



# A Review of Deep Transfer Learning Approaches for Class-Wise Prediction of Alzheimer's Disease Using MRI Images

Pushpendra Singh Sisodia<sup>1</sup> · Gaurav Kumar Ameta<sup>2</sup> · Yogesh Kumar<sup>3</sup> · Neelam Chaplot<sup>4</sup>

Received: 3 August 2022 / Accepted: 14 December 2022 / Published online: 3 January 2023

© The Author(s) under exclusive licence to International Center for Numerical Methods in Engineering (CIMNE) 2023

## Abstract

Alzheimer's disease is an irreversible, progressive neurodegenerative disorder that destroys the brain and memory functionalities. In Alzheimer's disease, the brain starts shrinking, and over time it converts into dementia. The diagnosis of dementia takes an ample amount of time, around 2.8 to 4.4 years after the first clinical symptoms arise. Alzheimer's disease cannot be cured by any pharmacologic therapies (drugs) now on the market. Alzheimer's disease can only be avoided by early detection and prompt treatment. This paper proposes deep transfer learning models and MRI (Magnetic Resonance Imaging) images to detect the multiple stages of Alzheimer's disease such as "Very-Mild -Demented," "Mild-Demented," "Moderate-Demented," and "No-Demented." Data preprocessing and augmentation process are applied, enabling the model to detect the correct class of Alzheimer's disease. Then further deep transfer learning models (Resnet50, VGG19, Xception, DenseNet201, and EfficientNetB7) are used to classify and predict the early stages of Alzheimer's disease. It is observed that the DenseNet201 model performs the best, with a validation accuracy of 96.59%. The performance of Resnet50, VGG19, Xception, and EfficientNetB7 models was also recorded with validation accuracy of 93.52%, 95.08%, 89.77%, and 83.20%, respectively. The probability curve is then measured and the class-wise prediction of Alzheimer's disease is recorded using the area under curves and receiver operating curve (AUC-ROC) in order to analyze it more deeply.

## 1 Introduction

Alzheimer's disease is an irreversible, progressive neurodegenerative disorder that destroys memory and brain functionalities [1, 2]. It is comprised of several symptoms

like loss of memory & vision, problem in speaking, loss of motivation, difficulty in making critical decisions, mood swings, etc. [1–3]. It is estimated that approx. 6.5 million Americans, over 65 and older have Alzheimer's disease. According to the research, there may be 13.8 million cases of Alzheimer's by 2060. In 2019, the official death recorded due to Alzheimer's disease was 121,499. Between 2020 and 2021, Alzheimer's disease will rank as the 6<sup>th</sup> leading cause of mortality in the US. Alzheimer's disease increased by more than 145% between 2000 and 2019 [4]. The diagnosis of dementia is problematic because it takes an ample amount of time, around 2.8 to 4.4 years, after the first clinical symptoms arise [4, 5]. Patients with advanced Alzheimer's disease may suffer a variety of symptoms that evolve over time. The severity of the nerve cell injury in various areas of the brain is reflected in these symptoms. Alzheimer's disease can be categorized on the basis of the degree of damage to nerve cells that arise over time as mild, very mild, and moderate. Most persons with moderate Alzheimer's disease can operate independently, but they will likely need help with daily tasks [6–12]. The moderate stage of Alzheimer's disease is characterized by personality and behavioral changes, including suspicion and agitation, as well as communication

✉ Pushpendra Singh Sisodia  
pushpendra.muji@gmail.com

Gaurav Kumar Ameta  
gauravameta1@gmail.com

Yogesh Kumar  
yogesh.arora10744@gmail.com;  
Yogesh.Kumar@sot.pdpu.ac.in

Neelam Chaplot  
neelam.chaplot@gmail.com

<sup>1</sup> Indus Institute of Technology & Engineering, Indus University, Ahmedabad, Gujarat, India

<sup>2</sup> Department of CSE, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India

<sup>3</sup> Department of CSE, School of Technology, Pandit Deendayal Energy University, Gandhinagar, India

<sup>4</sup> Department of CSE, Manipal University Jaipur, Jaipur, Rajasthan, India

and everyday task difficulties. When Alzheimer's disease is advanced, people need assistance with daily tasks and probably need round-the-clock care. Researchers are still struggling to find the proper treatment for Alzheimer's disease. Alzheimer's disease is not curable with any of the pharmacologic therapies (drugs) now on the market. The only means of preventing Alzheimer's disease is by early detection and prompt treatment [13]. To address the issue of early stages detection of Alzheimer's disease, we used the MRI (Magnetic Resonance Imaging) image and deep transfer learning models in this paper. The Convolution Neural Network (CNN) architectures are used to entirely avoid the feature extraction process and rely on the CNN feature learning process. The transfer learning process is used to reusing the already trained model's weights [14, 15]. The research study has provided accurate, efficient, and cost-effective solutions by using a better approach and techniques in the current research. The proposed system has been designed to detect the multiple stages of Alzheimer's disease, such as "Very-Mild-Demented," "Mild-Demented," "Moderate-Demented," and "No-Demented." Data pre-processing and augmentation techniques are applied, enabling the system to detect the correct class of Alzheimer's disease. This study used the deep learning CNN models that helped us to automate the feature extraction process. The study used Resnet50, VGG19, Xception, DenseNet201, and EfficientNetB7 deep learning models to classify and predict all four stages of Alzheimer's disease. It is observed that the DenseNet201 model performs the best, with a validation accuracy of 96.59%. The performance of Resnet50, VGG19, Xception, and EfficientNetB7 models are also recorded with validation accuracy of 93.52%, 95.08%, 89.77%, and 83.20%, respectively. The rest of this paper is laid out as follows: Sect. 2 is devoted to related work, which includes a review of the available literature and a comparative study. The research challenges and gaps in the available literature are reviewed in Sect. 3. Section 4 describes the proposed system, including the study's objectives, data preprocessing, data augmentation, and the methodology employed in the study. The experimental results acquired from the study with various performance indicators are the focus of Sect. 5. Finally, Sect. 6 discussed the study's pertinent findings and potential future improvements.

## 2 Background Study

The striking similarities between structural brain imaging of a healthy individual, and an Alzheimer's patient, initial stages Alzheimer's disease prediction is a challenging issue. Initial stages detection of Alzheimer's disease can help old people stop the progression of the disease at the initial stage. Recently, Machine learning algorithms have been used to

predict the early stages of Alzheimer's disease. The machine learning classification techniques such as Decision Tree, Support Vector Machine, Random Forest, Gradient Boosting, and Voting with Open Access Series of Imaging Studies (OASIS) data are used to predict Alzheimer's disease [16]. Brain imaging techniques such as Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and MRI biomarkers are used to predict Alzheimer's disease [17]. A recent study suggested that deep learning techniques perform better than conventional machine learning algorithms. Deep learning techniques are more efficient in identifying complex and high-dimensional structured data. A hybrid stacked auto-encoder along with machine learning is used for feature selection in Alzheimer's disease. Deep Learning based models such as Convolution Neural Networks (CNNs) are practical and efficient in early disease diagnosis [18, 19]. CNN model is used along with MRI (Magnetic Resonance Imaging) and clinical test data to classify Alzheimer's disease stages. Stack de-noising auto-encoders are used to extract features from clinical data, and 3D-Convolutional Neural Network (3D-CNN) for image data analysis [20]. A deep learning model is developed to diagnose Alzheimer's disease and mild cognitive impairment. This model works on the cross-sectional images of the brain [21]. The VGG-19 model is used along with the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset for the detection of Alzheimer's disease [22]. Deep learning models like ResNet18 and DenseNet201 are also used to perform the task of Alzheimer's disease multiclass classification [23]. An improved State-of-the-art deep learning-based pipeline technique was used to classify Alzheimer's disease or not [24]. The deep triplet network with a conditional loss function is used to overcome the lack of data samples and improve the accuracy of the model as well [25]. Recently, Transfer learning techniques based on ResNet50, VGG16, etc., was used to predict Alzheimer's disease [26]. Convolution Neural Network-CNN is the primary deep learning architecture for the prediction and detection of Alzheimer's disease till now. CNN models are successfully implemented in the clinical field because of less processing time, high accuracy, and the ability to generalize so that they could be applicable to other healthcare applications [27]. Chances of Alzheimer's disease identification at preliminary stages can be boosted by incorporating advanced deep learning practices by using a combination of the different datasets, i.e., Alzheimer Disease Neuroimaging Initiative (ADNI), Open Access Series of Imaging Studies (OASIS) [28]. A detailed background study of Alzheimer's disease is mentioned in Table 1.

Despite several studies on Alzheimer's disease prediction, the following research gaps are identified in the background study.

**Table 1** Background study of Alzheimer's disease

Authors	Dataset	Methodology used	Challenges	Results (%)
Venugopalan et al. (2021) [20]	SNP, MRI imaging, clinical, and neurological test data	KNN, SVM, Random Forests and Auto-Encoder	Small dataset size. Models must be tested on larger, more comprehensive datasets	Precision: 83 Recall: 83 F1 Score: 83 Accuracy: 84
Mehmood et al. (2021) [29]	ADNI dataset of 300 patients data with six binary classes	CNN, and VGG-19 Architecture	To apply the proposed model on breast cancer and lung cancer detection required some additional models for better comparisons	Accuracy: 83.72 Sensitivity: 82.09 Specificity: 85.13
Murugan et al. (2021) [30]	6400 MRI images from Kaggle repository and ADNI dataset	Dementia Network (DEMNET) and CNN model, SMOTE technique	The dataset's main flaw is a class imbalance. So, adequate data is required for fine-tuned model	Accuracy: 95.23 AUC: 97 Kappa: 0.93
Janghel and Rathore (2020) [31]	fMRI, and PET images	CNN VGG-16 used for feature extraction And SVM, Linear Discriminate, K-means clustering, and Decision tree are used for classification	Average accuracy of PET dataset is low	Accuracy: 99.95 fMRI dataset Accuracy: 73.46 PET dataset
Togacar et al. (2021) [32]	MRI images	VGG-16, SVM and LR	The performance of the proposed method improved by incorporating attention modules	Accuracy: 99.94
He et al. (2021) [33]	ADNI (MRI)	3D CNN and Multimodal Fusion Method	The models is compress for more efficiency, speed and accuracy	Accuracy: 97
Kim et al. (2021) [34]	FDG PET, ADNI KBASE, University Hospital Dataset	2.5D Deep Learning Model, 3D Deep Learning Model, and Amyloid PET positive categorization algorithm based on deep learning	To incorporate developed model into a crucial aid in medical decision-making for categorizing amyloid PET positive	Accuracy: 75 Sensitivity: 76 Specificity: 74 F1-score: 75 AUC: 86
Huggins et al. (2021) [35]	Raw EEG Signals	AlexNet Deep Learning Architecture, k-fold cross validation	The recording of EEG signals from scalp should be done in a controlled environment; it needs proper clinical space and experts. It makes the process time consuming and expensive	Accuracy: 98
Li et al. (2021) [36]	ADNI, (MCI)	ResNet Framework, Deep-learning genomics (DLG) model	The only input for the DLG categorization was gene sequences. Future brain imaging techniques might combine gene sequences with clinical information	Accuracy: 98:78 Sensitivity: 98:39 Specificity 99:44
Saratxaga et al. (2021) [37]	OASIS-1 (436), OASIS-2 (373) (Open Access Series of Imaging Studies)	BrainNet2D, BrainNet3D, and ResNet 18	Main drawback is scarcity of data for making comparisons. It needs additional datasets to do valid comparisons. Skull removing issue in image also should be carefully considered	OASIS-1 BrainNet2D-Accuracy: 79 BrainNet3D- Accuracy:80 OASIS-2 BrainNet2D- Accuracy:82 BrainNet3D- Accuracy:84 ResNet18:93

