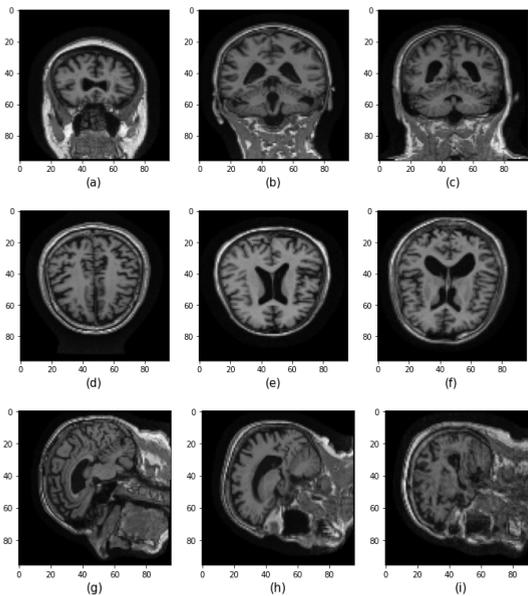
FIGURE 1. Brain MRI images from the ADNI database depicting various stages of Alzheimer's disease.(a)-(c) Normal controls (CN); (d)-(f) Mild Cognitive Impairment (MCI); (g)-(i) Alzheimer's Disease (AD).

models, 3D CNN models are computationally intensive and time-consuming to train due to their high-dimensional input. Another problem is that most current medical datasets are quite small. Due to the scarcity of data, it is not easy to train a deep network that can be generalized to high levels of complexity. Therefore, our study used 489 3D MRI scans of dimensions $192 \times 192 \times 160$ which cannot be directly fed into a 2D CNN model. First, the 3D MRI scans were downsized to $96 \times 96 \times 96$. The images were then divided into axial, coronal, and sagittal slices. The slices at the beginning and end were discarded since they did not contain any helpful information. Also, the slices were normalized with a mean of zero and a standard deviation of one. The randomly selected axial, coronal, and sagittal patch slices were then used for training the 2-D CNN model. Some examples of slices from MRI scans of CN, MCI, and AD patients are shown in Figure 1.

D. NETWORK ARCHITECTURE

The CNN architecture used in this study was inspired by the human visual cortex. The human eye receives information in its receptive field, which is similar to the convolution operation that folds the input image and creates the feature map by working with the receptive field of the input. The convolution involves multiple layers, including ReLU activation features, max-pooling layers, and fully connected layers.

Each input is processed through these operations to produce a final output in the form of a binary or multiclass classifier. A set of neurons interconnects the convolutional operation, shared hyperparameters, local connectivity, and shift-invariance, to increase the network's performance. Based on this motivation, we proposed an end-to-end deep CNN architecture for multi-label AD biomarker identification using the entire image volume as input.
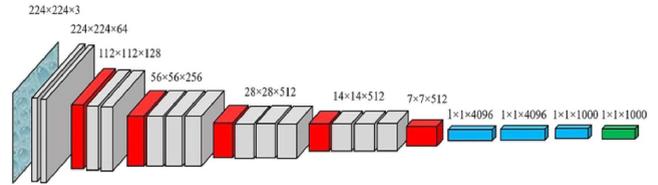
FIGURE 2. Basic architecture of VGG Net model [49].

Furthermore, we have also used the popular CNN architectures such as VGG-Net and ResNet for classification tasks compared to the proposed method. These models have been successfully used in various applications, including image classification, recognition, image labeling, and pose detection [47]. In this section, we discuss in detail our proposed CNN architecture as well as related methods used in this work.

1) VGG-NET MODEL

The VGG network is a brand name for the pre-trained CNN model proposed by Simonyan and Zisserman [48] at Oxford University in early 2014. VGG (Visual Geometry Group) was trained on the Image Net ILSVRC dataset, which consisted of images belonging to 1000 classes, with 1.3 million images used for training and 50000 for validation. VGG-19, a variant of VGG architectures has 19 deeply linked layers and has consistently performed better than other state-of-the-art models. The model consists of highly linked convolutional layers and fully linked layers, which allows better feature extraction and the use of max-pooling (instead of average-pooling) for down-sampling before classification with the SoftMax activation function. The VGG-19 model is used as a baseline model in this study, and the results are improved with the ADNI dataset to classify multiple stages of Alzheimer's disease. The architecture of VGG-19 is shown in Figure. 2.

2) ResNet MODEL

Residual Network [50] won first place in classification, localization, and detection at ILSVRC-2015. The researchers wanted to explore if learning better meant the addition of more layers on top of each other in the network. They discovered the degradation problem, where traditional models similar, to VGG, did not improve their performance after a certain number of layers rather get worse. To solve this problem, they proposed the residual function, which is the basic building block of a residual network (ResNet). In this work, we directly adopted ResNet from the non-bottleneck 50-layer architecture, where links with increasing dimensions were either (A) identity links, i.e., zero padding, or (B) projection links, i.e., convolutions with 1×1 filter (kernel) size, and used it as the basic model to classify AD using the ADNI dataset. Figure. 3 shows the basic architecture of ResNet-50 (only 34 layers are shown for simplicity).

