Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16026-0

Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease

Srividhya L[1] · Sowmya V[1] · Vinayakumar Ravi[2] · Gopalakrishnan E.A[1] · Soman K.P[1]

Received: 25 August 2022 / Revised: 27 April 2023 / Accepted: 11 June 2023

Abstract
Alzheimer's Disease (AD) is a common neurological brain disorder that causes the brain cells to die and shrink (Atrophy) gradually, resulting in a continuous decline in one's ability to function independently. Early diagnosis increases the possibility of preventing or delaying the advancement of this mental disorder. Magnetic Resonance Imaging (MRI) offers the potential of non-invasive longitudinal monitoring and plays a vital role as a biomarker of the disease progression. Structural Magnetic Resonance Imaging (sMRI) helps to measure Atrophy, which is considered to be the most dependable biomarker to assess the exact stage and severity of the neuro-degenerative aspect of AD pathology. There are five stages associated with AD, which include Normal Control (NC), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). In this work, we have used the Alzheimer's Disease Neuroimaging Initiative (ADNI2) sMRI image dataset to measure and classify the stage of AD. In recent years, Convolutional Neural Networks (CNNs) are widely used for medical image analysis. This work focuses on applying different Deep Learning algorithms for the multi-class classification of AD MRI images and proposes the best pre-trained model that can accurately predict the patient's stage. It is observed that ResNet-50v2 gives the best accuracy of 91.84% and f1-score of 0.97 for AD class. Visualization techniques such as Grad-CAM and Saliency Map are applied on the model that gave the best accuracy to understand the region of focus in the image which led to predicting its class.

Keywords Alzheimer's Disease (AD) · Alzheimer's Disease Neuroimaging Initiative (ADNI) · Convolutional Neural Networks (CNN) · Deep learning · Mild Cognitive Impairment (MCI) · Structural Magnetic Resonance Imaging (sMRI) · ResNet-50v2 · Grad-CAM · Saliency map

✉ Vinayakumar Ravi
vravi@pmu.edu.sa

Extended author information available on the last page of the article

Published online: 18 July 2023

1 Introduction

Dementia is a term coined to explain the set of brain disorders pertaining to cognitive decline. It impacts the memory, thinking, and general functional abilities severely enough to intervene with one's daily life. Dementia typically affects older adults, but its occurrence cannot be accounted for due to regular aging. According to the World Health Organization's (WHO) reports, there are over 55 million people affected by Dementia across the globe, and nearly 10 million fresh cases are identified each year. Today, Alzheimer's Disease (AD) is the most prevalent type of progressive Dementia which accounts for 60%-80% of total Dementia instances in the world. In the early stages, loss of memory may be minimal, but in the advanced stages, individuals lose the ability to carry on a conversation and respond to the circumstances. AD first starts with the destruction of neurons and its connections in that region of the brain that impacts learning (including the entorhinal cortex and hippocampus) and then advance through the other parts leading to more severe symptoms that include bafflement, mood swings, behavioral changes, deepening confusion about time, place and events, critical memory loss, difficulty in communication, etc.,

AD is a multifarious disease in which accruing pathological, physical or mental injury to the brain results in progressive cognitive decline. Some of the well-known pathological hallmarks of AD are Amyloid plaques, neuro-fibrillary tangles (NFTs), neurodegeneration, and inflammation. There are many Neuroimaging techniques such as PET, MRI and EEG used to diagnose AD. Especially, Magnetic Resonance Imaging (MRI), has been playing a significant role for the last couple of decades to get a clear picture of the functions of the brain and diagnose disorders if any [32]. The degeneration of the neurons associated with cognitive decline can be ascertained by Structural Magnetic Resonance Imaging (sMRI). sMRI measures brain morphometry and are considered to be the most influential AD biomarker. So, in the current study, sMRI biomarkers are considered in detecting the progression of AD. The Structural images from MRI can detect atrophic changes that influence the hippocampus and entorhinal cortex at the nascent phase of mild cognitive impairment, which may spread to temporal and parietal lobes and affect the frontal lobes at the final phases of AD [9, 12, 22]. The early occurrence of amyloid deposition leads to the development of AD [48]. The brain degeneration is an unpreventable progressive component of AD [44]. However, the radiologists and analysts strongly suggest that it is essential to develop processes that automate the way of extracting disease-specific information from the medical images and integrate them with other existing biomarkers for clinical usage. Figure 1 depicts the comparison of a normal brain and a brain impacted by AD.

The progression of AD can be classified into three general stages: preclinical, mild cognitive impairment, and Dementia, but it can also be more comprehensively described into a 5-Stage model - no impairment or Normal Control (NC), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI) and Alzheimer's Disease (AD) [37]. Various neurocognitive and neuropsychological tests are used to differentiate EMCI, LMCI and AD from NC stages. The multiple diagnostic tests make it cumbersome for clinicians to arrive at an objective clinical conclusion. It is possible to minimize the number of tests needed to make a fairly accurate evaluation about the severity of the disease using Machine Learning and Deep Learning algorithms [30]. An early diagnosis of AD is made at a stage when the individual suffers from MCI (Stage 2) but can function independently. This second phase of AD can be detected well before the onset of Dementia symptoms. Since AD is irreversible with no validated treatment, it is important to develop methods for early detection. Early and accurate diagnosis of Alzheimer's disease

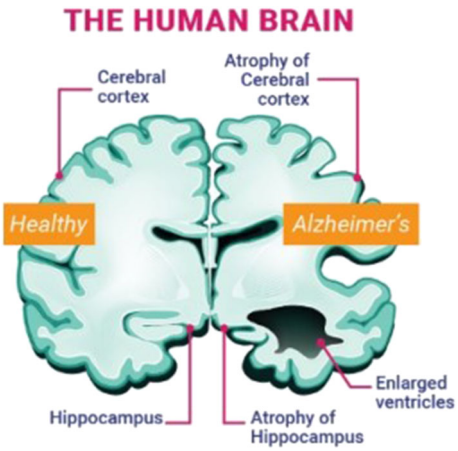

Fig. 1 Progress of Alzheimer's Disease from Mild Cognitive Impairment to severe stage [2]

is necessary to provide proper treatment to the patients [31]. Since no drugs are available for the inexorable progression of the disorder that is diagnosed during the later stage of the disease, the developed methods will help to diagnose AD at its pre-symptomatic stages [42].

The success of the Deep Learning approach over traditional Machine Learning in identifying complex structures and pattern recognition applications has brought immense enthusiasm and high expectations that Artificial Intelligence (AI) can play a remarkable role in health care [40]. AD classification using Deep Learning algorithms has recently gained enormous attention as rapid progress in neuroimaging techniques helps to detect the disease progression at an early stage [7, 20]. This work is carried out to comprehensively study the multi-class classification by applying all the 26 Keras pre-trained models and record the metrics of the best-performed model [14]. The sMRI of MCI and LMCI patients shows very minor visual differences compared to AD which motivated us to design a vigorous mechanism that can predict even the intermediate stages of MCI patients (EMCI and LMCI) precisely for the early diagnosis of AD. It is essential to train the model with more images belonging to various classes for getting accurate predictions. For research experiments, the 5 stage sMRI data from ADNI repository is less used when compared to the two-class (binary) data or the 3 class Open Access Series of Imaging Studies (OASIS) MRI dataset. Hence, this work would be a benchmark in applying Deep Learning techniques for this dataset. The results also support and demonstrate the scope for the use of Artificial Intelligence (AI) as a decision support system [47].