Table 1 (continued)

Authors	Dataset	Methodology used	Challenges	Results (%)
Koga et al. (2021) [38]	2522 digital slide images	YOLOv3 object detection algorithm, and Random forest classifiers	In laboratory variation may occur in the results due to thickness of the paraffin-embedded tissues	Accuracy: 92
Hu et al. (2021) [39]	ADNI and NIFD dataset	CNN based architecture	The performance on the multi-classification tasks in the lesser datasets	Accuracy: 92
Yang et al. (2021) [40]	ADNI -1257 subjects and Japanese-ADNI 286 subjects dataset	VGG-16, Ensemble Learning, Cox model	The proposed framework is based on exploration of the optimal prognostic signatures	Accuracy: 83 Sensitivity: 76 Specificity: 90 AUC: 91
Shen et al. (2021) [41]	More than 180 million clinical notes for 2.9 million individuals were gathered from the University of Minnesota's Clinical Data Repository (CDR) (UMIN)	NLP models and transformer-based bidirectional encoder representations	BERT models need very large dataset into the training phase and required large computation power	For Physical Activity: Precision: 96 Recall: 96 F1-Score: 96 For Excessive diet: Precision: 100 Recall: 100 F1-Score: 100
Buwaneswari and Gayathri (2020) [42]	ADNI dataset	The SegNet model is used for Feature Extraction, and ResNet-101 for Classification	The classification with medium amount of dataset requires additional information for more accurate classification	Normal Case: Accuracy: 93 Precision: 98 Specificity: 96 Sensitivity: 96 MCI Case: Accuracy: 95 Precision: 97 Specificity: 94 Sensitivity: 96 AD Case: Accuracy: 96 Precision: 97 Specificity: 94 Sensitivity: 97
Chen and Xia (2021) [43]	ADNI and MCI dataset	Iterative Space and Deep Learning Model (ISDL)	Examining the sparsity in multi-modality imaging data and selecting the sparse features within each anatomical structure is a challenging operation	For AD Classification- Accuracy: 95 Sensitivity: 91 Specificity: 94 For pMCI and sMCI Accuracy: 78 Sensitivity: 72 Specificity: 76

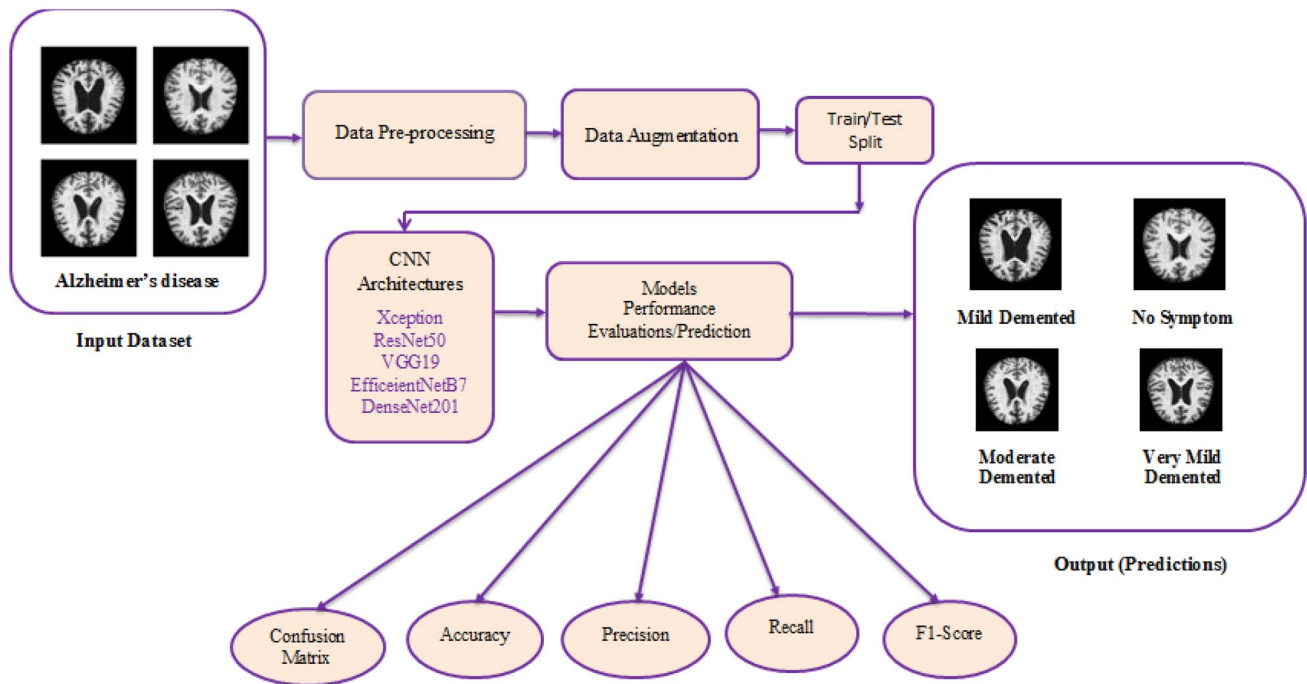


Fig. 1 System design for the early classification of Alzheimer's disease

1. Developing the Deep Learning model from scratch is required substantial training datasets.
2. Studies on transfer learning techniques are in initial stage.
3. Very few studies discussed the early stages of prediction, such as "Very-Mild-Demented," "Mild-Demented," "Moderate-Demented," and "No-Demented," of Alzheimer's disease.
4. Most of the studies classified Alzheimer's disease in binary classification that classifies the disease as "Alzheimer's disease" or "No Alzheimer's disease."

To address research gaps 1 and 2, we have used the transfer learning technique that requires less training dataset. We have classified and predicted Alzheimer's disease in "Very-Mild-Demented," "Mild-Demented," "Moderate-Demented," and "No-Demented," to address the research gaps 3 and 4.

The detailed framework is mentioned in Sect. 3.

### 3 Proposed Framework

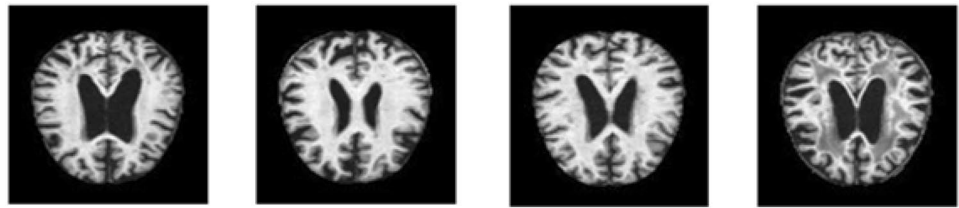
In this section, a brief description of the dataset and the suggested framework has mentioned. Figure 1 mentions the sequential framework that was employed in this investigation. A dataset is gathered and prepared using the suggested methods from the publicly available Kaggle dataset repository. The in-depth explanation of the dataset is mentioned in

Sect. 3.1. In the Sect. 3.2, the study described the deep transfer learning technique. Further, data augmentation and data split techniques for model validation purposes are elaborated in Sect. 3.3. The effective deep learning models ResNet50, VGG19, Xception, EfficientNetB7, and DenseNet201 are used to classify the subsequent stages of Alzheimer's disease. Section 3.3.1 mentions a thorough discussion of each deep learning model's design. The details of trainable and non-trainable parameters are discussed in Sect. 3.3.2. Further, accuracy, precision, recall, and f1-score metrics are used to gauge and compare the deep learning models' performance (ref. Section Results).

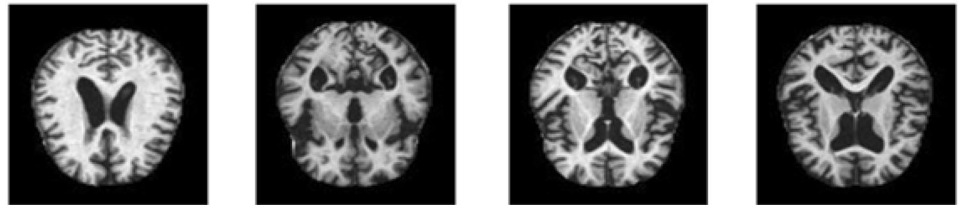
#### 3.1 Dataset Description

The dataset utilised in this study is composed of 6400 brain MRI images of Alzheimer's disease that were retrieved from the Kaggle repository. This dataset is further divided into four stages, "Mild-Demented," "Moderate-Demented," "Non-Demented," and "Very-Mild-Demented," which consists of 896, 64, 3200, and 2240 MRI images, respectively. The width and height of all the images are  $128 \times 128$  pixels. The Alzheimer's disease dataset is further divided with 80% and 20% ratios for the training and validating of the deep learning models, respectively. In Fig. 2, samples of the Alzheimer's disease dataset are mentioned with their respective categories. The distribution of the dataset is drawn using a bar graph and mentioned in Fig. 3.

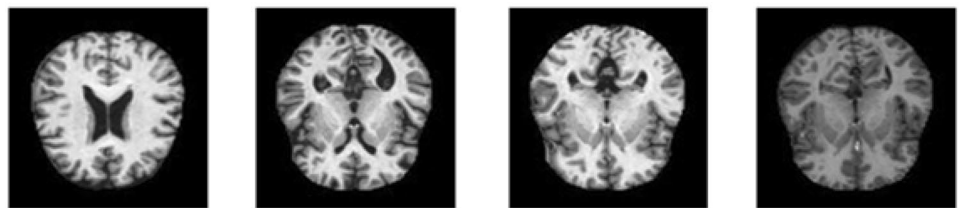
**Fig. 2** Sample dataset of Alzheimer's disease used in this study



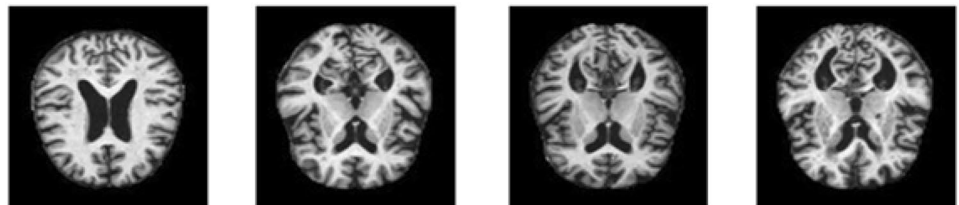
(a) Mild-Demented



(b) Moderate-Demented



(c) Non-Demented



(d) Very-Mild-Demented

The data set is accessible to everyone at <https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset>.

### 3.2 Dataset Augmentation and Split

The image augmentation technique helps to add additional training datasets by using minor perturbations to the input dataset. The purpose of image augmentation is to increase the size of the input image dataset by generating new synthetic training images. Data augmentation improves the model training efficiency by adding perturbations to input

training data, and it also makes the model generalized. All the perturbations added to the input dataset retained the same labels as the input training data. Figure 4 illustrates the augmented dataset. The data augmentation includes the image rotation, crop, flip, affine and AdditiveGaussianNoise etc. Data splitting is the necessary step to make sure the proper training and validation of the learning models. In this study, a total of 6400 brain MRI images related to Alzheimer's disease are used. The ratio of these MRI images is further split into training and test purposes with 80% and 20%, respectively.