3) PROPOSED MODEL

Convolutional layers are the fundamental building blocks for any deep CNN that uses complex activation functions
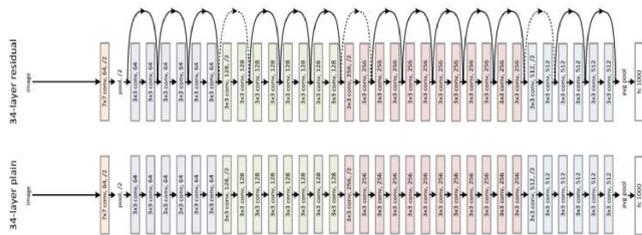
FIGURE 3. ImageNet network architecture examples. A residual network with 34 parameter layers is shown at the top. The dotted shortcuts expand the dimensions. At the bottom is a simple network with 34 parameter layers [52].

to achieve the best performance. To diagnose Alzheimer's disease, the proposed model uses a deep convolutional neural network to automatically extract features from the whole-brain MRI scans. Figure. 4 shows the proposed pipeline consisting of three main steps: brain volume resizing, 3D volume slicing, and CNN processing. Inspired by the architectural pattern of ResNet and ConvMixer [51], we proposed a simple yet effective convolutional method that simultaneously performs standard convolution, depthwise convolution, and point-wise convolution, followed by a skip convolution layer to learn multi-level features from MRI scans of the brain.

a: STANDARD CONVOLUTION

Figure. 5 illustrates the standard folding process. A standard convolutional layer takes a $D_F \times D_F \times M$ feature map I as input and generates a $D_G \times D_G \times N$ feature map output O, where D_F denotes the spatial width and height of the square input feature map, M denotes the number of input feature map channels, and N denotes the number of the output feature map channels. The feature is extracted from the $D_k \times D_k$ sized convolution kernel of the standard convolutional layer. D_k denotes the spatial width and height of the convolution kernel. The formula for the computation process of the standard convolution from feature map I to feature map O is given by:

$$G_{k,l,n} = \Sigma_{i,j,m} K_{i,j,m,n} . I_{k+i-1,l+j-1,m}, \qquad (1)$$

where I are the input feature maps, G is the output feature maps and k are the convolution kernels. The position of the convolutional kernel, elements is given by i and j. k and l determine the position of the element in the input feature map and the output feature map, respectively, while m is the channel of the input feature map, and n is the channel of the output feature map.

The parameters of standard convolution are calculated as follows:

$$F = M \times N \times D_k{}^2 \qquad (2)$$

The cost of computing standard convolution is shown by:

$$G = M \times N \times D_F{}^2 \times D_k{}^2 \qquad (3)$$

where F stands for the total number of model parameters, G represents the computational cost, M represents the number

of channels of the input feature map, N represents the number of channels of the output feature map, D_F represents the spatial width and height squared of the input features of the object map, and D_k represents the spatial width and height of the convolution kernel.

b: DEPTH-WISE AND POINT-WISE CONVOLUTION OPERATION

Convolution in depth is a convolution operation performed separately for each channel of the input image. It is used to extract spatial features in each dimension. Point-by-point convolution is a standard convolution operation for the output feature map. Figure. 6 shows a depthwise and point-wise structure, where the size of the input image is $D_f \times D_f \times M$, where D_f is the height and width of the input image and M is the number of channels of the map, and the size of the output feature map obtained by the convolution is $D_g \times D_g \times M$ (D_g is the height and width of the output image), which is equal to the number of channels of the input image. It is used as the input for the next convolution. The size of the convolution kernels for point-by-point convolution is 1×1, and the number of channels on each convolution kernel must be equal to the number of channels of the input feature map. When the number of convolution kernels is N, the output feature map after convolution is $D_g \times D_g \times N$.

The feature map for the depthwise convolution output is written as:

$$\bar{G}_{k,l,m} = \Sigma_{i,j} K_{i,j,m} \cdot I_{k+i-1,l+j-1,m}, \qquad (4)$$

where I denote the input feature maps, $\bar{G}$ denotes the output feature maps, and K denotes the convolution kernels. The element position of the convolution kernel is determined by I and j. The values k and l determine the position of the input feature map element and the output feature map, respectively, while m represents the input feature map channel.

The calculation of the convolution parameters in depth and the cost function are denoted as follows:

