SN Computer Science (2024) 5:507
https://doi.org/10.1007/s42979-024-02868-4

REVIEW ARTICLE

Alzheimer Disease Detection Using MRI: Deep Learning Review

Pallavi Saikia[1] · Sanjib Kumar Kalita[1]

Received: 16 June 2023 / Accepted: 3 April 2024 / Published online: 30 April 2024

Abstract
Deep learning for Alzheimer disease detection using MRI is an emerging area of research in medical image processing. With the advent of new technologies based on methods of Deep Learning, medical diagnosis of certain diseases has become possible. Alzheimer's is a disease which till date has no cure but the progression of the disease can be slowed down or a person who might develop Alzheimer in future can be treated if early prediction of it is possible. Early prediction of the disease benefits medical professionals a lot for the correct diagnosis. Medical professionals label Alzheimer patients based on the progression of the disease as AD (Alzheimer's), CN (cognitive impairment) and MCI (mild cognitive impairment). In literature many Deep Learning models are used for the early detection of Alzheimer's disease. Though there are many image modalities, MRI images being non-invasive are considered best for these types of medical experiments. In the present study, we have studied the evolution of Alzheimer's disease over time, research gaps, challenges towards building advanced models, possible recommendations to overcome those challenges and determining the best performance model. We have focussed on an exhaustive and comprehensive survey of very deep learning-based research papers on Alzheimer's disease detection. The present work will benefit researchers by providing a clear direction for future scope in Alzheimer disease detection and analysis.

Keywords Alzheimer · DL · CNN · Classification

Introduction

Medical image analysis using computer has been emerging area of research as research in medical image can help in diagnosing, detection and prediction of diseases at a very early stage which not only benefits patients but also medical professionals. A large amount of data is being generated in every healthcare sector on daily basis and studies of these medical data bring out many new approaches to deal a particular disease. Analysis of medical data shows the trend in which the disease is progressing and also it's relevant information like age, gender, food habit, lifestyle which can be of great help to predict whether a certain disease might affect a person in future. Among other machine learning techniques, Deep Learning (DL) plays an active role in this field with lots of contribution to the medical research. The most important characteristic of Deep Learning is its ability to extract features automatically without any human effort. The feature extraction and selection tasks has reduced significantly. Convolutional neural network (CNN) is one such DL model which is used extensively in medical image processing. In this paper a review has been done on the role of Deep Learning in classification and detection of Alzheimer disease. Alzheimer disease is a subtype of dementia. A person suffering from dementia usually have difficulty in memory, co-ordination and reasoning. Sometimes it occurs due to the blockage of blood flow to the brain and in such case the brain doesn't receive enough oxygen and nutrients. As a result the brain cells starts dying. Depending on the severity of the disease, dementia has been divided into many categories out of which Alzheimer is the most common type of dementia. Alzheimer's diseaseis a neurological disorder which occurs when two abnormal proteins: tau and amyloid accumulate in the brain and due to which disruption of communication takes place in between the nerve cells. The accumulation of tau and beta amyloid are the biomarkers of Alzheimer's disease. It is a specific progressive disease of the brain that slowly causes impairment in memory and

✉ Pallavi Saikia
saikia.pallavi@gmail.com

1 Department of Computer Science, Gauhati University, Guwahati, India

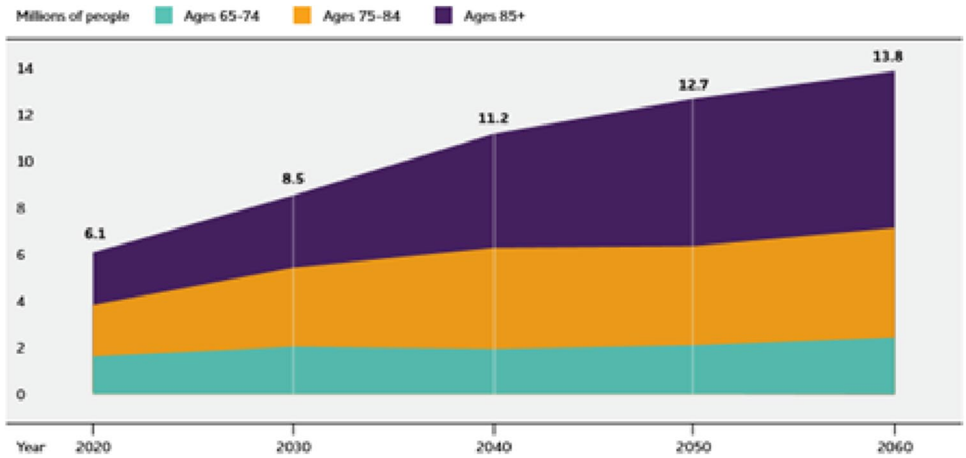


Fig. 1 Projected number of people age 65 and above in the US population with Alzheimer's dementia, 2020–2060

cognitive function and finally increases over time that a person suffering from Alzheimer's disease cannot even carry out the simple daily activities. The person finds difficulty in recalling the incidents that happened recently and also gets confused while ordering the events. It happens because of the surface layer cerebral cortex, that covers the cerebrum, which is the largest part of the brain, withers and shrinks. This damage causes the brain's inability to plan ahead, recall and concentrate. It also affects the hippocampus area which plays a very important role in memory [1–4]. Usually most of the elderly population are prone to Alzheimer. Alzheimer's disease phase are further categorized as mild, moderate and severe dementia. According to the report [5] by 2050, the number of people age 65 and older with Alzheimer's dementia is projected to reach 12.7 million in US [6] as depicted in Fig. 1.

In India also a lot of people are being affected by Alzheimer's disease and day by day the numbers are increasing. According to the Dementia India Report prepared for the Alzheimer's and Related Disorders Society of India (ARDSI) [7] the persons with dementia in younger age groups, 60–75 years, are expected to increase steadily over time; and a steep increment amongst age groups over 75 years can be predicted after 2030 as shown in Fig. 2.

The disease is not curable but with proper and early diagnosis, progression may be reduced. Due to advancement in technology, many research works are conducted to detect the disease at an early stage so that the treatment of such patients are done properly at the right time. Though till now there is no cure for Alzheimer, early detection and classifying them into different classes depending upon the severity of the disease helps the medical professionals to diagnose the same. Figure 3 depicts that a human brain shrinks due to Alzheimer's disease [8].

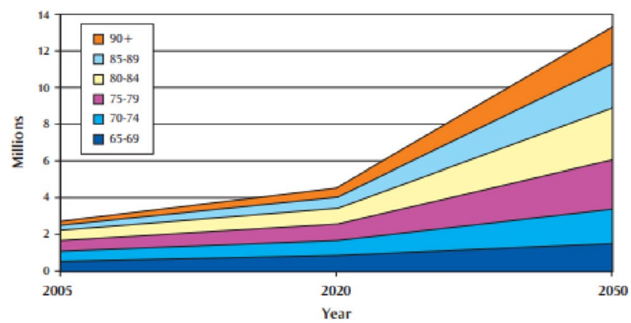


Fig. 2 Trend in dementia prevalence by age over time (2010–2050)

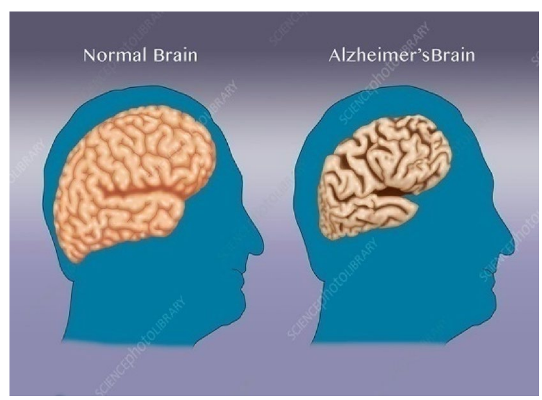


Fig. 3 **a** Normal brain **b** Alzheimer's brain

Dataset

Dataset play an important role in Machine Learning algorithm for efficient training. Development of a large quantity benchmark dataset is the primary challenge as any state of art classifier demands a large amount of data for feature extraction and for appropriate learning of the classifier. Since there is a deficiency of large scale corpus, much research is still working for large dataset. In our exhaustive research, we have found only three benchmark datasets. They are ADNI, OASIS, Kaggle.