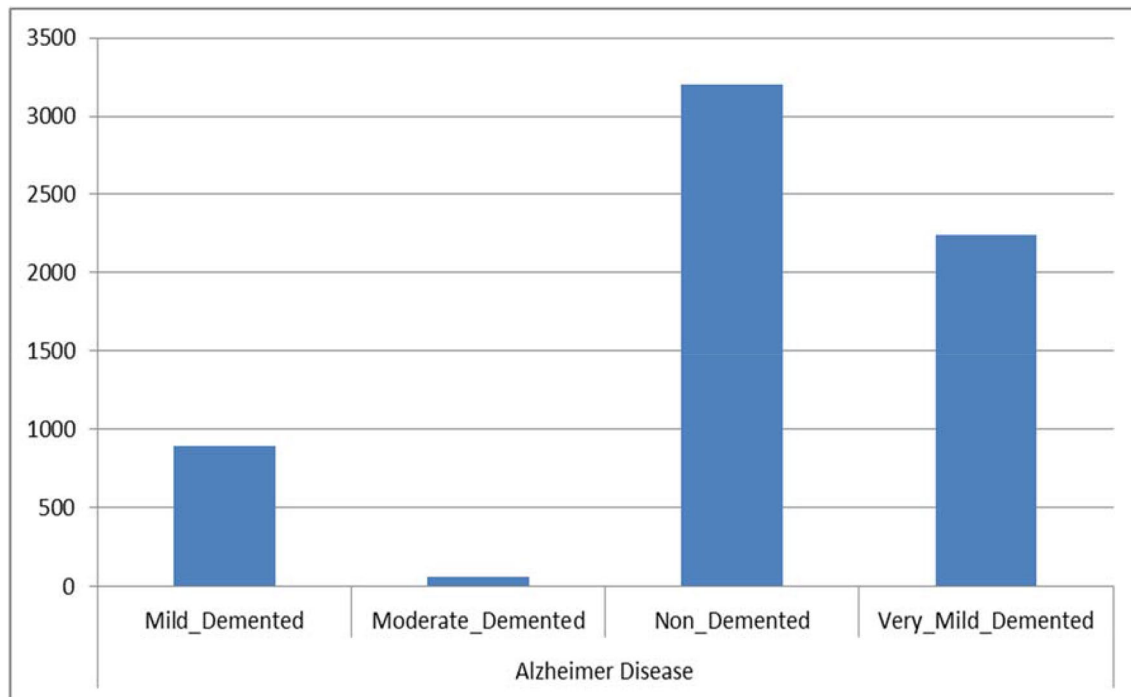


Fig. 3 Data distribution of Alzheimer's disease stages

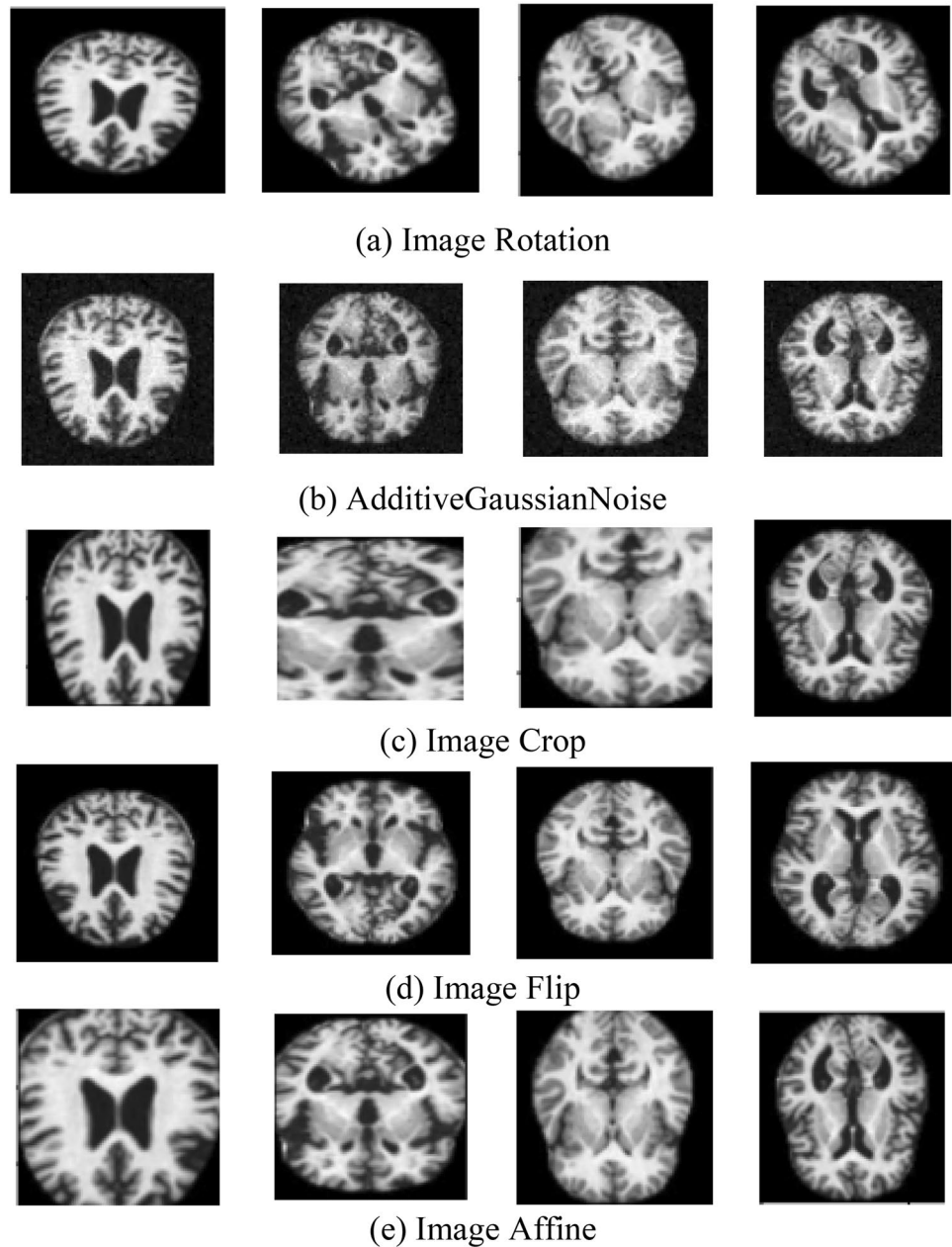
### 3.3 Deep Transfer Learning Techniques

Deep learning models have a strong dependency on large training datasets because they need a huge amount of dataset to understand the latent structure of a dataset. Insufficient training dataset leads to poor training, validation, and prediction of the models. However, the collection of the large-scale dataset is also a complex and expensive process. Transfer learning process helps us to address the problems like insufficient training data, identical train test split, and model trained from scratch. By transferring information from the source domain to the target domain, transfer learning also enables us to shorten the training period. A domain in transfer learning is represented by the expression  $D = (X, P(X))$ , where  $X$  is the feature space and  $P(X)$  denotes the edge probability distribution. The task can be written as  $T = y, f(x)$ , where  $y$  is label space and  $f(x)$  is the target prediction function.  $X$  can be written as  $x_1, x_2, \dots, x_n$ . Another way to think of the function  $f(x)$  is as a conditional probability function  $P(y | x)$ . So the transfer learning process can be defined as the learning task is  $L_t$  on a given domain  $D_t$ . We can take the help from source domain  $D_s$  for the learning task  $L_s$ . We have predictive function  $F_t$ , the goal of transfer learning is to increase the effectiveness of the predictive function  $F_t$  on a given task  $L_t$ . Transfer learning transfer the latent knowledge from  $D_s$  to  $L_s$ , where  $D_s \neq D_t$  and  $L_s \neq L_t$  whereas, size of  $D_s$  is larger than the  $D_t$  and  $N_s \neq N_t$ . Where  $N_s$  is the number of

images of source learning task  $L_s$  and  $N_t$  number of images in targets leaning task  $L_t$  [44].

For feature extraction in this work, we employed the transfer learning method; hence, we rely on the CNN architecture to automate the feature extraction process. CNN is a very versatile automatic feature extractor. If enough images are provided to the CNN models, they can learn the set of features and solve the given problem [45]. CNN is beneficial in getting generic features from a given set of images, but we need to tune the network parameters and training strategies to get better results. Figure 5 is illustrated the transfer learning process. The network's initial component is trained using a sizable training dataset on the source domain. The trained network first part is used to train the target domain with some fine-tuning in the second half. We used Imagenet weights to train the original ResNet50, VGG19, Xception, EfficientNetB7, and DenseNet201 models, with over 14 million images from 1000 different classes. It is a standard method for training a deep learning model to learn generic features from input images. It can also learn fundamental geometric features, corners, textures, etc. Following model training, we updated the input layer to support  $224 \times 224 \times 3$  inputs, where  $224 \times 224$  denotes the input images' width and height and 3 denotes the number of channels. The Alzheimer's disease dataset is then used to retrain the models. This step helps the models pick up characteristics that can later be used to identify stages of Alzheimer's disease. The convolutional layer weights were improved in the last stage such

**Fig. 4** Augmentation for MRI images



that they remained constant throughout the training of the model. To expedite the training process, we kept the features gathered from the convolutional layers up until the first fully connected layer. Finally, the model is adjusted using hyperparameters. The convolution layer employed in the investigation had a pool size of  $7 \times 7$ . The final layer activates using "ReLU" and "Softmax." The model optimizer is Stochastic Gradient Decent (SGD), with learning rate set to 0.0001, momentum set to 0.9, and decay set to  $1e-4/\text{epoch}$ . A batch size of 64 images has been set for training.

### 3.3.1 CNN Model Architectures

Image recognition is the well-known application of deep learning models, and since the advent of deep learning, they have been utilized in several picture identification applications. The adaptability, predictive power, and growing accessibility of deep learning models are what lead to their success in image identification [46]. We have used five deep learning models as "Xception," "EfficientNetB7," "DensNet201," "ResNet50," and "VGG19" in the study and discussed as follows.