The contents of the paper are as follows: In section 2, the existing research works pertaining to the diagnosis of AD using Deep Learning methods are discussed. The methodology followed for this work is described in detail in section 3. In section 4, the experiments and results of the best performing 2D CNN model are elaborated. In section 5 we have discussed the advantages and limitations of the proposed methodology. This paper concludes with the scope of future research work in this area. The summary of the work is given below:

- sMRI is one of the most common medical imaging techniques that help to detect MCI or brain shrinkage. But clinical assistance is required for the appropriate diagnosis to ascertain the stage of AD. This work aims to build a robust model using sMRI 2D images to detect the early onset and automated classification of AD.
- Majority of the approaches in the recent works perform binary classification (i.e., NC vs AD). The model that we have developed can be used to diagnose and differentiate

the various stages of AD (5 stages- NC, EMCI, MCI, LMCI & AD). This enables early diagnosis of AD which is the novelty of our work. We performed ample number of experiments to demonstrate that our proposed model outperformed comparative references in terms of accuracy on the ADNI 5 stage dataset.

- Most of the recent works focus on applying any one of the Keras pre-trained model and comparing its performance with the proposed model. In this work, we have conducted experiments with all the 26 Keras pre-trained models on the sMRI images and analyzed the results to determine the best performing model on this dataset.
- Synthetic Minority Oversampling Technique (SMOTE) technique is applied to overcome the class imbalance issue. This method randomly duplicates the images of the class that has the least number of samples (minority class). This reduces the chances of model getting over-fit [29].
- We performed brief experiments related to the impact of applying global-average pooling or flattening layer on accuracy before the Softmax classifier. The results (accuracy) of all the 26 Keras pre-trained models is recorded in Table 1.
- We have also applied Saliency map and Gradient-weighted Class Activation Mapping (Grad-CAM) to measure the spatial support of a particular class in each image. This

Table 1 Accuracy comparison for the 26 Keras pre-trained models Note: GAP - Global Average Pooling FL - Flatenning Layer

Model	Accuracy (%)
EfficientNet-B4 (GAP)	72.84
ResNet-50 (FL)	74.72
EfficientNet-B7 (GAP)	75.80
EfficientNet-B6 (FL)	77.64
ResNet-101 (GAP)	58.00
EfficientNet-B0 (GAP)	80.15
EfficientNet-B3 (GAP)	80.43
XceptionNet (FL)	80.91
EfficientNet-B2 (GAP)	81.72
VGG-19 (FL)	82.45
EfficientNet-B1 (GAP)	82.51
InceptionNet-v3 (FL)	82.52
Inception-ResNet-v2	82.96
NasNetMobile (FL)	82.98
ResNet-152 (FL)	83.27
DenseNet-121 (FL)	84.71
VGG-16 (FL)	83.76
EfficientNet-B5 (FL)	84.22
MobileNet-v2 (FL)	84.40
ResNet-101-v2 (FL)	88.86
DenseNet-169 (FL)	89.34
DenseNet-201 (FL)	89.63
ResNet-152-v2 (FL)	90.57
MobileNet (FL)	91.29
ResNet-50-v2 (FL)	91.84

enabled us to understand and explain the model's predictions, supervised localization, segmentation and the region of focus/interest in the image.

2 Related works

Recently, several research works on developing Deep Learning models have been proposed as diagnostic tools for AD. This assists doctors to make better medical diagnosis and improve the perception of the disease processes. AD biomarkers include clinical symptoms (such as cognitive impairment, memory loss), neurological tests and scores such as Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) etc., which are augmented with imaging, genetic, and protein biomarkers [49]. Some of the other works [23, 36] use the cross-sectional neuroimaging and demographic data as reference to study the prediction of MCI to AD conversion by applying multi modal Deep Learning approach that combines various imaging modalities such as sMRI (T1 weighted, T2 weighted), functional MRI (fMRI), Positron Emission Tomography (PET) and imaging genetics [20]. The anatomical properties of the brain can be visualized and analyzed using Structural Imaging approach, whereas functional imaging (fMRI) is used to identify metabolic functions of the brain. The limitation in using fMRI scanning is that it makes it difficult to interpret due to the strong influence of the contrast agents and also less affordable than the usual MRI. Most of the existing works apply Machine Learning or Deep Learning algorithms on the OASIS 4 stage MRI data set (which includes non-demented, very mild-demented, mild-demented or moderately demented MRI images) or targets binary classification such as AD vs. NC, AD vs. MCI, or 3 stage classification of ADNI MRI data (NC vs MCI vs AD) or using fMRI [52] or PET image data for diagnosis of AD.

In this work [19], the Inception-V4 network model is trained on the Oasis T1-weighted MRI dataset (https://www.oasis-brains.org/). The dataset consists of 418 subjects. They have applied 5-fold cross-validation since the dataset is small. For each fold 70% of data is used for training, 10% for validation and 20% for testing. An accuracy of 73.75% is attained.

The results of Deep Learning architectures such as InceptionNetv3 and XceptionNet on the AD binary classification problem using T1-weighted sMRI images on the three different datasets of OASIS are compared and analyzed in this work [45].

This paper [41] presents a CNN based multi modal AD classification framework discriminating between AD and normal control subjects. This work recorded 99% accuracy, 98% sensitivity, 100% specificity and Area under the Receiver Operating Characteristic Curve (AUC) of 1 across all test folds.

This work [38] compares the performance of LeNet-5 and GoogleNet pre-trained model to classify structural MRI data of AD subjects from NC. The model was trained with shift and scale invariant features extracted from different layers of CNN architecture and achieved an accuracy of 98.84% using GoogleNet.

[25], this work compares the performance of two Deep Learning algorithms- MobileNet network model with Visual Geometry Group-16 (VGG-16) as the baseline on the sMRI images (AD vs NC). They have recorded the accuracy of VGG-16 and MobileNet models as 92% and 94% respectively.

In this work [13], multi-class classification between NC, MCI, and AD patients are performed on multi-categorical data using Deep Learning method. The aim of this work is to demonstrate the advantages of using multi-categorical data for classification and to compare Artificial Neural Network (ANN) with CNN. The overall accuracy of 87.197% is

accomplished for the ANN classifier and 88.275% for the 1D Convolutional Neural Network classifier.

In this work [24], Deep Learning approach has been designed to accurately predict the MCI-to-AD conversion with MRI data. As part of data pre-processing, they have assembled local patches into 2.5 dimensions from the MRI images. The CNN model is then trained with these patches to identify Deep Learning features of MCI subjects. This is fed into the classifier to predict the AD conversion. By applying leave-one-out cross-validations, this model achieves an accuracy of 79.9% and AUC of 86.1%.

The primary goal of this paper [35] is to propose 2D CNN model to tackle the 3-class AD classification problem using T1-weighted MRI images from ADNI dataset. The feature extractor is done using the first two layers of ResNet-34 and then the classifier is trained using 64×64 sized patches from the 2D MRI slices. The proposed model achieved an accuracy of 68.6% for the multi-class problem.

This work [36] aims to classify the MCI diagnosed subjects as sMCI (not converted to AD) and pMCI (converted to AD). The designed architecture consists of 15 layers, with 3 residual blocks in each layer and each residual block comprising of 2 basic blocks with 2 convolutional layers each. To avoid the over-fitting problem, they have employed ridge regression technique. They have also applied domain learning in order to increase the separability of the pMCI versus sMCI classification. Since this domain (AD conversion from MCI) has limited data, to attain efficiency in training, the most informative features and related domains are extracted from the target (i.e., image sample). The test classification accuracy increased to 83%, over the actual cross-validated accuracy of 75% on the test data by this method.

A recent attempt of multi-class classification of 5 AD stages has been found in paper [33]. They have proposed fine-tuned ResNet-18 network model that can predict MCI, EMCI, LMCI and AD MRI images. This model was evaluated on ADNI fMRI dataset consisting of 138 subjects. They achieved a classification accuracy of 99.99%, 99.95%, and 99.95% on EMCI vs. AD, LMCI vs.AD, and MCI vs. EMCI MRI data, respectively. This model is trained with 78,735 images of which 70% are considered for training and 30% for validation.