$$F_2 = M \times{}_k^2 \qquad (5)$$

and

$$G_2 = M \times D_i^2 \times{}_k^2 \qquad (6)$$

The number of parameters is proportional to the number of feature mapping channels, and convolution kernels entered. The computational cost varies with the number of input feature mapping sources, convolution kernel, and quadratic input feature mapping function. The convolution depth parameters and computational cost need not consider the output feature mapping N. Compared with relations (2) and (3), equations (5), and (6) clearly show the simplicity of depthwise convolution. However, unlike a traditional convolutional layer, depth-wise convolution only filters the input channels and does not simply combine them to create new features. Therefore, to generate new features, an additional layer of 1×1 standard convolution is required [54]. The new feature map is generated by the adding the depth-wise and
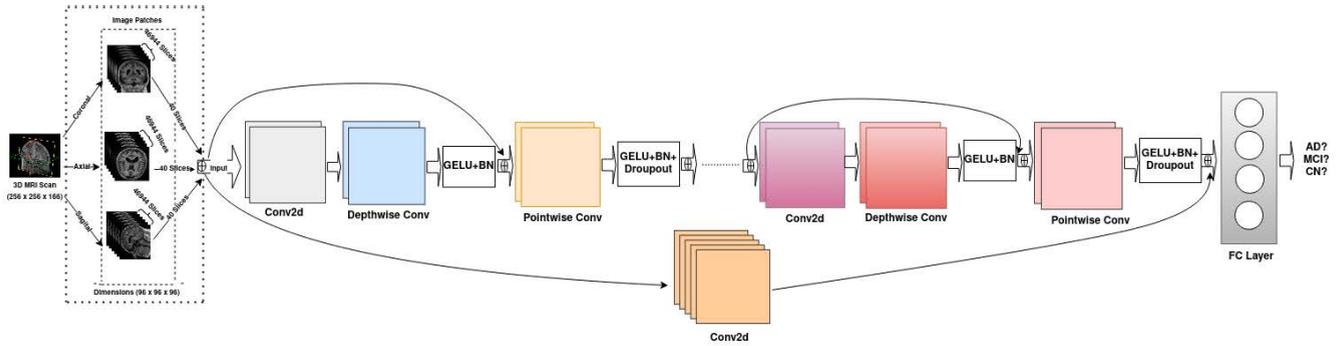
FIGURE 4. The layout of the proposed model. The **first part** is based on slice selection, whereas the **second part** introduces an isotropic architecture that allows the network to mix multiple feature maps that are more useful for classification.
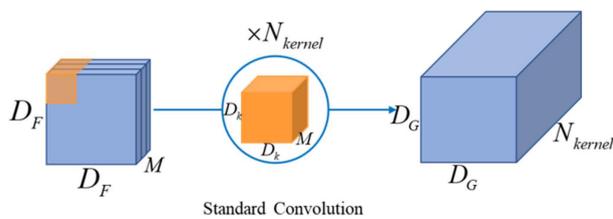
FIGURE 5. Demonstration of architecture of the standard convolution [53].
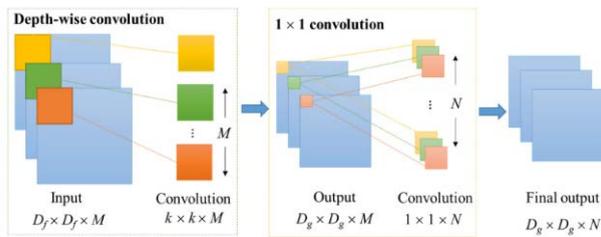
FIGURE 6. Depiction of depthwise convolution module and a pointwise convolution module [53].

1×1 pointwise convolution. Therefore, the parameters and the cost function can be calculated as shown in equations (7) and (8), respectively:

$$F_3 = M \times D_k^2 + M \times N \qquad (7)$$

and

$$G_3 = M \times D_i^2 \times D_k^2 + M \times N \times D_i^2 \qquad (8)$$

E. IMPLEMENTATION DETAILS
In this work, we have built our 2D model using 3D structural MRI scans of 489 patients (163 AD, 163 MCI, and 163 CN). The original MR images were resampled to $96 \times 96 \times 96$ resolution. A total of 40 brain fields were generated from each part of the subject (axial, coronal, and sagittal) for training and testing, resulting in 58680 feature fields, from which 19560 belonged to the AD group, 19560 to the MCI, and 19560 to the NC group.

During experiments, each group was randomly divided into three parts: Training (60% of the subjects), Validation (10% of the subjects), and testing (30% of the subjects). The proposed classification method implemented in Python 3.8.10 using the Keras library based on Tensorflow 2.5.0, and then run on a computer with NVIDIA RTX3090 GPU tested in the Ubuntu 20.04-x64 environment. Initially, the network parameters were initialized randomly and the Adaptive Moment Estimation (Adam) optimizer was used with an initial learning rate of 0.001 and a decay rate of 0.9. The batch size was configured to 32, and the dropout layer was added to avoid overfitting. We used the proposed algorithm to discriminate between AD patients, cognitively normal (CN) and mild cognitive impairment in a multiclass classification task (MCI), and to binary classify between AD and CN, AD and MCI, and CN and MCI.

IV. EXPERIMENT AND RESULTS
A. PERFORMANCE EVALUATION
In the diagnostic tasks, the predicted results are given as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP refers to a positive sample predicted correctly as a positive sample. TN refers to a negative sample correctly predicted as negative. The symbol FP means that a negative sample is falsely classified as a positive sample. The symbol FN means that a positive sample is falsely predicted as a negative sample. To evaluate our diagnostic model, we use the following widely used indicators: accuracy, specificity, sensitivity, precision, F1 score, and receiver operating characteristic curve (ROC curve). As shown in relation (9), accuracy is the proportion of correctly diagnosed samples among all test samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (9)$$

As shown in equation (10), specificity represents the percentage of samples correctly diagnosed.

$$Specificity = \frac{TN}{TN + FP} \qquad (10)$$

As shown in equation (11), sensitivity represents the ability of a model to identify AD patients in all positive samples.

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \quad (11)$$

Precision in equation (12) is defined as the percentage of correctly predicted positive observations out of all predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

The F1 score is calculated as the weighted average of Precision and Recall. As shown in equation (13).

$$F1 - score = 2(\frac{Precision \times Sensitivity}{Precision + Sensitivity}) \quad (13)$$

An ROC curve (receiver operating characteristic curve) is a graph that depicts a classification model's performance across all classification thresholds. This graph depicts two parameters:

First parameter is True Positive Rate (TPR) also referred as a recall (14).

$$TPR = \frac{TP}{TP + FN} \quad (14)$$

whereas, the second parameter is False Positive Rate (FPR) defined in equation (15).

$$FPR = \frac{FP}{FP + TN} \quad (15)$$

B. COMPARISON WITH BASELINE MODELS

The performance of the proposed method is evaluated using the ADNI database. The proposed method showed good performance in a multi-class classification task (AD versus MCI versus CN) with the highest accuracy of 96.41%. Comparison of the proposed work with other models trained on the same dataset is presented in Table 2. A comparison shows that the proposed model outperforms the benchmark models in terms of accuracy despite having less number of parameters.