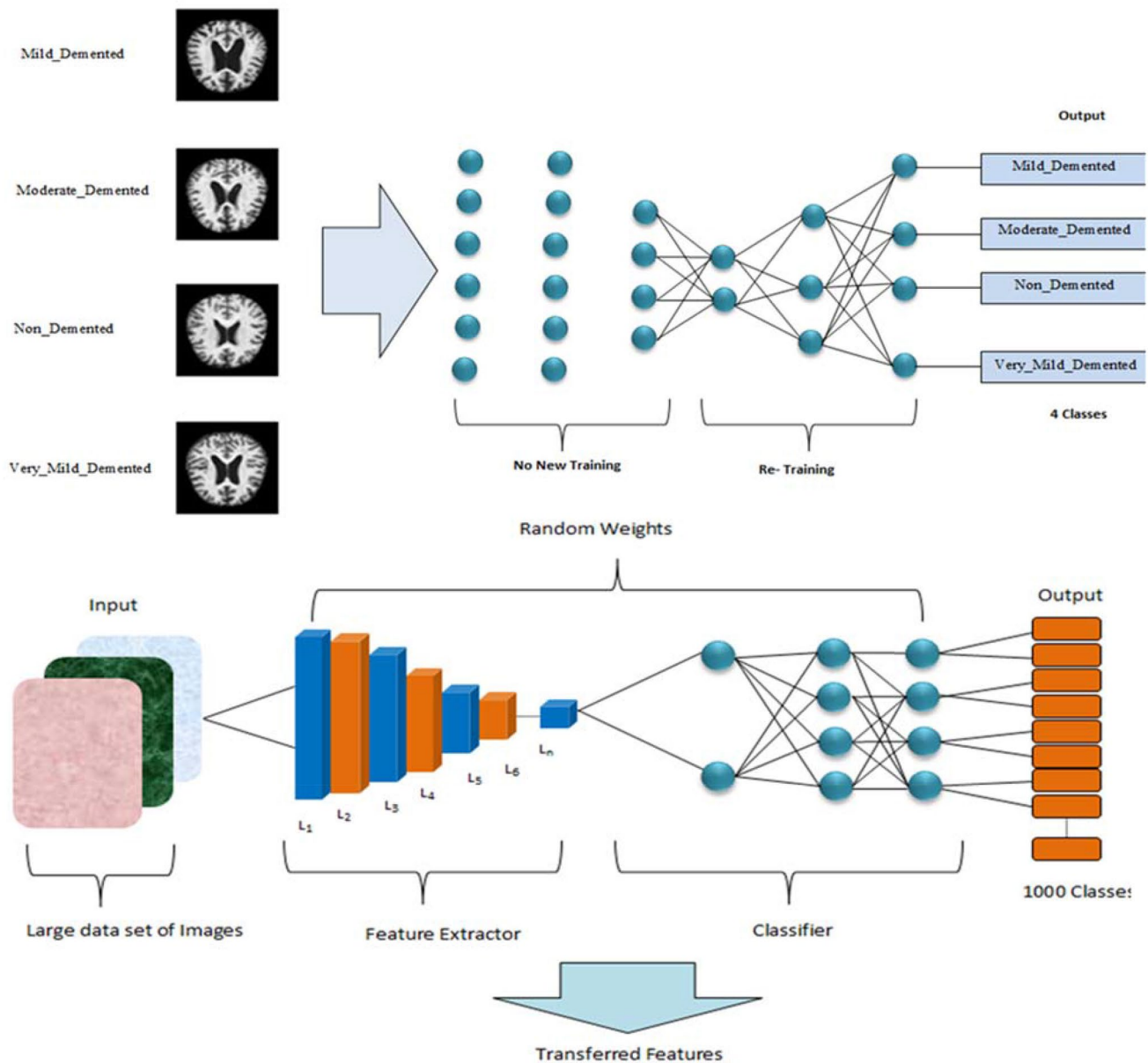


Fig. 5 Illustrate the transfer learning process

**3.3.1.1 Xception** Xception is also known as the extreme version of the Inception model. Xception model outperforms InceptionV3 by using a depth-wise divisible convolutional network. Convolution in this model is accomplished by first performing a point-wise convolution, followed by a depth-wise convolution [47]. Convolution is a process that refers to the dot product between the set of learnable parameters known as the kernel and the input image’s channel. If we have three channels in a given input image, then we will have three  $N \times N$  spatial and a  $1 \times 1$  point-wise convolution to modify the dimension of the input image. This is accomplished by performing a point-wise convolution first,

followed by a depth-wise convolution. This process is inherited from the InceptionV3 model that performs a  $1 \times 1$  convolution before doing  $N \times N$  spatial convolution. It performs better than the InceptionV3 model as a result. The Xception system’s architecture is shown in Fig. 6.

**3.3.1.2 EfficientNetB7** The EfficientNetB7 model is developed by improving its previously available versions of the EfficientNet family. Mobile inverted bottleneck convolution serves as the fundamental building piece of the initial EfficientNet design (MBCnv). The family of EfficientNet architecture included EfficientNetB0 to Efficient-

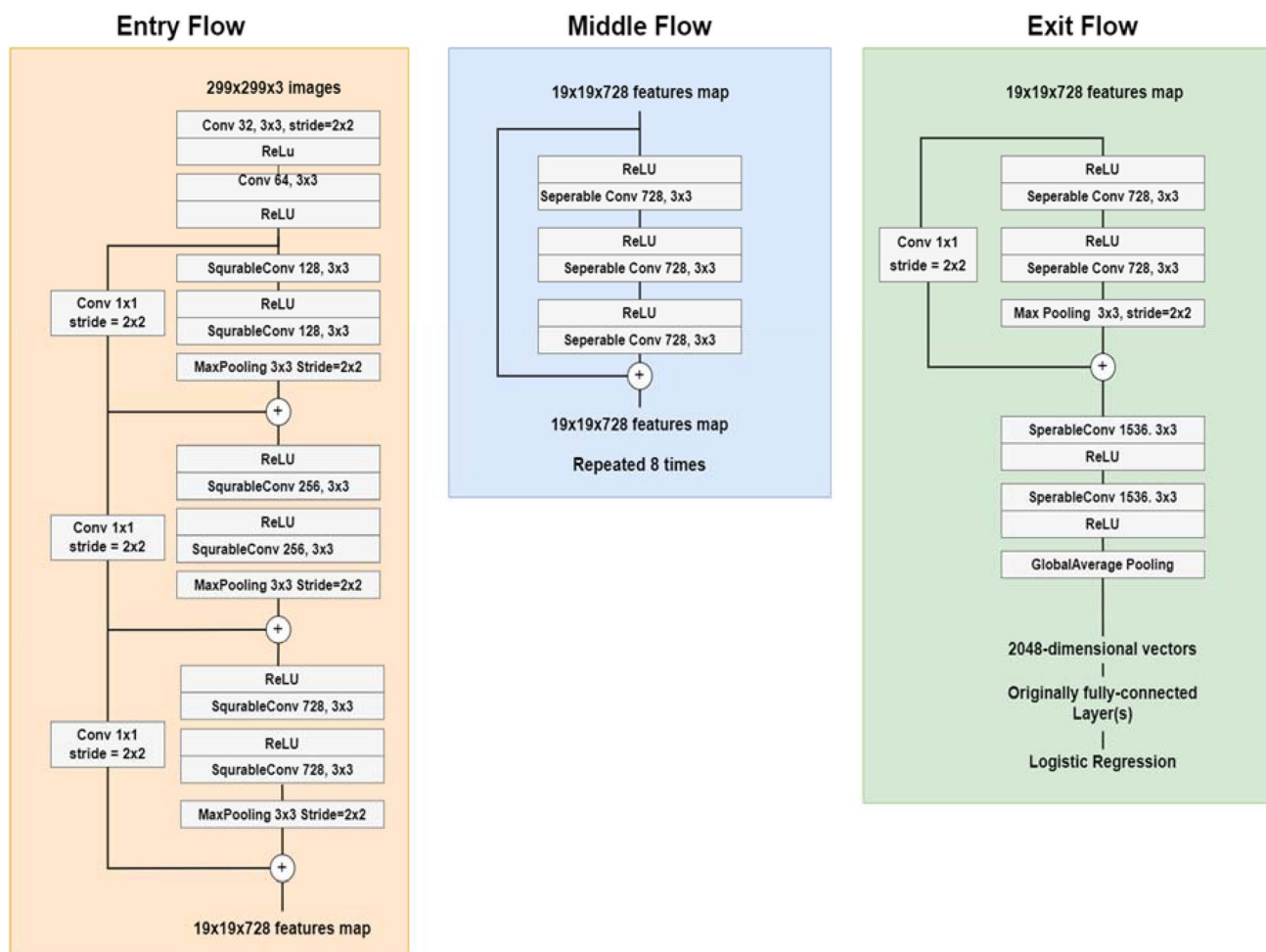


Fig. 6 Architecture of Xception model

NetB7. These models have a different number of MBCov blocks for each EfficientNet. The idea behind these models is to scale up the standard model and modify its architecture. The modification is done by increasing the model depth, width, resolution, and size to improve the model efficiency. The best performing model from the Efficient-Net family is EfficientNetB7 which outperforms on ImageNet and is also eight times smaller and six times faster than state-of-the-art CNN models [48]. EfficientNetB7 is divided into seven blocks based on the filter's size, stride, and number of channels. EfficientNetB7 architecture is illustrated in Fig. 7.

**3.3.1.3 DenseNet201** The Convolutional Neural Network is modified to have 201 depth layers in the DenseNet201 deep learning model [23]. DenseNet201 architecture received additional inputs at each layer from its preceding layers. Through the concatenation method it sends its own feature-maps to higher layers that follow. The collective

knowledge is received by each layer from the above levels. Densenet201 is thinner and more compact with fewer channels because all preceding layers receive feature maps from each layer. The architecture of DenseNet201 is depicted in Fig. 8.

**3.3.1.4 ResNet50** ResNet50 belongs to the residual network family that uses a method called "residual mapping." Residual networks use residual mapping to overcome the degradation problem. When the depth is increased in neural architecture, the accuracy of the Network will also fall, called the degradation problem [49]. Instead of assuming that the residual network fits a few layers to its proposed fundamental mapping, the residual net specifically enables a few layers to match a residual mapping. The ResNet50 architecture is depicted in Fig. 9. There are 48 Convolutional layers in the ResNet50 design, one Max-Pool layer, and one Average-Pool layer. Residual networks are used in numerous applications because their architecture is simple