A very similar and recent approach is followed in [21] where they have adopted transfer learning approach for classifying the 4 stage ADNI data. Tissue segmentation had been applied as pre-processing technique and the gray matter has been used to fine-tune the VGG-16 and VGG-19 architectures while freezing the other layers/features of the ImageNet database. 75 to 80 subjects have been used per class. Around 4500 samples have been used per class but EMCI has the least number i.e., around 1800 resulting in class imbalance issue. The dataset has been split into training (70%), testing (15%), and validation data (15%).They have attained an accuracy of 97.12% and 98.47% for VGG-16 and VGG-19 respectively.

Most of these research works focused on either binary classification (NC vs AD)[25, 38, 41, 45] or multi-class classification for 3 stages (NC, MCI & AD) [13, 24]. It is necessary to study and classify all the intermediate stages to find out the severity of the disease i.e., MCI stage. Although this work [33] has achieved high accuracy on fMRI data, only traditional structural imaging is currently advocated for routine use in clinical settings. Hence sMRI dataset has been utilized for this use case [10]. The limitation in this work [24] is that the classifier could precisely mis-classify MCI non-converters as AD, but the conversion may happen in future. This is the reason why it becomes significant to include samples from "EMCI" and "LMCI" stages. Other related works also suggest that longitudinal data provide more meaningful information in predicting the exact stage or the time of conversion. Hence, we have used the sMRI longitudinal data from ADNI [51]. Majority of the work [13, 24, 33] have not compared the performance of applying different CNN models to predict the progression of AD. Therefore, as part of the experiments, we have applied and studied the performances of

Table 2 Summary of the related works on the multi-stage diagnosis of Alzheimer's MRI data

Reference	Methodology	Dataset	Results (Accuracy)	Limitations
Khan and Akbar [21]	The ADNI 4 stage image data is tissue segmented and fed as input to fine-tune pre-trained VGG-16 & VGG-19 models.	4 stage ADNI sMRI of 315 subjects.	VGG-16 - 97.12% VGG-19 98.47%	This work does not clearly mention how they have handled the class imbalance problem though they have stated that they used different number of subjects.
Odusami et al. [33]	Fine-tuned ResNet-18 network model that predicts the 5-stage images is proposed. This model is trained with 78,735 images.	5 stage ADNI fMRI of 138 subjects.	99.99% - EMCI vs. AD, 99.95% - LMCI vs.AD, 99.95% - MCI vs. EMCI	Although this work has achieved high accuracy on fMRI data, only traditional structural imaging is currently advocated for routine use in clinical settings.
Pereira et al. [35]	2D CNN model for the 3-class AD classification of 64 × 64 sized patches of T1-weighted MRI images is proposed by extracting features from the first two layers of ResNet-34.	3 stage ADNI sMRI (AD, MCI, NC)	68.6%	In this work, the dataset had very less samples for the positive class, hence it achieved less accuracy than the benchmark results.
Lu et al. [25]	This work compares the performance of MobileNet with VGG-16 model on the sMRI images of AD & NC classes.	2 stage ADNI sMRI (AD vs NC)	VGG-16 - 92% MobileNet - 94%	This work focuses on the binary classification but it is important to study and classify all the intermediate stages to diagnose the severity of the disease.
Spasov et al. [41]	Presents a CNN based multi-modal classification framework for the diagnosis of AD & NC stages.	2 stage ADNI sMRI (AD vs NC)	99%	Although it has achieved high accuracy on this binary class data, working with multi-class data helps to accurately determine the disease progression in the early stage.

Table 2 continued

Reference	Methodology	Dataset	Results (Accuracy)	Limitations
Islam and Zhang [19]	In this work, the Inception-V4 network model is trained on the Oasis T1-weighted MRI dataset.	3 stage OASIS MRI (very mild, mild, moderate)	73.75%	They have not applied & compared the results with the other transfer learning approaches. Also the model's performance is not assessed by updating the hyper parameters.
Sarraf and Tofighi [38]	This work compares performance of LeNet-5 and GoogleNet model for the classification of sMRI data of AD and NC subjects.	2 stage ADNI sMRI (AD vs NC) 302 subjects	GoogleNet - 98.84%	The train data consisting of AD & NC classes were imbalanced (5 - 1 ratio) in this work. They had under-sampled and claimed that there is no much impact in the accuracy. But under-sampling leads to discarding potential useful data for classification.

different transfer learning models to understand the merits of the best-performing model. All these works intended to develop deep neural networks for bio-image analysis require a large amount of informative data, which in most cases is difficult to obtain. Due to the insufficient samples for the positive class, these works [24, 35, 36] achieved less accuracy than the benchmark results. Data augmentation process is considered for this issue to customize the initial data and increase sample space. To overcome the class imbalance problem in our work, we have applied SMOTE (another data augmentation method) to synthetically generate observations of unbalanced classes which are similar to the existing samples using the Nearest Neighbors classification. The summary of the most significant literature pertaining to this research area on the multi-stage diagnosis of Alzheimer's MRI data is given in Table 2.

3 Methodology

The block diagram in Fig. 2 shows the methodologies adopted for the classification of the AD MRI data. Each of the steps are detailed in subsequent sub-sections.

3.1 Dataset source and description

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/) [4], is a multi-center study launched with an objective to examine AD biomarkers at an earlier stage [6]. ADNI GO phase (an extension of ADNI1) added 200 new subjects with EMCI in 2009. As part of the ADNI2 initiative, the datasets from the ADNI1/ADNI GO phases were included. In addition to that, the following new subject groups were also added: 580 normal controls, 240 EMCI, 72 LMCI, and 171 mild AD patients. The ADNI dataset used for this work comprises of 1296 Structural MRI 2D images divided into Training and Testing set. Figure 3 shows samples of data from each class. The number of images and description of each stage is specified in Table 3.

3.2 Data pre-processing

Generally, the visual datasets used for classification have a repository of millions of images like the ImageNet database. However, neuroimaging datasets generally include a few hundred images only. A large image database with a sufficient number of positive samples will enable to build an efficient CNN model. But in reality, there is a paucity of such databases for medical image analysis, especially for neuro disorders. Hence it becomes essential to develop models that can learn useful features from a smaller dataset [18]. Since the number of positive samples in most of the medical image datasets usually be very less in number, it results in a data imbalance problem, thereby hampering the performance of the model. Therefore, it becomes inevitable to apply data augmentation process to the existing image database. Table 3 shows the number of samples that every class had before any pre-processing. As mentioned earlier, the challenge we encountered was the less number of samples in the positive class i.e., LMCI and AD. In order to resolve the class imbalance problem we decided to adopt an oversampling method. Random Over Sampling, Smote, BorderLine Smote, KMeans Smote, SVM Smote, ADASYN, Smote-NC are the commonly used data augmentation techniques to overcome imbalances in the dataset today. In this work, we have applied SMOTE (Synthetic Minority Over-sampling Technique) which is a superior oversampling option that helps to produce synthetic data points rather than duplicates that only slightly deviate from the actual