ResNet and VGG Net were selected as the benchmark models for comparison. Using common indicators of classification performance such as sensitivity (SEN), specificity (SPE), precision, F1 score, and accuracy (ACC), results show that ResNet achieves an accuracy of about 95.34%, which is 2.08% higher than the VGG (93.26%) due to feature propagation enhancement and skip connection, but lower than the proposed work. Our model achieved the highest accuracy of 96.41%, which is 1.07% higher than ResNet and 3.15% higher than the VGG network. Our model also outperformed the baseline models in terms of other performance metrics. For example, it achieved a specificity of 97.73% and precision of 95.50%, which are higher than baseline models. With a sensitivity and F1 score of 95.31% and 95.32% respectively, ResNet was marginally better. Figure. 7 (a) shows the corresponding ROC for the ADNI dataset as well as the corresponding confusion matrix in Figure. 7 (b). Figure. 7 (c) and (d) depicts the accuracy

and loss graph respectively for both training and validation dataset. We have demonstrated the effectiveness of our CNN model developed in this study for both binary classifications and multiclass classifications. Training for binary classification is performed for three different scenarios: AD versus CN, AD versus MCI, and MCI versus CN.

C. COMPARISON WITH OTHER EXISTING METHODS

Using the ADNI database, we compared the classification results of our model with those from previous studies (see Table 3). We started by comparing our proposed model with traditional machine learning methods. Liu et al. [55] proposed ROI-based methods named as a whole-brain hierarchical network to extract brain features, which were then classified using machine learning methods such as multiple kernel boosting (MKBoost) algorithm. It achieved 94.65% for AD vs CN, 89.63% for AD vs MCI, and 85.79% accuracy for MCI vs CN classification using a single structural MRI modality dataset. Sun et al. [56] achieved similar results. In their study, proposed a new SVM-based learning method to extracted spatial-anatomical information and also introduce a group lasso penalty to induce the structure sparsity. Their proposed method achieved 95.1% for AD vs CN, 70.8% for MCI vs CN, and 65.7% accuracy for AD vs MCI classification. In addition, we compared our model with existing deep learning methods. In their work, Hosseini et al. [7] proposed a 3D convolutional auto-encoder-based method. To capture anatomical shape variations in structural MRI scans of the brain used a pre-trained model. Later, they tested their model with CAD Dementia MRI dataset without preparatory skull removing testing and achieved 89.1% accuracy in multi-class classification. Furthermore, on binary classification tasks, their model achieved 97.6% for AD vs NC, 95% for AD vs MCI, and 90.8% for MCI vs NC. Basaia et al. [40] proposed a deep learning algorithm for predicting individual Alzheimer's disease diagnosis based on a structural cross-sectional MRI scan. Their proposed model was 99.2% accurate for AD vs CN, 87.1% accurate for MCI vs CN, and 75.4% accurate for AD vs MCI. Whereas Liu.J et al. [57] used the OASIS dataset to construct a CNN-based architecture that achieved 78.02% accuracy for multiclass classification, 84.65% for MCI versus CN, 72.96% for AD versus MCI, and 75.2% accuracy for MCI versus CN classification using the ADNI data set. Later, in the same paper they improved their work to reduce the number of parameters using a deep separable convolution model and achieved 77.79% accuracy by reducing the parameters of the CNN model by 87.94%. To learn the features from the segmented part of the hippocampal, Liu.M et al. [57] developed an architecture combining 3D Densely connected convolutional networks (3D Dense Net) and multi-task CNN. They achieved 88.9% accuracy for AD vs NC and 76.2% accuracy for MCI vs NC. Moreover, Xu et al. [58] proposed a modified Tresnet architecture in their paper, to extract features from MRI images of the gray matter. They achieved 86.9% accuracy in classifying AD vs CN and 63.2% in classifying
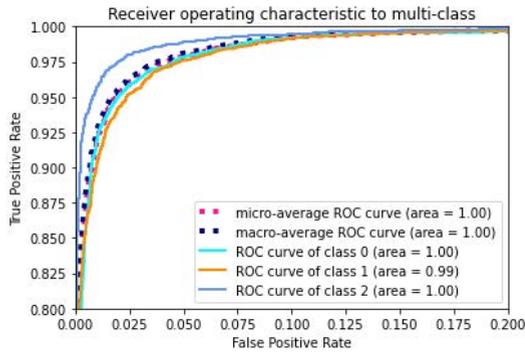
(a) Receiver operating characteristic (ROC) for multi class classification.
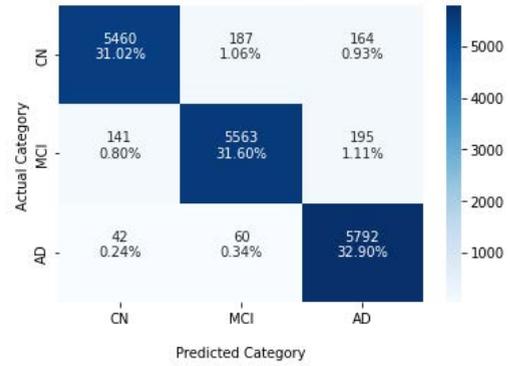
(b) Confusion Matrix for multi class classification.
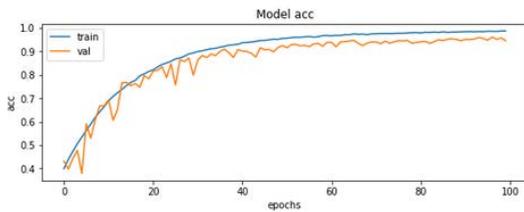
(c) Accuracy plot between training and validation data.
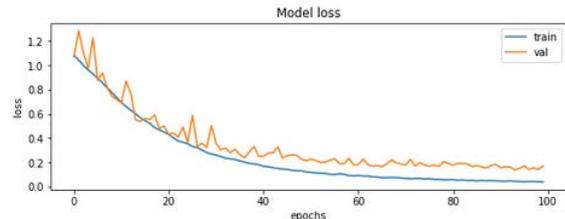
(d) Loss graph between training and validation.