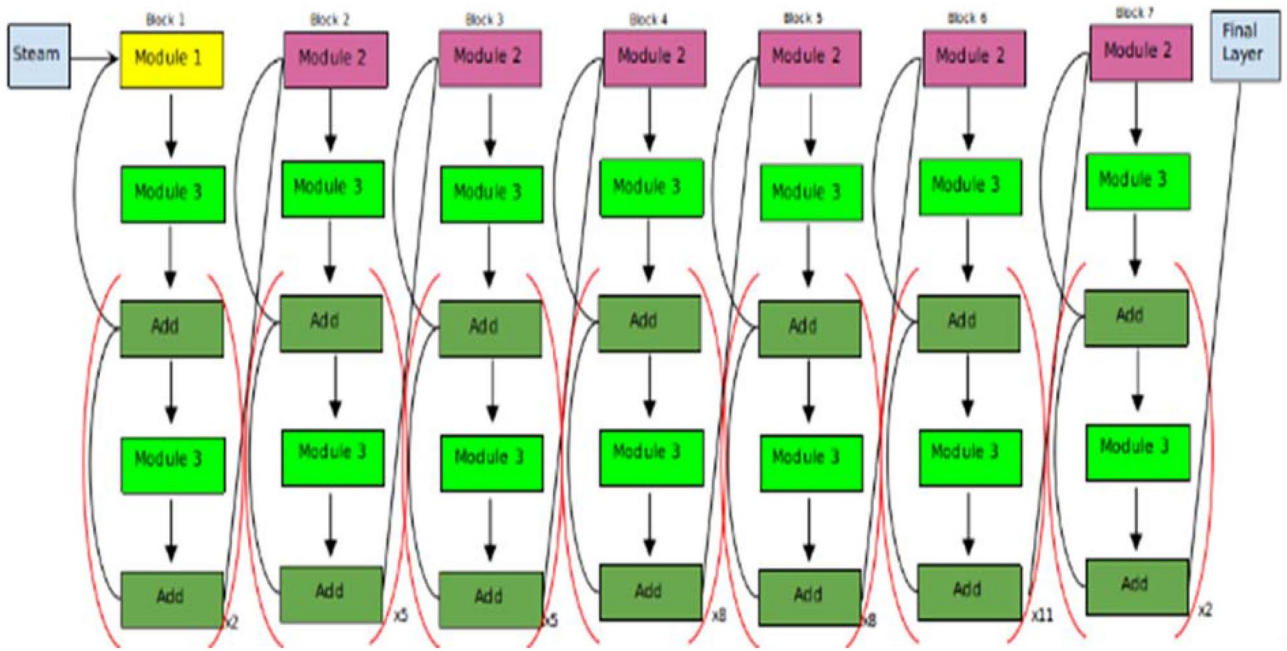


Fig. 7 Architecture of EfficientNetB7 model

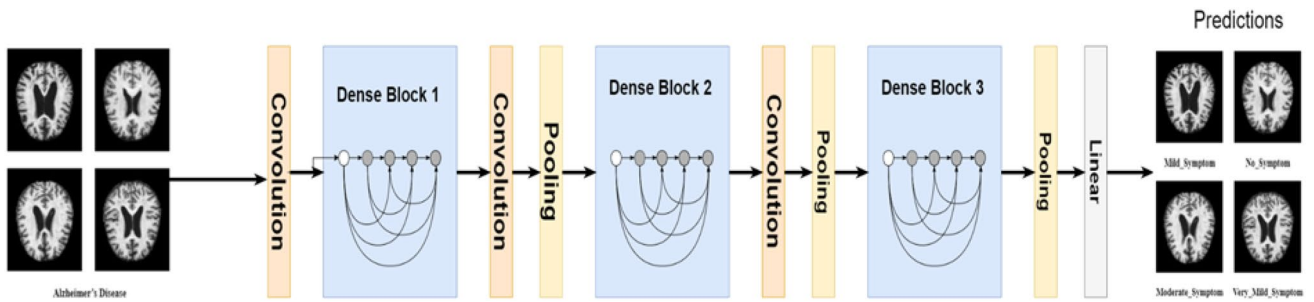


Fig. 8 Architecture of DensNet201 model

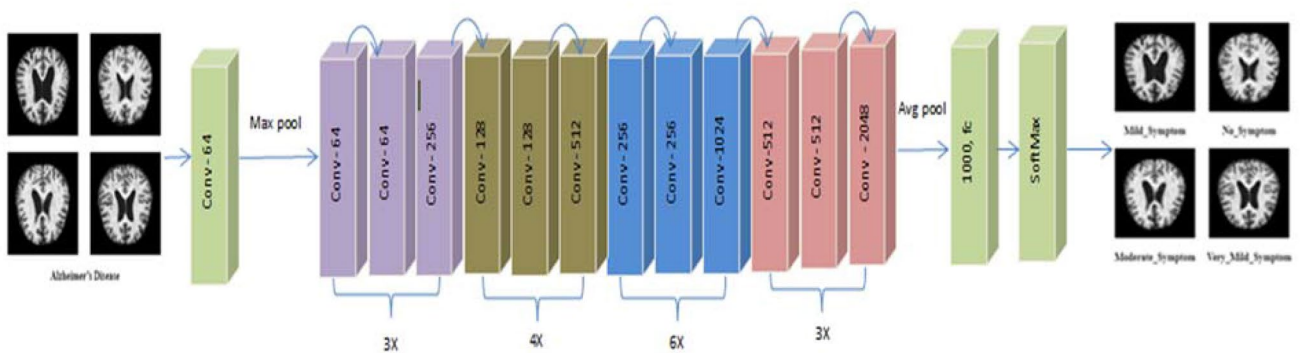
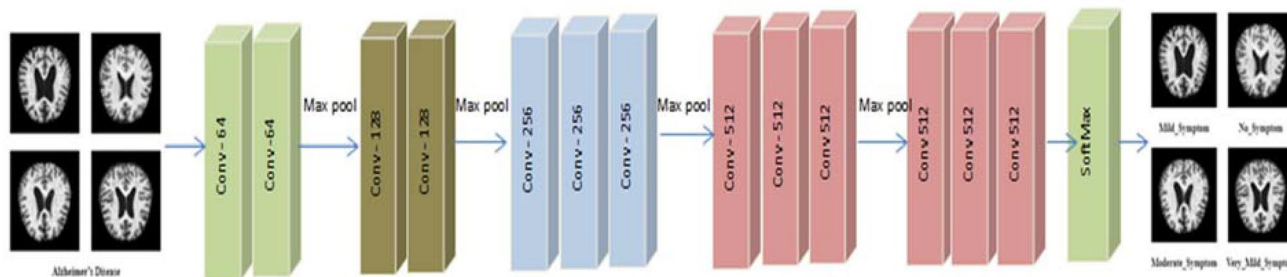


Fig. 9 Architecture of ResNet50 model



**Fig. 10** Architecture of VGG19

**Table 2** Illustrated the trainable and no-trainable parameters

Models	Depth	Total parameters	Trainable parameters	Non-trainable parameters
ResNet50	107	27,784,580	27,731,460	53,120
VGG19	19	21,075,524	21,075,524	0
Xception	81	25,058,348	25,003,820	54,528
EfficientNetB7	438	69,343,131	69,032,404	310,727
DenseNet201	402	22,256,708	22,027,652	229,056

to optimize, and with the increase in the Network's depth, they acquire more accuracy.

**3.3.1.5 VGG19** VGG19 is created by the Visual Geometry Group at Oxford university and hence the name VGG. VGG19 is the improved version of its predecessors [50]. VGG19 has used the fixed size of RGB images as an input, which means the shape matrix is  $(224, 224, 3)$ . Where  $224 \times 224$  is the width and height, and three represents the number of channels. The mean RGB value from each pixel was preprocessed by the VGG19 and calculated for the entire training set. To maintain the spatial resolution of the input image, the VGG19 used  $3 \times 3$  size kernels with a stride size of 1 pixel. Using a stride size of 2, the max pooling process is carried out over a  $2 \times 2$  matrix. To better classify the VGG19 model, nonlinearity and the softmax function are handled using the rectified linear unit (ReLU). Figure 10 illustrates the architecture of VGG19.

### 3.3.2 Model Parameters

Deep learning model parameters are the weights that are learnt during the model training. Parameters are weight matrices that help deep learning models to train well during the model training. Trainable parameters are those that are continuously updated during back-propagation process whereas, non-trainable parameters are those that are not updated during back-propagation process. Depth of a network considered as number of layers in the network.

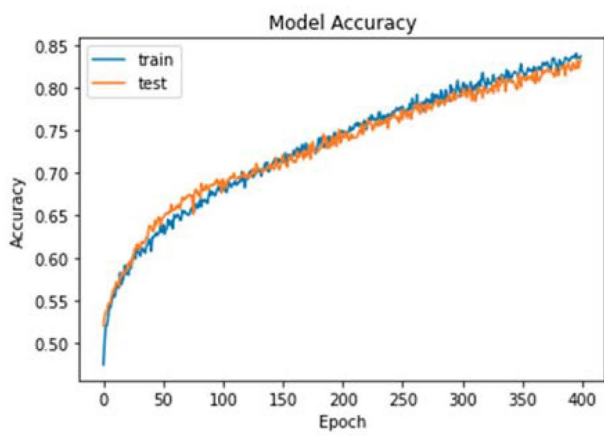
Table 2 illustrated the trainable and no-trainable parameters of deep learning models.

## 4 Results and Analysis

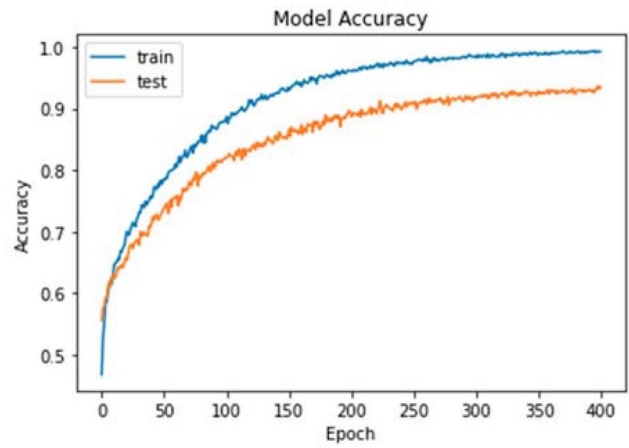
The performance of the prediction models was rigorously assessed using the accuracy and weighted loss curves.

### 4.1 Accuracy

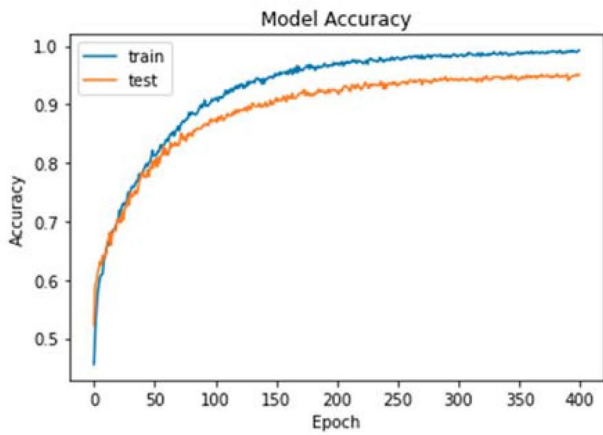
The performance metric most frequently used to assess the model's effectiveness in an understandable way is accuracy. The specification of the model parameters is used to determine a model's accuracy, which is typically reported as a percentage (percent). The model's accuracy measures how closely the predicted results matched the actual data for a certain set of parameters. Figure 10 lists the accuracy curves for each deep learning model. The Figure displays a line graph that compares the accuracy of several deep learning models using train and validation data over time. Figure 11's line graph of training and validation indicates that both model training and model validation accuracy are consistent across the majority of epochs. While VGG19 at 250 and DenseNet201 at 125 epochs are stable and have achieved 100% accuracy, ResNet50 and Xception model training accuracy stabilizes after 250 epochs. While after 100 epochs, the EfficientNetB7 training accuracy curve is not steady. Except for EfficientNetB7, all models exhibit



