IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 July 2022, accepted 29 August 2022, date of publication 12 September 2022, date of current version 23 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3206389

RESEARCH ARTICLE

A Deep Convolutional Neural Networks Based Approach for Alzheimer's Disease and Mild Cognitive Impairment Classification Using Brain Images

RUHUL AMIN HAZARIKA[1,2], DEBDATTA KANDAR[1], (Member, IEEE), AND ARNAB KUMAR MAJI[1], (Senior Member, IEEE)

[1]Department of Information Technology, North Eastern Hill University, Shillong, Meghalaya 793022, India
[2]Department of CSE, Gandhi Institute of Technology and Management, Doddaballapura, Karnataka 561203, India

Corresponding author: Arnab Kumar Maji (akmaji@nehu.ac.in)

ABSTRACT Alzheimer's disease (AD) is a hazardous neurological disorder of people aged in the early 60s. The main symptoms of AD is significant memory loss. Mild Cognitive Impairment (MCI) is a state of dementia in which a patient exhibits the early symptoms of AD. Since brain is the most impacted region, the disorders can be classified by analyzing factors from brain tissues in different subjects. Machine Learning (ML) is a widely utilised concept that aids in the decision-making process. Deep Convolutional Neural Network (DNN) is a type of ML techniques that uses artificially connected neurons to mimic the human brain. In this work, we have proposed a novel DNN-based model for distinguishing AD and MCI patients from Cognitively Normal individuals. Inspired by the original VGG-19, we have created 19 deep layers in the network. In Back Propagation, deeper models suffer from the problem of vanishing gradient and information loss. As a solution, we borrowed the Dense-Block notion from the original DenseNet architecture, which provides a path of information exchange amongst all the layers. Furthermore, we have implemented depth-wise convolutional procedures to make the model computationally faster. Outcome of the proposed model is compared with some prominent DNN models and observed that, the proposed approach performs most convincingly with an average performance rate of 95.39%.

INDEX TERMS Alzheimer's disease (AD), mild cognitive impairment (MCI), deep convolutional neural network (DNN), cognitively normal (CN), machine learning (ML), DenseNet, VGG-19.

I. INTRODUCTION

Alzheimer's disease (AD) is one of the deadliest neurological diseases, impairing memory, decision-making, mood control, and other brain functions [1], [2]. In AD, grey cells in the brain that regulates the cognitive and emotional activities, such as the amygdala and hippocampus are significantly impaired [3], [4], [5]. Memory nerve cells are harmed initially, followed by the destruction of further grey cells, rendering the patient incapable of performing even the most basic tasks and suffering from extensive memory loss [6]. Mild Cognitive Impairment (MCI) is a state of schizophrenia in which a patient's cognitive deterioration is greater than that of a Cognitively Normal (CN) person of the same age [7], [8]. Though MCI patients have difficulties with communication, cognition, and reasoning skills, their problems are not as serious as those of AD. According to a study, most persons with MCI get AD after few years [8]. As a reason, it is imperative to diagnose MCI, and effective neurological therapy may help the patient prevent from developing AD.

Manual diagnostic process of AD is a challenging job. It's challenging for a psychologist to determine if a patient is

The associate editor coordinating the review of this manuscript and approving it for publication was Szidonia Lefkovits.

experiencing AD by the cognitive tests, since notable cognitive impairment is typical in normal ageing as well [9]. Furthermore, the entire process takes a long time, and psychiatrists may also need to take the help of neuro-imaging investigation after completing the manual tests. As a result, it is better way to classify AD relying on bio-markers found in neuro cells. Using neuro-imaging and extracting features out of each class, classification of AD can be done quite successfully [10]. Classical imaging modalities sometimes may not produce good findings in distinguishing characteristics from brain scans due to the intricate cell architecture. Hence, we have used the concept of Deep Convolutional Neural Networks or in short Deep Neural Networks (DNN) to classify AD, MCI, and CN subjects.

Artificial Neural Networks (ANN) which is a subclass of ML, creates some interconnected artificially created nerve cells that aids the system in assimilating fresh information out of its surroundings [11], [12]. An effective learning algorithm is applied to assign a precise weight to each of the neurons, and neurons are then begin to function as a set of processing units by sharing information with each other [13]. The network is expected to categorise the unfamiliar data after it has been properly trained. A DNN is a type of ANN in which numerous hidden layers assist the network in efficient training and producing desired results [14]. In the domain of computer vision, DNN is a commonly used technique, specially to deal with complex image processing applications such as MRIs [15], [16].

In this paper, we present a new architecture for the classification of CN, MCI, and AD from brain MR images using a DNN-based model. Our work's contribution can be summarised as follows:

- By taking VGG-19 as a reference model, we have created a model having 19 deep layers. The main motivation behind taking VGG as a reference model is that, i) VGG is a well-known model with a simple and effective architecture. VGG is a well performing classification model that has comparatively less number of layers than most of the highly performing models, ii) In the ''The ImageNet Large Scale Visual Recognition Challenge'' (ILSVRC)-2014, VGG won first place for localizing and second place for classifications [17], iii) VGG is a commonly used model in image processing applications (particularly as a medical image analysis framework) [18].
- Deeper models, such as VGG-19, have a number of difficulties, including gradient vanishing, information losses, and increased processing time. To mitigate gradient and information losses (particularly in back propagation), we used the Dense-Block notion from the DenseNet design, which provides a conduit for information sharing across all layers.
- To minimize the computational time required in the network, we have replaced all the convolutional operations by depth-wise convolutional operations. The use of depth-wise convolutions resulted in a significant reduction in computational time.
- We compared the proposed model's performance to those of other widely used models such as LeNet, AlexNet, VGG-16 & 19, Inception-V3,ResNet152-V2, InceptionResNet, MobileNet-V2, EfficientNet-B7, Xception, NasNet-C, and DenseNet-121. It is observed from the performance evaluations that the proposed model performs better and faster (in terms of training/testing time required per epoch).

The rest of the paper is organised as follows: a) in section 2, we discussed some of the recently published related state of the art, b) in section 3, we discussed the materials and methods used in this work, c) in section 4, we discussed briefly about the proposed architecture, d) in section 5, we discussed and compared the results of different DNN models, e) in section 6, we discussed and concluded the work, and f) in the last section, we have the references.

II. RELATED STUDY: AD CLASSIFICATION USING ARTIFICIAL NEURAL NETWORK

ANN is considered amongst one of the most effective approaches for AD categorization for its ability to learn from prior iterations and improve predictions in subsequent rounds [19]. This section discusses some of the recent publications on AD classifications utilising ANN-based techniques.