FIGURE 7. Figure(a): Receiver operating characteristics (ROC) curve for CN (Class 0), MCI (Class 1) and AD (Class 2) Figure(b): Confusion matrix for testing data with true and predicted labels, Figure(c): Accuracy for training and testing data over 100 epochs Figure(d): Training and validation data los.

TABLE 2. Comparison of methods performance for multiclass classification.

Methods	# Parameters (Millions)	#FLOPs (Billions)	ACC	SEN	SPE	Precision	F1-score
ResNet50	23.587	0.65	95.27	**95.31**	97.63	95.39	**95.32**
VGG19	38.940	3.59	92.77	93.15	96.49	92.77	92.75
Proposed Model	**3.514**	1.42	**96.12**	94.99	**97.73**	**95.50**	95.23

FLOPs: Floating points operations;

Maximum value of each column is denoted by bold figures;

ACC: Accuracy; SEN: Sensitivity; SPEC: Specificity.

multiple classes. And lastly, Aderghal *et al.* [59] proposed a transfer learning scheme based on convolutional neural networks to automatically classify brain scans, relying on small ROI (few slices of the hippocampal region). Their evaluation shows that AD vs CN is 91.86%, AD vs MCI is 69.95%, and MCI vs CN is 68.52%. Table 3 summarizes the comparison of different studies, whereas Table 4 compares several proposed methods in terms of parameters. Both these tables show the discrimination tendency of the proposed method with a reduced number of parameters.

V. DISCUSSION
Effective and accurate diagnosis of Alzheimer's disease is critical for early intervention and treatment of the disease. Therefore, researchers have focused on developing computer-based systems to detect Alzheimer's disease at an early stage. Since then, CNN-based image classification has

been widely used in medical disease diagnosis. However, it is not possible or realistic to create an efficient CNN model capable of producing good results. In this study, we investigated the classification of MRI images using CNN features with improved accuracy and less number of parameters.

Previously, the contemporary models focused on increasing the depth and complexity of the network to improve classification performance. In this study, however, a new approach to reducing the number of parameters and the computational complexity of a CNN is presented. Earlier proposed models with increased depth suffered from the issues of vanishing gradient. In this regard, we attempted to mitigate the vanishing gradient problem, promote feature reuse, and greatly reduce the number of parameters by proposing a modified convolutional network. Where the network consisted of three different layer types. The first type was an input layer, into which the N gray level image X_n, $n \in [1, N]$ patches were

TABLE 3. Classification results of the proposed method and published methods on AD/MCI/CN, AD vs CN, MCI vs CN, and AD vs MCI.

References	Subjects	No. of Samples	Method	AD/MCI/CN ACC (%)	SEN (%)	SPEC (%)	AD vs CN ACC (%)	SEN (%)	SPEC (%)	MCI vs CN ACC (%)	SEN (%)	SPEC (%)	AD vs MCI ACC (%)	SEN (%)	SPEC (%)
Liu et al., 2018 [55]	MRI	AD-200, MCI-280, CN-230	MKBoost + SVM	-	-	-	94.65	95.03	91.76	84.79	88.91	80.34	88.63	91.55	86.25
Sun et al., 2018 [56]	MRI	AD-137, sMCI-134, pMCI-76, CN-162	Group Lasso, SVM	-	-	-	95.10	93.80	83.80	70.80	72.10	69.10	65.70	63.20	67.30
Hosseini et al., 2016 [7]	MRI	AD-70, MCI-70, CN-70	3D Convolutional Auto Encoder	89.10	-	-	97.60	-	-	90.80	-	-	95.00	-	-
Basaia et al., 2019 [40]	MRI	AD-294, MCI-253, CN-352	3D CNN	-	-	-	**99.20**	**98.90**	**99.50**	87.10	87.80	86.50	75.40	74.50	76.40
Liu J et al., 2021 [45]	MRI (OASIS)	AD-90, MCI-136, CN-266	Multi Layer Neural Network	78.02	83.21	75.32	82.50	-	-	84.65	82.35	79.50	72.96	78.34	82.15
Liu J et al., 2021 [45]	MRI (ADNI)	AD-99, MCI-353	Depth separable convolution	-	-	-	-	-	-	-	-	-	75.32	80.13	65.32
Liu M et al., 2020 [57]	MRI	AD-97, MCI-233, CN-119	3D DenseNet	88.90	86.60	90.80	-	-	-	76.20	79.50	69.80	-	-	-
Xu et al., 2021 [58]	MRI	AD-85, MCI-244, CN-133	Tresnet_L + SK_module	63.20	84.00	85.40	86.90	82.10	88.30	-	-	-	-	-	-
Aderghal et al., 2020 [59]	MRI	AD-294, MCI-672, CN-627	Le-Net	-	-	-	92.30	93.95	90.65	78.48	77.72	81.44	79.16	82.72	78.36
This Study	MRI	AD-163, MCI-163, CN-163	Proposed Method	**96.12**	**94.99**	**97.73**	97.00	94.78	98.40	**96.00**	**96.29**	**96.22**	**90.00**	**88.00**	**91.71**