ADNI

The Alzheimer's disease neuroimaging initiative (ADNI) is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease (AD). It is a consortium of universities and medical centers in the United States and Canada [9]. Since its launch more than a decade ago, the landmark public–private partnership has made major contributions to AD research, enabling the sharing of data between researchers around the world. ADNI dataset has been used in almost 3722 total publications since it's inception.

Three overarching goals of the ADNI study are:

1. To detect AD at the earliest possible stage (pre-dementia) and identify ways to track the disease's progression with biomarkers.
2. To support advances in AD intervention, prevention, and treatment through the application of new diagnostic methods at the earliest possible stages (when intervention may be most effective).
3. To continually administer ADNI's innovative data-access policy, which provides all data without embargo to all scientists in the world. The dataset provides different types of data which includes Structural MRI as well as Functional MRI [9]

The dataset consists of the three views of the brain: axial, coronal and saggital. ADNI dataset is updated from time to time adding more data each time. The details of the dataset is given in the Table 1 below:

OASIS

The Open Access Series of Imaging Studies (OASIS) is a project aimed at making neuroimaging data sets of the brain freely available to the scientific community. The dataset provides both longitudinal and cross-sectional MRI images. The OASIS datasets hosted by central.xnat.org provide the community with open access to a significant database of neuroimaging and processed imaging data across a broad demographic, cognitive, and genetic spectrum an easily accessible platform for use in neuroimaging, clinical, and cognitive research on normal aging and cognitive decline [10]. The specifications of the dataset is depicted Table 2:

Kaggle

Alzheimer's disease dataset can be obtained from the Kaggle website. The Data is hand collected from various websites with each and every labels verified. The main inspiration behind sharing this Dataset is to make a very highly accurate model predict the stage of Alzheimer's disease [11]. The dataset consists of MRI images of the axial view of the brain. The labels of Alzheimer's disease dataset available in Kaggle dataset are: Mild Demented, Moderate Demented, Non-Demented and Very Mild Demented. The specifications of the Kaggle dataset is depicted in Table 3:

Imaging Technique

The image data acquired from the dataset may have different imaging technique. Magnetic Resonance Imaging (MRI), most commonly used in analysing medical images, is a

Table 1 Specifications of ADNI Dataset with specified number of images in each category

Dataset name	Class labels with number of images	Total number of images	MRI acquisition
ADNI-1	Early Mild cognitive Impairment(EMCI)—477 Significant Memory Concern(SMC)—165 Alzheimer's Disease(AD)—527 Late MCI(LMCI)—235 Cognitive Normal(NC/CN)—1050 Mild Cognitive Impairment(MCI)—751	200 elderly controls 400 MCI 200 AD	1.5 T scanners using T1- and dual echoT2-weighted sequences
ADNI GO		Existing ADNI-1 + 200 early MCI	3 T scanners using T1-weighted and 2D FLAIR as well as T2*-weighted imaging was added
ADNI-2		Existing ADNI-1 and ADNI-GO + 150 elderly controls 100 early MCI 150 late MCI 150 AD	
ADNI-3		Existing ADNI-1, ADNI-GO, ADNI-2 + 133 elderly controls 151 MCI 87 AD	3 T scanners was used and nearly all of the imaging sequences from ADNI2 have been updated for inclusion in ADNI 3

Table 2 Specifications of OASIS Dataset

Dataset name	Description of the images	Number of images	MRI acquisition
OASIS-1	Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults	416	T1 weighted
OASIS-2	Longitudinal MRI Data in Nondemented and Demented Older Adults	150	T1 weighted
OASIS-3	Longitudinal Multimodal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer's Disease	1379	T1w, T2w, FLAIR, ASL, SWI, time of flight, resting-state BOLD, and DTI sequences
OASIS-3_TAU	A subset of OASIS-3 images that have also undergone TAU (AV1451) PET imaging	451	PET Imaging
OASIS-4	This clinical cohort was evaluated for memory disorders and dementia including clinical, csf, neurospychometric, and neuroimaging assessments. This is a unique dataset and not an update to the OASIS-3 Longitudinal Multimodal Neuroimaging dataset	664	

Table 3 Specifications of Kaggle dataset

Dataset name	Total number of images	Mild demented	Moderate demented	Non demented	Very mild demented
Kaggle	6400	896	64	3200	2240

Fig. 4 Transverse view of the brain

non-invasive method. Magnetic resonance imaging (MRI) is one of the best techniques for imaging the brain because it gives very high quality images, with excellent contrast between the different types of tissues. MRI image of a normal brain has been depicted in Fig. 4 [52].

MRI is based on the phenomenon of NMR (nuclear magnetic resonance) [12]. Atomic nuclei have the ability to absorb and re-emit radio frequency energy when exposed to a magnetic field. These atomic nuclei behave like tiny magnets which tends to align with the magnetic field exposed to it. Now if a radiofrequency is pulsed through it, the atomic nuclei are bound to change from their current alignment in a direction opposite to it and as a result it rotates slowly and steadily around a fixed point radiating it's own radio frequency. The radio frequency generated is captured and detected by the MRI scanner. After a certain period of time these radio frequencies ceases to exist and finally disappears. The amount of time required by the radio frequencies to finally disappear depends on the body tissue. This time is called T2 and the process is termed as T2 relaxation. As soon as the radio frequency dies off, the nuclei is realigned to it's original position as with the magnetic field. The amount of time required for doing this also depends on the body tissue and is called T1 relaxation. Different tissues take different T1 and T2 time. This process of realignment is repeated over and over so that enough radio frequency signals are detected for analysing diagnostically a particular image. By varying the sequence of radio frequency pulses applied & collected, different types of images are created. Repetition Time (TR) is the amount of time between successive pulse sequences applied to the same part of the tissue. Time to Echo (TE) is the time between the delivery of the radio frequency pulse and the receipt of the echo signal. The most common MRI sequences are T1-weighted and T2-weighted scans.T1-weighted images are produced by using short TE and TR times. The contrast and brightness of the image are predominately determined by T1 properties of tissue. Conversely, T2-weighted images are produced by using longer TE and TR times. In these images, the contrast and brightness are predominately determined by the T2 properties of tissue. Figure 5 depicts T1 and T2 weighted images.

MRI provides exquisite detail of brain, spinal cord and vascular anatomy. MRI has the advantage of being able to visualize anatomy in all three planes: sagittal, axial and coronal. MRI can detect flowing blood and cryptic vascular malformations. Figure 6 depicts the human brain in all the three planes [13].

There are mainly two types of Imaging Techniques: Structural MRI and Functional MRI [14].

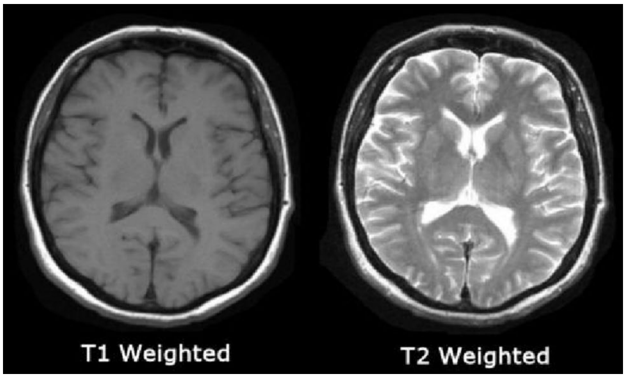


Fig. 5 T1 and T2 weighted Brain MRI

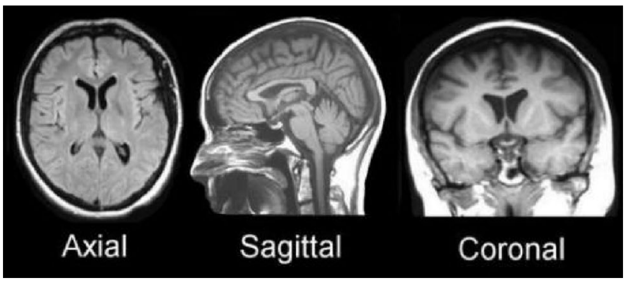


Fig. 6 Human Brain in all the three planes: axial, sagittal and coronal

(a) *Structural Imaging Technique*: This technique is usually used for visualizing and analysing the anatomical properties of the brain. It helps in detecting any brain damage or any other abnormalities related to brain. The two main structural techniques include Computed Tomography (CT) and Magnetic Resonance Imaging (MRI).