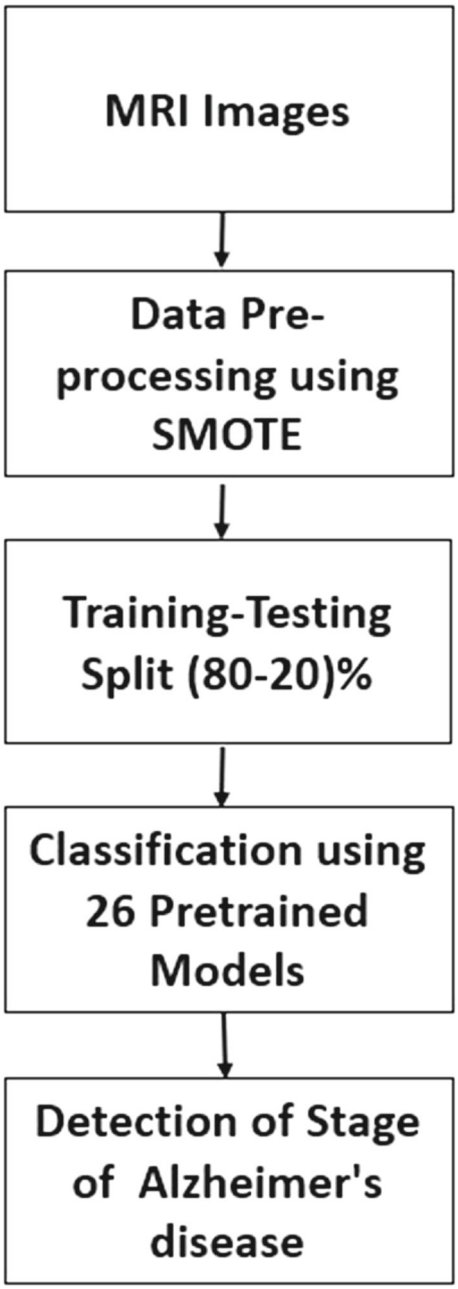

Fig. 2 The proposed Deep Learning pipeline for the 5 stage classification of AD MRI images into NC, EMCI, MCI, LMCI and AD

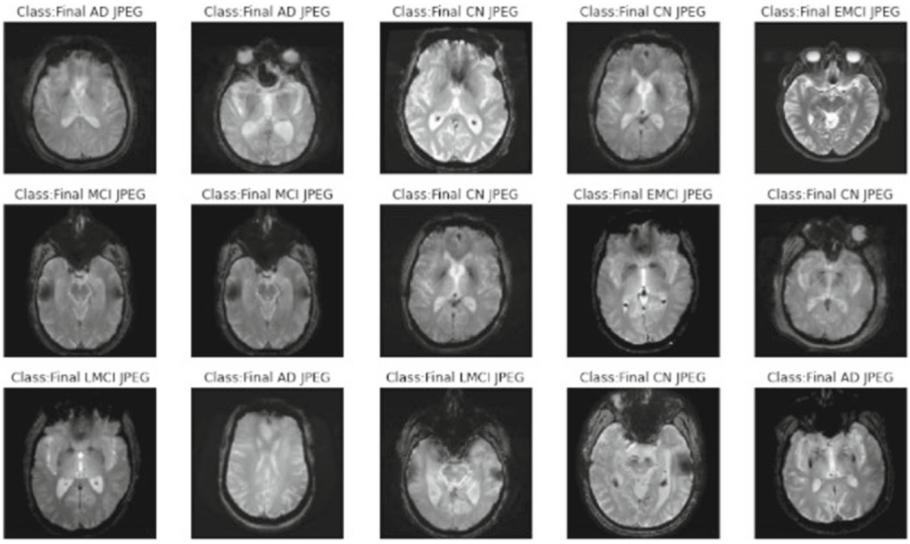

Fig. 3 MRI Images of the 5 stages

data points. The simplest way to achieve this is by duplicating the samples in the minority class. SMOTE generates the synthetic data using the k-nearest neighbor technique. By oversampling the minority class of the imbalanced dataset the over-fitting problem can be overcome. Thus, the augmented samples do not furnish any new information to the model. The SMOTE algorithm works as follows. A random sample is chosen from the minority group. The k nearest neighbors of the observations in this sample will be determined. The vector between the current data point and the selected neighbor is then calculated. The vector is then multiplied by an integer chosen randomly from 0 to 1. In order to get the synthetic data point, this difference is added to the feature vector under consideration. This procedure is equivalent to pushing the data point closer to its neighbor i.e it causes the selection of a random point along the line segment between two specific features. By performing this the synthetic data point is not an exact duplicate of an existing data point and is also not very different from known observations in the minority class. The major advantage of applying SMOTE is that the synthetic samples are generated in a less application-specific manner by operating in the "feature space" rather than "data space" [11]. By applying this data augmentation technique for the minority class the number of samples is increased to 2900 in the training data with 580 samples belonging to each class.

Table 3 AD Stages and number of samples

STAGE	DESCRIPTION	SAMPLES
NC	Normal Control	580
EMCI	Early Mild Cognitive Impairment	240
MCI	Mild Cognitive Impairment	233
LMCI	Late Mild Cognitive Impairment	72
AD	Alzheimer's Disease	171

3.3 Proposed AI framework

For Bio-medical analysis, Convolutional Neural Networks has become the widely used Deep Learning technique [3]. Although Deep CNN networks were introduced more than two decades ago, it could not be easily trained until recently as they are computationally expensive and demands more hardware and software capabilities. Only with the enhancements in hardware resources and network architectures truly deep CNN has become relatively easier to train [1]. From the previous experiments conducted using deep networks, it is understood that very deep model results in performance degradation as they start converging. Generally, the degradation of training accuracy is said to occur due to the over-fitting or increased number of layers added to the deep model [16]. But in this case, it is understood that degradation occurs due to the optimization function, initialization of the network, and vanishing/exploding gradient problem. So, this problem can be solved by constructing a deeper model with layers of identity mapping and copying from the learned shallower model which forms the core idea of residual networks. This neural network layer called "The Residual Block" alleviates the deeper network training. This solution proved that the training error for the deeper networks is no higher than the shallower model. In a favorable scenario, they give much better accuracies than the shallower model.

In this work, ResNet-50v2 performed the best when compared to all other 2D CNN architectures. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 classification challenge, ResNet performed the best and it is claimed to be one of the most powerful architecture. To ease the training of deeper networks introduced earlier such as DenseNet, ResNet architecture uses a residual learning framework. Each of the ResNet architecture come with a varied number of layers i.e., ResNet-18, , ResNet-50, ResNet-101, ResNet-152, etc. In every ResNet architecture, two or more digit number is mentioned to denote the number of layers present in the model. Similar to the conventional Keras models, Residual Network also consists of the strided convolution layer, pooling layer, activation, and fully connected layers which are placed one after the other. The only distinction in the residual network construction is that the identity connection is introduced between the layers. The explicit reformulation of the layers enables the residual functions to learn better with reference to the layer inputs. This paves the way to optimize the training and gain higher accuracy from considerably increased depth. Therefore, an optimal identity mapping makes pushing the residual to zero easier than fitting the model by adding a stack of nonlinear layers. The identity mapping(as shown in Fig. 4) just adds the output from the previous layer

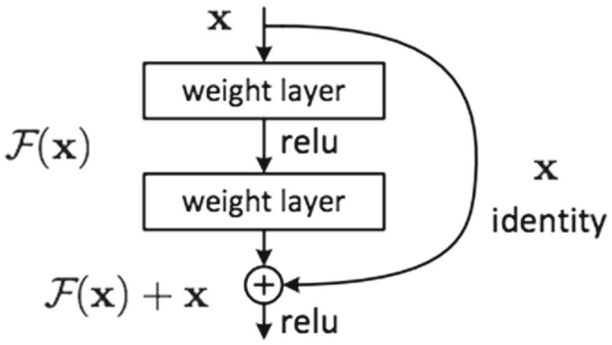

Fig. 4 ResNet-50 Architecture denoting Skip Connection

to the next layer and does not add any parameters. However, the problem arises when the dimension of x (input vector) and F(x, Wi) (residual mapping function to be learned) are not the same (refer Equation: 1). There are two ways by which this dimensional mismatch can be solved. As mentioned in the ResNet paper [16], one approach to retain the size of the spatial dimensions is by multiplying the identity mapping by a linear projection Ws (refer Equation 1 [16]) i.e., by performing 1x1 convolutions to increase or decrease the depth (i.e., it impacts the number of channels). This way the channels are expanded to align the residual mappings thereby allowing the input x and F(x) to be added and fed to the next layer ('y' in Equation: 1 represents the output vector of a layer which is given as input to the next layer). Another method is to use the padding approach. This is done by downsampling the input by using 1*1 convolution with any stride (standard stride size of 2 is followed) and zero channels are padded to increase the depth. This helps to match the dimensions without increasing the number of trainable parameters across the skip connections. Therefore, these Skip or Shortcut Connections between layers increase the ability to train much deeper networks by adding the outputs from previous layers to the outputs of the stacked layers [16]. Thus the biggest advantage of Residual Networks is that the gradient can flow directly through the identity function from later layers to the previous layers [17].