Abbreviations: AD = Alzheimer's disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; MCI = Mild Cognitive Impairment; CN = Cognitive Normal; Maximum value of each column is denoted by bold figures;
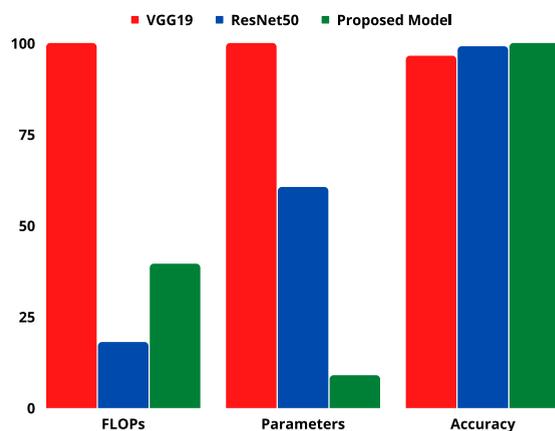
FIGURE 8. Comparison of models in terms of FLOPs, parameters and accuracy.

taken, then their pixels were scaled to 96×96 and normalized to the interval [0, 1] and fed into the network. The second type was a convolutional layer; Table 5 shows the proposed body architecture used in this work. The proposed CNN model was built by the interaction of four standard convolutional layers, and three convolutional blocks, each of which consists of a skip connection based depth-wise convolution (i.e., grouped convolution with groups equal to the number of channels) followed by point-wise (i.e., kernel size 1×1). Each convolution in the block is followed by the GELU activation function, Batch Normalization, and dropout layer. This model also uses a residual convolutional layer inspired by the skip connection model. The size of the convolution filters was set to 5×5, and the filter numbers were set to 256. This was maintained throughout the model to ensure that the same weights are shared across the different sets of pixels in an image.

The fully connected layer was the third type of layer and consisted of a set of input and output neurons that produced the learned linear combination of all neurons from the previous layer after passing through a non-linearity. The inputs and outputs of the fully connected layer were no longer spatially arranged but represented a 1D vector.

We used three CNN architectures in this paper: VGG, ResNet, and our proposed model. VGG networks are known for their uniformity, making them relatively easy to customize, verify, and use for a wide range of tasks; however, this property also makes them significant in terms of the number of parameters and hardware requirements. These drawbacks have been addressed in ResNet architectures, which also use concrete building blocks and allow the extraction of more complex patterns from data while increasing the number of layers and decreasing the number of parameters. Taking into account the above drawbacks, our architecture tries to address them through the concept of feature mixing as described in [60], [61]. We have chosen depth-wise convolution to mix spatial locations, point-wise convolution to mix channel locations, as well as the skip convolution layer to mix global

TABLE 4. Summarized results of the proposed method and published methods in terms of parameters.

Methods	Deep Learning Architecture	Parameters	Reduction(%)
Hosseini et al., 2016 [7]	3D Convolutional Auto Encoder	76.22	95.39%
Liu. J et al., 2021 [45]	Multi Layer Neural Network	29.80	88.22%
Liu. J et al., 2021 [45]	Depth separable convolution	11.15	68.52%
Xu et al., 2021 [58]	Tresnet_L + SK_module	52.0	93.25%
Aderghal et al., 2020 [59]	Le-Net	62.50	94.38%
Base Line Model	Resnet50	23.58	85.11%
Base Line Model	VGG19	38.94	90.98%
This Study	**Proposed Method**	**3.51**	-

*Parameters are measured in millions

TABLE 5. Proposed model layers.

Layer Type	Size	Input	Output
Conv1	5 x 5 x 256	96 x 96 x 1	48 x 48 x 256
DW Conv1(GELU + BN)	5 x 5 x 256	48 x 48 x 256	48 x 48 x 256
add1 (Conv1 + DW Conv1)		48 x 48 x 256	48 x 48 x 256
PW Conv1(GELU + BN)	1 x 1 x 256	48 x 48 x 256	48 x 48 x 256
Dropout1	0.5		
Conv2	5 x 5 x 256	48 x 48 x 256	24 x 24 x 256
DW Conv2(GELU + BN)	5 x 5 x 256	24 x 24 x 256	24 x 24 x 256
add2 (Conv2 + DW Conv2)		24 x 24 x 256	24 x 24 x 256
PW Conv2(GELU + BN)	1 x 1 x 256	24 x 24 x 256	24 x 24 x 256
Dropout2	0.5		
Conv3	5 x 5 x 256	24 x 24 x 256	12 x 12 x 256
DW Conv3(GELU + BN)	5 x 5 x 256	12 x 12 x 256	12 x 12 x 256
add3 (Conv3 + DW Conv3)		12 x 12 x 256	12 x 12 x 256
PW Conv3(GELU + BN)	1 x 1 x 256	12 x 12 x 256	12 x 12 x 256
Dropout3	0.5		
RL Conv	5 x 5 x 256	96 x 96 x 1	12 x 12 x 256
add4 (Dropout3 + RL Conv)		12 x 12 x 256	12 x 12 x 256
FC	3	256	3