(b) *Functional MRI*: Functional imaging technique is used to identify brain areas and underlying brain processes that are associated with performing a particular cognitive or behavioral task. The two main structural techniques are Functional Magnetic Resonance Imaging fMRI and Positron Emission Tomography (PET).

Table 4 Labels used in Alzheimer's disease

HC: Healthy control	VMD: Very mild demented
NL: Cognitive healthy	EMCI: Early MCI
MID: Mild demented	LMCI: Late MCI
MOD: Moderate demented	SMCI: Stable MCI
ND: Non-demented	PMCI: progressive MCI
MCI.C: MCI convertible	MCI.NC: MCI non-convertible
NC: Normal control	

(c) *Hybrid MRI*: Recently one more imaging technique is used called Hybrid imaging. Hybrid imaging technique combines two imaging modalities to generate more powerful images. It provides detailed anatomical properties thus helping localization of a particular part of a body. Some of the hybrid imaging modalities are: PET-CT, SPECT-CT, MRI-PET, MRI-SPECT, ultrasound and MRI, ultrasound and CT, MRI and CT.

Literature Review

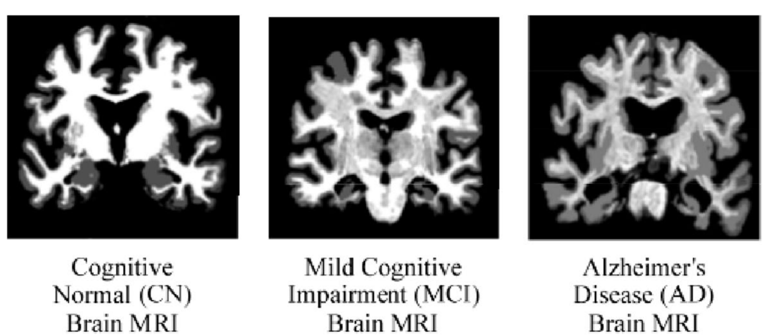
Researchers have been using different deep learning models with different values of hyperparameters to classify Alzheimer's disease. The class labels used for Alzheimer's disease are (a) Alzheimer's disease (AD), (b) mild cognitive impairment (MCI) and (c) normal control group (NC). The MRI of the three class labels [15] has been depicted in Fig. 7:

Besides these other class labels are also used depending upon the dataset applied. The other class labels have been depicted in Table 4:

Classification

Image classification is the task of assigning a label to an image from a predefined set of categories. The main task is to analyse an input image and return a label that categorizes

Fig. 7 Brain MRIs of CN, MCI and AD patient



the image. The label is always from a predefined set of possible categories [51].

We have divided the literature review in two sections: binary classification and multiclass classification.

Binary Classification

Binary Classification is the most common and simplest problem that is tackled in the field of image classification. Deep learning techniques provide a very high degree of accuracy in case of binary classification. The number of classes in case of binary classification is always two. The attributes of the image be it an object, or an animal or a medical MRI,is being examined and if it matches with the attributes defined, it is placed in one class or else it is placed in the other class. In case of AD classification also, binary classification is being done on different class labels like AD vs HC, MCI vs EMCI, EMCI vs LMCI etc. Binary classification task is less complex than multi-classification task. Here are a few research papers that have conducted binary classification tasks and achieved good accuracy values.

Binary classification [16] was conducted using the ADNI dataset. The data set consists of MRI, PET, and CSF data from 51 AD patients, 99 MCI patients (43 MCI patients who converted to AD (MCI.C), and 56 MCI patients who did not progress to AD in 18 months (MCI.NC)) as well as 52 healthy normal controls. The class labels used in this experiment were: AD, Healthy Control (HC), MCI, MCI.C and MCI.NC. ROI's (Region of Interest) were obtained from the MRI and PET images and these features were subjected to PCA (Principal Component Analysis) as well as Lasso to identify the most effective features. The deep learning architecture was pre-trained with RBM which helps the model to study the data first and then to extract the useful features from it. Dropouts were used for generalization as deep learning models suffer from overfitting when there is no sufficient data in the training set. They claimed that the model achieved 91.4%, 77.4%, 70.1% and 57.4% accuracies for AD vs HC, MCI vs HC, AD vs MCI, and MCI.c vs MCI.

Binary classification [17] of AD vs NC was performed using 28 AD patients and 15 elderly NC images (24 female and 19 male) were selected from ADNI dataset. The fMRI data downloaded from ADNI was used in the experiment. A number of pre-processing were done using FMRIB, BET followed by motion correction, skull striping, spatial smoothing and high pass temporal filtering was used for low level noise removal. They used LeNet5 model as the model was easy to implement. It has three convolution and 3 max pooling layers followed by the Flattening and the fully connected layer. Cross validation with 5 folds was used. Deep Learning LeNet model successfully recognized the Alzheimer's data from Normal Control and claimed an average accuracy of 96.85%. For binary classification the LeNet 5 model may perform well but performance may degrade or the model might find difficult to converge or generalize as only few features might get extracted due to it's less number of layers. LeNet 5 model's accuracy was found better than SVM which was 84%.

In [18] a model was proposed to differentiate AD images from HC. OASIS was used as the dataset and the selection of data was done by calculating the entropy of an image. They considered 416 images whose ages are between 18 and 96 from the dataset. In their experiments, they have randomly picked 200 images, 100 of whom were picked from the AD group, while the other 100 from the HC group. Cross sectional MRI images were used though longitudinal data is also present in OASIS. 5 cross validation was used to obtain the results. VGG16 was used for this experiment and was compared with VGG16 with transfer learning as well as Inception V4 with transfer learning. Batch size of 40 with 100 epochs was used for VGG16 and used RMSprop as an optimizer. In case of InceptionV4, batch size of 8 with 100 epochs and SGD with 0.0001 was used as the optimizer. Pre-trained weights of ImageNet was used for transfer learning. Using transfer learning VGG16 acquired 92.3% and Inception V4 acquired 96.25% accuracy respectively.

The studies performed in [19] was to observe CNN model's performance using different image segmentation methods with different datasets. ADNI was used and they considered 101 AD images, 234 MCI images and 169 were NL. Initially Support Vector Machine was used for detection of Alzheimer's disease but as the performance was not good, CNN model was used. Two experiments have been conducted using CNN. In the first experiment six evaluation processes were used depending upon the different image segmentation methods.

- Full image without detecting any edge
- Full image after detecting the edges
- Use of ROI with detecting edges
- Use of ROI without detecting edges
- Use of limited ROI with detecting edges
- Use of limited ROI without detecting edges

They found that the evaluation process was best considering extended ROI without detecting edges with classification accuracy 96%. The second experiment was used on two different dataset. The first dataset consisted of 36 images (AD—7, MCI—14, NL—15). After converting those 3D MRI films into 2D images, 1615 images generated. Out of those images, 1292 images have been randomly selected to train the CNN model and remaining 323 images to test the model. The second dataset also used 36 images (AD—9, MCI—16, NL—11). 17432D images were created using those 36 MRI films. Batch size, the number of classes and the number of epochs were defined as 32, 3 and

20 respectively. The number of convolutional filters, the size of the convolutional kernel and the size of the pooling area were declared as 32, 3 and 2 respectively. They claimed that CNN remains unbiased to the dataset. CNN model's performance can be further improved by increasing the amount of data used to train the model and also by fine tuning. This study mainly focused on the Coronal view of MRI.

A dataset was created in [20] using images that included all ADNI1, ADNI2 and ADNI-GO images that had baseline 3D T1-weighted scans. A total of 1409 images (294 images with probable AD, 763 images with MCI, and 352 healthy controls) have been considered. The dataset obtained was named as Milan. The six binary classifications namely AD vs HC, c-MCI vs HC, s-MCI vs HC, AD vs c-MCI, AD vs-MCI, c-MCI vs s-MCI were done on both the ADNI dataset and also the ADNI + Milan dataset. Image normalization and segmentation was done as a pre-processing step. The architecture of the network contains: 12 repeated blocks of convolutional layers (2 blocks with 50 kernels of size $5\times5\times5$ with alternating strides 1 and 2 and 10 blocks with 100 to 1600 kernels of size $3\times3\times3$ with alternating strides 1 and 2); a Rectified Linear Unit, a fully-connected layer; and one output layer. The network used in this study differs from the standard CNNs as max-pooling layers were replaced by standard convolutional layers with stride of 2. They claimed that accuracy,98% was obtained in AD vs HC classification which is the best out of all classifications. They have made an important point in this paper that to develop procedures with better accuracy which can be used in preclinical phase as screening tool to identify those people who might have a greater chance of developing dementia.