$$y = F(x, W_i) + W_s x \tag{1}$$

The 5 stage Resnet-50 model includes convolution and identity blocks. There are 3 convolution layers in every convolution and identity block. The three layers are 1×1, 3×3, 1×1 convolution. The 1×1 convolution layer helps to reduce and restore the dimensions. The 3×3 layer acts as a narrow layer with smaller input or output dimensions. ResNet takes input image of height and width 150*150 and 3 as channel width. In our experiment, the number of trainable parameters was over 107 million. The preliminary convolution and max pooling layers uses Kernel sizes of 7×7 and 3×3 respectively. The first stage of the network consists of 3 Residual blocks with 3 layers each. The number of Kernels used in each convolution for every layer of the residual block of the first stage are 64, 64, and 128 respectively and

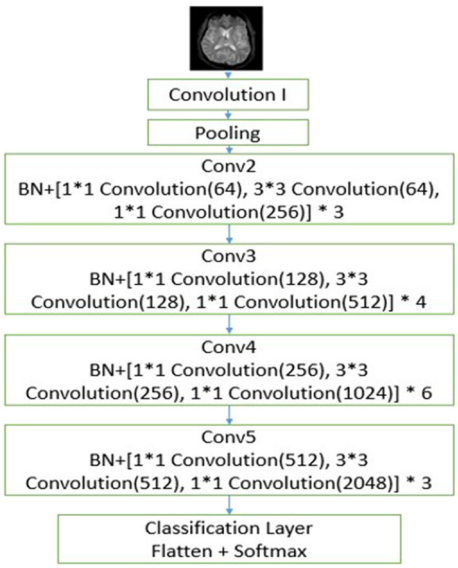

Fig. 5 ResNet-50v2-Architecture

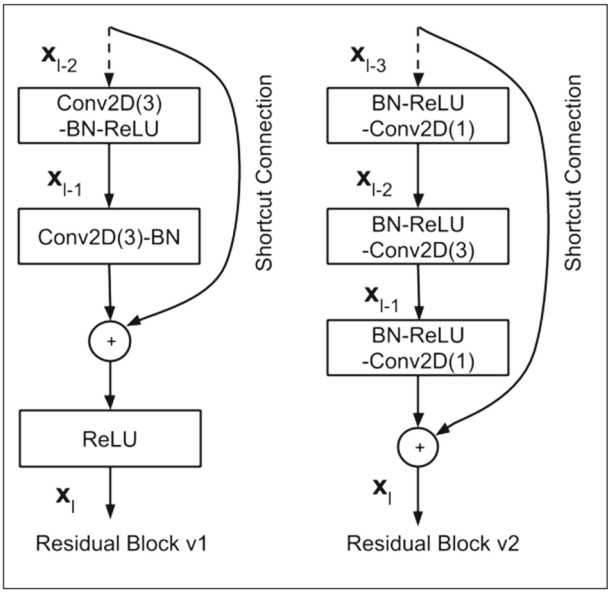

Fig. 6 Difference between ResNet-50 and ResNet-50v2 architectures

this is doubled for the subsequent layers. The convolution is performed with stride 2 thereby reducing the height and width of the previous layer input to half and doubling the width of the channel. In the end, the network has a Flattening layer followed by a Fully Connected (FC) layer. In this case, the FC layer has 5 neurons (as the image should be classified into 5 target classes) with Softmax as the activation. The general architecture block diagram of ResNet-50v2 is given in Fig. 5. The upgraded version of ResNet-50v1 i.e., ResNet-50v2 uses batch normalization before each weight layer, as shown in Fig. 6. The curved arrows in Fig. 6 refer to the skip connection. The dashed downward arrow represents the convolution operation in the Residual Block. The categorical-crossentropy loss function is used for the multi-class classification and early-stopping is used for regularization. The learning rate controls the step size at each iteration of an optimization algorithm as it advances toward a minimum of loss function. The default learning rate of 0.001 is applied and batch size is 50. The model is trained for 15 epochs.

4 Experiments and results

The train and test data of the ADNI dataset consists of 1101 and 195 image samples respectively which are combined to form a single repository. We observed that the number of samples in every class is not the same leading to a class imbalance problem. The challenge in working with an imbalanced dataset is that most of the Deep Learning techniques result in poor performance on the minority class. Hence SMOTE technique is adopted for oversampling the minority class data. After balancing the data, the number of samples is increased to 2900 images. The train and validation images are then split into an 80%-20% ratio, which results in 2320 train images (includes 464 test images) and 580 validation images.

Table 4 Training and Validation Accuracy of the top performed pre-trained models

Model	Training Accuracy (%)	Validation Accuracy (%)
ResNet-101v2	99.99	94.23
DenseNet-169	99.63	95.09
DenseNet-201	99.21	94.78
ResNet-152v2	99.91	96.26
MobileNet	99.94	96.49
ResNet-50V2	99.99	96.01

Research work related to transfer learning is motivated by the fact that new problems can be solved efficiently at a very less amount of time by intelligently utilizing the related knowledge learned previously [34]. All the 26 pre-trained models from Keras Deep Learning framework were applied to understand which model performed the best for this classification problem. All the experiments related to this project were carried out using Google Colaboratory and Kaggle GUI. Table 4 shows the Receiver Operating Characteristic (ROC) curve accuracy for the performance of top-6 models and Table 5 compares the accuracy with the existing works.

Experiments were also done to understand the impact of the Global Average Pooling and Flattening layer before the Softmax classifier. "Global Average Pooling" for EfficientNet greatly reduced the number of trainable parameters. This enabled relatively much faster training. In Table 1, the models for which Global Average Pooling was used is highlighted. But in cases where the maximum representable features are to be extracted from the input image, average pooling failed as it returns less accuracy. This happens for a binary or multi-class classification problem where images in the classes look almost alike [27]. For instance, ResNet-101 yielded an accuracy of 58% with global average pooling but with flattening layer gave 78.90%. Since the primary goal of this work is to classify the images with good accuracy without over-fitting and not dimensionality reduction, flattening layer was applied for all other models. Table 1 shows the accuracy comparison of all the 26 Keras pre-trained models. In this case, EfficientNet-B1 performed the best with 82.51%. The other EfficientNet models such as EfficientNet-B4 and EfficientNet-B7 gave accuracy around 72%-75% as they were relatively difficult to be trained due to the large memory size requirement and were computationally expensive due to very high number of trainable parameters. The accuracy of these models was recorded for the same training time of 70 epochs.

Confusion matrix is a way to assess the overall performance of a classification algorithm by comparing the predicted samples with the ground truth. For bio-medical data, it is imperative to determine how many predicted samples fell into true positive, true negative, false positive and false negative categories (i.e., false positive and negative for every class should be very minimum). This acts as a foundation evaluation metric to define the performance of the model. The confusion matrix for the best performed model ResNet-50v2 is shown in Fig. 7. It is observed that this model has correctly predicted for most of the test cases and misclassification for each of the classes is very minimal. The AD stage is predicted with 98.95%acccuracy (Out of 96 samples considered, 95 are classified correctly) as denoted in Fig. 7.