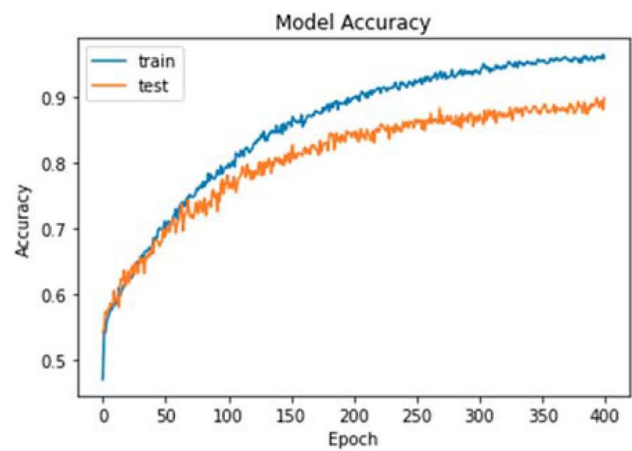
(a) EfficientNetB7 Accuracy



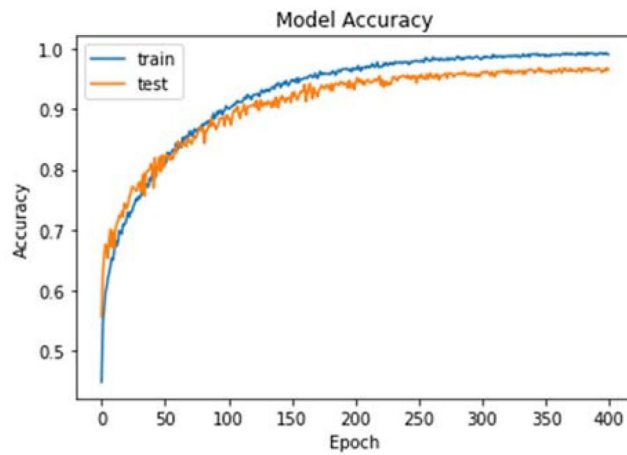
(b) ResNet50 Accuracy



(c) VGG19 Accuracy

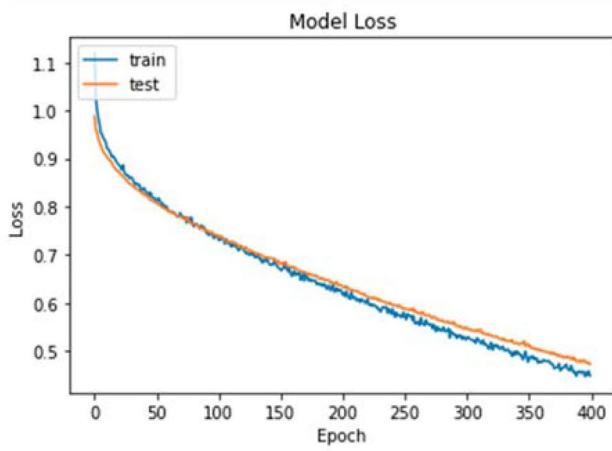


(d) Xception Accuracy

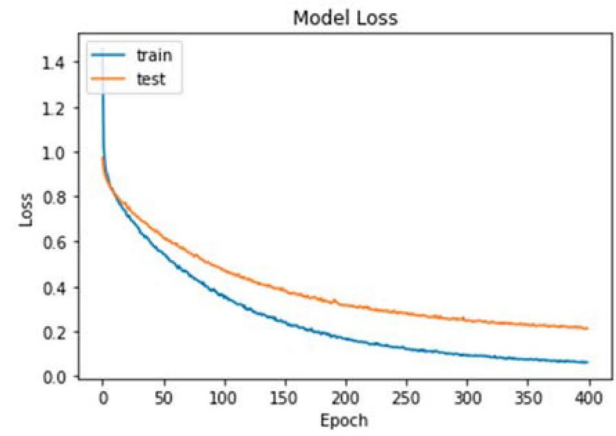


(e) DenseNet201 Accuracy

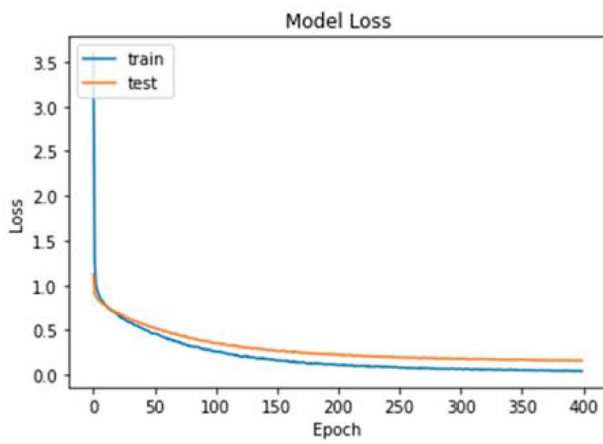
**Fig. 11** Accuracy curves of **a** EfficientNetB7, **b** ResNet50, **c** VGG19, **d** Xception and **e** DenseNet201



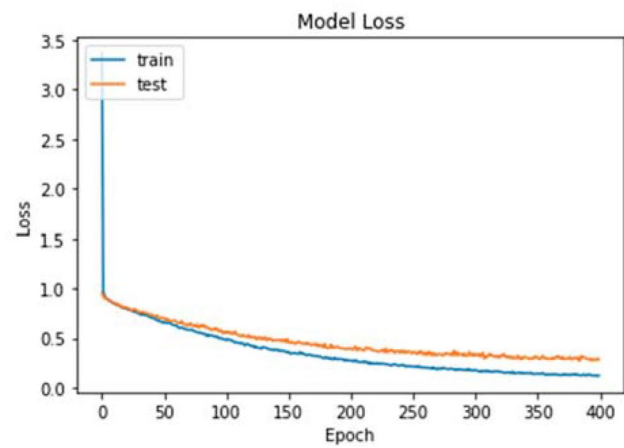
(a) EfficientNetB7 Loss



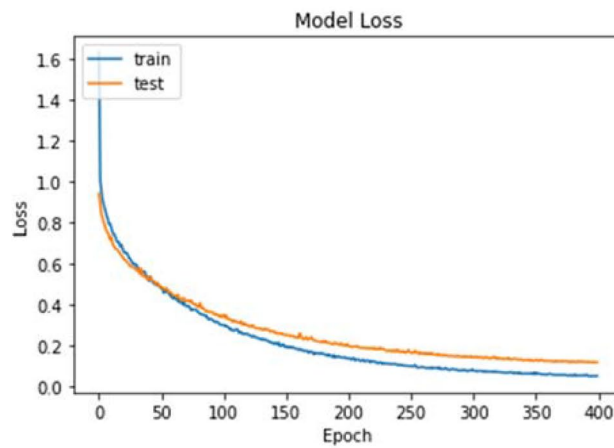
(b) ResNet50 Loss



(c) VGG19 Loss



(d) Xception Loss



(e) DenseNet201 Loss

**Fig. 12** Loss curves of **a** EfficientNetB7, **b** ResNet50, **c** VGG19, **d** Xception and **e** DenseNet201

**Table 3** Deep learning model's average training and validation accuracy and loss score

S. no.	Model	Avg. training accuracy (%)	Avg. training loss (%)	Avg. validation accuracy (%)	Avg. validation loss (%)
1	ResNet50	100	3	93	21
2	VGG19	100	1	94	17
3	Xception	100	4	89	27
4	EfficientNetB7	92	34	82	48
5	DenseNet201	100	2	96	13

**Table 4** Depicts the confusion matrices of deep learning models (a) ResNet50, (b) VGG19, (c) Xception, (d) EfficientNetB7, and (e) DenseNet201 that obtained during performance evaluation process

Classes	Mild-Demented	Moderate-Demented	Non-Demented	Very-Mild-Demented
<b>(a) ResNet50</b>				
Mild-Demented	156	0	9	14
Moderate-Demented	0	12	1	0
Non-Demented	3	0	606	31
Very-Mild-Demented	1	0	24	423
<b>(b) VGG19</b>				
Mild-Demented	156	0	13	10
Moderate-Demented	0	12	0	1
Non-Demented	4	0	614	22
Very-Mild-Demented	0	0	13	435
<b>(c) Xception</b>				
Mild-Demented	148	0	11	20
Moderate-Demented	1	11	1	0
Non-Demented	7	0	588	45
Very-Mild-Demented	11	0	35	402
<b>(d) EfficientNetB7</b>				
Mild-Demented	107	0	29	43
Moderate-Demented	1	5	1	6
Non-Demented	10	0	573	57
Very-Mild-Demented	12	0	56	380
<b>(e) DenseNet201</b>				
Mild-Demented	175	0	3	1
Moderate-Demented	0	13	0	0
Non-Demented	1	0	622	17
Very-Mild-Demented	4	0	21	423

stability after 250 epochs when taking validation accuracy into account. After 125 epochs, DenseNet201's training and validation accuracy stabilized, showing improved prediction performance.

### 4.2 Loss

The model's performance loss is calculated using the training and validation sets, and this loss is equal to how well the

model performs on these two sets. The loss of the model is not stated as a percentage, in contrast to the accuracy of the model. It represents the total number of errors made in either the training or validation datasets. Figure 12 shows the loss curves for each learning model. The Fig. 12 shows a line graph of the loss recorded by the deep learning models on the train and validation data over each iteration. Figure 12 shows that deep learning models suffered more losses from validation data than from training data. The training loss

**Table 5** Precision, Recall, F1-score of (a) EfficientNetB7, (b) ResNet50, (c) VGG19, (d) Xception and (e) DenseNet201

Classes	Precision	Recall	f1-score
(a) ResNet50			
Mild-Demented	0.97	0.87	0.92
Moderate-Demented	1	0.92	0.96
Non-Demented	0.95	0.95	0.95
Very-Mild-Demented	0.90	0.94	0.92
(b) VGG19			
Mild-Demented	0.97	0.87	0.92
Moderate-Demented	1	0.92	0.96
Non-Demented	0.96	0.96	0.96
Very-Mild-Demented	0.93	0.97	0.95
(c) DenseNet201			
Mild-Demented	0.97	0.98	0.97
Moderate-Demented	1	1	1
Non-Demented	0.96	0.97	0.97
Very-Mild-Demented	0.96	0.94	0.95
(d) EfficientNetB7			
Mild-Demented	0.83	0.60	0.69
Moderate-Demented	1	0.38	0.56
Non-Demented	0.87	0.90	0.88
Very-Mild-Demented	0.78	0.85	0.81
(e) Xception			
Mild-Demented	0.89	0.86	0.86
Moderate-Demented	1	0.85	0.92
Non-Demented	0.93	0.92	0.92
Very-Mild-Demented	0.86	0.90	0.88

decreased significantly for ResNet50, Xception, VGG19, and EfficientNetB7, although some fluctuations and loss can be noticed in the validation data. The model has suffered the least loss on the training and validation data when compared to DenseNet201.

### 4.3 Average Training and Validation Accuracy and Loss

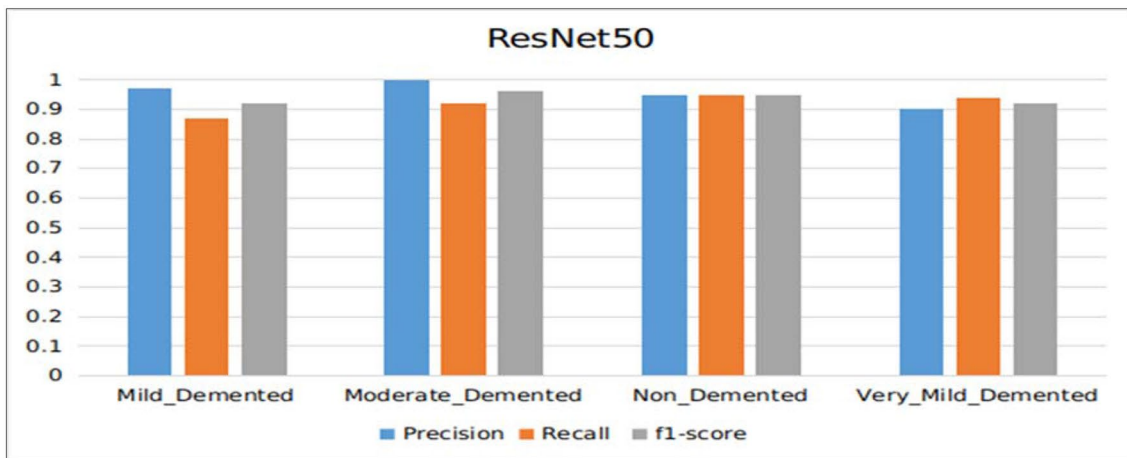
Five deep learning models' performance is shown in terms of model accuracy and loss. In terms of model training and validation, the model accuracy and loss scores are calculated in a five-fold method. Table 3 lists the deep learning models' training and validation accuracies as well as losses sustained over various folds. The table shows that, rather than the validation dataset, the training dataset is where all of the models have achieved the target average accuracy score. Comparing the training dataset to the validation dataset, models have experienced smaller average losses. The DenseNet201 model performs best when comparing the classification models based on the average training accuracy, with the training set achieving 100% accuracy. However, all of the classification models perform admirably (90%) in terms of

training data. DenseNet201 outperforms EfficientNetB7 by 96% in terms of average validation accuracy, whereas EfficientNetB7 produced the least accurate findings at 82%. All the models showed great prediction performance on training data, taking into account the loss sustained by the classification modes, however the prediction losses progressively rose on the validation sets. DenseNet201 had the least validation loss, just 13%, whereas EfficientNetB7 suffered the largest validation loss, 48%, i.e., poor performance. Thus, it is clear that DenseNet201 had the best overall performance in terms of prediction results.

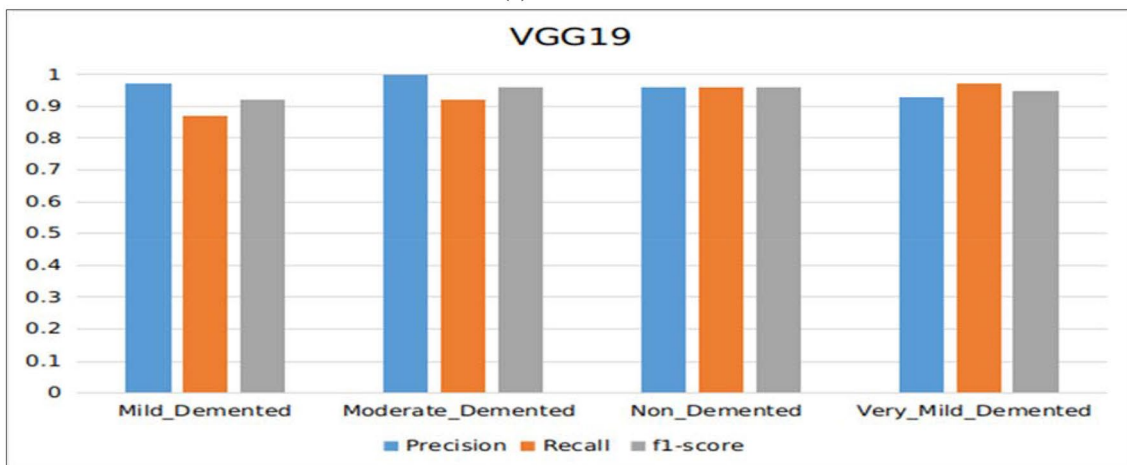
### 4.4 Confusion Matrix

In the confusion matrix, N stands for the size of the target class and NxN is a matrix used to evaluate the performance of deep learning models. In the matrix, actual goal values are contrasted with predictions made by categorization models. To fully comprehend the classification outcomes, we also map the average confusion matrices for each of the five classification models in Table 4.