In [21] a CNN method was proposed for AD binary classification and the class labels used are demented and non-demented. OASIS dataset was used and considered 100 AD images (ranging from very mild demented to moderate level) and 20 Non-demented images. Image resizing and denoising was performed as pre-processing step. The proposed model is a 12-layerCNN that consisted of 4 Conv2D layer, 4 MaxPool2D layers, a flatten layer, 2 dense layers and one output layer. LeakyReLU activation was used. The proposed model was compared with four pre-trained models, namely InceptionV3, Xception, MobileNetV2, and VGG19 and it was found to be the best with 97.75%, whereas the accuracy of InceptionV3, Xception, and MobileNetV2 are 90.62%, 84.37%, and 81.24%, respectively.

T1 weighted images of ADNI dataset was used in [22] to conduct six binary classification: NC vs AD, NC vs EMCI, NC vs LMCI, EMCI vs LMCI, EMCI vs AD and LMCI vs AD using ADNI database. In this paper, 85 NC images, 70 EMCI, 70 LMCI, and 75 AD images have been considered. Data-preprocessing was done using SPM12 and the grey matter of the brain was obtained through segmentation as it is helpful for predicting AD at an early stage. MNI space was used for spatial normalization. They have implemented the CNN model with and without data augmentation. Transfer learning concept was used with the help of VGG19 model. The proposed model was divided into two blocks Group A and Group B with the following specifications:

Group A: 8 Convolutional layer with 3 Max Pooling layer were kept frozen.
Group B: 12 Convolutional layer with 4 Max Pooling layer were kept frozen.

They have claimed that AD vs NC classification showed the best accuracy of 93.83% and 95.33% in Group A and Group B respectively when the data without augmentation have been considered. On the other hand when data augmentation was considered, again AD vs NC showed the best result with an accuracy of 95.38% and 98.73% in Group A and Group B respectively. They claimed that data augmentation improves the classification accuracy.

To differentiate between AD images from HC have been carried out in [23]. The work investigated the effects of many parameters on the model which was named as AlzNet, such as the number of layers, number of filters, and dropout rate. OASIS dataset have been used in this study and 2D MRI slices were fed to the CNN model. A total of 240 images were used for this experiment. The CNN architecture used in this study is composed of 5 convolutional layers which take an input image with a size of 200*200. All five convolutional layers were followed by a max-pooling layer of a kernel size set on 2*2. The 64 filters with a kernel size of 9*9, 7*7 and 5*5 were considered for the first, second and third convolutional layer respectively. The 32 filters with a kernel size of 5*5 and 3*3 were considered for the fourth and fifth convolutional layer respectively. Training and testing accuracy was found to be 97.06 and 98.92 respectively.

Kaggle dataset was used in [24] to conduct Alzheimer's disease classification. Though the dataset consisted of four classes of image, only two classes of images: 717 MD and 2560 ND were considered for this experiment. Two models CNN from the scratch and pre-trained VGG16 were used for the task. Image transformation and resizing of the images was performed as pre-processing step. Three experiments were conducted with three configurations. The first and second experiments are for figuring out how to classify AD when splitting data with different ratios using the CNN model. In experiment 1, the data was splitted into 20% testing and 80% training, whereas in experiment 2, the data was splitted into 30% testing and 70% training. And both of these experiments used different ratios of dropout (0.2 and 0.5) to demonstrate the efficacy of the dropout in the proposed model. The third experiment observed at how well the VGG16 model worked with different optimizers and transfer learning using pre-trained VGG16. An accuracy of (100%) is achieved when dropout was not considered with an increasing number of epochs and decreasing batch size. The best average accuracy of training results, 99.95%, is achieved at 32 batch sizes and without applying dropout. In experiment 2 same results as in experiment 1 was achieved. Experiment 3 VGG16 model was compared with different optimizers like Adam, SGD, Adadelta, RMSprop, Adagrad. Adam optimizer was found to be the best with an accuracy of 97.44% over other optimizers. This result was achieved after 512 epochs and with 64 batch sizes.

Multi Classification

Multi classification as the name suggests consists of multiple classes and the objective is to place the right image in the right class. It is complicated compared to binary classification as binary classification has to answer only yes/no but in case of multi- class classification the model needs to learn the features of each class so well that all the images are being rightly classified into the classes it belongs. When in any classification tasks, there are more than two class labels, it is termed as Multi-class classification task. In case of AD classification, multi-class classification is done on class labels like: AD vs MCI vs CN, MCI vs LMCI vs EMCI, MID vs MOD vs VMD vs ND. Here are a few research papers which have conducted multi-class classification with promising results.

T1 weighted [25] sMRI data of 150 images: 50 AD, 50 CN and 50 MCI were considered to construct a mathematical model for AD classification. Motion correction and conform, Non-uniform intensity normalization, Talairach transform computation, intensity normalization,skull stripping and selecting most informative slice based on entropy were performed using the FreeSurfer software. Above processing results in 4800 images and were used for three way (AD vs CN vs MCI) and two way (AD vs CN, AD vs MCI and CN vs MCI) classification. VGG16 was considered as the base model for feature extraction and later the fully connected layers were removed. The output of the last convolutional layer is flattened into one column vector of size 18,432 and new fully connected layers are added at the end of the base model in which the first layer consists of 256 neurons, the second layer is the dropout layer with a value of 0.5 and at the end is the softmax layer whose output is 3 class scores and two class scores for 3-way and 2-way classification respectively. The three-way classification achieved 95.73% accuracy. The two-way classification achieved accuracies of 99.14%, 99.30% and 99.22% for AD vs CN, AD vs MCI and CN vs MCI classifications respectively.

Some researchers tried to classify AD, MCI and NL using ADNI multi-modality MRI data such as PET, functional MRI, MRI and genetic data in [26]. CNN model have been fed with 2D slices of MRI images. A total of 1726 images consisting of 340 AD, 812 MCI and 574 CN have been considered for the experiment. Image processing tasks such as intensity normalization, noise removal, bias correction, contrast adjustment and rescaling was done. They have constructed Deep CNN model with 6 convolutional layers and the Kernel size for all the convolutional layers was set to be the same, however, 4 to 128 size filters were used to extract complex and multi-scale information which makes a feature map to pass these features to the next layers for extracting further complex features. The input image with 300×300 size is followed by 5 fully connected layers with 1024, 512, 256, 128, 64 and 32 neurons respectively. MaxPooling with 2×2 stride and RMSProp optimizer with a learning rate of 0.0001was used for this experiment. The proposed model was named as DeepConvNet and was compared withVGG16 and AlexNet. DeepConvNet achieved the highest accuracy of 99.89%.

3-way classification AD, MCI and NC group using ADNI dataset was addressed [27]. Deep learning models such as Deep Neural Network (DNN), Convolutional Neural Networks (CNN), Deep Automatic Encoder (DA), Deep Boltzmann Machine (DBM) have been used for this classification task. The dataset considered from ADNI comprised of 1409 MRI, 302 Functional and MRI, 338 MRI data. They have claimed that highest accuracy 99.2% was found considering DNN and MRI images.

A hybrid structure model was proposed in [28] for classifying AD. Instead of training a network from scratch, they aimed to benefit from the current knowledge of Resnet50 architecture. They modified the ResNet50 model by removing the last five layers and ten new layers were added in place of these removed layers. Thus the number of layers increased from 177 to 182. They used the Kaggle dataset with 80% data for training and 20% data for testing. The hybrid model was compared with other CNN architectures like DenseNet201, VGG16, AlexNet and ResNet50. The hybrid model achieved 90% accuracy and was found to be highest as other models accuracy was in the range 78%-87%.

An AlexNet structure for AD classification based on MRI images processing using have been used in [29]. The model consisted of five hidden layers and three fully connected layers. They have used varying learning rates with Adam as optimizer The Alzheimer dataset that is available in the Kaggle website is used for this experiment. A total of 1409 samples (294 AD, 763 MCI, and 352 healthy controls) were considered in this study. They claimed that the proposed model using Adam optimizer with learning rate 0.0001 performs best with an accuracy of 95%. Determining the optimal learning rate is a challenge in any deep learning model.