Jae Y Choi, *et al.* presented a new AD classifying method focused on the fusion of numerous DNN by accumulating the generalization Loss [20]. Taking brain MRIs as inputs, the authors proposed to combine several DNN based models together. For the combinations of DNNs, For brain images, several perspectives (axial, sagittal, and coronal) are combined to form assemblages for various DNN. A deep ensemble oriented generalisation loss towards discovering the most optimal weights among the neurons is used, which also aids in connecting and collaborating for the optimal weight searching. The referenced models for constructing the ensemble model are VGG-16, GoogLeNet, and AlexNet. The authors employed ADNI data and attained a 93.84% percent average classification performance rate.

In an article, Jong B Bae, *et al.* presented a novel ANN-based approach for classifying AD [21]. The training data-set is divided in five batches, each spanning the medial temporal lobe (TL) of thirty coronal slices. Shrinkage of the TL areas in various subject groups is investigated. FreeSurfer toolkit is used to do the pre-processing tasks, such as TL separation. For classifying AD, a CNN architecture based on the Inception-v4 model is constructed. To train the model, more than 100 AD and CN individuals are used. For validating the accuracy of the model, approximately 40 individuals from different categories are used. The average classification performance of the model is approximately 88.5%.

B S Rojas, *et al.* introduced a CNN-based strategy for classifying AD using DenseNet as a reference model [22].

Initially, the authors have acquired 3D MRIs, but in pre-processing, only 42 most suitable slices are selected for using in the model. DenseNet's Bottleneck-Compressed concept was used in the model. The M3d-Cam tool is combined with a Guided Gradient weighted Class Activation Mapping (Grad-CAM) technique to improve feature selection strategy. The procedure is known as attention mapping, and it aids in the discovery of undesired elements. The classification performance of the model is approximately 88.6%.

C Lian, *et al.* developed a network for combined shrinkage localisation and AD classification utilizing brain MRIs [23]. To determine the most discriminatory regions in the brain, a hierarchy CNN (H-CCN) based model is created. The most essential information are retrieved from the detected regions and are utilised in training the model. A voxel-wise anatomy is created for all linearly aligned images to get the approximated positions for extracting the features. An idea of using hybrid loss function is used to improve the outcomes. The average performance rate of the model is around 82.63%.

Y Zhao, *et al.* developed an CNN-based paradigm for identifying and estimating the development of Alzheimer's disease [24]. A 3D Multi-information Generative Adversarial Network is used in order to determine brain transformations as people grow older. A DenseNet-based design is implemented for classification tasks, that fundamentally optimises a focal degradation of brain scans to predict Alzheimer's phases. The model takes into account a variety of factors, including age, gender, and so on. The Voxel based morphometry toolkit is used in pre-processing that accomplish skull removal as well as segmentation of MRI scans into 3 sections (Grey Matter, White Matter, and Cerebrospinal fluid). The suggested model is designed to distinguish between distinct phases of dementia, including CN, MCI, Progressive MCI, Stable MCI, and AD. The average performance rate of the model is approximately 86.34%.

A unique framework integrating CNN and ensemble learning has been discussed in the literature for advanced diagnosis of Alzheimer's disease [25].Firstly, a variety of CNNs is built for diverse sagittal, coronal, and longitudinal nerve cells training dataset. For classifying, all of the CNNs are integrated into a single net. All of the brain cells have been transformed into space at the Montreal Neurological Institute (MNI). The zones in which majority of the pixels are overlapped are taken as the key bio-markers. There are 2 methods used to the ensemble learning process. For all entities in MNI space, a group of 123 various CNNs is built in step 1. For following procedures, the 5 highest ranked CNNs on every slices inclination are chosen. For the overall categorization, all 3 CNNs then integrated in phase 2. The overall performance of the model is around 75%.

A DNN based framework for AD classification is proposed by DH Chaihtra, et al [26]. The authors have used the transfer learning based on 4 pre-trained DNN models, including DenseNet121, MobileNet, InceptionV3 and Xception. From Kaggle dataset, the authors have acquired data for four different subject groups are, Mild Dementia, Moderate Dementia, Non Dementia, and Very Mild Dementia. After comparing the outcomes from all the 4 models, it is concluded that, DenseNet achieved the highest performance rate of around 91%.

Abol Basher, *et al.* proposed a unique technique to Alzheimer's disease categorization based on tissue-wise hippocampus characteristics extracted from brain scans [27]. A dual ensemble Hough CNN is used to choose the optimum slices for localising the hippocampus lobes. All 3D slices are converted to 2D and hippocampus patches are then extracted. The anthropometric detail from 2D slices is obtained using a DiscreteVolumeEstimationCNN (DCNN). Convolutional Layers, Rectified Linear Unit, Batch-Normalization layer, and Hidden Layers are utilised in the Hough CNN. Also in the DCNN, a set of Convolutional layers, Batch normalization layers, and a ReLu activation layer is used. The model achieved an average performance of 93%.

A new CNN based AD classification paradigm is presented by Liu *et al.* [28]. As per the authors claim, they only used a minimal number of MR scans for training and validation and yet got good results. For faster training, Depth-wise Separable Convolution (DwSC) is used in place of normal convolutional layers. The authors applied transfer learning from 2 popular DNN models, AlexNet and GoogLeNet, to achieve better and more precise categorization. The overall performance of the model is around 92.21%.

Jingwen Sun, *et al.* presented a new CNN model for dementia categorization [29]. A revised functional 3-D DNN is designed for conducting two tasks simultaneously: hippocampal separation and AD diagnoses using MRI scans. The authors combined the concept of V-Net and DenseNet models, where the lower elements in V-Net are replaced by the bottle neck of DenseNet. Following the acquisition of separated hippocampal areas, the separated data are transmitted to a 3D CNN for categorization of Alzheimer's disease. Regional hippocampus characteristics and also structural similarity from brain imaging are exploited for individual classification. Furthermore, the authors proposes a new loss estimation method that aided in the production of persuasive findings. The average performance of the model is around 86.2%.

III. MATERIALS AND METHODS

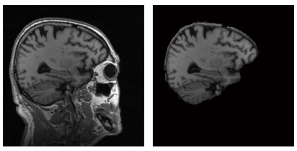
A. DATA AND TOOLS

All data utilised in this study are obtained from the Alzheimer's Disease Neuroimaging Initiative's public data sets (ADNI) [30]. The acquired data are in the form of volumetric T1-weighted, Magnetization Prepared Rapid Gradient Echo (MP-RAGE) MRIs. More than 2500 images are collected for three patient groups (CN, MCI, AD). The images are taken from 240 people (CN=80, MCI=80, AD=80), 120 of whom are male and 120 of whom are female.

