# Alzheimer's disease classification: a comprehensive study

Ayoub Assmi[1] · Khaoula Elhabyb[2] · Achraf Benba[1] · Abdelilah Jilbab[1]

## Abstract

Alzheimer's disease (AD) is a neurodegenerative disease that is well-known for causing continuous loss of memory, cognition, and other higher brain functions. AD is not a single disease, but rather a group of related diseases with similar characteristics. The use of deep neural network-based pattern classification techniques, such as convolutional neural networks, is effective in classifying patients into different sub-types of AD and in distinguishing the different stages of severity of the disease. in the medical field, early detection of its start can be quite beneficial. This article focuses on the early detection of various stages of cognitive aging and AD using neuroimaging and transfer learning (TL). Images of imagery via resonance magnetic (IRM) obtained from a Kaggle database called Alzheimer's Dataset ( 4 class of Images) with several classes of non-dementia (NONDEM), very mild dementia(VERDEM), mild dementia(MILDEM), moderate dementia(MODDEM) are classified using a transfer learning approach. In this work, we compare the classification performance of six pre-trained networks, which are VGG-19, VGG-16, ResNet-50, InceptionV3, Xception, and DenseNet169. They were enthralled and tested using 6400 images from the Kaggle data pool. The confusion matrix and its parameters are used to assess the classification performance of these six networks. VGG-19, VGG-16, Inception-V3, Xception, ResNet-50, and DenseNet169 all have 92.86%, 92.83%, 91.04%, 90.57%, 85.99%, and 88.64% overall precision in MA detection, respectively.

✉ Ayoub Assmi
ayoub_assmi@um5.ac.ma

Khaoula Elhabyb
elhabyb.khaoula@inpt.ac.ma

Achraf Benba
achraf.benba@um5s.net.ma

Abdelilah Jilbab
abdelilah.jilbab@ensam.um5.ac.ma

1　Electronic Systems, Sensors and Nanobiotechnologies, National School of Arts and Crafts(ENSAM), Mohammed V University in Rabat, Rabat, Morocco

2　Networks, Architectures, Services Engineering and Security, National Institute of Posts and Telecommunications(INPT), Rabat, Morocco

# 1 Introduction

The neurological illness known as Alzheimer's disease (AD) affects 88% of the older population and is the primary cause of dementia in this group [1]. It affects approximately 50 million people worldwide. It has an impact on the cognitive abilities of patients and consequently on their quality of life. The main signs of the disease are forgetting past events or conversations or remembering recent conversations. There is no definitive solution for this disease, but early diagnosis and subsequent treatment are important to help slow its progression.

The mortality rate of Alzheimer's disease is reduced when it is diagnosed early. There was a time in the history of Alzheimer's treatment when the disease could not be detected until after death. In contrast, medical imaging techniques now play a major role in the diagnosis and treatment of Alzheimer's disease. Some imaging methods used to identify Alzheimer's disease include X-ray scanning, magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), and positron emission tomography (PET). Because it is non-invasive and generates high-resolution 3D pictures, magnetic resonance imaging (MRI) is one of the most widely utilized techniques. Manually processing these images is of high cost and often time-consuming. This has led to the creation of algorithms capable of extracting the most relevant information from the data to support a diagnosis. Supervised deep learning-based approaches are used to classify MRI images of patients with different stages of Alzheimer's disease, which varies from non-dementia (NONDEM), very mild dementia (VERDEM), mild dementia (MILDEM) to moderate dementia (MODDEM). In particular, convolutional neural networks (CNNs) [2] showed the highest performance in this task, however, they require a huge volume and variety of data and are time-consuming. However, recent works have shown that CNNs can be successfully adapted to overcome difficulties such as large data sets and network complexity.

Karim Aderghal et al [3] present a technique of cross-modal transfer learning across structural MRI to diffusion tensor imaging modality to overcome the issue of a lack of publicly available big datasets to train on. The model parameters are initialized by the authors using models pre-trained on a structural MRI dataset with domain-dependent data augmentation, and it is then trained using Mean Diffusivity data. The results of this work show a reduction in over-fitting issues and an improvement in the model performance in classifying normal control, Alzheimer's patients, and moderate cognitive impairment.

Buvaneswari et al [4]. use a deep learning-based segmentation approach using SegNet [5] to sense features of Alzheimer's disease-related parts of the brain based on structural magnetic resonance imaging (MRIs) and then accurately classifying Alzheimer's disease and dementia using ResNet-101 [6].For AD classification and recognition.

Mosleh Hmoud Al-Adhaileh [7] uses transfer learning of two pre-trained deep neural network architectures, namely AlexNet [8] and ResNet50 [6], the experimental findings showed that the newly suggested approach has a higher detection accuracy than the current methods.

All of the above research has the common goal of diagnosing Alzheimer's disease as quickly and accurately as possible. Since there is a possibility to be spared from this irreversible disease by adopting preventative treatments that might considerably slow its progression, doctors advise diagnosing Alzheimer's disease as early as possible. Deep learning has proven to be the perfect solution to address all these challenges and produce fascinating results, which is the focus of this research. Deep learning can be used to efficiently classify brain scans at different stages of the disease. With the help of the high-performance GPU computing platform, it is feasible to learn a lot of data in a short amount of time.

The principal objective of this research is to develop and conduct a comparative analysis of classification models designed for Alzheimer's disease classification, primarily leveraging MRI data. We aim to achieve this by employing a range of deep-learning CNN-based algorithms. A central focus of our study lies in meticulously assessing the efficacy and accuracy of the chosen methodology in the precise classification of Alzheimer's disease.

Furthermore, our investigation extends beyond the confines of algorithmic comparison. We also delve into the examination of state-of-the-art methods for Alzheimer's disease stage classification. This comprehensive evaluation allows us to discern which of these methods excels in terms of accuracy, precision, and F1 score. This endeavor is instrumental in enhancing our capacity to predict Alzheimer's disease in its early stages, thereby aiding in the prevention of its progression. The format of this article is as follows: Section 1 introduces the topic of Alzheimer's disease, its characteristics, and its types. Section 2 provides the academic approaches for the classification of Alzheimer's disease based on transfer learning. Section 3 focuses on the field of machine learning, especially convolutional neural networks, their architecture, and their applications for classification. Section 4 presents the definition and operation of the algorithms applied to transfer learning and contributes to the implementation of classification models. Sections 5 and 6 present and discuss the results of the implemented models and, finally, Section 7 concludes the paper.

## 2 Existing works on AD classification using transfer learning

Artificial intelligence's subset of machine learning (ML) includes all methods that attempt to mimic human behavior [9]. Machine learning concerns the utilization of Statistical methods and algorithm development that allow the computer to learn automatically from the data and to improve over time [10, 11].

From a single cross-sectional structural brain MRI, Silvia Basaia et al [12]. developed and validated a deep learning system to predict individual diagnoses of Alzheimer's disease (AD) and moderate cognitive impairment that would evolve into AD (c-MCI). On 3D images, convolutional neural networks (CNN) were used to distinguish between AD, c-MCI, and s-MCI(stable MCI). CNNs were able to identify c-MCI patients from s-MCI patients with up to 75% accuracy. CNNs are a valuable tool for automatically diagnosing patients across the Alzheimer's disease spectrum. Despite the heterogeneity in imaging techniques and scanners, the method functioned effectively without prior feature engineering. CNNs can help to speed up the use of structural MRI as a tool for patient assessment and management.

Manhua Liu et al [13] use convolutional neural networks (CNNs) to propose a multi-model deep learning framework for autonomous joint segmentation of the hippocampus and categorization of Alzheimer's disease using structural MRI data. First, a deep multi-task CNN model for hippocampus segmentation and illness classification is created. Second, based on the outcomes of the hippocampus segmentation, a 3D dense connection convolutional network (3D DenseNet) [14] is built to learn the features of retrieved 3D patches. Finally, the classification of the illness condition is performed utilizing the multi-task CNN and DenseNet models' learned features. The Alzheimer's Disease Neuroimaging Initiative (ADNI) database [15] contains baseline T1-weighted structural MRI data from 97 Alzheimer's disease individuals, 233 MCI subjects, and 119 normal control (NC) subjects. For hippocampus segmentation, the combined technique achieves an 87.0 percent similarity coefficient. Furthermore, when compared to NC subjects, the composite technique obtains an accuracy of 88.9% and an

AUC of 92.5 percent for the classification of AD subjects, and an accuracy of 76.2 percent and an AUC of 77.5 percent for the classification of MCI participants.