ADNI dataset was used for classification of AD in [30] and out of 6000 images they considered 240 images for this experiment comprising of 80 AD, 80 CN and 80 MCI images. ResNet101 architecture was used for classification. ResNet includes full 3×3 convolution layer scheme of

VGG. The residual block contains 3×3 convolution layers of two in number with the equal amount of output lines. Every convolution layer subsequently comprises a layer of batch-normalization and ReLU as activating function. Subsequently, these convolution performance are skipped and sum the input straightaway prior the ultimate ReLU activating function. When likely to alter the size of output or the stride, an extra 1×1 convolution layer must be introduced to change the input into required size to perform addition operation. Skull striping is carried out as pre-processing step. The white matter, grey matter and the pial regions are extracted from brain using SegNet. The proposed ResNet101 acquired an accuracy of 96.3% accuracy and was compared with Multimodal and multiscale DNN and Hierarchical fully CNN whose accuracy was found to be only 75.4% and 90.3%accuracy respectively.

The dataset for this study [31] was obtained from ADNI dataset and 58 AD, 48 MCI and 73 CN images was considered. Fully automated surface-based cortical segmentation and sub-cortical volume based segmentation was done for feature extraction using FreeSurfer software. After feature extraction image normalization was performed. For classification, the proposed DNN consists of 4 layers. The first layer (input layer) has 62 nodes and the input layer is followed by 2 hidden layers. For the first hidden layer was created with 30 neurons and similarly, the second hidden layer with 15 neurons. The DNN model is built using the Keras library. Model is tested with a combination of Leaky ReLU, PreLU, and ELU in hidden layers. The accuracy is obtained for 100, 200, 300, and 400 epochs. The listed accuracy scores are obtained from a fivefold cross validation score. Using Leaky ReLU in the first hidden layer, PReLU in the second hidden layer, and 400 epochs produced the best result in the experiment. The accuracy score of this model is 79.81%.

Four-way classification such as MD, VMD, ND and MOD using ADNI dataset have been considered in [32]. Pre-trained CNN using Imagenet dataset were trained on the Xception model to extract features from MRI images. The Xception architecture consists of 36 convolutional layers forming the feature extraction base of the network and each of them is organized into 14 modules, all containing residual linear connections about them, except for the first and last units, which make up the feature extraction rule for the network. A total of 5000 images have been considered for this experiment. The model attained an accuracy of 97.74%.

AD classification on the Kaggle dataset [33] was conducted using four classes of images labeled as MID, MOD, VMD and ND. The research work consisted of comparing two models VGG19 and DenseNet169 performance on the dataset. The training set consisted of 3048 MRI images and were tested on a total of 2067 images. DenseNet169 model provided an accuracy of about 87% in the train data and about 80% in the test data using 40 epochs. In VGG19, the training accuracy was found to be 88% and 82% was found in the test set using 50 epochs. Both the models used batch size of 128 and AUC as evaluation metric. AUC for DenseNet169 is found to be about 88% in train data and about 82% in the test data. For VGG19 AUC in training data is 94% and test data is 86.7%.

In [34] they have used Kaggle dataset images which are pre-processed using ImageDataGenerator from keras and they have used the models VGG16 and VGG19. They have claimed testing accuracies of 71.02%, 77.04%, 77.66% for CNN, VGG16 and VGG19 models respectively.

A CNN model was proposed in [35] to extract the discriminative features by effectively improving accuracy in AD classification. The experiment was done consisting of the following steps: data pre-processing, balancing dataset using SMOTE and classification using DEMNET. Kaggle dataset was used for this study. A total of 6400 samples were used for this purpose. Initially the image was 176×208 but it was resized to 176×176.SMOTE technique was used as the dataset was class imbalanced. The proposed architecture consists of two convolutional layers with Rectified Linear Unit (ReLU) activation function, 1 Max-pooling layer, four DEMNET blocks, two dropout layers, three dense layers and a SoftMax classification layer. The model without using SMOTE technique has a training accuracy of 96% and validation accuracy of 78% due to the class imbalance and overfitting problem. The model with SMOTE technique achieves an overall training accuracy of around 99% and validation accuracy of 94%.

AD classification into three labels AD, CN AND MCI using 20 different CNN architectures like LeNet, AlexNet, VGG-16, VGG-19, Inception-V1 (Googlenet), Inception-V2, Inception-v3, ResNet-50, ResNet-101, ResNet50-V2 ResNet152-V2, InceptionResNet, MobileNet, MobileNet-V2, EfficientNet-B0, EfficientNet-B7, Xception, NasNet-A, NasNet-C and DenseNet-121 was performed [36]. MR images of more than 200 images from ADNI was selected for this experiment. Data-Generator function was used to increase the number of training images with many possible parameters such as rotation, mirror reflection, etc. The number of images increased to more than 15,000. The five most commonly used skull segmentation method was performed as the pre-processing step. Out of the five skull segmentation techniques namely Region Growing, Region Splitting—Merging, K—Means Clustering, Histogram Based Thresholding, and Fuzzy C Means, Histogram Based Thresholding produced the most convincing result. Classification was performed with and without skull striping. DenseNet-121 outperformed other models by achieving an accuracy of 85% without skull striping and 88% with skull striping.

Kaggle dataset was used for AD classification and it was divided into two different datasets [37]. In the first dataset a total number of 6400 images; was taken and divided into two folders, namely, train and test containing 5112 and 1279 images, respectively. The class labels used in the first dataset are: MID, MOD, VMD and ND. In the second dataset a total of 6330 images were taken and the class labels used are: ND, MID and VMD. VGG16 was used for feature extraction and after that it was fed into a neural network. The input image of size 224*224 is passed through the initial stack of two convolutions with a receptive area of 3 × 3. Each of these layers contains 64 filters. The padding is always kept as 1 pixel, whereas the stride value is fixed at 1.A flattening layer is placed between the three fully connected layers, which are used after a stack of convolutions. The results have shown that the proposed model for dataset 1 has achieved an accuracy of 90.4% whereas precision, recall, F1-score, and AUC are 0.905, 0.904, 0.904, and 0.969, respectively. The proposed model has achieved accuracy, precision, recall, F1-score, and AUC as 71.1%, 0.710, 0.711, 0.710, and 0.850, respectively, for dataset 2. After performance analysis it was concluded that performance of dataset 1 was better than dataset 2 due to the following factors: number of images in dataset1 is higher than the number of images in dataset2, and also both training and test dataset are balanced in case of dataset1.

VGG16 model was proposed [38] to classify AD into four class labels MD, MOD, ND and VMD using the Kaggle dataset. A total of 10,432 images were used for this experiment. VGG16 was used with Adam as optimizer and softmax as activation function. Five convolution layers and two dense layers were used. The images were resized to 128X128 pixels. This model achieved 100% training Accuracy, 0.0012 training loss, 97% Validating Accuracy, and 0.0832 Validating loss.

AD classification based on five categories AD, CN, EMCI, MCI and LMCI using the ADNI dataset was performed [39]. Linear contrast stretching was performed as pre-processing step. They have implemented the dataset with transfer learning models AlexNet, GoogleNet and ResNet-18. The overall accuracy of the GoogleNet, AlexNet and ResNet-18 in detecting AD are obtained as 96.39%, 94.08% and 97.51% respectively.

OASIS3 dataset consisting of four class labels ND, MoD, MD and VMD was considered [40]. The number of images for each class is 3200, 64, 896, and 2240 for ND, MoD, MD, and VMD, respectively. Image normalization and data augmentation was used as pre-processing step. The customized CNN model consisted of 2D convolutional layers with a kernel size 5 × 5 for the first layer, and 3 × 3 for the second and third layers. For the last two layers, kernel size of 2 × 2 was used. For all the layers, the ReLU activation function is used and no Batch Normalization (BN) was used. Each block's second convolutional layer used a stride of two to conduct downsampling. The initial block had 64 filters, and each succeeding block had double the number of filters. Following the last convolutional layer, a dropout layer (p = 0.3) was added before being linked to one FC dense layer with ReLU activation values of 1024. Between those thick layers, a drop-out layer (p = 0.3) was also placed. Finally, a multi-dense neuron with softmax activation supplied the model output. Batch size of 40 and learning rate of 0.001 with Adam optimizer was applied to the CNN model. The proposed CNN model was compared with other models like VGG16, DenseNet121, ResNet50, Inception V3 and EfficientNetB7 and was found to be best with 99.68% accuracy for testing and 100% for training with validation accuracy 99.27%.