Python is a popular programming language tool that is often used in clinical computer vision tasks [31]. Python is better than several other tools in terms of implementation

TABLE 1. Average performance of various skull stripping methods.

Algorithm	Accuracy	Sensitivity
Region-growing	0.62	0.68
Histogram-based	**0.85**	**0.90**
Fuzzy C-means	0.53	0.77
K-Means	0.64	0.75
Region-Splitting & Merging	0.61	0.74



a b

FIGURE 1. Example of an input Brain MR image, and it's skull stripped image.

because of its simple as well as user-friendly features [32]. We used Python 3.0 for implementing all the models. For expanding the training dataset, we utilized data generating methods including rotations, contrast adjustment, inverting, and so on. Detail data distribution is presented in Table 2.

B. PREPROCESSING

Acquired brain MRIs also include some non-brain elements, called skull. Existence of the skull may enhance the dimensionality of data and participation of the skull component can be ignored in AD categorization. As a result, we separated the MRI scans from the skull components. In a previous work, we evaluated several of the most widely used separation approaches & select the most appropriate functioning approach (Histogram Based Thresholding approach) [33]. The average performances of all the implemented segmentation methods are presented in Table 1.

A visual outcome of the segmentation approach is shown in Figure 1.

C. DATA DISTRIBUTION

The structure of brain cells changes with ageing [34], [35]. In a previous research, the cumulative hippocampal anomalies in brains for various patients were evaluated with the help of an experienced physician and a radiologist [36]. Additionally, the average Grey Matter (GM) volume differences in brains for AD, MCI, and healthy controls are investigated in a different study [37]. The ages of the individuals have been used for analyzing the differences in brain patterns are of between 60 to 90 years. If we divide the subjects in various groups such as, CN1 (CN patients aged 60-69 years), CN2 (CN patients aged 70-79 years), CN3 (CN patients aged 80+ years), MCI1 (MCI patients aged 60-69 years), MCI2 (MCI patients aged 70-79 years), MCI3 (MCI patients aged 80+ years), AD1 (AD patients aged 60-69 years), AD2 (AD patients aged 70-79 years), and AD3 (AD patients aged 80+ years), then from the study we can observe the following

TABLE 2. Data distributions.

Subjects	No. of individuals	classes	No. of training images	No. of testing & validation images	Total images
CN	80	CN1	1300	500	5400
		CN2	1300	500	
		CN3	1300	500	
MCI	80	MCI1	1300	500	5400
		MCI2	1300	500	
		MCI3	1300	500	
AD	80	AD1	1300	500	5400
		AD2	1300	500	
		AD3	1300	500	
Total	**240**		**11700**	**4500**	**16200**

outcomes:

1. GM volume/size: $CN1 > CN2 > CN3, MCI1 > MCI2 > MCI3, AD1 > AD2 > AD3$
2. Hippocampus volume/size: $CN1 > CN2 > CN3, MCI1 > MCI2 > MCI3, AD1 > AD2 > AD3$
3. GM atrophy: $CN3 > CN2 > CN1, MCI3 > MCI2 > MCI1, AD3 > AD2 > AD1$
4. Hippocampus atrophy: $CN3 > CN2 > CN1, MCI3 > MCI2 > MCI1, AD3 > AD2 > AD1$

Hence, to analyze the model in more effective way, we have further distributed the data in different sub-classes namely CN1, CN2, CN3, MCI1, MCI2, MCI3, AD1, AD2, and AD3. Table 2 depicts the data organisation.

IV. ARCHITECTURE OF THE PROPOSED DNN MODEL FOR AD CLASSIFICATION

The architecture of the proposed DNN model is presented in Figure 2. Inspired by the original VGG-19 model, in the proposed architecture, we have taken 19 deep layers (Convolution layers: 16, Dense layers: 3). For pooling operations we have used the MaxPooling layers. For information sharing, shortcut bridges are created from output of each pooling layers to the input of all next convolution layers in forward direction. If we divide the whole architecture in different blocks, then each block of the architecture will look like as Figure 3.

In the input layer, brain images of different subject groups are used for training as well as testing the model. As discussed in Table 1, 11700 no.s of training and 4500 no.s of testing and validations images are used.

Next layer in the model is for performing convolutional operations. Convolution layers comprised of a series of characteristic layouts that are used to retrieve crucial image features including boundaries, curves, and so on. Kernels are a series of quadratic arrays of the same parameters. Convolution is the process of rolling and overlaying filters across the whole image pixels. In the suggested model, normal convolution processes are computationally expensive since the model shares information from each output layer to next all the layers in forward direction. To overcome this issue, we have used the Depth_wise_Convolutional (DwC) operations which is a popular method for reducing executing cost and enhancing

representation effectiveness [38]. For all channels within the frames, DwC utilises separate kernels and a point-wise 1×1 convolutional procedure is used to integrate all of the outcomes across distinct channels. Equation 1 expresses the DwC mathematically.

$$\hat{D}_{a,c,d} = \sum_{x,y} \hat{P}_{x,y,d} \cdot Q_{a+x-1,c+y-1,d} \tag{1}$$

where, $\hat{P}$ denotes a DwC filter of size $R_P \times R_P \times A$ (R_P denotes the spatial dimension, and A denotes the sum of input-channels). Moreover, d^{th} kernel in $\hat{P}$ is employed in d^{th}-channel of Q, for producing the d^{th} channel for the filtered feature map $\hat{D}$. Equation 2 can be used to estimate the DwC's computational load.

$$R_P \times R_P \times A \cdot R_Q \cdot R_Q \tag{2}$$

The regular convolution process has a computation time of $R_P \times R_P \times A \cdot B \cdot R_Q \cdot R_Q$ (B stands for the number of output channels), that is higher than the DwC. The whole cost of DwC, involving point convolution operation, can be written as Equation 3.

$$R_P \times R_P \times A \cdot R_Q \cdot R_Q + A \cdot B \cdot R_Q \cdot R_Q \tag{3}$$

Equations 4 and 5 can be used to represent the total cost drop.

$$\frac{R_P \times R_P \times A \cdot R_Q \cdot R_Q + A \cdot B \cdot R_Q \cdot R_Q}{R_P \cdot R_P \cdot A \cdot B \cdot R_Q \cdot R_Q} \tag{4}$$

$$\frac{1}{B} + \frac{1}{R_P^2} \tag{5}$$

The model's subsequent layer is utilised to reduce the density of feature arrays. The step is called pooling operation, in which less significant features are eliminated. In the proposed model, MaxPooling layer is used where only the highest valued items in a kernel get chosen and transmitted towards the next layer. Equation 6 is a mathematical representation of the Max pooling process.

$$Max_{x,y} = max(K_{x-m,y-n}, \forall 1 \leq x \leq m \text{ and } 1 \leq y \leq n) \tag{6}$$

Loss function is determined once all the data-set has been redistributed, usually totally or in chunks. The sum of faults specified at the predicted outcomes is known as loss function or Forward Propagation [39]. Out of many popular loss functions, one of the most widely used technique is the Binary Cross-Entropy (BiCE). We have used BiCE in our model which can be presented as Equation 7.

$$BiCE : \varphi(w) = \frac{1}{m} \sum_{j=1}^{m} p_j \log[f(p_j; w)] + (1 - p_j) \log[(1 - f(p_j; w)] \tag{7}$$

where, p_j represents real and $f(p_j; w)$ represents the anticipated parameters.