Buvaneswari et al [4]. use a deep learning-based segmentation approach using SegNet to sense features of Alzheimer's disease-related parts of the brain based on structural magnetic resonance imaging (MRIs) and then accurately classify Alzheimer's disease and dementia using ResNet-101. To classify, It has been shown that the ResNet-101, trained using features taken from SegNet and the ADNI dataset, can achieve a high degree of automated classification. ResNet is trained to classify images using the 7 morphology features-white matter, gray matter, cortical thickness, cortical area, sulcus contour, gyri hippocampus, and cerebrospinal fluid space-that were extracted from 240 MRIs using SegNet. This classifier has achieved 96 percent sensitivity and 95 percent accuracy from 240 ADNI MRIs other than the training ones.

Sadiq et al [16] using the Open Access magnetic resonance imaging (MRI) dataset from the Series of Imaging Studies and the Alzheimer's Disease Neuroimaging Initiative (ADNI). The authors suggest creating a very small feature vector for each MRI using the 3D Shearlet transform (3D-ST) [17]. For AD classification, the feature vectors from 3D-ST and CNN are combined. The feature vectors are combined and then used to train a classifier. The customized CNN model is used, in which all the descriptors are processed end-to-end to obtain the classification model. The experimental findings demonstrate that the performance of classification is enhanced by the combination of deep features with shearlet-based descriptors.

Haifeng Wu et al [18] This article proposes a 3D transfer network that is built on a 2D transfer network to classify the illness and the normal groups with MRI and uses machine learning as an additional diagnostic of Alzheimer's disease. The method uses a 2D transfer network to extract characteristics from 2D MRI slices, which are subsequently subjected to dimension reduction. Then, for categorization, all of a subject's 2D slices' features are combined. The experiment's findings show that the recommended 3D network's classification precision is higher than the present 2D transfer network's, increasing by around ten percentage points, and that its classification time is just one-fourth of the current time.

The work of Jayanthi VenkatramanShanmugam et al [19] focuses on utilizing neuroimaging and transfer learning to identify early signs of cognitive change and AD (TL). A transfer learning strategy is used to classify the magnetic resonance imaging (MRI) pictures of early mild cognitive impairment (EMCI), moderate cognitive impairment (MCI), and late mild cognitive impairment (LMCI). three pre-trained networks, which are GoogLeNet [20], AlexNet and ResNet-18 [6], are used for this classification. 6000 photos from the ADNI database are used to train and test the models. The confusion matrix and associated metrics are used to analyze the classification performance of the three networks. In terms of AD detection, GoogleNet, AlexNet, and ResNet-18 have worldwide accuracies of 96.39%, 94.08%, and 97.51%, respectively. The parameters of the confusion matrix are also used to analyze the performance of the pre-trained networks by class.

Mosleh Hmoud Al-Adhaileh [7] utilizes in this article, two deep neural network studies, called AlexNet and Restnet50 [6], which have been applied for AD classification and recognition. The data from magnetic resonance imaging (MRI) brain scans obtained from the Kaggle website is the dataset utilized in this article to assess and test the proposed model. To accurately categorize AD, the convolutional neural network (CNN) method has been used. The AlexNet and Restnet50 transfer learning models were used to pretrain the CNN. The experimental findings showed that the newly suggested approach has a higher detection accuracy than the current methods. For brain MRI datasets, the AlexNet demonstrated remarkable performance on all five assessment measures (accuracy, F1 score, precision, sensitivity, and specificity). In comparison to Restnet50, AlexNet outperformed it with an

accuracy of 94.53%, specificity of 98.21%, F1 score of 94.12%, and sensitivity of 100%. Table 1 provides a summary of the literature review on Alzheimer's detection.

# 3 Background for Alzheimer's disease

## 3.1 Machine learning

A machine learning-based image segmentation approach is [22] typically used to classify the area of interest, such as the diseased or healthy region. Preprocessing, which may include the use of a filter to remove any noise or increase contrast, is the first step in creating such an application. After the preprocessing step, the image is segmented using techniques such as thresholding, clustering, and edge segmentation. After segmentation, features are collected from the color, texture, contrast, and size information of the ROI. After that, the most important components are identified using statistical analysis or feature extraction techniques like principal component analysis (PCA). After that, a machine learning (ML) classifier like SVM or CNN is used with the selected attributes. After being trained, the ML classifier may be used to categorize brand-new, unknown data.

## 3.2 Convolutional neural networks

CNN is the most extensively used deep learning architecture since it is extremely comparable to a traditional NN. For greater performance and efficiency, CNN uses local connectors known as local receptive field and weight sharing, which are stacked on top of each other. Because of the deep design, these networks can learn a wide range of complicated features that a simple neural network can not [22]. Computer vision, which has many uses such as autonomous driving, robotics, and therapies for the blind, is powered by convolutional neural networks. A three-dimensional network of neurons in the CNN uses an image as its input but only links to a . Layers in a CNN include convolutional layers [23], non-linear activation layers (such the rectified linear unit (ReLU) layer, pooling layers [24], and fully connected layers [25]. To create feature map volumes made up of features retrieved by the filter, the convolutional layer performs a convolution operation between pixels of the input image and a filter. ReLU is a nonlinear activation layer that applies the feature to the input values to hasten learning and increase nonlinearity. The pooling layer subsamples the input values to reduce the spatial dimension of the picture and prevent overfitting [26] since calculations depend on nearby pixels. Additionally, it has translational invariance. The last layer of a CNN is often a fully linked layer, similar to the hidden layers of standard NNs, where every neuron is related to every other neuron in the layer before it. As previously stated, CNNs are commonly employed to solve classification difficulties, Fig. 1 shows CNN architecture.

### 3.2.1 Convolutional layer

The convolutional layer [23] may determine the convolution for a pixel by using the following equation [7]:

$$net(i, j) = (x * w)[i, j] = \sum \sum mnx[m, n]w[i - m, j - n], \qquad (1)$$

Where w is the noyau or filter matrix, x is the entering data made up of a collection of pictures, the asterisk marks the convolution process, and net(i,j) is the result of the convolutive layer

**Table 1** Previous work in AI-based Alzheimer detection

| Researchers Works | Data set | Architecture | Concept | Evaluation |
|---|---|---|---|---|
| • Basaia et al. [12] | • ADNI MRI data | • CNNs | • A cross-sectional brain structural MRI is used to build and verify a deep learning algorithm that can predict a person's diagnosis of Alzheimer's disease (AD) and moderate cognitive impairment that would advance to AD (c-MCI). | • Accuracy: 75 % |
| • Liu et al. [13] | • ADNI MRI data | • 3D DenseNet | • A multi-task deep CNN model is developed for the categorization of diseases and combined hippocampus segmentation. DenseNet uses the findings of its segmentation to categorize the Disease stage. | • Accuracy:92.5%<br>• Precision:88.9% |
| • Buvaneswari et al. [4] | • ADNI MRI data | • ResNet-101<br>• SegNet | • Presents a deep learning-based segmentation technique based on SegNet to find characteristics of MRI-relevant brain regions for AD, followed by ResNet-101 classification of AD. | • Accuracy: 95%<br>• Sensitivity: 96% |
| • Sadiq et al. [16] | • ADNI MRI data<br>• OASIS MRI data [21] | • SqueezeNet<br>• MobileNet<br>• Xception<br>• Inception | • used a mix of deep features and 3D shearlet-based descriptors to classify Alzheimer's illness. | **Accuracy:**<br>• SqueezeNet:84%<br>• MobileNet: 90%<br>• Xception: 90%<br>• Inception: 86% |
| • Wu et al. [18] | • ADNI MRI data | • MobileNet<br>• VGG16<br>• Inception | • Presents a two-dimensional (2D) transfer network-based three-dimensional (3D) transfer network for the classification of AD using magnetic resonance imaging (MRI). | **Accuracy:**<br>• MobileNet:74%<br>• VGG16: 77%<br>• Inception: 78% |
| • Shanmugam et al. [19] | • ADNI MRI data | • GoogLeNet<br>• AlexNet<br>• ResNet-18 | • Using neuroimaging and transfer learning, early stages of cognitive impairment and AD can be identified (TL). | **Accuracy:**<br>• GoogleNet: 96.3%<br>• AlexNet: 94.08%<br>• ResNet-18: 97.51% |
| • Mosleh Hmoud Al-Adhaileh [7] | • MRI brain images Kaggle website | • AlexNet<br>• ResNet50 | • AD was successfully classified using the convolutional neural network (CNN) technique. AlexNet and Restnet50 transfer learning models were used to train the CNN. | • AlexNet: 94,53%<br>• ResNet50: 58,7% |
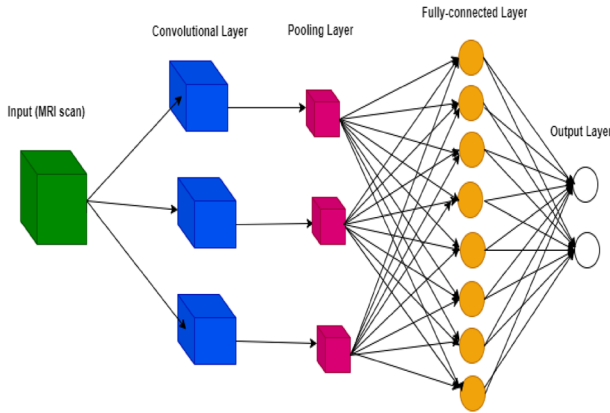
**Fig. 1** CNN architecture