The Scikit Learn or sklearn classification report function is used for getting the class-wise performance results for the metrics such as accuracy, sensitivity, specificity, recall, F1-score, macro and weighted averages. Accuracy can be defined as the ratio of the number of cases classified correctly to the total number of cases under evaluation. This model gave the best accuracy of 91.84%. Recall of positive class, also termed as sensitivity gives the ratio of the

Table 5 Comparison of accuracy of the proposed model with the existing works

Reference	Dataset	Reference Model	Reference Model Accuracy (%)	Proposed Model Accuracy (%)
Sarraf and Tofighi [38]	ADNI 2-stage (AD vs NC)	GoogleNet	98.84	91.16
Islam and Zhang [19]	OASIS 3-stage (very mild, mild, moderate)	InceptionNet-v4	73.75	(AD + MCI) vs NC - 90.30
Lin et al. [24]	ADNI 2-stage (AD vs NC)	Own Model	79.9	91.16
Spasov et al. [41]	ADNI 2-stage (AD vs NC)	Own Model	99	91.16
Toshkhujaev et al. [43]	ADNI 4-stage (AD vs NC) (EMCI vs LMCI)	RBF- SVM	AD vs NC - 91.57 EMCI vs LMCI - 87.50	AD vs NC - 91.26 EMCI vs LMCI - 83.42
Lu et al. [25]	ADNI 2-stage (AD vs NC)	VGG-16, MobileNet	VGG-16 - 92% MobileNet - 94%	91.16
Bae et al. [5]	ADNI 2-stage (mild AD vs NC) & SNUBH	InceptionNet-v4	88	91.16
Odusami et al. [33]	ADNI 5-stage fMRI (EMCI vs LMCI vs MCI vs AD vs NC)	ResNet-18	99.99% - EMCI vs. AD, 99.95% - LMCI vs.AD, 99.95% - MCI vs. EMCI	Overall - 91.84 EMCI vs AD - 83.42 LMCI vs AD - 89.06 MCI vs EMCI - 94.17

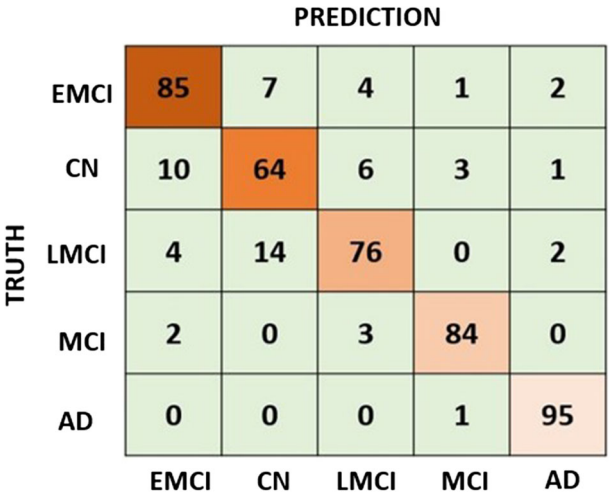

Fig. 7 Confusion Matrix for 5 stage classification

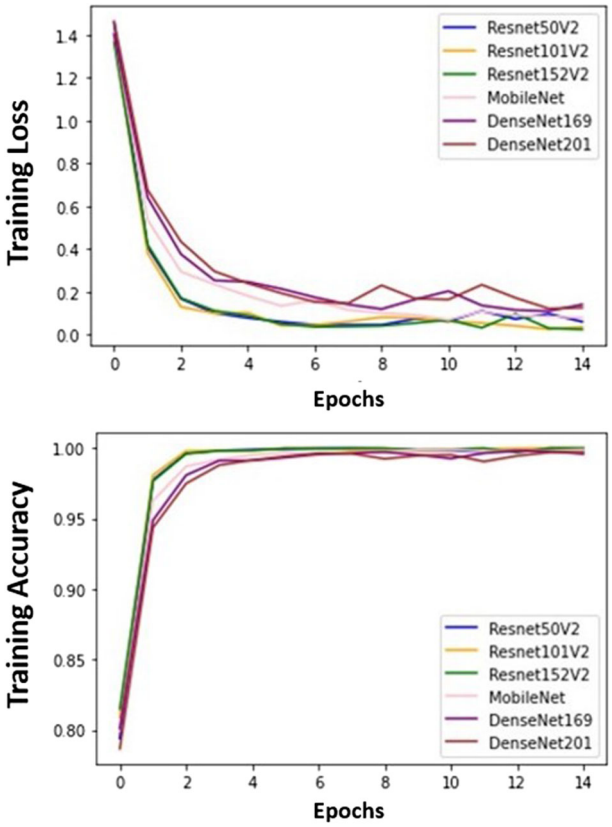

Fig. 8 Training accuracy and loss curves for the best performed models

Table 6 Performance Metrics for ResNet-50v2 Model

Class	Precision	Recall	F1-Score	Support
EMCI	0.84	0.86	0.85	99
NC	0.75	0.76	0.76	84
LMCI	0.85	0.79	0.82	96
MCI	0.94	0.94	0.94	89
AD	0.95	0.99	0.97	96
Accuracy is 0.9184				

Table 7 Confusion Matrix for 3 stage classification

3 stage (NC vs MCI vs AD) Class	NC	MCI	AD
NC	64	19	1
MCI	21	259	4
AD	0	1	95
Accuracy(%)	90.30		

Table 8 Confusion Matrix and Performance Metrics for binary classification

(a) AD vs NC

Class	NC	AD
NC	64	20
AD	21	359

Performance Metrics	
Precision	0.95
Recall	0.95
F1-score	0.95
Specificity	0.761

(b) EMCI vs AD, LMCI vs AD, MCI vs EMCI, EMCI vs LMCI

EMCI vs AD

Class	EMCI	AD	Accuracy(%)
EMCI	85	14	83.41
AD	0	96	

LMCI vs AD

Class	LMCI	AD	Accuracy(%)
LMCI	76	20	89.06
AD	1	95	

MCI vs EMCI

Class	MCI	EMCI	Accuracy(%)
MCI	84	5	94.18
EMCI	20	13	

EMCI vs LMCI

Class	EMCI	LMCI	Accuracy(%)
EMCI	85	14	83.41
LMCI	1	76	

true Positive to the number of actual positive cases. For each of the classes the sensitivity values are high indicating that the number of samples predicted as False Negative are very less, especially predicting for AD class where the sensitivity is 99% indicating that there was only single mis-classification or false negative case, as highlighted in Fig. 7. The precision of positive class is innately the ability of the classifier not to predict a positive sample as negative. The best value of precision is 1 and the worst value is 0. The precision for the NC class is minimum as 10 samples were predicted as EMCI but only one sample was predicted as AD. But for all other classes the precision score is also observed to be good. Regardless of the class imbalance problem, F1-score is considered as one off the best metrics for classification models. F1-score is the metric computed by taking the weighted average of recall and precision of each of the classes. It is observed to be more than 75% for all 5 classes. The micro-average for all the 5 classes is 0.87 indicating equal number of samples in each class. The Performance metrics for the ResNet-50v2 model is shown in Table 6. It is also essential to understand how our model predicts in the case of binary or tertiary class data in order to draw fair comparison of our model's performance (in terms of accuracy) with the existing works as shown in Table 5 which compares the accuracy with the existing works. Table 7 shows the confusion matrix generated by considering all EMCI, MCI & LMCI predicted samples into MCI class thereby deriving the accuracy for tertiary classification. Since we worked on multiclass data the accuracy of the same ResNet-50v2 model for binary classification can be derived by combining all other categories into AD class which is a fair comparison to be done in this case (see Table 8). In Table 8a, the performance metrics for NC vs AD is computed by considering all data belonging to MCI, EMCI, LMCI & AD as positive class i.e AD & Table 8b shows the confusion matrix & accuracy for the binary classification between intermediate stages, for instance, EMCI vs LMCI. These computed accuracies are included in Table 5 for comparison with the existing works.