In [41] MobileNet was used as a pretrained model for AD classification. MobileNet is a lightweight network used for mobile applications. This is the first time MobileNet has been used for medical data analysis and achieved an accuracy of 96.6% which is better than the results obtained using VGG16 and ResNet50. ADNI dataset was used and AD was classified into five categories CN, MCI, EMCI, LMCI and AD. Upsampling of the images were done to address the problem of unbalanced dataset which leads to overfitting. The data are refined, standardized, scaled, denoised, formatted and normalized appropriately for pre-processing. RMSprop optimizer with learning rate of 0.00001was used for this experiment.

A multiclass categorization into AD, LMCI, MCI, and NC on the ADNI dataset was performed in [42]. For pre-processing, they have used SPM12 for obtaining 2D Grey Matter slices and then they have used the histogram threshold technique for skull striping. Segmentation, normalization, rescaling and smoothing was also done later. The classification was done using with and without transfer learning. The proposed CNN model consisted of a 7 × 7 convolution layers, 3 × 3 max-pooling layers, four dense blocks with a different number of convolution layers (6, 12, 32, 32) in respective blocks, three transitional layers and a fully connected classification layer. The convolution layer with a kernel matrix of size 7 × 7 and a stride value set as 2, was used for the extraction of macro-features. These extracted macro-level features were then forwarded to dense blocks, where more features were extracted. Each dense block contained 1 × 1 convolutions, which were used to reduce the number of input feature maps, followed by 3 × 3 convolution layers. In each dense block, the extracted features were passed to all the next layers. Between the dense blocks, there were transition layers, along with the activation functions, like batch normalization and ReLu, etc. The transition layer contained 1 × 1 convolution and a 2 × 2 max pooling layer with a stride value set as 2. Once the transition layer reduced the dimensions of the features, they were forwarded to the fully connected classification layer where Softmax was used for the classification. Initially, all the initial blocks of layers were kept frozen and the last two blocks were retrained.

DenseNet-169 was used as the base model for transfer learning and attained an accuracy of 97.84%.

Both binary and multiclass classification was conducted in [43] using the Kaggle dataset. In this work a total of 6400 images were used with 5121 training and 1279 test images. Data augmentation techniques were applied as a pre-processing step. For binary classification AD vs Normal, the CNN model consisted of 2 Conv2d layers and 2 MaxPooling2d layers followed by a flattening layer and finally the two dense layers including one as the output layer. 99.22% of accuracy was found in the binary classification. For multiclass classification of MD, MOD, ND and VMD, the proposed method acquired an accuracy of 95.93% (Table 5).

The use of different deep learning models for the detection and classification of Alzheimer's disease mentioned in the literature survey is depicted in Fig. 8:

Challenges

- *Dataset*: Alzheimer Disease research depends mainly on ADNI, OASIS and Kaggle dataset. It is really a challenging task for research on real life dataset. Development of large quantity Alzheimer Disease image balanced benchmark dataset is the primary challenge as state-of-art classifier demands a large amount of data for feature extraction and classification work. Since there is a deficiency of large-scale Alzheimer Disease dataset, much research is still waiting for Alzheimer Disease classification. As small size dataset is a barrier for improvement of accuracy of deep learning-based models, so it is a challenge for Alzheimer Disease classification. In most cases uneven distribution of dataset causes biased learning of the model which is a challenge to researchers. Researchers use batch normalization to solve this issue.
- *Pre-processing*: Literature review shows that researcher's uses pre-processing toolkits and some others have used different predefined algorithms on the Alzheimer disease dataset. Some researchers have performed different pre-processing like image slicing, entropy of an image, segmentation, skull striping etc. to improve model accuracy.
    - *Image Slicing*: One of the challenges faced during brain slicing is that, it may affect the morphology and physiological properties of the tissues thus compromising the brain tissues. The longevity of brain slices is a limiting factor and this makes difficult for the researchers to conduct their studies on the brain's neuronal properties [44]. And lastly the variables that are important in making healthy brain slices has not yet been studied [45].
    - *Entropy of an image*: Entropy is a measure of information and its concept has been employed in many image processing methods but has not been used to assess the effects of information change in fused images. The reason is that entropy sees information as a frequency of change in the digital numbers in images. It cannot distinguish between information belonging to the scene and noise. In such cases nothing much important information is gained from image entropy. [46] In case of grayscale images, the higher the variance of gray intensities, the higher the contrast. Contrast is defined as the difference between the light and dark areas of the image. When the difference between the maximum and minimum intensities of an image is very small, the image has low entropy and poor contrast. [47]
    - *Segmentation*: In brain segmentation, intensity inhomogeneity is a challenge which occurs due to the intensity of non-homogeneity of homogeneous tissues during contrast injection and the variations of spatial intensity over each dimension. The bias field is another challenge faced during the process of brain segmentation in MR images, which is caused by the defects in the acquisition sequences or radiofrequency coil imperfections. The various biases associated with MR images include shading, noise, artifacts, and partial volume effects. Non-standardized intensity is also an issue because the intensity of MR modalities depends on the magnetic fields and radio wave parameters, which are, in turn, influenced by the MR system hardware requirements [48].
    - *Skull Striping*: A number of works on skull striping have been reported in literature. [49, 50] MRI brain image contains some non-brain tissues like skull, skin, fat, neck etc. These non-brain tissues are considered as a cause of difficulty in further analysis. So it is a challenging task to remove these non-brain tissues before analysis. It is also a challenging task to find a threshold for brain tissues, as the brain and non-brain tissues are homogenous.
    - On the other hand, some researchers have also toolkits like MANGO, FLIRT, FSL and BET, BrainSuite, SPM12, DARTEL while some researchers have not opted for any pre-processing methods. So the main aspect here is that when we apply Deep Learning, do we need to do any kind pre-processing, does it affect the performance of the model.

- *Model*: Selection of model is one of the challenging tasks for classification. Deep Learning has a wide variety of models. Every model has some advantages and disadvantages compared to the earlier ones in terms of efficiency, accuracy, computation time etc.
- *Parameters*: A number of parameters like number of neurons, filters, filter size, pooling size, activation func-

Table 5 Summary of AD classification using deep learning

Author	Model	Dataset	Class labels	Number of samples	Evaluation measure
F. Li et al. [16]	Deep Model initialized with RBM	ADNI	AD, HC,MCI,MCI.C,MCI.NC	258	91.4% for AD vs HC 77.4% for MCI vs HC 70.1% for AD vs MCI 57.4% for MCI.C vs MCI
S. Saraff and G. Tofighi [17]	LeNet5 Model	ADNI	AD,NC	43	84%
M. Hon and N. M. Khan [18]	Transfer learning with VGG16 Inception V4	OASIS	AD, HC	200	92.3% for VGG16 96.25% for Inception V4
K.A.N.N.P Gunawardena et al. [19]	CNN	ADNI	AD, MCI, NL	3358	96%
S. Basaia et al. [20]	CNN	ADNI	AD, HC, cMCI, sMCI	1409	98% for AD vs HC
Hussain et al. [21]	CNN	OASIS	AD, ND	128	97.75%
A. Mehmood et al. [22]	VGG19	ADNI	NC, AD, EMCI, LMCI	300	95.33% without data augmentation 98.73% with data augmentation
F.E. Alkhuzaie et al. [23]	CNN	OASIS	AD, HC	240	98.92%
D.A. Arafa et al. [24]	CNN VGG16	Kaggle	MD, ND	3277	99% when Dropout not considered 97.4% for Adam optimizer when compared with SGD, Adadelta, RMSProp and Adagrad
R. Jain et al. [25]	VGG16	ADNI	AD,CN,MCI	150	95.73% for AD vs CN vs MCI 99.14% for AD vs CN 99.30% for AD vs MCI
A. Nawaz et al. [26]	Deep CNN	ADNI	AD, MCI, NL	1726	98.89%
E. Altinkaya et al. [27]	Deep Neural Network (DNN), Convolutional Neural Networks (CNN), Deep Automatic Encoder (DA), Deep Boltzmann Machine (DBM)	ADNI	AD, MCI	2049	99.2% for DNN
Muhammed Yildirim and Ahmet Cinar [28]	ResNet50	Kaggle	MD, MOD, VMD, ND	6400	90%
Y.N. Fuadah et al. [29]	AlexNet	Kaggle	MD, MOD, VMD, ND	1409	95%
P.R. Buvaneswari et al. [30]	ResNet101	ADNI	AD, CN, MCI	240	96.3%
R. Prajapati et al. [31]	DNN	ADNI	AD, MCI, CN	179	79.81%
H.R. Almadhoun and S.S. Abu Naser [32]	Xception	Kaggle	MD, VMD, ND, MOD	5000	97.74%
A. Pradhan et al. [33]	DenseNet 169 VGG16	Kaggle	MD,MOD,VMD,ND	5115	82% for DenseNet169 86.7% for VGG16
S.A.Ajagbe et al. [34]	CNN VGG16 VGG19	Kaggle	MD, MOD, VMD, ND	6400	71.02% for CNN 77.04% for VGG16 77.66% for VGG19
S. Murugan et al. [35]	CNN	Kaggle	MD, MOD, VMD, ND	6400	94%