Once determining the loss estimate, next task is to compute the gradients for all of the essential parameters and enhance them using an effective approach. Equations 8 and 9 would be used to represent the process.

$$Gradient = \frac{\partial \varphi(w)}{\partial w} \tag{8}$$

$$W_{new} = W_{current} - \beta \frac{\partial \varphi(w_{current})}{\partial w_{current}} \tag{9}$$

In Equation 9, β represents learning rate.

Since the proposed model is deeper in nature, it has a number of difficulties, including gradient vanishing, information losses. To overcome the issue, we used the Dense-Block like notion from the DenseNet design, which provides a conduit for information sharing across all layers, which can be seen in Figure 3. Equation 10 can be deduced from Figure 3.

$$L_j = L_j + \sum_{m=1}^{j-1} f(L_m, w_m) \tag{10}$$

In Equation 10, $L_0, L_2, .., L_j$ represent the j^{th} block, and $w_1, w_2, \ldots, w_j$ represents the weight of the block. Back propagation of a block can be presented as Equation 11.

$$\frac{\partial \varphi(w)}{\partial L_1} = \frac{\partial \varphi(w)}{\partial L_j} \cdot \frac{\partial L_j}{\partial L_1} = \frac{\partial \varphi(w)}{\partial L_j} \left(1 + \frac{\partial}{\partial L_1} \sum_{m=1}^{j-1} f(x_m, w_m)\right) \tag{11}$$

In Equation 11, φ is the loss function.

After a series of DwC and pooling layers, we have dense or fully connected layers. All neurons in dense layers are interconnected to one another. For example, 'k' is a neuron in dense layer, and $l_1, l_2, l_3, \ldots\ldots, l_i$ are the input weights from n_i different neurons. Equation 12 can be used to express the outcome of 'k'.

$$k_{out} = \sum_i l_i \times n_i + bias \tag{12}$$

Following computing 'k_{out}' the cell 'k' must make a resolution of the dispute. The activating factor *phi* utilized in the model determines the response 'Out' of 'k' which may be expressed as Equation 13. Some of the popular activating factors includes the ReLu, Softmax, Sigmoid, etc.

$$Out = \phi(k_o ut) \tag{13}$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

Experimental setup: We used a CPU of i-7 processor with 16 GB RAM, 500 GB SSD storage, 2 GB graphics, and Windows-10 OS to execute the model. Python is a widely utilised tool in computer vision for its client-friendly interface as well as quick execution capabilities [31], [32]. Python 3.0 is used in this study. With a data batch size of 32, the model is trained for a total of 60 epochs.

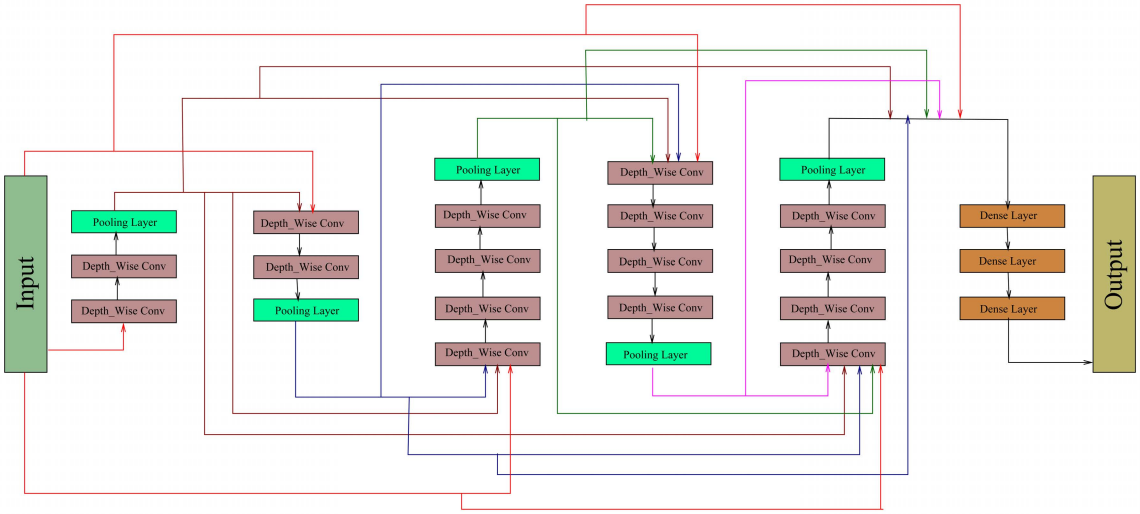


FIGURE 2. Proposed DNN model for AD classification.

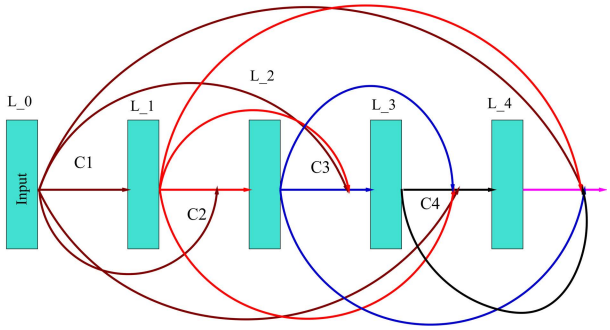


FIGURE 3. A block of the proposed model.

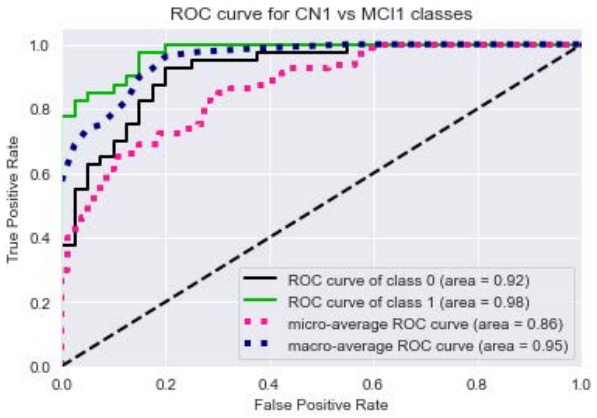


FIGURE 4. ROC curve for CN1 vs MCI1 classes.