Conv: Convolution Layer; DW Conv: Depthwise Convolution;

PW Conv: Point-wise Convolution ; RL Conv: Residual convolutional layer;

FC: Fully connected layer.

features from the data while reducing model complexity. In addition to that, from a practical point of view, hardware performance must be regarded as an important task. To accomplish this, floating-point operations (FLOPs) are also considered as an important predictor of energy usage and latency along with the number of parameters [62]. Generally speaking, the network with lower parameters and FLOPs requires less memory to save the model, requires less hardware memory, and thus it is more friendly to the embedded end. To this end, a comparison of accuracy and the computational costs of the benchmark and proposed model is depicted in Figure. 8. The advantage of the proposed model is that the number of FLOPs is significantly reduced as compared to the VGG19 model while reducing the number of parameters by 90.98%. The number of FLOPs is comparable with ResNet50; however, in terms of parameter count, the proposed method has 85.11% fewer parameters than ResNet50. Comparative experiments show that the proposed

model achieves higher accuracy results with a reduced number of parameters and the computational cost.

Furthermore, distinguishing individual groups, particularly MCI who were at high risk of developing AD, is critical for clinical control and management of the disease. We obtained an accuracy of 97.00% for AD vs. CN, 96.29% for MCI vs. CN, and 88.00% for AD vs MCI using the proposed method, indicating an improvement in performance compared to some popular network models and indicating the potential of the proposed model to detect subjects in prodromal dementia.

VI. CONCLUSION

We optimized a CNN for 3D whole-brain images using ADNI data and obtained the best accuracy with an isotropically repeated convolutional block network architecture. Contemporary state-of-the-art systems were outperformed by the proposed method. Moreover, our approach is fully automatic (i.e., no additional information input or manual intervention

is required) and extremely fast. The proposed method can be used to find meaningful patterns in data, confirm previous findings by specialists, aid in diagnostic scenarios, and eventually identify patterns for diseases other than Alzheimer's. Future research could look at obtaining similar or better results for images that have already been pre-processed for skull alignment and subtraction. Finally, it would be interesting to include patient history data to enrich the information available in the MRIs, drive the decision-making process, and link it to the patient's background.

REFERENCES

[1] S. Gauthier, P. Rosa-Neto, J. A. Morais, and C. Webster, "World Alzheimer report 2021: Journey through the diagnosis of dementia," Alzheimer's Disease Int., London, U.K., 2021. [Online]. Available: https://www.alzint.org/u/World-Alzheimer-Report-2021.pdf

[2] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, "A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification," *Magn. Reson. Imag.*, vol. 78, pp. 119–126, May 2021.

[3] M. Nuvolone, N. Schmid, G. Miele, S. Sorce, R. Moos, C. Schori, R. R. Beerli, M. Bauer, P. Saudan, K. Dietmeier, I. Lachmann, M. Linnebank, R. Martin, U. Kallweit, V. Kana, E. J. Rushing, H. Budka, and A. Aguzzi, "Cystatin f is a biomarker of prion pathogenesis in mice," *PLoS ONE*, vol. 12, no. 2, Feb. 2017, Art. no. e0171923.

[4] M. A. DeTure and D. W. Dickson, "The neuropathological diagnosis of Alzheimer's disease," *Mol. Neurodegeneration*, vol. 14, no. 1, pp. 1–18, Aug. 2019.

[5] D. P. Veitch, M. W. Weiner, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. C. Morris, R. C. Petersen, A. J. Saykin, L. M. Shaw, A. W. Toga, and J. Q. Trojanowski, "Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's disease neuroimaging initiative," *Alzheimer's Dementia*, vol. 15, no. 1, pp. 106–152, 2019.

[6] S. Lahmiri and A. Shmuel, "Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease," *Biomed. Signal Process. Control*, vol. 52, pp. 414–419, Jul. 2019.

[7] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 126–130.

[8] J. Islam and Y. Zhang, "Brain mri analysis for alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, pp. 1–14, 2018.

[9] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 295–308, Oct. 2018.

[10] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, "Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, Dec. 2018.

[11] T. Zhou, K.-H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Hum. Brain Mapping*, vol. 40, no. 3, pp. 1001–1016, 2019.

[12] S. Sarraf, D. D. DeSouza, J. Anderson, and G. Tofighi, "DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI," *BioRxiv*, 2017, Art. no. 070441.

[13] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018.

[14] J. Samper-González, N. Burgos, S. Bottani, S. Fontanella, P. Lu, A. Marcoux, A. Routier, J. Guillon, M. Bacci, J. Wen, A. Bertrand, H. Bertin, M.-O. Habert, S. Durrleman, T. Evgeniou, and O. Colliot, "Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data," *NeuroImage*, vol. 183, pp. 504–521, Dec. 2018.

[15] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[16] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[17] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018.

[18] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, Nov. 2015.

[20] Z.-Z. Wu, T. Weise, Y. Wang, and Y. Wang, "Convolutional neural network based weakly supervised learning for aircraft detection from remote sensing image," *IEEE Access*, vol. 8, pp. 158097–158106, 2020.

[21] R. K. Mishra, S. Urolagin, J. A. A. Jothi, A. S. Neogi, and N. Nawaz, "Deep learning-based sentiment analysis and topic modeling on tourism during covid-19 pandemic," *Frontiers Comput. Sci.*, vol. 3, Nov. 2021.