that sends it to the subsequent layer. In the layer below, the outcome of the entrance and nose is computed, aggregated, and expressed as an analog point. The tangled sofa is seen in the figure.

### 3.2.2 Nonlinearity

The mathematical procedure was completed by the convolutional layer, and the results were sent to the nonlinear layer, which came next [27] You may trim or adjust the discharge that has been produced using this couch. However, this layer is employed to saturate or limit the output. The convolution layer is inextricably linked to the nonlinearity layer [7]. Deep learning approaches have mostly depended on the two activation functions sigmode and tanh during the past 20 years. The following two equations, however, show that the rectified linear unit (ReLU) has simpler representations of the functions and gradient:

$$ReLU(x) = max(0, x); \tag{2}$$

$$\frac{d}{dx} ReLU(x) = \begin{cases} 1 \text{ si } 1 \text{ if } x > 0 \\ 0 \text{ otherswise} \end{cases} \tag{3}$$

In (2), the ReLU function returns the maximum of 0 or the input value x. In simple terms, it returns x if x is positive and 0 otherwise.

Equation (3) is the derivative of ReLU concerning x, i.e. 1 for positive x and 0 for non-positive x. This derivative is crucial in neural network learning, as it enables network weights to be updated during backpropagation.

In summary, the ReLU function introduces non-linearity into a neural network by producing the input for positive values and zero for non-positive values, and its derivative simplifies to 1 for positive input and 0 otherwise.

### 3.2.3 Pooling layer

CNN uses pooling for two reasons [24]. First, a fixed size for the output feature map of pooling is needed for classification. No matter how big the filters are, using max pooling on each of the 256 filters will provide a 256-dimensional output, for instance. Since it enables
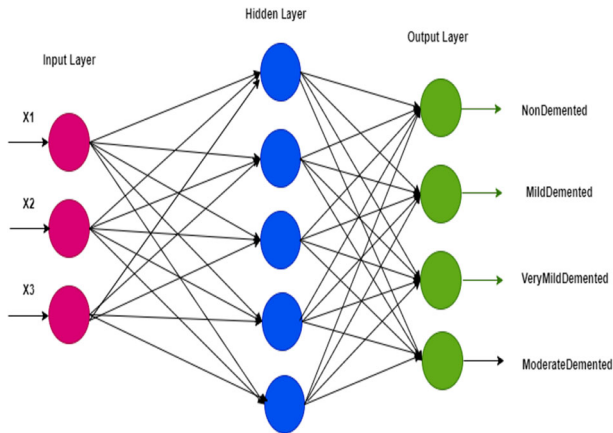
**Fig. 2** Fully connected deep learning

a decrease in the dimensionality of the data and a reduction in the time needed for data training for upcoming layers in the network, downsampling is a crucial step in the layer pooling process. The pooling layer can be researched and utilized in image-handling systems to reduce resolution. The pooling does not take the number of filters into account.

### 3.2.4 Fully connected layer

After the pooling layer comes the fully connected layer [25], It is utilized to link and arrange every neuron in a conventional neural network. As a result, every neuron in a layer is directly coupled to every other neuron in the layer above it and the layer below it. The CNN method's most often utilized parameter is this layer. The time required for CNN data training may be cut down with the usage of this layer. The fundamental disadvantage of a fully connected layer is that it requires several parameters that need a more complex calculation of the training samples. A fully linked neural network is displayed in Fig. 2.
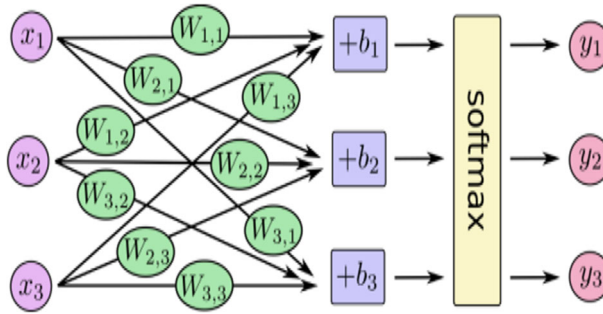
### 3.2.5 Softmax layer

The probability distribution of N-dimensional vectors for the input pictures is computed using the softmax layer [28], which is the last layer of the model being shown. The main use of softmax in this model's output layer is multiclass classification in deep learning-based models. The properly adjusted target class for the input image can be chosen with the aid of an accurate output probability calculation. The softmax layer represents a specific probability of the output and is differentiable. Additionally, the exponential component raises the likelihood of reaching maximum values. The softmax function is represented by the following equation:

$$Oi = \frac{e^{xi}}{\sum_{l=1}^{M} e^{xi}} \tag{4}$$

where xi is the output I before the softmax, Oi is the softmax output number I, and M is the total number of output neurons [7]. The position of the softmax layer in the network is depicted in Fig. 3:

**Fig. 3** Example of a Softmax layer utilization. X is the feature vector of a training sample, W is the weight vector, B is the bias unit, and Y is the model output

## 4 Transfer learning

Transfer learning [29] is an effective way to train CNNs when there is a lack of training data and appropriate computing resources. Transfer learning refers to a procedure in which a trained model on one problem is used for refining the other one. It helps to shorten the time to train a model and to build a better-generated model [30]. The training of deep convolutional neural networks (CNNs) from the ground up is challenging because it demands a large quantity of training data. In turn, CNN models are time intensive, at times taking many days or weeks to build. These restrictions can be surpassed by reusing the pre-trained model weights. The models are trained on over a million IRMs and can categorize the IRM into 100 classes. The high-performance models can be used straight away or, for a new task, can be mixed. The pre-trained models may be applied in a variety of ways. They may be used to categorize IRM from a fresh dataset, to start. Secondly, the pre-trained model is used in picture preprocessing and feature extraction. The state-of-the-art architecture templates used for transfer learning are as follows: Visual Geometry Group (VGG16), (VGG19), Residual Networks (ResNet), Inception CNNs, Xception and the procedure of functioning is cited in Fig. 4.