We have implemented some of the commonly used DNN models (LeNet, AlexNet, VGG-16, VGG-19, Inception V3, ResNet152-V2, InceptionResNet-V1, MobileNet-V2, Efficient-B7, Xception, NASNet-C, and DenseNet-121) for performance comparison with the proposed model. The average performance of some existing DNN models is presented in Table 3.

Table 3 shows the average performance based on the values obtained from several performance measurement criteria such as Accuracy, Precision, Recall, and F1-Score. From Table 3, it can be observed that, DenseNet achieved the highest average performance rate amongst all the compared existing DNN models.

One of the key reasons behind DenseNet-121's improved performance is the use of DenseBlock, which results in the least information loss over the network [17].

Architecture of the proposed model is shown in Figure 2 and Figure 3. Layer-wise connectivity is presented in Figure 2, whereas a sample block of the model is presented.

The average performance of the proposed model is presented in Table 4. From Table 4, it can be observed that the average performance of the proposed model is approximately 95.39%, which is more than all the implemented existing models. The proposed model outperforms all of the previously discussed state of the art models. from Table 4, it can also observed that the average time taken per epoch is 125 seconds which is computationally faster than original DenseNet-121 (865 seconds), and VGG-19 (288 seconds).

The ROC (Receiver Operating Characteristic) curve for all the classes are also obtained as shown in Figure 4 to Figure 12.

From Figure 4, it can be observed that, while classifying the classes CN1 vs MCI1, the average ROC score is 0.95. The mean micro & macro average ROC score is 0.905.

As presented in Figure 5, while classifying CN2 vs MCI2 classes, the average ROC score of both the classes is achieved as 0.92, and the mean micro & macro average ROC score is 0.915.

As shown in Figure 6, the proposed method achieved an average ROC score of 0.955 while classifying CN3 vs MCI3

TABLE 3. Average performance of some existing DNN models for AD classifications.

Sl. No.	**Model**	Average performance	Average training/testing time (per epoch)
1	LeNet	0.8158	73 seconds
2	AlexNet	0.7211	92 seconds
3	VGG-16	0.8124	195 seconds
4	VGG-19	0.8722	288 seconds
5	Inception-V3	0.8544	225 seconds
6	ResNet152-V2	0.8811	828 seconds
7	InceptionResNet-V1	0.8723	1092 seconds
8	MobileNet-V2	0.8821	525 seconds
9	Efficient-B7	0.7632	915 seconds
10	Xception	0.8831	877 seconds
11	NASNet-C	0.88	862 seconds
12	DenseNet-121	0.8985	865 seconds

TABLE 4. Performance evaluation table of the proposed model.

Model	Class	Accuracy	Precision	Recall	F1 Score	Average performance	Average time required per epoch
Proposed Model	CN1 vs MCI1	0.97	0.96	0.95	0.95	0.9539	125 seconds
	CN2 vs MCI2	0.95	0.96	0.96	0.95		
	CN3 vs MCI3	0.96	0.94	0.95	0.92		
	MCI1 vs AD1	0.93	0.94	0.95	0.95		
	MCI2 vs AD2	0.92	0.95	0.94	0.96		
	MCI3 vs AD3	0.95	0.97	0.96	0.95		
	CN1 vs AD1	0.95	0.97	0.96	0.97		
	CN2 vs AD2	0.97	0.96	0.95	0.98		
	CN3 vs AD3	0.93	0.97	0.96	0.97		

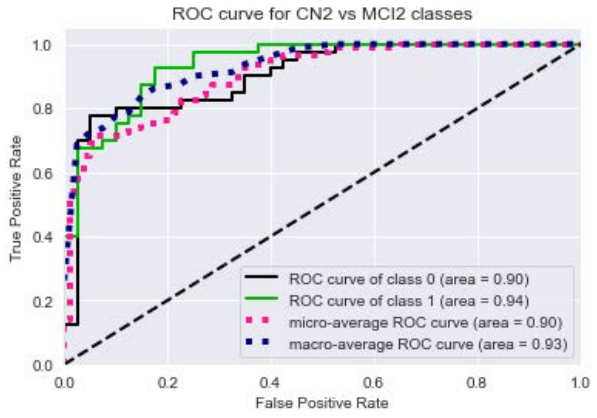


FIGURE 5. ROC curve for CN2 vs MCI2 classes.

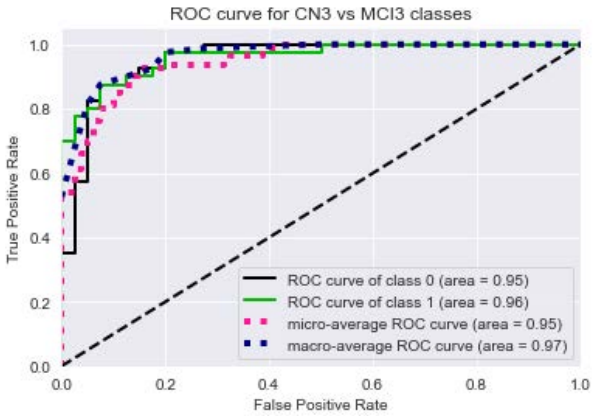


FIGURE 6. ROC curve for CN3 vs MCI3 classes.

classes. The micro & macro average ROC score for the same classes is achieved as 0.96.

The ROC curve of MCI1 vs AD1 is presented in Figure 7. It can be observed from Figure 7 that the average ROC score is achieved as 0.98. For the same classes, the mean micro & macro average ROC score is 0.985.

Figure 8 represented the ROC curve obtained by using the proposed model for MCI2 vs AD2 classes. It is observed from Figure 8 that the average ROC score is 0.91 and the mean micro & macro average ROC score is 0.91.

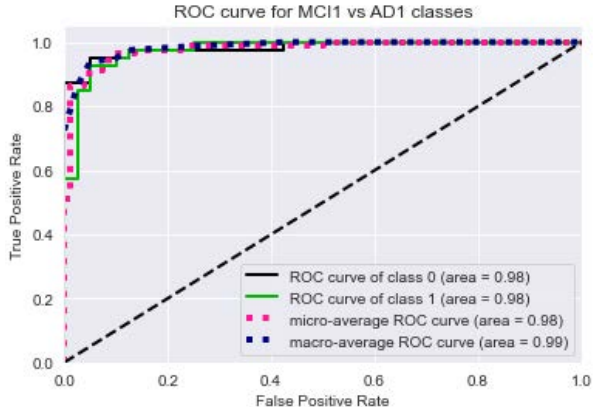


FIGURE 7. ROC curve for MCI1 vs AD1 classes.