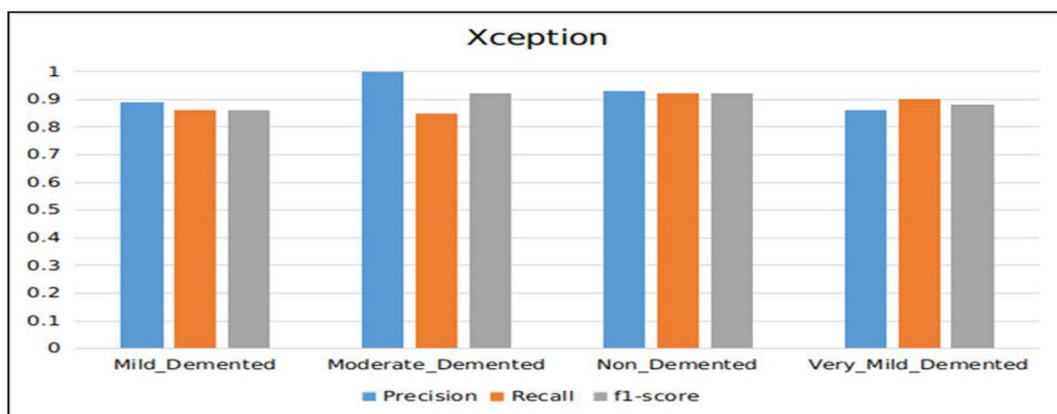




(a) ResNet50

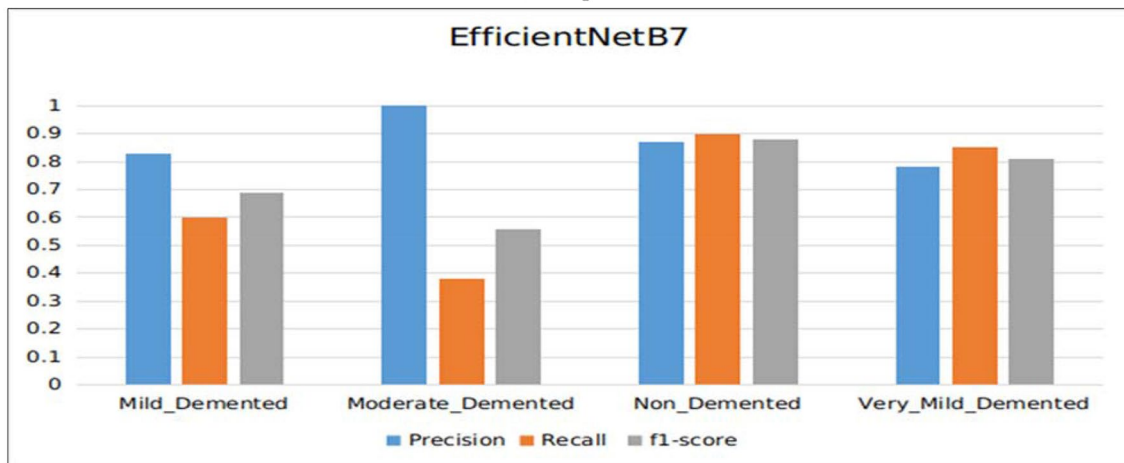


(b) VGG19

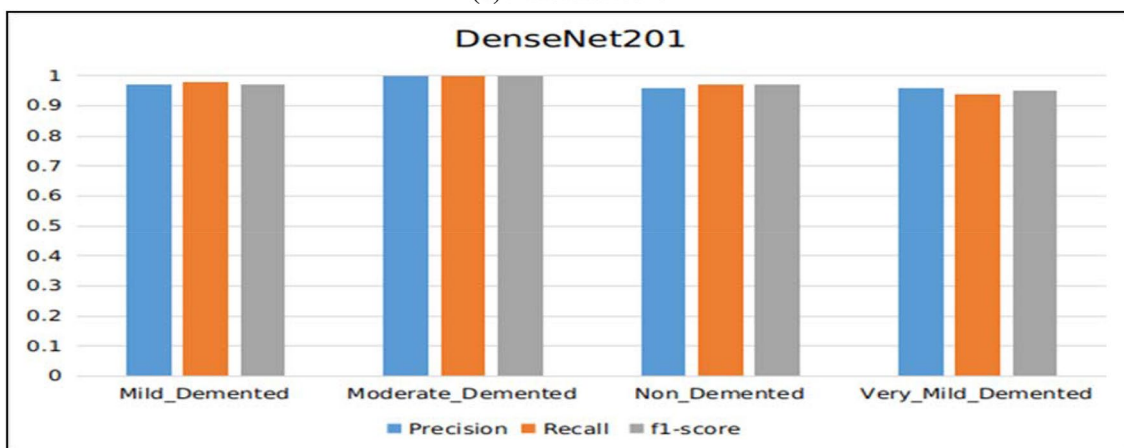


(c) Xception

**Fig. 13** Precision, Recall, f1-score of deep learning models of **a** EfficientNetB7, **b** ResNet50, **c** VGG19, **d** Xception and **e** DenseNet201



(d) EfficientNetB7



(e) DenseNet201

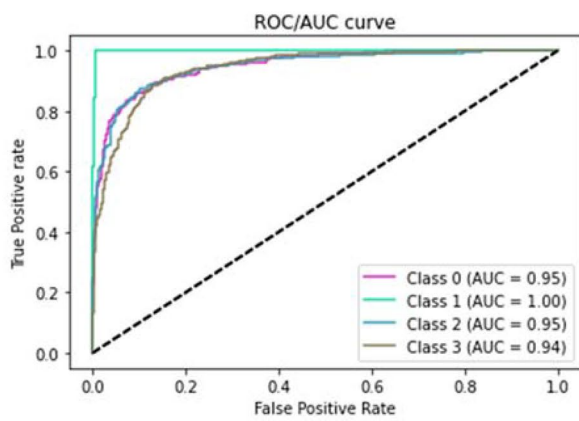
Fig. 13 (continued)

#### 4.5 Comparison of Prediction Performance

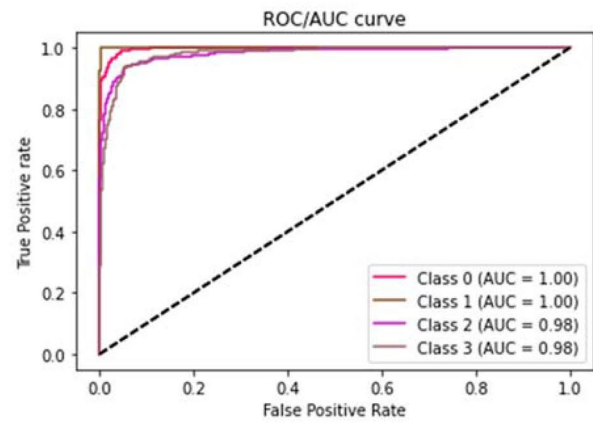
In Table 5, we have provided a thorough evaluation of the performance of five deep learning models for predicting Alzheimer's disease. Precision, recall, and f1-score have all been taken into consideration while comparing results. On the stages of Alzheimer's disease known as "Mild-Demented," "Moderate-Demented," "Non-Demented," and "Very-Mild-Demented," the DenseNet201 model has shown the best performance. VGG19 and ResNet50 model results have been recorded as the second-best performance. In our study, the Xception and EfficientNetB7 performed the worst at identifying the stages of Alzheimer's disease. Figure 13 shows the full bar graphs that were created to show all model results in terms of recall, precision, and f1-score.

#### 4.6 ROC/AUC Curve

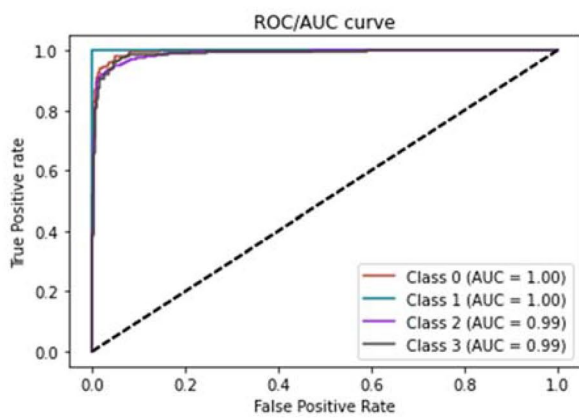
The performance of deep learning models is evaluated using the Receiver Operating Characteristic (ROC) curve. Instead of using measurements to show model performance, ROC uses a graphical representation. The ROC curve is the resultant of the True Positive Rate (TPR) against the False Positive Rate (FPR) for a classifier at a variety of thresholds. In contrast, Area Under the Curve (AUC) stands for the area under the (ROC) curve. The higher value of AUC is generally considered as good for a classifier to perform the given task. The detailed ROC/AUC graphs have been presented in Fig. 14. In the graphs, "Class 0," "Class 1," "Class 2," and "Class 3" represents the stages of Alzheimer's disease as "Mild-Demented," "Moderate-Demented," "Non-Demented," and "Very-Mild-Demented" respectively. The DenseNet201 model has shown the highest AUC values as "1," "1," "0.99," and "1" for the Alzheimer's disease



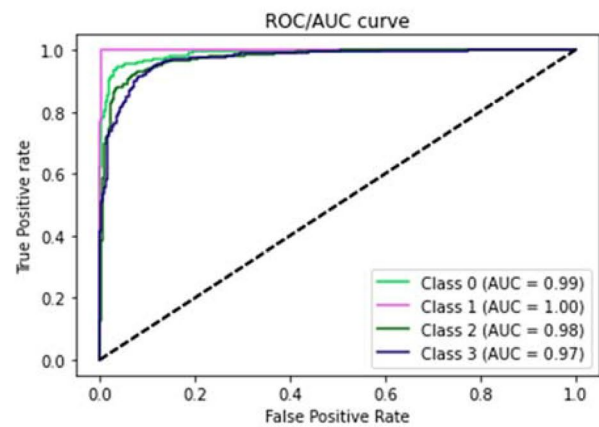
(a) EfficientNetB7



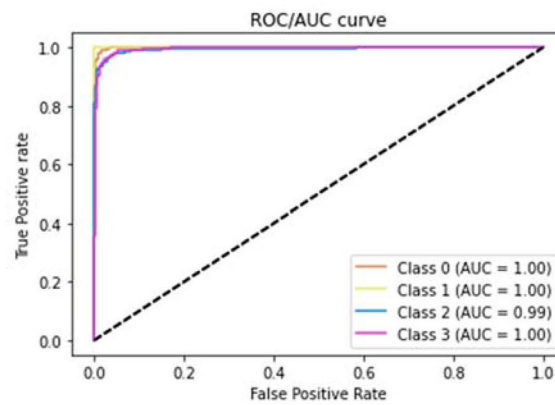
(b) ResNet50



(c) VGG19



(d) Xception



(e) DenseNet201

Fig. 14 ROC/AUC curve of a EfficientNetB7, b ResNet50, c VGG19, d Xception and e DenseNet201

stages “Mild-Demented,” “Moderate-Demented,” “Non-Demented,” and “Very-Mild-Demented” respectively.