In Fig. 4, the diagram shows that we have divided the MRI data set into training and test sets for the evaluation of our models, in particular using transfer learning. The MRI data set represents 100%. For training purposes, 80% of this data has been allocated. Of the training data, 20% of the MRI data is used as a validation set, and the remaining 80% is used to train models using transfer learning techniques. On the other hand, 20% of the total MRI data has been reserved exclusively for model performance testing, where we evaluate parameters such as accuracy, specificity, and sensitivity. This data partitioning strategy guarantees a comprehensive evaluation of model performance on previously unseen data while optimizing its parameters during the learning process.
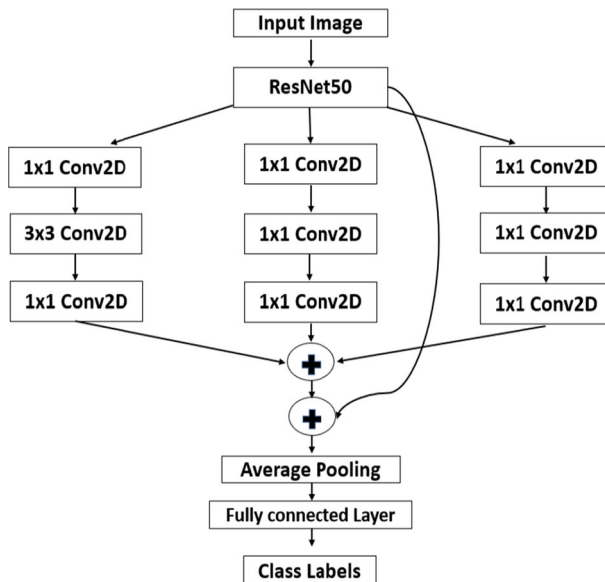
### 4.1 ResNet50

ResNet [6], a reliable deep learning architecture, won the 2015 Imagenet classification competition. Many computer vision jobs are saved by this design. As the number close to ResNet indicates how many deep levels are present, we picked ResNet50, which has 50 layers for processing. The presence of jump connections, which assist in the resolution of the evanescent gradient descent problem, is the primary advantage of the ResNet50 design.
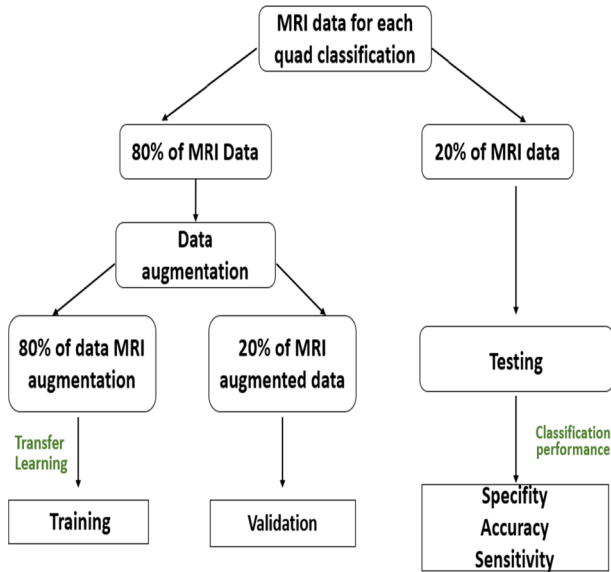
**Fig. 4** Diagram of the system used

256-dimensional input is supported by the ResNet50. The initial procedures, which include convolution and max-pooling and mention [31], are carried out by each ResNet, followed by stacked convolutions. A completely linked layer with 4 labels-health, scab, rust, and numerous diseases-follows the network's final averaging layer. The ResNet50's overall architecture is displayed in Fig. 5.



**Fig. 5** ResNet50 architecture
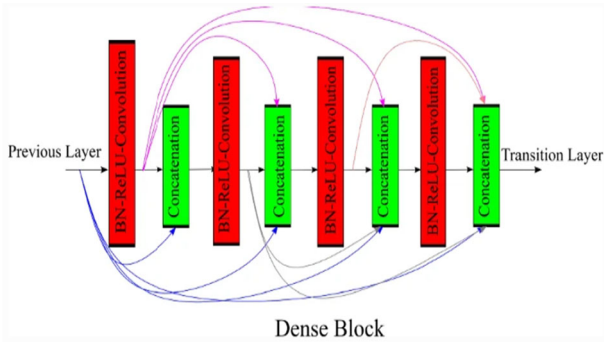
**Fig. 6** DenseNet-169 architecture

## 4.2 DenseNet-169

In DenseNet, which introduced dense connectivity in which each layer gets signals from all preceding layers merged by channel concatenation, resulting in a minor information bottleneck, Huang et al [32]. initially presented dense blocks. DenseNets are effective characterization extractors because they combine the characteristics of identity maps, deep supervision, decreased characterization redundancy, and diversified depth for feature reuse. Figure 6 introduces DenseNets of various depths, such as DenseNet-121, 169, 209, and 264.

For the characterization extraction procedure, the convolutional neural network DenseNet-169 (DenseNet-169) was used. Huang and al.(2016) proposed this model. DenseNet-169's architecture consists of an initial convolution and commonality layer, three transition layers, and four dense blocks. The categorization layer, cited as [33], comes next after these layers. A maximum commonality factor of 3x3 is employed with a factor of two in the first convolution layer, which then conducts 7x7 convolutions with a factor of two. The network then consists of three sets, each consisting of a transition layer before a dense block.

## 4.3 Inception-V3

The development of CNN classifiers was aided by the Inception network [34]. The basic goal of the Inception model is to figure out how a convolutional vision framework's optimal sparse local design can be related to and protected by readily available dense components. Here, a layer-by-layer construction is used, with the final layer's relationship statistics dissected and grouped into sets of units with strong connections. These groups serve as the following layer's units and are linked to the preceding layers' units. Each unit in the previous layer is supposed to be compared to a specific position in the input image before being merged into filter banks. The linked units in the lower tiers are focused on adjacent local regions. In comparison to previous CNN models, Inception v3 has an extremely complicated CNN architecture. The Tensorflow machine learning framework was used to train this model using the ImageNet dataset. The Inception module is made up of a series of convolution layers of various sizes that run in parallel. This network contains 48 layers and around 23 million parameters and it is shown in Fig. 7. The 2014 ILSRVC was won by this network.
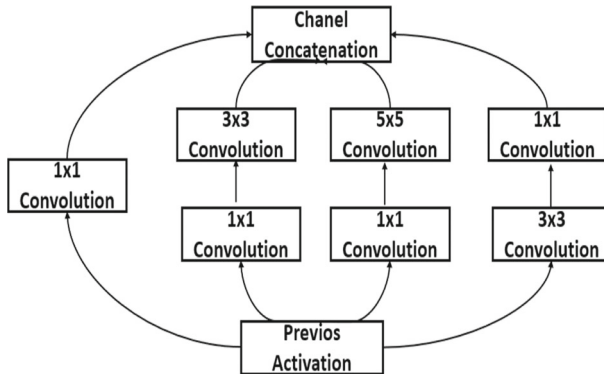
**Fig. 7** Inception-V3 architecture

## 4.4 Xception

Chollet presented the Xception architecture (2017) as a more advanced variant of the Inception model [35]. The architecture of Inception inspired the model. It is made out of a stack of residual connections from depth-separable convolutions. The Xception model uses inverted depth-separable convolutions. In other words, the input is first subjected to a point convolution process, followed by a depth convolution operation. Although it has almost the same number of parameters as Inception-V3, Using this approach, Xception outperformed Inception-V3 on the ImageNet dataset. The 36 convolution layers of the 14 modules that make up the Xception architecture are seen in Fig. 8.

## 4.5 VGG-16

The VGG-16 model [36], which employs fewer hyperparameters than other CNN designs, is one of them. In addition to a stride in the convolution layer and a stride 2 [37] in the shared layer with the same padding, this model uses a 3x3 size filter throughout the design. It is named VGG-16 because it is made up of 16 couches, including a softmax convolution couch. Instead of 10 couches FC with softmax, the last couch of the VGG-16 is replaced by two
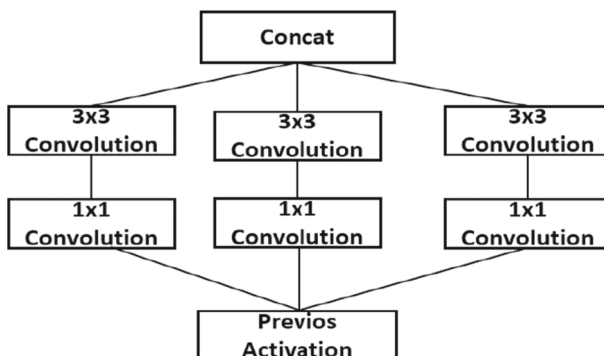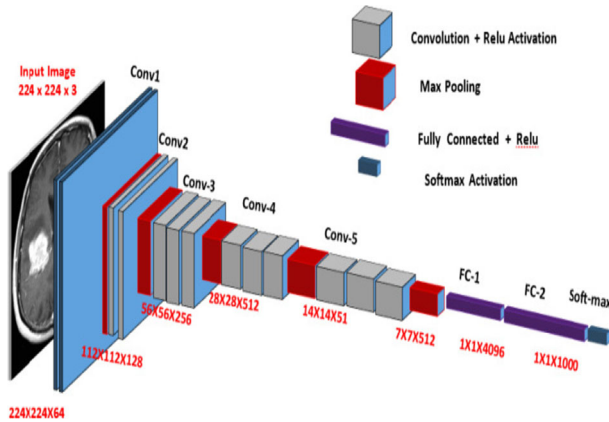


**Fig. 8** Xception architecture

**Fig. 9** VGG-16 architecture

FC with softmax activation. Klymentiev, 2019, and Rehman et al., 2020, explain the model VGG-16's whole construction by couche, and the model VGG-16's detailed couche-by-couch setup is illustrated in the following Fig. 9.