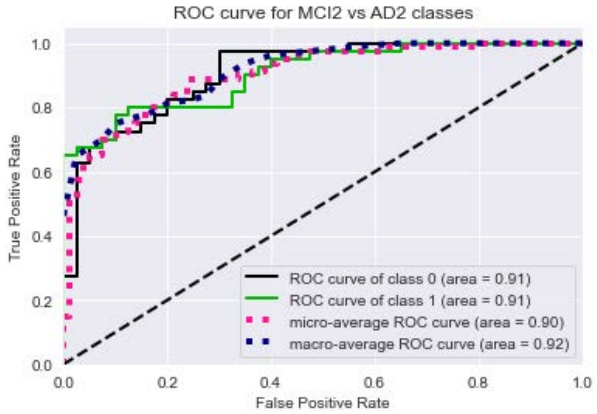


FIGURE 8. ROC curve for MCI2 vs AD2 classes.

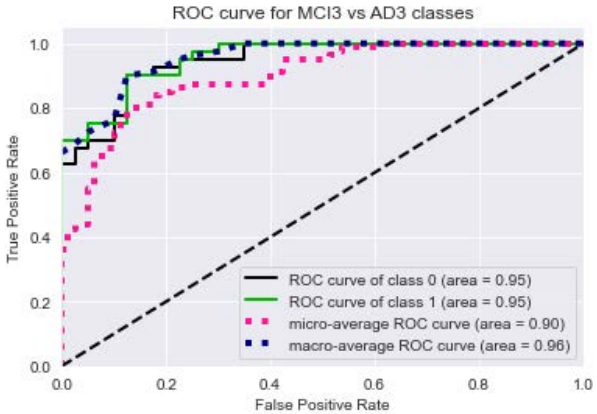


FIGURE 9. ROC curve for MCI3 vs AD3 classes.

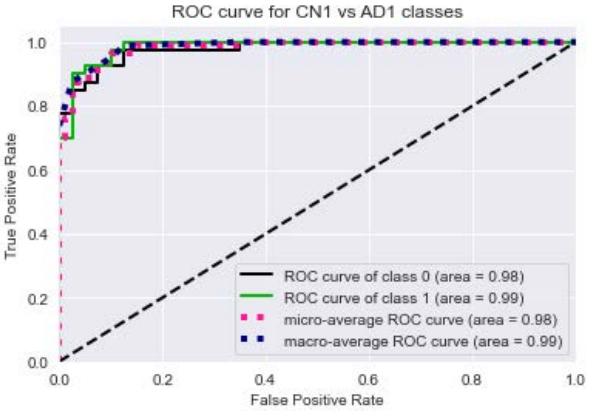


FIGURE 10. ROC curve for CN1 vs AD1 classes.

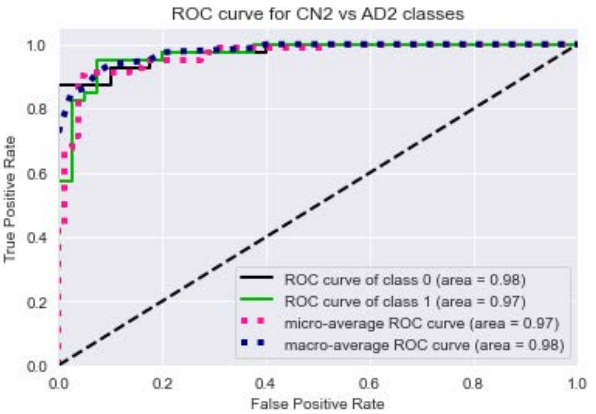


FIGURE 11. ROC curve for CN2 vs AD2 classes.

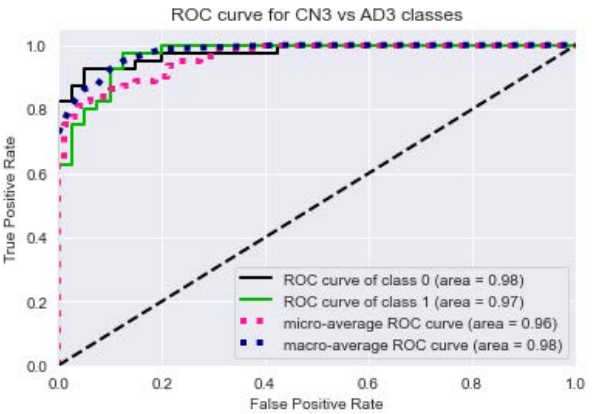


FIGURE 12. ROC curve for CN3 vs AD3 classes.

ROC curve of MCI3 vs AD3 is presented in Figure 9. The average ROC score is found as 0.95 for the proposed model. The mean micro & macro average ROC score is 0.93.

In figure 10, the ROC curve of CN1 vs AD1 is presented. It is observed that, the proposed model achieved average ROC score of 0.985 and the mean micro & macro average ROC score is 0.985.

As presented in Figure 11, the average ROC score achieved by the proposed model for CN2 vs AD2 classes is 0.975 and the mean micro & macro average ROC score is 0.975.

In Figure 12, the ROC curve of CN3 vs AD3 classes are presented. It is observed that the average ROC score of the proposed model for CN3 vs AD3 classes is 0.975 and the mean micro & macro average ROC score is 0.97.