Table 5 (continued)

Author	Model	Dataset	Class labels	Number of samples	Evaluation measure
R.A. Hazarika et al. [36]	LeNet, AlexNet, VGG-16, VGG-19, Inception-V1 (Googlenet), Inception-V2, Inception-v3, ResNet-50, ResNet-101, ResNet50-V2 ResNet152-V2, InceptionResNet,MobileNet, MobileNet-V2, EfficientNet-B0, EfficientNet-B7, Xception, NasNet-A, NasNet-C DenseNet-121	ADNI	AD, CN, MCI,	15,120	DenseNet-121 outperformed other models by achieving an accuracy of 85% without skull striping and 88% with skull striping
S. Sharma et al. [37]	VGG16	Kaggle	MD,MOD, VMD, ND	6391 for Dataset1 6330 for Dataset 2	90.4% for Dataset1 71.1% for Dataset2
L.F. Samhan et al. [38]	VGG16	Kaggle	MD, MOD,VMD,ND	10,432	97%
J.V. Shanmugam et al. [39]	AlexNet, GoogleNet ResNet-18	ADNI	AD, CN, MCI, EMCI, LMCI	7800	96.39% in AlexNet 94.08% in GoogleNet 97.51% in ResNet-18
Marwa El-Geneedy et al. [40]	CNN	OASIS	ND,MOD,MD,VMD	6400	99.68%
A. Bhagat et al. [41]	MobileNet	ADNI	CN, MCI, EMCI, LMCI, AD	300	96.6%
Noman Raza et al. [42]	CNN with Dense-Net169 as base model	ADNI	AD, LMCI, MCI,NC	5016	97.84%
AAA. El-Latif [43]	DNN	Kaggle	AD, MOD, ND, VMD	6400	AD vs ND: 99.22% MD vs MOD vs ND vs VMD: 95.93%

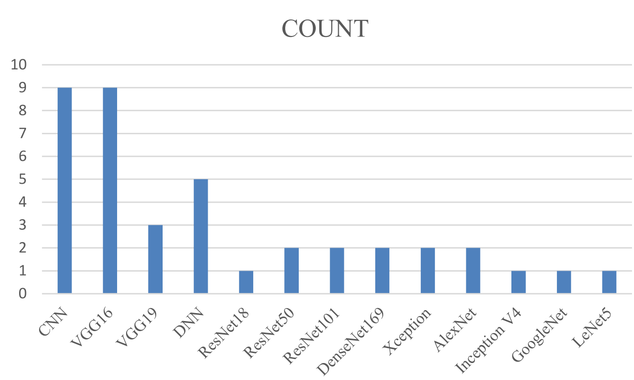


Fig. 8 Use of DNN models in the literature survey done

tions, loss functions, optimization method, dropout, class imbalance, classifier etc. affects the performance of the model. It is observed that epochs and batch size play a very important role in the performance of the model. So determining these parameters for a particular model is a challenging area of research.

- *Evaluation Metric*: The evaluation metric is a very important issue. As most of the experiments have used accuracy, use of metrics like F1 score, precision, recall, AUC etc. can be used further in the research. Sensitivity and specificity are also used in some of the experiments.
- *Collaboration*: In medical research domain expert is very much essential. In literature it is observed that there is no domain expert in the research team. Collaboration of a team consisting of a medical expert, radiologist and the researcher might help in carrying out the experiments taking in all the factors affecting the disease to get good results. This can happen if a proper mutual agreement is being done between the institution of the researcher and the medical professional.

Deep Learning Challenges

- *Requirement of Huge Amount of Data*: Deep Learning requires huge corpus of data to train on then only it can give good performance. A human brain learns new things and deduce information from a very large experience to take a decision. As Deep Learning mimics a human brain, same applies to it also. In order to arrive to a conclusion, Deep learning algorithms requires millions and millions of data. In some cases the data related to a certain domain may be less, then it becomes a challenge for the researcher to get good results. If limited amount of data is available then the Deep learning algorithm cannot generalize and results in poor performance. Due to scarcity of data Deep learning may be subjected to spoofing in certain cases.

- *Hyperparameter Optimization*: Optimization of hyperparameters plays a very important role in the performance of a Deep Learning algorithm. These are the parameters that needs to be decided before the algorithm starts training/learning. A small change can have a huge impact on the final results. So to tune the parameters according to some good proven method might help the model in achieving better results.
- *Overfitting*: Overfitting occurs when the model works very well in the training data but shows poor performance when unseen data/test data is being fed to it. It so happens because after several iterations, the model learns so well that it gains information of the noise as well as also the inaccurate data entries in the dataset. This results in high variance and low bias. As a result the performance of the model becomes very poor in test data. Overfitting usually occurs in supervised learning. There are many techniques to handle overfitting like regularization, cross validation, early stopping etc.
- *Vanishing Gradient Problem*: Vanishing Gradient problem occurs when the neural network halts its learning process as updating weights and biases using back propagation cause the gradients that are used to update the weights to shrink exponentially. This happens because chain rule is used in back propagation. As a result the weights are not updated anymore and hence the performance of the model becomes poor. Use of activation functions like ReLU, Residual Neural Network(RNN), Long short-term memory are some of the techniques that are used to handle Vanishing Gradient problem.
- *Exploding Gradient Problem*: Exploding Gradient problem is just the opposite of the Vanishing Gradient problem. Here the gradients that are used to update the weights increases exponentially which results in larger updates of the weights thus making the model unstable and unable to learn anything. This problem is handled by gradient clipping, proper weight initialization etc.
- *Requirement of High performance Hardware Equipments*: A very good computation power is required by Deep Learning model. This computing power cannot be accomplished by the CPU alone. So the hardware equipments like the RAM, GPU as well as the memory capacity are the important factors to decide the performance of a Deep Learning model.

Conclusion

Literature review reveals use of different Deep Learning model in Alzheimer disease detection and classification. With the advent of many new models and also further study about the different parameters affecting the model performances, accuracy can be improved. The main challenge is to keep the accuracy high considering the size of the training set as well as also on the distribution of the image data both in training and test data. The data in testing set may be highly different from the training data set if the number of samples is less and the data from each category may contain variety of images. Deep Learning performs well when we have a huge amount of data. Also multiclass classification is much challenging than binary classification due to the variation of images in each category. In this review study we have tried analysing the factors affecting the performance and tried to gain an insight on the probable areas of future research regarding Alzheimer's disease detection and classification.

Data availability The data that support the findings of this study are openly available in

1. Alzheimer's Disease Neuroimaging Initiative (ADNI) at https://adni.loni.usc.edu/about/.

2. Open Access Series of Imaging Studies (OASIS) at https://sites.wustl.edu/oasisbrains/.

3. Alzheimer's Dataset (4 class of Images) at https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images.

Declarations

Conflict of Interest 1
On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Yetman D. Dementia and Alzheimer's: what are the differences? 2021. https://www.healthline.com/health/alzheimers-disease/difference-dementia-alzheimers.
2. Leonard W. Symptoms of Dementia. 2018. https://www.healthline.com/health/dementia-symptoms#takeaway.
3. Dementia. https://my.clevelandclinic.org/health/diseases/9170-dementia.
4. Overview—Alzheimer's disease. https://www.nhs.uk/conditions/alzheimers-disease/.
5. Gaugler J, James B, Johnson T, Reimer J, Solis M, Weuve J, Hohman TJ. 2022 Alzheimer's disease facts and figures. Alzheimers & Dementia. 2022;18(4):700–89.
6. Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population estimate of people with clinical AD and mild cognitive impairment in the United States (2020–2060). Alzheimers Dementia. 2021. https://doi.org/10.1002/alz.12362.
7. Shaji KS, Jotheeswaran AT, Girish N. Alzheimer's & related disorders Society of India. The dementia India report—prevalence, impact, costs and services for Dementia: Executive summary. 2010.
8. Alzheimer's and normal brains, comparison [Photograph]. https://www.sciencephoto.com/media/1137925/view/alzheimer-s-and-normal-brains-comparison.
9. ADNI Alzheimer's disease neuroimaging initiative. https://adni.loni.usc.edu/about/.