### 4.6 VGG-19

The VGG 19 [36] for widespread visual recognition, was created. The primary benefit of this technique is that its source code is available for public use, allowing us to quickly implement transfer learning and adapt the network to other designs. To enable the system to learn complicated features, the method also learns small collective kernels rather than a single massive kernel. The resilience of this design in comprehending complicated features is the basis for its use in the classification of neurological diseases. Figure 10 depicts the overall architecture of ACV 19.

### 4.7 Performance analysis

The accuracy of the trained network is computed and then compared to that of other pretrained networks. To increase classification accuracy, the network design, preprocessing procedures, and training settings are changed. The confusion matrix parameters for the TP, TN, FP, and FN are crucial for assessing this performance.

The ratio of true to total estimations is used to compute accuracy, which refers to the network's correct predictions. Indeed, the precision is untrustworthy if the network's performance for all classes is equal. The simple notion is that error is the polar opposite of precision in the sense, $Error = 1 - Accuracy$. It means that the total number of forecasts is divided by the number of instances that were incorrectly categorized.

The false positive rate, also known as fallout, is a comment on the network's rate of making positive predictions. It's calculated as the proportion of erroneous positive predictions to total negations.

Precision reveals the number of truly positive forecasts within the total number of positive predictions. As a result, it's also called the positive predictive value (PPV). The best accuracy number is 1.0 (100%), while the poorest is 0. The true positive rate, also known as recall or
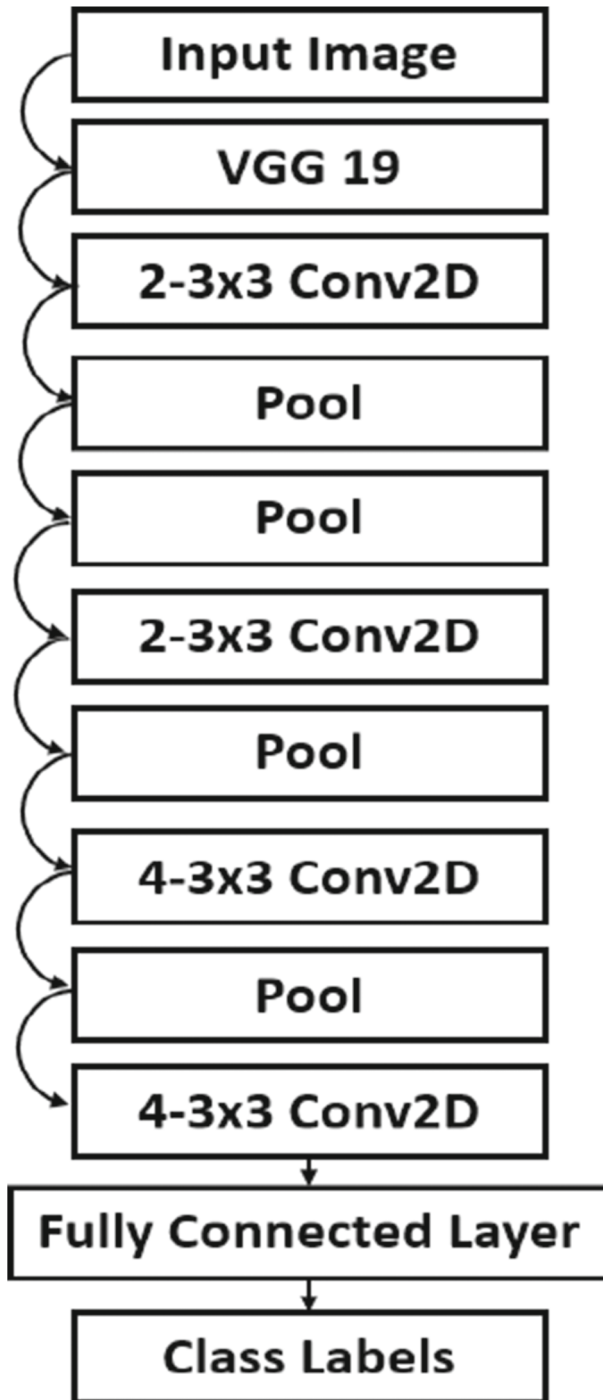
**Fig. 10** VGG-19 architecture

sensitivity, is expressed as follows $True positive rate = 1 - False positive ratre$. It is the proportion of appropriately classified positives in the overall number of positives [38].

$$F1 - score = \frac{2.Precision.Recall}{Precision + Recall} \quad (5)$$

The F1 score [39], also known as the harmonic mean of precision and recall [19] Alzheimer's, is beneficial when the model has poor accuracy and good memory or vice versa. This is the performance parameter that is most frequently used when network performance differs by class.
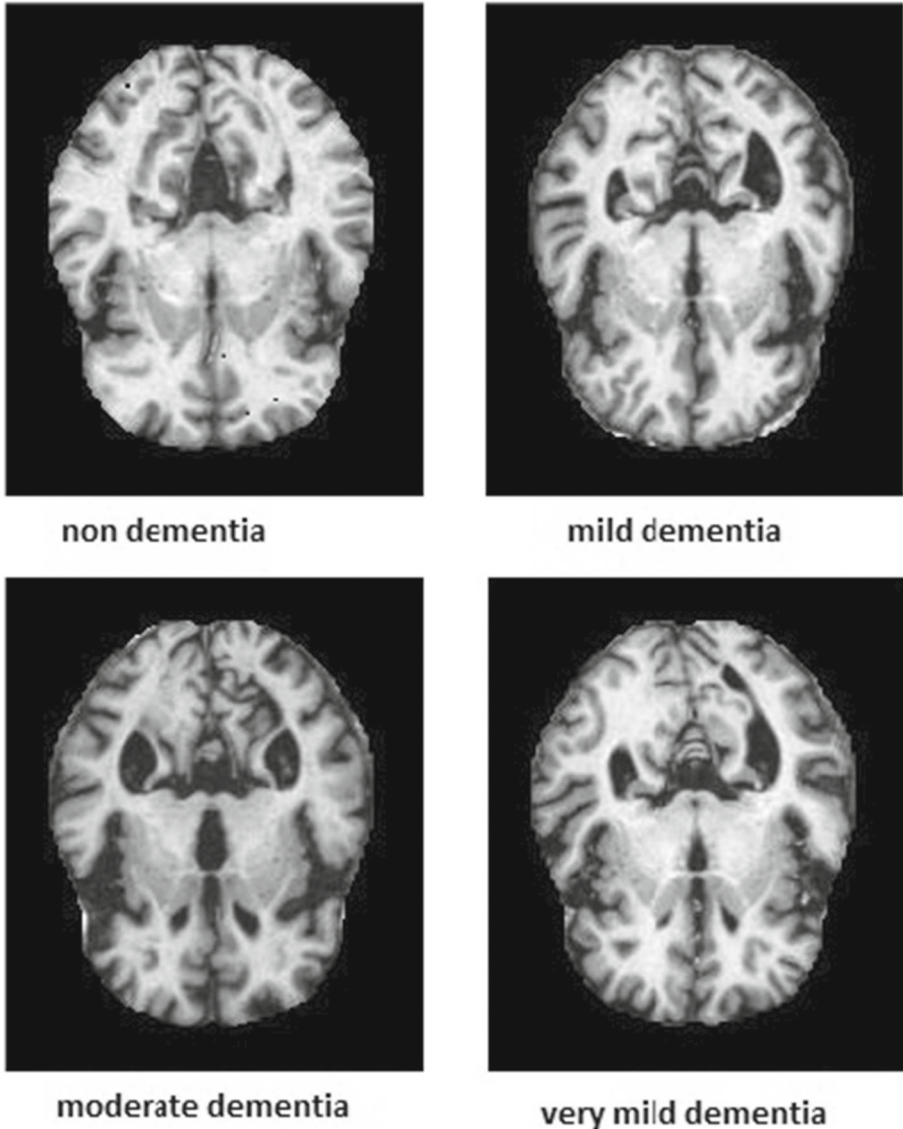


**Fig. 11** Examples of the four Alzheimer's stages (classes) sampled from the Kaggle Alzheimer's dataset

The real negative rate, which gauges the network's specificity or selectivity, is the proportion of accurately identified negatives to actual negatives.