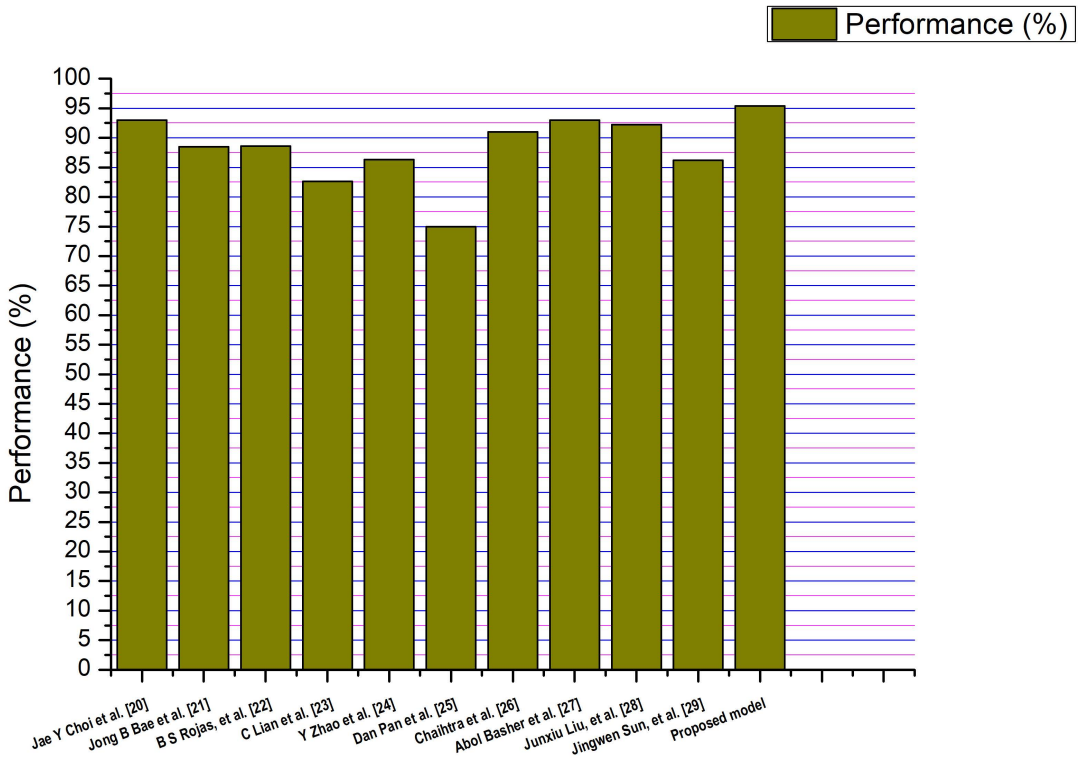


FIGURE 13. Performance comparison graph amongst the discussed state-of-arts.

Performance of the proposed model is compared with the discussed state-of-arts as presented in Figure 13.

From the performance comparison graph in Figure 13, it is observed that the proposed model has the highest average performance in classifying CN, MCI, and AD as compared to some of the recently published state-of arts.

VI. CONCLUSION AND FUTURE WORK

A novel DNN-based model for AD classification is proposed in this paper. For classification, three separate dementia phases are considered: CN, MCI, and AD. All essential data for training and testing the model is obtained from the ADNI online public data set. VGG-19 is one of the most effective models for image classification. Hence, the proposed architecture (with 19 deep layers) of the model is designed by taking VGG-19 as a reference model. However, one of the primary issues with deeper models like VGG-19 is gradient vanishing and information loss. To minimize this issue, we adopted the Dense_Block notion from the original DenseNet model and used a forward-direction shortcut bridge connection from each of the output layers to all of the next input layers. However, a model with too many bridge connections may become computationally expensive due to the enormous number of convolutional calculations. The depth_wise convolutional operation is an effective way to get the task done better and faster. As a result, we employed depth_wise convolutions in the model instead of regular convolutional layers. The proposed model's performance is compared to that of some of the most widely used DNN models, as well as some of the most current state-of-the-art models. Based on all of the performance comparisons, the proposed architecture appears to be the most convincing.

Despite the fact that the proposed model performs admirably, there is still some work may be done in the future. Firstly, this model holds a significant number of features due to too many bridge connections. An appropriate feature minimization method can be utilised in the future to minimize extraneous features from feature maps. GradCam/ScoreCam visualisation can be also added to analyse the model's performance in future. Secondly, data from several sources can be accumulated in the future to enhance performance evaluations. Furthermore, several more dementia stages, including the early stages of Alzheimer's disease, such as Progressive MCI, Stable MCI, and so on, can be added as classification classes.

VII. AUTHOR CONTRIBUTIONS

All authors are responsible for analysis, conceptualization, and writing the original manuscript. All authors have read and agreed to the published version of the manuscript.

VIII. CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

REFERENCES

[1] I. O. Korolev, "Alzheimer's disease: A clinical and basic science review," *Med. Student Res. J.*, vol. 4, no. 1, pp. 24–33, 2014.

[2] Alzheimer's Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 14, no. 3, pp. 367–429, Mar. 2018.

[3] R. Donev, M. Kolev, B. Millet, and J. Thome, "Neuronal death in Alzheimer's disease and therapeutic opportunities," *J. Cellular Mol. Med.*, vol. 13, nos. 11–12, pp. 4329–4348, 2009.

[4] S. W. Moon, B. Lee, and Y. C. Choi, "Changes in the hippocampal volume and shape in early-onset mild cognitive impairment," *Psychiatry Invest.*, vol. 15, no. 5, pp. 531–537, 2018.

[5] J. Barnes, J. L. Whitwell, C. Frost, K. A. Josephs, M. Rossor, and N. C. Fox, "Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration," *Arch. Neurol.*, vol. 63, no. 10, pp. 1434–1439, 2006.

[6] R. A. Hazarika, A. K. Maji, S. N. Sur, B. S. Paul, and D. Kandar, "A survey on classification algorithms of brain images in Alzheimer's disease based on feature extraction techniques," *IEEE Access*, vol. 9, pp. 58503–58536, 2021.

[7] Y. Varatharajah, V. K. Ramanan, R. Iyer, and P. Vemuri, "Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics," *Sci. Rep.*, vol. 9, no. 1, pp. 1–15, Dec. 2019.

[8] National Institute on Aging. *What is Mild Cognitive Impairment?* Accessed: Nov. 15, 2021. [Online]. Available: https://www.nia.nih.gov/health/what-mild-cognitive-impairment

[9] J. A. Kaye, "Diagnostic challenges in dementia," *Neurology*, vol. 51, pp. S45–S52, Jul. 1998.

[10] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, 2019.

[11] S.-C. Wang, "Artificial neural network," in *Interdisciplinary Computing in Java Programming*. Boston, MA, USA: Springer, 2003, pp. 81–100.

[12] J. F. Pagel and P. Kirshtein, *Machine Dreaming and Consciousness*. New York, NY, USA: Academic, 2017.

[13] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inform.*, vol. 35, nos. 5–6, pp. 352–359, 2002.

[14] V. V. Raghavan, V. N. Gudivada, V. Govindaraju, and C. R. Rao, *Cognitive Computing: Theory and Applications*. Amsterdam, The Netherlands: Elsevier, 2016.

[15] Y. Xue, Y. Wang, J. Liang, and A. Slowik, "A self-adaptive mutation neural architecture search algorithm based on blocks," *IEEE Comput. Intell. Mag.*, vol. 16, no. 3, pp. 67–78, Aug. 2021.