The training accuracy and loss is compared for the top performing models in Fig. 8. These models were trained for 15 epochs. Almost all the top performing models took less training time. Very deep networks like DenseNet-201 and models with large memory requirement like Efficient (all 7 versions) consumed a lot of training time. It is generally observed that as the training time increases, the model's validation accuracy increases and loss decreases. The ResNet-50v2 model attained an accuracy of 99.99% on the training set and an accuracy of 96.01% on the validation set. Thus, the model has not over-fitted and has performed equally well on both the training and validation sets.

For understanding and visualization of the region of focus, Grad-CAM and Saliency Map is used. Grad-CAM is one of the widely used visualization tools for producing explanations for the predictions from the CNN model [39]. Grad-CAM makes use of the gradients flowing into the final convolution layer to produce a rough localization map thereby projecting the critical regions in the image for predicting the image into its class. In computer vision, a Saliency map is an image that highlights the region which differentiates a sample from one class from another. The gradual increase in the highlighted portion in structural MRI shown in Fig. 9 indicates the shrinking of brain from normal to AD stage. Another data exploration and visualization technique that is applied in this work is t-Distributed Stochastic Neighbor Embedding (t-SNE) which is non-linear, unsupervised technique mainly used to give an idea about the data organization in high-dimensional space [46]. It can be observed that each majority of the samples are classified into their well-defined clusters as shown in shown Fig. 10. We can also observe that NC and EMCI are closely related. Similarly, only one instance of LMCI is mis-classified into the AD class.

5 Conclusion

Overall, from the literature survey, we could understand that 5 stage ADNI sMRI 2D Image data set has not been widely used for AD detection. Hence, we applied all the 26 pre-trained Keras Deep Learning models to predict the classification of AD. The success of using MRI to detect brain abnormalities associated with brain disorders motivated us to use the ADNIsMRI dataset for prediction.

Currently, the analyses of medical imaging scans are done manually by skilled radiologists which leads to over-dependence on the analyst's skills. This may potentially affect the efficiency and accuracy of the results. So, there is a need to create an automated system for the AD diagnosis [15]. After applying all the Keras 26 pre-trained models, we observed that ResNet-50v2 architecture performed the best with a benchmark accuracy of 91.84%. The efficiency of the model is reported with respect to performance metrics like accuracy, F1-score, precision and recall. The main limitation of this work is that this can assist only in the diagnosis of AD but not tested with other Dementia data such as Fronto-temporal Dementia (FTD) [26], Lewy-Body Dementia (LBD) [50] and Vascular Dementia (VD) [8]

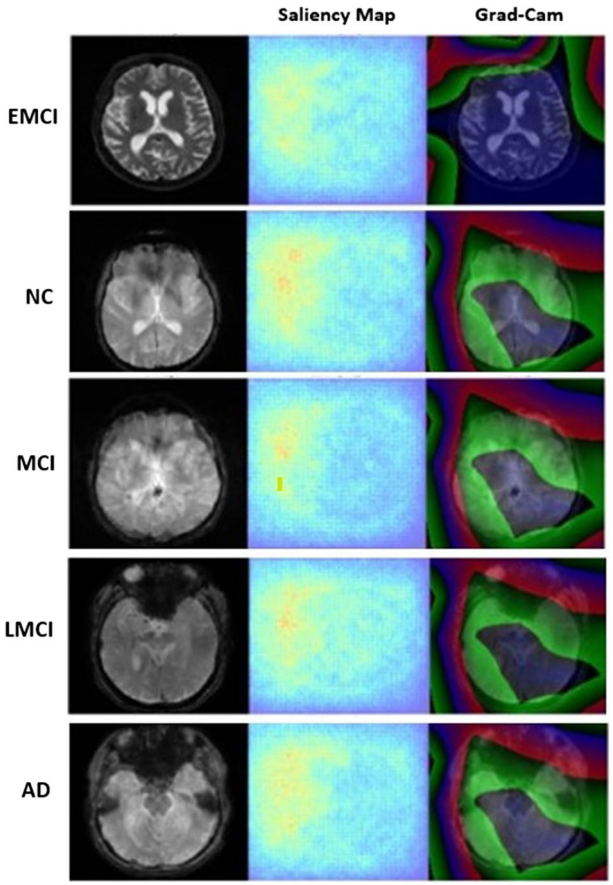

Fig. 9 Saliency map for 5 stages using ResNet-50v2 model

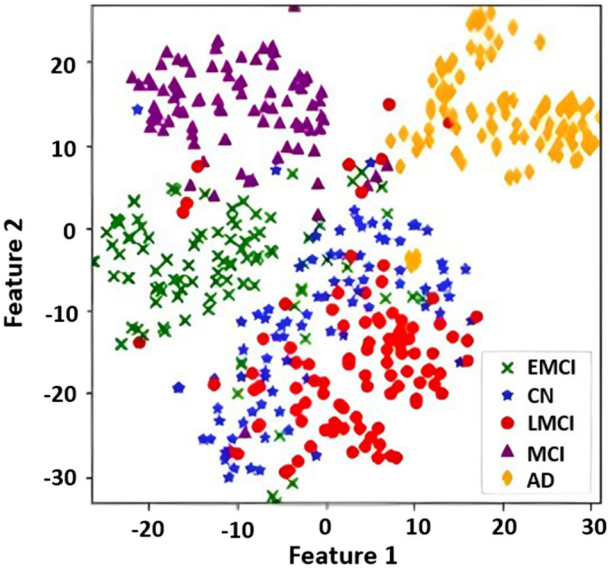

Fig. 10 t-SNE Plot

which has overlapping symptoms with AD. Also the demographic features of this data could have also been included to make the prediction more reliable. Another dis-advantage of the propsed work is that more than using the individual pre-trained models, multimodal or ensemble models could be used so that the aggregate results are less noisy than the individual models. This leads to model stability, robustness and also helps to capture linear as well as non-linear relationships in the data. Hence to overcome these limitations, we plan to apply Ensemble modelling approach to boost the accuracy of the trained model and experiment with pre-processing technique called Generative Adversarial Networks (GAN) as part of our future work [28]. We also intend to experiment by applying explainable AI which helps to assess the expected impact and potential biases for the differential diagnosis of AD with other types of Dementia. The proposed approach could also be extended on fMRI or PET images to compare and understand which neuroimaging technique helps for a more accurate AD diagnosis.

Data availability The data that support the findings of this study are available from the first author upon reasonable request.

Code availability The code is available from the first author upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Aloysius N, Geetha M (2017) A review on deep convolutional neural networks. In: 2017 International Conference on Communication and Signal Processing (ICCSP). IEEE, pp 0588–0592