# 5 Implementation and results

## 5.1 Data acquisition and preprocessing

The images come from the Kaggle database entitled "Kaggle Alzheimer's Dataset (4 image classes)".The data comprise four image classes, for both training and testing (Mild Demented, Moderate Demented, Non Demented, Very Mild Demented). All subjects were over 50 years of age, due to the age-related increased risk of Alzheimer's disease. Images were collected by MRI, and patients ranged in age from 55 to 100. Each class contains images of male and female subjects. The dataset contains a total of 6400 grayscale images of resolution 176×208. In this study, 5120 images are used for training the model, and 1280 images are used in the testing phase. the database is available online in [40]. Examples of the four classes sampled from the dataset are shown in Fig. 11.

## 5.2 Methodology

Our methodology is based on multi-class classification of scans from the Kaggle database to different stages of Alzheimer's disease, which are non-dementia (NONDEM), very mild dementia (VERDEM), mild dementia (MILDEM), moderate dementia (MODDEM). To perform this, we first Split the dataset to two folds, 5120 images for the training phase and 1280 images for the test, for a total of 6400 images. The sampled images, from both folds, are resized from (176x208) to (224x224) and transformed to grayscale. Figure 11 shows examples of the used data. Then, the prepared data is fed to six pre-trained classifier models, which are VGG-16, VGG-19, ResNet-50, DensNet-169, Inception, and Xception. We change the top layer of all models to fit the number of classes in our case. The training data is then used to refine the six models. The model parameters are optimized throughout 5000 iterations using gradient descent and the Adam Optimizer [41]. Categorical The adversarial loss is calculated using cross-entropy loss. Additionally, the input picture and video's pixel values are scaled to the range $[-1.1]$. The Keras Framework is used for the network's implementation. The performance of the model is finally evaluated for each class classification and for the total classification, using the confusion matrix, classification ratio, as well as loss and precision curves.

## 5.3 Algorithm of the proposed study

In this section, we present the algorithm employed to study the pre-trained models, with a specific focus on ResNet-50. It is important to note that this algorithmic approach is consistent across all other models within our study.

## 5.4 Results and comparison of performance

A performance evaluation of the six networks in each AD class is undertaken to verify the performance of the networks. The confusion matrix and accuracy curves of the six networks

---

**Algorithm 1** Algorithm of the proposed study

---

$img \leftarrow Kaggle - data$
$model \leftarrow pretrained - ResNet50$
$traing - data, testing - data \leftarrow split$
**for** epochNumber **do**
　　**for** batchSize **do**
　　　　$\hat{y} = model(features)$
　　　　$loss = crossEntopy(y, \hat{y})$
　　　　$optimization(loss)$
　　　　$accuracy()$
　　　　$best - Accuracy = max(bestAccuracy, accuracy)$
　　**end for**
**end for**

---

are used for this purpose. This enables us to assess the network's performance, check to see whether it performs consistently across all classes, and choose the right training parameters so that the network can successfully distinguish between the various AD classes.

### 5.4.1 Performance parameters of every model

The performance value is shown in this section:

- **Performance parameters of the VGG-16 :**

The performance of VGG-16 by class, shown in Table 2, demonstrates that the brain's IRM pictures are classified most correctly for the class MODDEM and least correctly for the class NONDEM. It is clear that the MODDEM class is extremely well classified, with 100% precision, and that the other classes' precision is less than that of the MODDEM class.

- **Performance parameters of the VGG-19:**

According to Table 3 performance of VGG-19 by class, brain MRI pictures are classified most accurately for the MODDEM class, with 92 percent accuracy, and least accurately for the NONDEM class.

- **Performance parameters of the ResNet-50:**

According to Table 4, which breaks out ResNet-50's performance by class, the classes MILDEM and NONDEM are the most accurate in classifying brain MRI images, with accuracy rates of 77% and 78% respectively. And the classes MODDEM and VERDEM least accurately.

- **Performance parameters of the Inception-V3:**

**Table 2** Performance parameters of the VGG-16

| Performance metric | MILDEM | MODDEM | NONDEM | VERDEM |
|---|---|---|---|---|
| Precision | 78% | 100% | 75% | 76% |
| Recall | 53% | 75% | 89% | 67% |
| F1-score | 63% | 86% | 82% | 71% |

**Table 3**  Performance parameters of the VGG-19

| Performance metric | MILDEM | MODDEM | NONDEM | VERDEM |
|---|---|---|---|---|
| Precision | 87% | 92% | 76% | 78% |
| Recall | 48% | 100% | 90% | 70% |
| F1-score | 62% | 96% | 82% | 74% |

**Table 4**  Performance parameters of the ResNet-50

| Performance metric | MILDEM | MODDEM | NONDEM | VERDEM |
|---|---|---|---|---|
| Precision | 77% | 38% | 78% | 49% |
| Recall | 26% | 25% | 60% | 79% |
| F1-score | 39% | 30% | 68% | 61% |

**Table 5**  Performance parameters of the Inception-V3

| Performance metric | MILDEM | MODDEM | NONDEM | VERDEM |
|---|---|---|---|---|
| Precision | 74% | 100% | 73% | 70% |
| Recall | 50% | 17% | 86% | 62% |
| F1-score | 60% | 29% | 79% | 66% |

**Table 6**  Performance parameters of the Xception

| Performance metric | MILDEM | MODDEM | NONDEM | VERDEM |
|---|---|---|---|---|
| Precision | 66% | 75% | 70% | 76% |
| Recall | 51% | 50% | 90% | 51% |
| F1-score | 58% | 60% | 79% | 61% |

**Table 7**  Performance parameters of the DensNet-169

| Performance metric | MILDEM | MODDEM | NONDEM | VERDEM |
|---|---|---|---|---|
| Precision | 74% | 100% | 71% | 68% |
| Recall | 36% | 33% | 89% | 58% |
| F1-score | 49% | 50% | 79% | 63% |

Based on the performance of Table 5 of Inception-V3 by class, brain MRI images are classified most accurately for the MODDEM class, with 100% accuracy, and least accurately for the other classes.

- **Performance parameters of the Xception:**

The performance of Xception by class, presented in Table 6, shows that the classification performance is almost similar for all classes.

- **Performance parameters of the DensNet-169:**

According to the class-wise performance of DenseNet-169 Table 7, brain MRI images are most accurately classified for the MODDEM class, with 100% accuracy, and the least accuracy for other classes.

### 5.4.2 Confusion matrix

In our case, the confusion matrix was utilized for the performance evaluations of the methods used after the classification. The following are the VGG-19, VGG-16, ResNet-50, Inception-V3 and Xception, and DenseNet169 confusion matrices Fig. 12.

### 5.4.3 Accuracy curves

The accuracy curves are very important for evaluating the performance of the methods used after classification. Here are the accuracy curves of VGG-19, VGG-16, ResNet-50, Inception-V3 and Xception, and DenseNet169 Fig. 13.

### 5.5 Comparison with other results

In our comparison with other studies, our Alzheimer's disease classification model stands out for its superior accuracy. Specifically, our model achieves an impressive accuracy of 92.83% for VGG-16 and 91.04% for Inception. In contrast, Haif Wu et al.'s research [18] reports lower accuracy values of 77% for VGG-16 and 78% for Inception. Furthermore, our ResNet-50 model attains an accuracy of 85.99%, surpassing the 58.7% accuracy reported by Mosleh Hmoud Al-Adhaileh et al. [7].