[22] J. Islam and Y. Zhang, "A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data," in *Proc. Int. Conf. Brain Inform.* Cham, Switzerland: Springer, 2017, pp. 213–222.

[23] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cogn. Syst. Res.*, vol. 57, pp. 147–159, Oct. 2019.

[24] H. T. Gorji and N. Kaabouch, "A deep learning approach for diagnosis of mild cognitive impairment based on MRI images," *Brain Sci.*, vol. 9, no. 9, p. 217, Aug. 2019.

[25] J. P. Kim, J. Kim, Y. H. Park, S. B. Park, J. S. Lee, S. Yoo, E.-J. Kim, H. J. Kim, D. L. Na, J. A. Brown, S. N. Lockhart, S. W. Seo, and J.-K. Seong, "Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease," *NeuroImage, Clin.*, vol. 23, Jan. 2019, Art. no. 101811.

[26] X. Long, L. Chen, C. Jiang, L. Zhang, and A. Disease Neuroimaging Initiative, "Prediction and classification of Alzheimer disease based on quantification of MRI deformation," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173372.

[27] H. Guo, F. Zhang, J. Chen, Y. Xu, and J. Xiang, "Machine learning classification combining multiple features of a hyper-network of fMRI data in Alzheimer's disease," *Frontiers Neurosci.*, vol. 11, p. 615, Nov. 2017.

[28] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and I. A. Illán, "Independent component analysis-based classification of Alzheimer's disease from segmented MRI data," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.* Cham, Switzerland: Springer, 2015, pp. 78–87.

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

[30] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G. R. Kwon, "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*, vol. 14, no. 10, Oct. 2019, Art. no. e0222446.

[31] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, Jan. 2017.

[32] S. Vieira, W. H. L. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neurosci. Biobehavioral Rev.*, vol. 74, pp. 58–75, Mar. 2017.

[33] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019.

[34] I. R. Silva, G. S. Silva, R. G. de Souza, W. P. dos Santos, and A. D. A. Roberta, "Model based on deep feature extraction for diagnosis of Alzheimer's disease," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–7.

[35] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, May 2014, pp. 1015–1018.

[36] S. Alinsaif and J. Lang, "3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data," *Comput. Biol. Med.*, vol. 138, Nov. 2021, Art. no. 104879.

[37] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling," *J. Med. Syst.*, vol. 42, no. 5, pp. 1–11, May 2018.

[38] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G.-R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of patch-based classifiers for diagnosis of Alzheimer diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019.

[39] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019.

[40] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage, Clin.*, vol. 21, 2019, Art. no. 101645.

[41] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 835–838.

[42] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, "Alzheimer's disease detection through whole-brain 3D-CNN MRI," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 1193, Oct. 2020.

[43] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease," *J. Neurosci. Methods*, vol. 339, 2020, Art. no. 108701.

[44] A. Basher, B. C. Kim, K. H. Lee, and H. Y. Jung, "Volumetric feature-based Alzheimer's disease diagnosis from sMRI data using a convolutional neural network and a deep neural network," *IEEE Access*, vol. 9, pp. 29870–29882, 2021.

[45] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, and Y. Bi, "Alzheimer's disease detection using depthwise separable convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 203, May 2021, Art. no. 106032.

[46] H. Wang, Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang, and X. Zhao, "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," *Neurocomputing*, vol. 333, pp. 145–156, Mar. 2019.

[47] X. Liu, Q. Xu, and N. Wang, "A survey on deep neural network-based image captioning," *Vis. Comput.*, vol. 35, no. 3, pp. 445–470, Mar. 2019.

[48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[49] M. Y. kamil, "A deep learning framework to detect COVID-19 disease via chest X-ray and CT scan images," *Int. J. Electr. Comput. Eng.*, vol. 11, pp. 844–850, Feb. 2021.

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

[51] A. Trockman and J. Z. Kolter, "Patches are all you need?" 2022, *arXiv:2201.09792*.

[52] A. R. Chetan, A. Rao, and P. K. J. Mohapatra, "COVID-19 face mask prediction using machine learning techniques," in *Proc. ICDSMLA*. Singapore: Springer, 2022, pp. 749–763.

[53] C.-F. Wang. (2018). *A Basic Introduction to Separable Convolutions*. Accessed: Apr. 12, 2022. [Online]. Available: https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728?source

[54] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[55] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 2, pp. 624–632, Mar. 2018.

[56] Z. Sun, Y. Qiao, B. P. F. Lelieveldt, and M. Staring, "Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer's disease classification," *NeuroImage*, vol. 178, pp. 445–460, Sep. 2018.

[57] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, and M. Xu, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, Mar. 2020, Art. no. 116459.

[58] Z. Xu, H. Deng, J. Liu, and Y. Yang, "Diagnosis of Alzheimer's disease based on the modified tresnet," *Electronics*, vol. 10, no. 16, p. 1908, Aug. 2021.

[59] K. Aderghal, K. Afdel, J. Benois-Pineau, and G. Catheline, "Improving Alzheimer's stage categorization with convolutional neural network using transfer learning and different magnetic resonance imaging modalities," *Heliyon*, vol. 6, no. 12, Dec. 2020, Art. no. e05652.

[60] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters–improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4353–4361.

[61] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-mixer: An all-MLP architecture for vision," 2021, *arXiv:2105.01601*.

[62] R. Tang, W. Wang, Z. Tu, and J. Lin, "An experimental analysis of the power consumption of convolutional neural networks for keyword spotting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5479–5483.

● ● ●