## 5 Conclusions

No known cure is available for Alzheimer’s disease, but the early prediction of Alzheimer’s disease stages can prevent Alzheimer’s disease from converting into dementia. Various deep learning modes such as ResNet50, VGG16, Xception, EfficientNetB7, and DenseNet201 are devised through transfer learning techniques to predict the early stages of Alzheimer’s disease. We also compared the performance of deep learning models. As per the results, DenseNet201 have been attended the highest level of prediction accuracy and proved that the transfer learning technique is suitable for the early prediction of Alzheimer’s disease. We have attended high-level prediction accuracy with less number of MRI images through transfer learning. In the future, we will also include clinical data to improve the prediction capabilities of deep learning models. Further, we will use proposed models to detect other disease using MRI images.

**Funding** The author is declared that they do not have any funding or grant for the manuscript.

**Data Availability** Not Applicable.

**Code Availability** Not Applicable.

## Declarations

**Conflict of interest** The authors declare that they do not have any conflict of interests that influence the work reported in this paper.

## References

- Feng C, Elazab A, Yang P, Wang T, Zhou F, Hu H, Xiao X, Lei B (2019) Deep learning framework for Alzheimer’s disease diagnosis via 3D-CNN and FSBi-LSTM. *IEEE Access* 7:63605–63618
- Yang K, Mohammed EA (2020) A review of artificial intelligence technologies for early prediction of Alzheimer’s disease. *arXiv preprint arXiv:2101.01781*
- Raees PM, Thomas V (2021) Automated detection of Alzheimer’s disease using deep learning in MRI. *J Phys: Conf Ser* 1921(1):012024
- Gaugler J, James B, Johnson T, Reimer J, Solis M, Weuve J, Hohman TJ (2022) 2022 Alzheimer’s disease facts and figures. *Alzheimers Dement* 18(4):700–789
- Bron EE, Klein S, Papma JM, Jiskoot LC, Venkatraghavan V, Linders J, Aalten P, De Deyn PP, Biessels GJ, Claassen JA, Middelkoop Neurodegenerative Diseases study group (2021) Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer’s disease. *NeuroImage Clinical* 31:102712
- Kumar Y, Gupta S, Singla R et al (2022) A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. *Arch Comput Methods Eng* 29:2043–2070. <https://doi.org/10.1007/s11831-021-09648-w>
- Bhardwaj P, Bhandari G, Kumar Y et al (2022) An investigational approach for the prediction of gastric cancer using artificial intelligence techniques: a systematic review. *Arch Comput Methods Eng*. <https://doi.org/10.1007/s11831-022-09737-4>
- Kaur I, Sandhu AK, Kumar Y (2022) Artificial intelligence techniques for predictive modeling of vector-borne diseases and its pathogens: a systematic review. *Arch Comput Methods Eng*. <https://doi.org/10.1007/s11831-022-09724-9>
- Kumar Y, Gupta S (2022) Deep transfer learning approaches to predict glaucoma, cataract, choroidal neovascularization, diabetic macular edema, DRUSEN and healthy eyes: an experimental review. *Arch Comput Methods Eng*. <https://doi.org/10.1007/s11831-022-09807-7>
- Koul A, Bawa RK, Kumar Y (2022) Artificial intelligence in medical image processing for airway diseases. In: *Connected e-health*. Springer, Cham, pp 217–254
- Bansal K, Bathla RK, Kumar Y (2022) Deep transfer learning techniques with hybrid optimization in early prediction and diagnosis of different types of oral cancer. *Soft Comput* 26:11153–11184. <https://doi.org/10.1007/s00500-022-07246-x>
- Koul A, Bawa RK, Kumar Y (2022) Artificial intelligence techniques to predict the airway disorders illness: a systematic review. *Arch Computat Methods Eng*. <https://doi.org/10.1007/s11831-022-09818-4>
- Alroobaea R, Mechti S, Haoues M (2021) Alzheimer’s Disease Early Detection Using Machine Learning Techniques. <https://doi.org/10.21203/rs.3.rs-624520/v1>
- Kumar Y, Patel NP, Koul A, Gupta A (2022) Early prediction of neonatal jaundice using artificial intelligence techniques. In: *2022 2nd international conference on innovative practices in technology and management (ICIPTM)*, pp 222–226. <https://doi.org/10.1109/ICIPTM54933.2022.9753884>
- Kumar Y, Singla R (2022) Effectiveness of machine and deep learning in IOT-enabled devices for healthcare system. In: Ghosh U, Chakraborty C, Garg L, Srivastava G (eds) *Intelligent internet of things for healthcare and industry*. Internet of things. Springer, Cham. [https://doi.org/10.1007/978-3-030-81473-1\\_1](https://doi.org/10.1007/978-3-030-81473-1_1)
- Kavitha C, Mani V, Srividhya SR, Khalaf OI, Romero CAT (2022) Early-stage Alzheimer’s disease prediction using machine learning models. *Front Public Health*. <https://doi.org/10.3389/fpubh.2022.853294>
- Al-Shoukry S, Rassem TH, Makbol NM (2020) Alzheimer’s diseases detection by using deep learning algorithms: a mini-review. *IEEE Access* 8:77131–77141
- Jo T, Nho K, Saykin AJ (2019) Deep learning in Alzheimer’s disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci* 11:220
- Pan D, Zeng A, Jia L, Huang Y, Frizzell T, Song X (2020) Early detection of Alzheimer’s disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning. *Front Neurosci* 14:259
- Venugopalan J, Tong L, Hassanzadeh HR, Wang MD (2021) Multimodal deep learning models for early detection of Alzheimer’s disease stage. *Sci Rep* 11(1):1–13
- Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M, Alzheimer’s Disease Neuroimaging Initiative (2019) Automated classification of Alzheimer’s disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clin* 21:101645
- Helaly HA, Badawy M, Haikal AY (2021) Deep learning approach for early detection of Alzheimer’s disease. *Cognit Comput* 14:1711–1727

23. Odusami M, Maskeliūnas R, Damaševičius R (2022) An intelligent system for early recognition of Alzheimer's disease using neuroimaging. *Sensors* 22(3):740
24. Sarraf S, DeSouza DD, Anderson J, Tofighi G (2017) DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv*. <https://doi.org/10.1101/070441>
25. Orouskhani M, Rostamian S, Zadeh FS, Shafiei M, Orouskhani Y (2022) Alzheimer's disease detection from structural MRI using conditional deep triplet network. *Neurosci Inform* 2:100066
26. Subramoniam M, Aparna TR, Anurenjan PR, Sreeni KG (2022) Deep learning-based prediction of Alzheimer's disease from magnetic resonance images. In: *Intelligent vision in healthcare*. Springer, Singapore, pp 145–151
27. Bae JB, Lee S, Jung W, Park S, Kim W, Oh H et al (2020) Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. *Sci Rep* 10(1):1–10
28. Salehi AW, Baglat P, Gupta G (2020) Alzheimer's disease diagnosis using deep learning techniques. *Int J Eng Adv Technol* 9(3):874–880
29. Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, Maqsood M, Yaqub M (2021) A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. *Neuroscience* 460:43–52
30. Murugan S, Venkatesan C, Sumithra MG, Gao XZ, Elakkiya B, Akila M, Manoharan S (2021) DEMNET: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. *IEEE Access* 9:90319–90329
31. Janghel RR, Rathore YK (2021) Deep convolution neural network based system for early diagnosis of Alzheimer's disease. *IRBM* 42(4):258–267
32. Toğaçar M, Cömert Z, Ergen B (2021) Enhancing of dataset using DeepDream, fuzzy color image enhancement and hypercolumn techniques to detection of the Alzheimer's disease stages by deep learning model. *Neural Comput Appl* 33(16):9877–9889
33. He Y, Wu J, Zhou L, Chen Y, Li F, Qian H (2021) Quantification of cognitive function in Alzheimer's disease based on deep learning. *Front Neurosci* 15:651920
34. Kim S, Lee P, Oh KT, Byun MS, Yi D, Lee JH et al (2021) Deep learning-based amyloid PET positivity classification model in the Alzheimer's disease continuum by using 2-[18F] FDG PET. *EJN-MMI Res* 11(1):1–14
35. Huggins CJ, Escudero J, Parra MA, Scally B, Anghinah R, De Araújo AVL, Basile LF, Abasolo D (2021) Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing. *J Neural Eng* 18(4):046087
36. Li L, Yang Y, Zhang Q, Wang J, Jiang J, Neuroimaging Initiative (2021) Use of deep-learning genomics to discriminate healthy individuals from those with Alzheimer's disease or mild cognitive impairment. *Behav Neurol*. <https://doi.org/10.1155/2021/3359103>
37. Saratxaga CL, Moya I, Picón A, Acosta M, Moreno-Fernandez-de-Leceta A, Garrote E, Bereciartua-Perez A (2021) MRI deep learning-based solution for Alzheimer's disease prediction. *J Personal Med* 11(9):902
38. Koga S, Ikeda A, Dickson DW (2022) Deep learning-based model for diagnosing Alzheimer's disease and tauopathies. *Neuropathol Appl Neurobiol* 48(1):e12759
39. Hu J, Qing Z, Liu R, Zhang X, Lv P, Wang M, Wang Y, He K, Gao Y, Zhang B (2021) Deep learning-based classification and voxel-based visualization of frontotemporal dementia and Alzheimer's disease. *Front Neurosci* 14:626154
40. Yang L, Wang X, Guo Q, Gladstein S, Wooten D, Li T, Robieson WZ, Sun Y, Huang X, Alzheimer's Disease Neuroimaging Initiative (2021) Deep learning based multimodal progression modeling for Alzheimer's disease. *Stat Biopharm Res* 13(3):337–343
41. Shen Z, Yi Y, Bompelli A, Yu F, Wang Y, Zhang R (2021) Extracting lifestyle factors for Alzheimer's disease from clinical notes using deep learning with weak supervision. *arXiv preprint arXiv:2101.09244*
42. Buvanewari PR, Gayathri R (2021) Deep learning-based segmentation in classification of Alzheimer's disease. *Arab J Sci Eng* 46(6):5373–5383
43. Chen Y, Xia Y (2021) Iterative sparse and deep learning for accurate diagnosis of Alzheimer's disease. *Pattern Recogn* 116:107944
44. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C (2018) A survey on deep transfer learning. In: *International conference on artificial neural networks*. Springer, Cham, pp 270–279
45. Demiris G, Rantz MJ, Aud MA, Marek KD, Tyrer HW, Skubic M, Hussam AA (2004) Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study. *Med Inform Internet Med* 29(2):87–94
46. Ghazal TM, Issa G (2022) Alzheimer disease detection empowered with transfer learning. *Comput Mater Contin* 70(3):5005–5019
47. Savas S (2022) Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. *Arab J Sci Eng* 47(2):2201–2218
48. Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*, PMLR, pp 6105–6114
49. ALSaeed D, Omar SF (2022) Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. *Sensors* 22(8):2911
50. Bansal K, Batla RK, Kumar Y, Shafi J (2022) Artificial intelligence techniques in health informatics for oral cancer detection. In: Mishra S, González-Briones A, Bhoi AK, Mallick PK, Corchado JM (eds) *Connected e-health. Studies in computational intelligence*, vol 1021. Springer, Cham. [https://doi.org/10.1007/978-3-030-97929-4\\_11](https://doi.org/10.1007/978-3-030-97929-4_11)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.