The outstanding performance of our model becomes even more evident when compared to the approach taken by Sadiq et al. [16], who used the Inception and Xception models, achieving accuracy rates of 90% and 86%, respectively. In contrast, our model achieves higher accuracy rates, with 90.57% for Xception and 91.04% for Inception, reinforcing its effectiveness in the classification of Alzheimer's disease. The Table 8 summarizes this comparison.

Our research uniquely classifies Alzheimer's disease into four stages (non-dementia, very mild dementia, mild dementia, and moderate dementia), a novel dimension not commonly explored. This distinctive approach enhances the overall comprehensiveness of our work, contributing to improved diagnostic accuracy in Alzheimer's disease classification. In summary, our results highlight the promising potential of our proposed model to advance the field of Alzheimer's disease classification research.
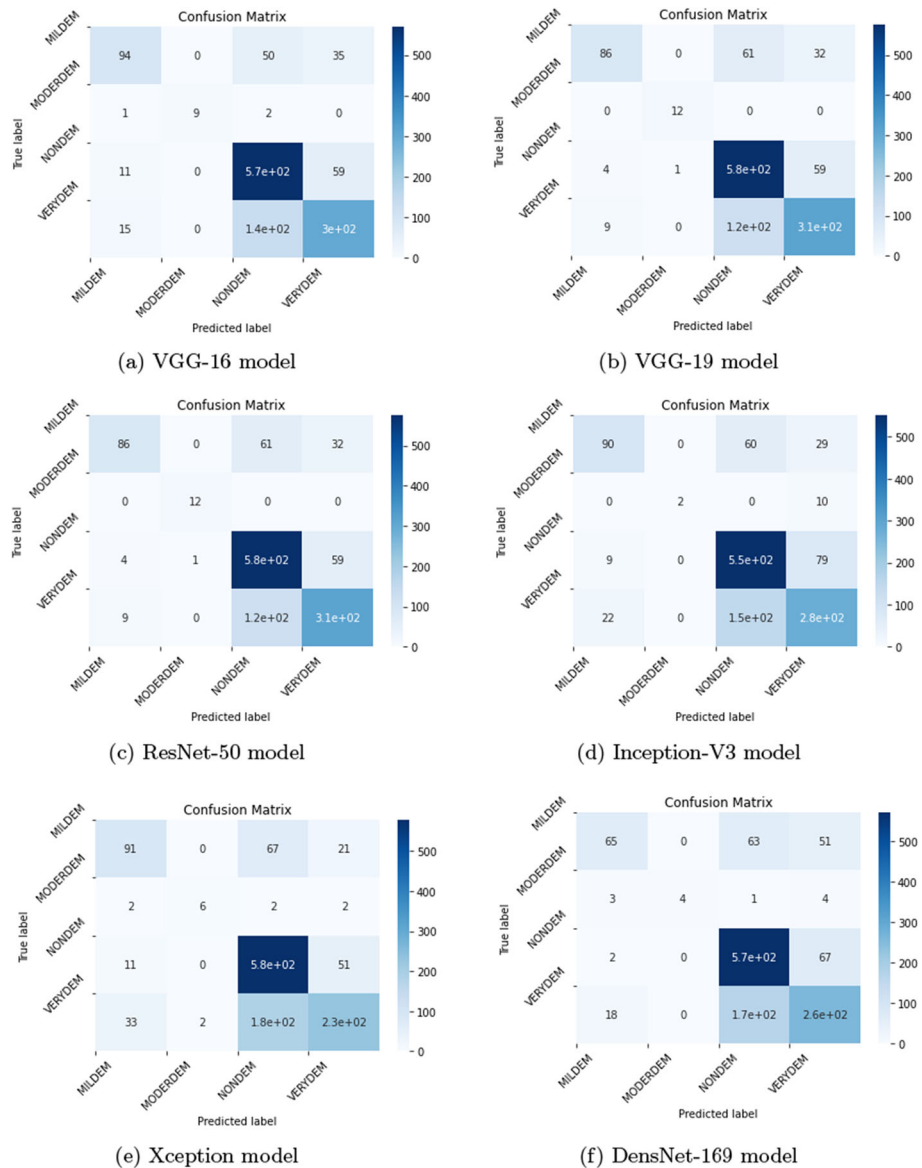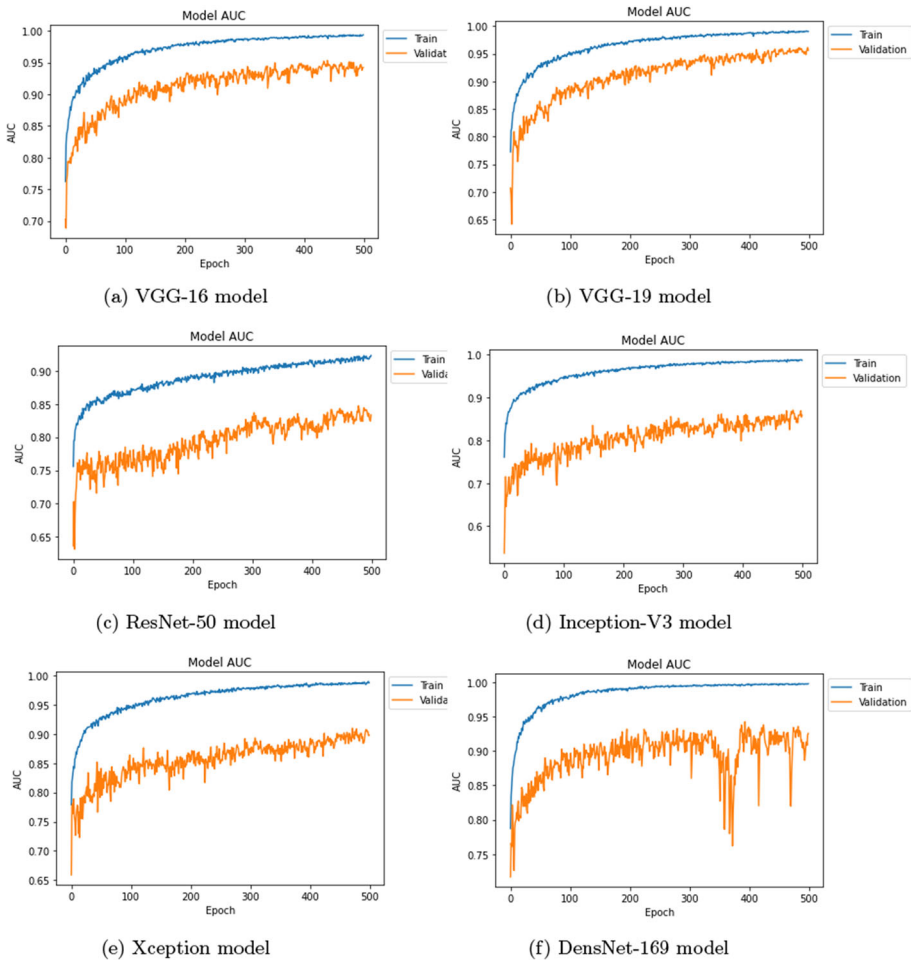
Fig. 12 Confusion matrix of every model

## 6 Discussion

Alzheimer's disease (AD) is a neurological condition that worsens with age and results in disconnected nerve cells. The hippocampus gets smaller, the brain shrinks, and the ventricles get bigger all as a result of AD. Memory, logic, and answers to issues in everyday activities are all impacted as AD advances. Many researchers' major objective is to provide a replacement for the AD diagnosis system.