[16] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[17] R. A. Hazarika, D. Kandar, and A. K. Maji, "An experimental analysis of different deep learning based models for Alzheimer's disease classification using brain magnetic resonance images," *J. King Saud Univ., Comput. Inf. Sci.*, Sep. 2021.

[18] T. Kaur and T. K. Gandhi, "Automated brain image classification based on VGG-16 and transfer learning," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Dec. 2019, pp. 94–98.

[19] C. Dumitru and V. Maria, "Advantages and disadvantages of using neural networks for predictions," *Ovidius Univ. Ann., Ser. Econ. Sci.*, vol. 13, no. 1, pp. 444–449, 2013.

[20] J. Y. Choi and B. Lee, "Combining of multiple deep networks via ensemble generalization loss, based on MRI images, for Alzheimer's disease classification," *IEEE Signal Process. Lett.*, vol. 27, pp. 206–210, 2020.

[21] J. B. Bae, S. Lee, W. Jung, S. Park, W. Kim, H. Oh, J. W. Han, G. E. Kim, J. S. Kim, J. H. Kim, and K. W. Kim, "Identification of alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, 2020.

[22] B. Solano-Rojas and R. Villalón-Fonseca, "A low-cost three-dimensional DenseNet neural network for Alzheimer's disease early discovery," *Sensors*, vol. 21, no. 4, p. 1302, Feb. 2021.

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

[24] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, and S. Li, "Prediction of Alzheimer's disease progression with multi-information generative adversarial network," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 711–719, Mar. 2021.

[25] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early detection of Alzheimer's disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning," *Frontiers Neurosci.*, vol. 14, p. 259, May 2020.

[26] D. Chaihtra and S. V. Shetty, "Alzheimer's disease detection from brain MRI data using deep learning techniques," in *Proc. 2nd Global Conf. Adv. Technol. (GCAT)*, Oct. 2021, pp. 1–5.

[27] 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.

[28] 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.

[29] J. Sun, S. Yan, C. Song, and B. Han, "Dual-functional neural network for bilateral hippocampi segmentation and diagnosis of Alzheimer's disease," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 3, pp. 445–455, Mar. 2020.

[30] ADNI. *Alzheimer's Disease Neuroimaging Initiative: ADNI*. Accessed: Jul. 13, 2020. [Online]. Available: http://adni.loni.usc.edu/data-samples/access-data

[31] R. Folks, "Using the Python programming language for image processing in nuclear medicine," *J. Nucl. Med.*, vol. 55, no. 1, p. 1322, 2014.

[32] S. G. Virupakshappa, R. Sequeira, A. Rastogi, and N. Jain, "Essence of Python programming language in medical image analysis: Enhancing workplace productivity," in *Proc. Eur. Cong. Radiol.*, 2018, pp. 1–15.

[33] R. A. Hazarika, K. Kharkongor, S. Sanyal, and A. K. Maji, "A comparative study on different skull stripping techniques from brain magnetic resonance imaging," in *Proc. Int. Conf. Innov. Comput. Commun.* Singapore: Springer, 2020, pp. 279–288.

[34] R. Peters, "Ageing and the brain," *Postgraduate Med. J.*, vol. 82, no. 964, pp. 84–88, 2006.

[35] L. L. Beason and B. Horwitz, "Aging brain," Tech. Rep., 2002.

[36] R. A. Hazarika, A. K. Maji, D. Kandar, P. Chakrabarti, T. Chakrabarti, K. J. Rao, J. Carvalho, B. Kateb, and M. Nami, "An evaluation on changes in hippocampus size for cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) patients using fuzzy membership function," North-Eastern Hill Univ., Shillong, India, Tech. Rep., 2021.

[37] R. A. Hazarika, A. K. Maji, S. N. Sur, I. Olariu, and D. Kandar, "A fuzzy membership based comparison of the grey matter (GM) in cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) using brain images," *J. Intell. Fuzzy Syst.*, vol. 43, no. 2, pp. 1779–1792, Jun. 2022.

[38] A. N. Gomez, L. M. Kaiser, and F. Chollet, "Depthwise separable convolutions for neural machine translation," U.S. Patent 10 853 590, Dec. 1, 2020.

[39] I. Mebsout. *Deep Learning's Mathematics*. Accessed: May 23, 2021. [Online]. Available: https://towardsdatascience.com/deep-learnings-mathematics-f52b3c4d2576

RUHUL AMIN HAZARIKA received the B.Tech. and M.Tech. degrees in information technology from North Eastern Hill University (NEHU), Shillong, India, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree in information technology. He is also working as an Assistant Professor with the Department of Computer Science Engineering, Gandhi Institute of Technology and Management, Doddaballapura, Karnataka, India. His research interests include AI and image processing.

DEBDATTA KANDAR (Member, IEEE) was born in Deulia, Kolaghat, Purba Medinipur, West Bengal, India, in 1977. He received the M.Tech. degree, in 2006, and the Ph.D.(Engg.) degree from the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, West Bengal, in 2011. He has more than 18 years of professional experience. He has completed his postdoctoral research at the Department of Electrical and Electronic Engineering Technology, University of Johannesburg, Johannesburg, South Africa. He is currently a Professor with the Department of Information Technology, North Eastern Hill University, Shillong, India. He has successfully guided several Ph.D., M.Tech., and B.Tech. students. Four of his students already awarded Ph.D. degree. He has national and international collaborations for carrying out research works. He has published approximately 100 research papers in different national and international journals, conference proceedings, book chapters, and edited book. His research interests include wireless mobile communications, artificial intelligence, soft computing, and radar operation.

ARNAB KUMAR MAJI (Senior Member, IEEE) received the B.E. degree in information science and engineering from Visvesvaraya Technological University (VTU), in 2003, the M.Tech. degree in information technology from Bengal Engineering and Science University, Shibpur (currently IIEST, Shibpur), in 2005, and the Ph.D. degree from Assam University, Silchar (A Central University of India), in 2016. He has approximately 18 years of professional experience. He is currently working as an Associate Professor with the Department of Information Technology, North Eastern Hill University, Shillong (A Central University of India). He has published more than 32 number of articles in different reputed indexed journals, and around 50 numbers of articles as book chapter and conference proceedings. He also edited three books with reputable publishers like Springer. He has also published two Indian patents and filed one patent. He has successfully guided four numbers of Ph.D. scholars, and 19 numbers of M.Tech. thesis. He has active collaboration with several reputed Indian and foreign institutes. His research interests include machine learning, image processing, and natural language processing. He is also a reviewer of several reputed international journals and a guest editor of one Springer journal.