10. OASIS open access series of imaging studies. https://www.oasis-brains.org/.
11. Cagler U. What is Kaggle? 2022. https://www.datacamp.com/blog/what-is-kaggle.
12. [ThePIRL]. How MRI works: Part 1 NMR Basics [Video]. https://www.youtube.com/watch?v=TQegSF4ZiIQ&t=262s
13. Preston DC. Magnetic resonance imaging (mri) of the brain and spine: basics. MRI Basics, Case Med. 2006;30:1–6.
14. Hirsch GV, Bauer CM, Merabet LB. Using structural and functional brain imaging to uncover how the brain adapts to blindness. Ann Neurosci Psychol. 2015;2:7.
15. Mukhtar G, Farhan S. Convolutional neural network-based prediction of conversion from mild cognitive impairment to Alzheimer's disease: a technique using hippocampus extracted from MRI. Adv Electr Comput Eng. 2020;20(2):113–22.
16. Li F, Tran L, Thung KH, Ji S, Shen D, Li J. A robust deep model for improved classification of AD/MCI patients. IEEE J Biomed Health Inform. 2015;19(5):1610–6.
17. Sarraf S, Tofighi G. Classification of Alzheimer's disease using FMRI data and deep learning convolutional neural networks. arXiv preprint, 2016. https://arXiv.org/1603.08631.
18. Hon M, Khan NM. Towards Alzheimer's disease classification through transfer learning. In: 2017 IEEE International conference on bioinformatics and biomedicine (BIBM). IEEE. 2017, pp 1166–1169.
19. Gunawardena KANNP, Rajapakse RN, Kodikara ND. Applying convolutional neural networks for pre-detection of Alzheimer's disease from structural MRI data. In: 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP). IEEE, 2017, pp. 1–7.
20. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Alzheimer's Disease Neuroimaging Initiative. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage: Clin. 2019;21:101645.
21. Hussain E, Hasan M, Hassan SZ, Azmi TH, Rahman MA, Parvez MZ. Deep learning based binary classification for Alzheimer's disease detection using brain MRI images. In: 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE. 2020, pp. 1115–1120.
22. Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, Yaqub M. A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. Neuroscience. 2021;460:43–52.
23. Al-Khuzaie FE, Bayat O, Duru AD. Diagnosis of Alzheimer disease using 2D MRI slices by convolutional neural network. Appl Bionics Biomech. 2021;2021:1–9.
24. Arafa DA, Moustafa HED, Ali HA, Ali-Eldin AM, Saraya SF. A deep learning framework for early diagnosis of Alzheimer's disease on MRI images. Multimed Tools Appl. 2023;83:1–33.
25. Jain R, Jain N, Aggarwal A, Hemanth DJ. Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn Syst Res. 2019;57:147–59.
26. Nawaz A, Anwar SM, Liaqat R, Iqbal J, Bagci U, Majid M. Deep convolutional neural network based classification of Alzheimer's disease using MRI data. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, 2020, pp. 1–6.
27. Altinkaya E, Polat K, Barakli B. Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: a comprehensive review. J Inst Electron Comput. 2020;1(1):39–53.
28. Yildirim M, Cinar A. Classification of Alzheimer's disease MRI images with CNN based hybrid method. Ing Syst Inf. 2020;25(4):413–8.
29. Fuadah YN, Wijayanto I, Pratiwi NKC, Taliningsih FF, Rizal S, Pramudito MA. Automated classification of Alzheimer's disease based on MRI image processing using convolutional neural network (CNN) with AlexNet architecture. J Phys Conf Ser. 2021;1844(1):012020.
30. Buvaneswari PR, Gayathri R. Deep learning-based segmentation in classification of Alzheimer's disease. Arab J Sci Eng. 2021;46:5373–83.
31. Prajapati R, Khatri U, Kwon GR. An efficient deep neural network binary classifier for Alzheimer's disease classification. In: 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, 2021, pp. 231–234.
32. Almadhoun HR, Abu-Naser SS. Classification of Alzheimer's disease using traditional classifiers with pre-trained CNN. Int J Acad Health Med Res (IJAHMR). 2021;5(4):17–21.
33. Pradhan A, Gige J, Eliazer M. Detection of Alzheimer's disease (AD) in MRI images using deep learning. Int J Eng Res Technol (IJERT). 2021;10:580–5.
34. Ajagbe SA, Amuda KA, Oladipupo MA, Afe OF, Okesola KI. Multi-classification of Alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches. Int J Adv Comput Res. 2021;11:53.
35. Murugan S, Venkatesan C, Sumithra MG, Gao XZ, Elakkiya B, Akila M, Manoharan S. DEMNET: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. IEEE Access. 2021;9:9.
36. Hazarika RA, Kandar D, Maji AK. An experimental analysis of different deep learning based models for Alzheimer's disease classification using brain magnetic resonance images. J King Saud Univ-Comput Inform Sci. 2022;34(10):8576–98.
37. Sharma S, Guleria K, Tiwari S, Kumar S. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Meas Sens. 2022;24:100506.
38. Samhan LF, Alfarra AH, Abu-Naser SS. Classification of Alzheimer's disease using convolutional neural networks. Int J Acad Inform Syst Res (IJAISR). 2022;6(3):18–23.
39. Shanmugam JV, Duraisamy B, Simon BC, Bhaskaran P. Alzheimer's disease classification using pre-trained deep networks. Biomed Signal Process Control. 2022;71: 103217.
40. Marwa EG, Moustafa HED, Khalifa F, Khater H, AbdElhalim E. An MRI-based deep learning approach for accurate detection of Alzheimer's disease. Alex Eng J. 2023;63:211–21.
41. Bhagat A, Ansarullah SI, Othman MTB, Hamid Y, Alkahtani HK, Ullah I, Hamam H. A novel framework for classification of different Alzheimer's disease stages using CNN model. Electronics. 2023;12(2):469.
42. Raza N, Naseer A, Tamoor M, Zafar K. Alzheimer disease classification through transfer learning approach. Diagnostics. 2023;13(4):801.
43. El-Latif AAA, Chelloug SA, Alabdulhafith M, Hammad M. Accurate detection of Alzheimer's disease using lightweight deep learning model on MRI data. Diagnostics. 2023;13(7):1216.
44. McKenna MC, Dienel GA, Sonnewald U, Waagepetersen HS, Schousboe A. Energy metabolism of the brain. In: Basic neurochemistry. Academic Press; 2012. p. 200–31.
45. Tashiro A, Aaron G, Aronov D, Cossart R, Dumitriu D, Fenstermaker V, Yuste R. Imaging brain slices. In: Handbook of biological confocal microscopy. Springer; 2006. p. 722–35.
46. Leung LW, King B, Vohora V. Comparison of image data fusion techniques using entropy and INI. In: 22nd Asian Conference on Remote Sensing 2001; vol 5, no. 9, pp. 152–157.
47. Mello Román JC, Vázquez Noguera JL, Legal-Ayala H, Pinto-Roa DP, Gomez-Guerrero S, García Torres M. Entropy and contrast enhancement of infrared thermal images using the multiscale top-hat transform. Entropy. 2019;21(3):244.

48. Fawzi A, Achuthan A, Belaton B. Brain image segmentation in recent years: a narrative review. Brain Sci. 2021;11(8):1055.
49. Das D, Kalita SK. Skull stripping of brain MRI for analysis of Alzheimer's disease. Int J Biomed Eng Technol. 2021;36(4):331–49.
50. Kalavathi P, Prasath VB. Methods on skull stripping of MRI head scan images—a review. J Digit Imaging. 2016;29(3):365–79.
51. Rosebrock A. Image classification basics. 2021 https://Pyimagesearch.com/2021/04/17/Image-Classification-Basics/.
52. Watts R. Transverse view of the brain. 2007. https://static.sciencelearn.org.nz/images/images/000/001/104/embed/75yo_male.jpg?1674165508.