(a) VGG-16 model

(b) VGG-19 model

(c) ResNet-50 model

(d) Inception-V3 model

(e) Xception model

(f) DensNet-169 model

**Fig. 13** Accuracy curve of every model

**Table 8** Comparison with other results

| Researchers Works | Architecture | Evaluation(Accuracy) |
|---|---|---|
| • Sadiq et al [16] | • Xception<br>• Inception | • Xception: 90%<br>• Inception: 86% |
| • Haif Wu et al [18] | • VGG-16<br>• Inception | • VGG16: 77%<br>• Inception: 78% |
| • Mosleh Hmoud Al-Adhaileh [7] | • ResNet-50 | • ResNet-50: 58,7% |
| • Ours | • VGG-19, VGG-16, Inception-V3, Xception, ResNet-50, and DenseNet169 | • VGG-19: 92.86%<br>• VGG-16: 92.83%<br>•<br>Inception-V3: 91.04%<br>• Xception: 90.57%<br>• ResNet-50: 85.99%<br>• DenseNet169: 88.64% |

Today, artificial intelligence methods have been successful in detecting and classifying neurological diseases and, of course, in classifying AD. In this work, we used CNN and specifically transfer learning for AD detection and prediction using MRI scans obtained from Kaggle. Based on the six learning models VGG-16, VGG-19, Inception-V3, Xception, ResNet-50, and DensNet-169, we classified the four types of AD as non-dementia (NON-DEM), very mild dementia (VERDEM), mild dementia (MILDEM), moderate dementia (MODDEM).The results obtained show that the VGG-16 and VGG-19 models give a good performance since the accuracy of both models is close to 93%. We rank second to the Xception and Inception-V3 models when we obtain an accuracy close to 91%, for the ResNet-50 and DensNet-169 models the accuracy decreases to 86% and 89% respectively. These results can be seen from the accuracy curves in Fig. 13 since the best curves are for the VGG-16 and VGG-19 models and more or less the curves for the Xception and Inception-V3 models, but the curves for the ResNet-50 and DensNet-169 show that models fall short in comparison to the others, especially when we look at the validation curves.

In the tables of performance parameters, we can see that with the Inception-V3, DensNet-169, and VGG-16 models, we obtained the best accuracy of 100% for the MODDEM class and close accuracies for the other classes, and with the ResNet-50 model we obtained a low accuracy of 38% for MODDEM. For the NONDEM and MILDEM classes, we have an accuracy between 66% and 87% for all learning models. For the VERDEM class, we have an accuracy between 70% and 78% for all models except the ResNet-50 model, for which we obtained an accuracy of 49%.

## 7 Conclusion

A correct diagnosis of AD is necessary for successful treatment. Because of this, several researchers have concentrated their efforts on developing a computer-aided system capable of providing early, accurate diagnoses of AD. In this study, six separate pre-trained deep learning models, VGG-19, VGG-16, ResNet-50, Inception-V3, Xception, and DensNet-169, were fine-tuned to fir the problem of AD classification by transfer learning. The 6400 MRI images in the Kaggle database are used to train and test these models to distinguish the four classes of maldysis, NONDEM, MILDEM, MODDEM, and VERDEM. Transfer learning is used to modify the architecture of the networks to classify the four different AD types. The results show that the learning models are quite effective in distinguishing AD patients from non-AD patients. The total performance of the networks is evaluated and each class separately. With these six networks, it is possible to clearly and accurately categorize the different stages of AD based on the performance analysis. From the results, it is clear that VGG-16 and VGG-19 networks have better performance than other networks. And the Xception and Inception-V3 are ranked second.

The primary goal of future study will be the creation of deep learning networks specifically for AD class categorization. According to the study's findings, deep learning methods and MRI technology can aid in the early identification of neurodegenerative disorders like Alzheimer's disease. Accurate identification of AD stages might speed up the creation of novel medications by giving a better way to measure the efficacy of targeted therapies that may be able to slow disease progression.

**Data availability statement** Data supporting the findings of this study are freely available in [Kaggle] at [https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images], reference number [40]

## Declarations

**Conflict of interest statement** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Sudharsan M, Thailambal G (2021) Alzheimer's disease prediction using machine learning techniques and principal component analysis (pca). Materials today: proceedings
2. O'Shea K, Nash R (2015) An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458
3. Aderghal K, Khvostikov A, Krylov A, Benois-Pineau J, Afdel K, Catheline G (2018) Classification of alzheimer disease on imaging modalities with deep cnns using cross-modal transfer learning. In: 2018 IEEE 31st International symposium on computer-based medical systems (CBMS). IEEE, pp 345–350
4. Buvaneswari P, Gayathri R (2021) Deep learning-based segmentation in classification of alzheimer's disease. Arab J Sci Eng 46(6):5373–5383
5. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495
6. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
7. Al-Adhaileh MH (2022) Diagnosis and classification of alzheimer's disease by using a convolution neural network algorithm. Soft Computing, 1–12
8. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
9. Sambath Kumar S, Nandhini M (2022) Automated classification of alzheimer's disease using mri and transfer learning. In: Mobile computing and sustainable informatics. Springer, ???, pp 663–686
10. Kumari A, Tanwar S, Tyagi S, Kumar N (2018) Fog computing for healthcare 4.0 environment: opportunities and challenges. Computers & Electrical Engineering 72:1–13
11. Tanwar S, Kumari A, Vekaria D, Kumar N, Sharma R (2022) An ai-based disease detection and prevention scheme for covid-19. Comput Electr Eng 103:108352
12. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M, Initiative ADN et al (2019) Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks. NeuroImage: Clinical 21:101645
13. Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, Xu M, Initiative ADN et al (2020) A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer's disease. Neuroimage 208:116459
14. Bui TD, Shin J, Moon T (2017) 3d densely convolutional networks for volumetric segmentation. arXiv preprint arXiv:1709.03199
15. Petersen RC, Aisen P, Beckett LA, Donohue M, Gamst A, Harvey DJ, Jack C, Jagust W, Shaw L, Toga A et al (2010) Alzheimer's disease neuroimaging initiative (adni): clinical characterization. Neurology 74(3):201–209
16. Alinsaif S, Lang J, Initiative ADN et al (2021) 3d shearlet-based descriptors combined with deep features for the classification of alzheimer's disease based on mri data. Comput Biol Med 138:104879
17. Patel VM, Easley GR, Healy DM (2009) Shearlet-based deconvolution. IEEE Trans Image Process 18(12):2673–2685
18. Wu H, Luo J, Lu X, Zeng Y (2022) 3d transfer learning network for classification of alzheimer's disease with mri. International Journal of Machine Learning and Cybernetics, 1–15
19. Shanmugam JV, Duraisamy B, Simon BC, Bhaskaran P (2022) Alzheimer's disease classification using pre-trained deep networks. Biomed Signal Process Control 71:103217
20. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9

21. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 19(9):1498–1507
22. Haque IRI, Neubert J (2020) Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked 18:100297
23. Wu J (2017) Introduction to convolutional neural networks. National Key Lab for Novel Software Technology. Nanjing University. China 5(23):495
24. Wu H, Gu X (2015) Max-pooling dropout for regularization of convolutional neural networks. In: International conference on neural information processing. Springer, pp 46–54
25. Sainath TN, Vinyals O, Senior A, Sak H (2015) Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 4580–4584
26. Ying X (2019) An overview of overfitting and its solutions. In: Journal of physics: conference series, vol 1168. IOP Publishing, p 022022
27. Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375
28. Banerjee K, Gupta RR, Vyas K, Mishra B et al (2020) Exploring alternatives to softmax function. arXiv preprint arXiv:2011.11538
29. 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, pp 270–279
30. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q (2021) A comprehensive survey on transfer learning. Proc IEEE 109(1):43–76. https://doi.org/10.1109/JPROC.2020.3004555
31. Subetha T, Khilar R, Christo MS (2021) A comparative analysis on plant pathology classification using deep learning architecture–resnet and vgg19. Materials today: proceedings
32. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
33. Varshni D, Thakral K, Agarwal L, Nijhawan R, Mittal A (2019) Pneumonia detection using cnn based feature extraction. In: 2019 IEEE International conference on electrical, computer and communication technologies (ICECCT). IEEE, pp 1–7
34. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
35. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
36. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
37. Rai HM, Chatterjee K (2020) Detection of brain abnormality by a novel lu-net deep neural cnn model from mr images. Machine Learning with Applications 2:100004
38. Gandhi ST (2020) Context Sensitive Image Denoising and Enhancement Using U-Nets. Rochester Institute of Technology, ???
39. Lipton ZC, Elkan C, Narayanaswamy B (2014) Thresholding classifiers to maximize f1 score. arXiv preprint arXiv:1402.1892
40. Sarvesh dubey, (2019, december). alzheimer's dataset (4 class of images) retrived march 2022 from https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images
41. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980