2. Al-Shoukry S, Rassem TH, Makbol NM (2020) Alzheimer's diseases detection by using deep learning algorithms: a mini-review. IEEE Access 8:77131–77141
3. Altinkaya E, Polat K, Barakli B (2020) Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: a comprehensive review. J Electr Comput Eng 1(1):39–53
4. Alzheimer ADNI dataset. (2004). [Online]. https://adni.loni.usc.edu/data-samples/access-data/
5. Bae JB, Lee S, Jung W, Park S, Kim W, Oh H, Han JW, Kim GE, Kim JS, Kim JH et al (2020) Identification of Alzheimer's disease using a convolutional neural network model based on t1-weighted magnetic resonance imaging. Sci Rep 10(1):1–10
6. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M, Alzheimer's Disease Neuroimaging Initiative 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
7. Burgos N, Bottani S, Faouzi J, Thibeau-Sutre E, Colliot O (2021) Deep learning for brain disorders: from data processing to disease treatment. Brief Bioinform 22(2):1560–1576
8. Castellazzi G, Cuzzoni MG, Cotta Ramusino M, Martinelli D, Denaro F, Ricciardi A, Vitali P, Anzalone N, Bernini S, Palesi F et al (2020) A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia fed by MRI selected features. Front Neuroinform 14:25
9. Cedazo-Minguez A, Winblad B (2010) Biomarkers for Alzheimer's disease and other forms of dementia: clinical needs, limitations and future aspects. Exp Gerontol 45(1):5–14
10. Chandra A, Dervenoulas G, Politis M (2019) Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. J Neurol 266(6):1293–1302
11. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
12. Chitradevi D, Prabha S (2020) Analysis of brain sub regions using optimization techniques and deep learning method in Alzheimer disease. Appl Soft Comput 86:105857
13. Cohen DS, Carpenter KA, Jarrell JT, Huang X, Initiative ADN et al (2019) Deep learning-based classification of multi-categorical Alzheimer's disease data. Curr Neurobiol 10(3):141
14. Farooq A, Anwar S, Awais M, Rehman S (2017) A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: 2017 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, pp 1–6
15. Feng W, Halm-Lutterodt NV, Tang H, Mecum A, Mesregah MK, Ma Y, Li H, Zhang F, Wu Z, Yao E et al (2020) Automated MRI-based deep learning model for detection of Alzheimer's disease process. Int J Neural Syst 30(06):2050032
16. 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
17. 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
18. Islam J, Zhang Y (2018) Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Inform 5(2):1–14
19. Islam J, Zhang Y (2017) A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data. In: International conference on brain informatics. Springer, pp 213–222
20. Jo T, Nho K, Saykin AJ (2019) Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front Aging Neurosci 11:220
21. Khan R, Akbar S, Mehmood A, Shahid F, Munir K, Ilyas N, Asif M, Zheng Z (2022) A transfer learning approach for multiclass classification of Alzheimer's disease using MRI images. Front Neurosci 16
22. Knight M, McCann B, Kauppinen R, Coulthard E (2016) Magnetic resonance imaging to detect early molecular and cellular changes in Alzheimer's disease. Front Sging Neurosci 8:139. https://doi.org/10.3389/fnagi.2016.00139
23. Lee G, Nho K, Kang B, Sohn K-A, Kim D (2019) Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci Rep 9(1):1–12
24. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, Guo G, Xiao M, Du M, Qu X et al (2018) Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front Neurosci 12:777
25. Lu X, Wu H, Zeng Y (2019) Classification of Alzheimer's disease in mobilenet. In: Journal of Physics: Conference Series, vol 1345, no 4. IOP Publishing, p 042012
26. Ma D, Lu D, Popuri K, Wang L, Beg MF, Initiative ADN et al (2020) Differential diagnosis of frontotemporal dementia, Alzheimer's disease, and normal aging using a multi-scale multi-type feature generative adversarial deep neural network on structural magnetic resonance images. Front Neurosci 14:853

27. Magnin B, Mesrob L, Kinkingnéhun S, Pélégrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehéricy S, Benali H (2009) Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51(2):73–83
28. McCrackin L (2018) Early detection of Alzheimer's disease using deep learning. In: Canadian Conference on Artificial Intelligence. Springer, pp 355–359
29. Murugan S, Venkatesan C, Sumithra M, Gao X-Z, Elakkiya B, Akila M, Manoharan S (2021) DemNet: a deep learning model for early diagnosis of Alzheimer diseases and dementia from mr images. IEEE Access 9:90319–90329
30. Nagaraj S, Duong TQ (2021) Deep learning and risk score classification of mild cognitive impairment and Alzheimer's disease. J Alzheimers Dis (Preprint):1–12
31. Nair JJ, Mohan N (2017) Alzheimer's disease diagnosis in MR images using statistical methods. In: 2017 International Conference on Communication and Signal Processing (ICCSP). IEEE, pp 1232–1235
32. Noor MBT, Zenia NZ, Kaiser MS, Al Mamun S, Mahmud M (2020) Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease. Parkinson's disease and schizophrenia. Brain Inform 7(1):1–21
33. Odusami M, Maskeliūnas R, Damaševičius R, Krilavičius T (2021) Analysis of features of Alzheimer's disease: detection of early stage from functional brain changes in magnetic resonance images using a finetuned resnet18 network. Diagnostics 11(6):1071
34. Oh K, Chung Y-C, Kim KW, Kim W-S, Oh I-S (2019) Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci Rep 9(1):1–16
35. Pereira MEDC et al (2019) An extended-2D CNN approach for diagnosis of Alzheimer's disease through structural MRI: Abordagem CNN 2D estendida para o diagnóstico da doença de alzheimer através de imagens de ressonância magnética estrutural
36. Rana SS, Ma X, Pang W, Wolverson E (2020) A multi-modal deep learning approach to the early prediction of mild cognitive impairment conversion to Alzheimer's disease. In: 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT). IEEE, pp 9–18
37. Rasmussen J, Langerman H (2019) Alzheimer's disease-why we need early diagnosis. Degener Neurol Neuromuscul Dis 9:123
38. Sarraf S, Tofighi G (2016) Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks. Preprint at http://arxiv.org/abs/1607.06583
39. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization, in Proceedings of the IEEE International Conference on Computer Vision. pp 618–626
40. Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19:221–248
41. Spasov SE, Passamonti L, Duggento A, Liò P, Toschi N (2018) A multi-modal convolutional neural network framework for the prediction of Alzheimer's disease. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp 1271–1274
42. Thushara A, Amma CU, John A, Saju R (2020) Multimodal MRI based classification and prediction of Alzheimer's disease using random forest ensemble. In: 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA). IEEE, pp. 249–256
43. Toshkhujaev S, Lee KH, Choi KY, Lee JJ, Kwon G-R, Gupta Y, Lama RK (2020) Classification of Alzheimer's disease and mild cognitive impairment based on cortical and subcortical features from MRI t1 brain images utilizing four different types of datasets, Journal of Healthcare Engineering, vol. 2020
44. Trojachanec K, Kitanovski I, Dimitrovski I, Loshkovska S (2017) Longitudinal brain MRI retrieval for Alzheimer's disease using different temporal information. IEEE Access 6:9703–9712
45. Tufail AB, Ma Y-K, Zhang Q-N (2020) Binary classification of Alzheimer's disease using sMRI imaging modality and deep learning. J Digit Imaging 33(5):1073–1090
46. Van der Maaten L, Hinton G (2008) Visualizing data using T-SNE. J Mach Learn Res 9(11)
47. Veetil IK, Gopalakrishnan E, Sowmya V, Soman K (2021) Parkinson's disease classification from magnetic resonance images (MRI) using deep transfer learned convolutional neural networks. In: 2021 IEEE 18th India Council International Conference (INDICON). IEEE, pp 1–6
48. Vemuri P, Jack CR (2010) Role of structural MRI in Alzheimer's disease. Alzheimers Res Ther 2(4):1–10
49. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD (2021) Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci Rep 11(1):1–13
50. Wada A, Tsuruta K, Irie R, Kamagata K, Maekawa T, Fujita S, Koshino S, Kumamaru K, Suzuki M, Nakanishi A et al (2019) Differentiating Alzheimer's disease from dementia with lewy bodies using a deep learning technique based on structural brain connectivity. Magn Reson Med Sci 18(3):219
51. Xu Z, Shen X, Pan W, Initiative ADN (2014) Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. PLoS ONE 9(8):e102312

52. Zhang T, Zhao Z, Zhang C, Zhang J, Jin Z, Li L (2019) Classification of early and late mild cognitive impairment using functional brain network of resting-state fMRI. Front Psych 10:572

Authors and Affiliations

Srividhya L[1] · Sowmya V[1] · Vinayakumar Ravi[2] · Gopalakrishnan E.A[1] · Soman K.P[1]

Srividhya L
srividhya.amrita@gmail.com

Sowmya V
v_sowmya@cb.amrita.edu

Gopalakrishnan E.A
ea_gopalakrishnan@cb.amrita.edu

Soman K.P
kp_soman@amrita.edu

1 Center for Computational Engineering and Networking (CEN), Amrita School of Engineering Coimbatore, Amrita Vishwa Vidyapeetham, Coimbatore, India

2 Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia