

# Convolutional neural networks for Alzheimer's disease detection on MRI images

Amir Ebrahimi<sup>ID</sup>\* and Suhuai Luo,  
for the Alzheimer's Disease Neuroimaging Initiative<sup>†</sup>

The University of Newcastle, School of Electrical Engineering and Computing,  
Callaghan, New South Wales, Australia

## Abstract

**Purpose:** Detection of Alzheimer's disease (AD) on magnetic resonance imaging (MRI) using convolutional neural networks (CNNs), which is useful for detecting AD in its preliminary states.

**Approach:** Our study implements and compares several deep models and configurations, including two-dimensional (2D) and three-dimensional (3D) CNNs and recurrent neural networks (RNNs). To use a 2D CNN on 3D MRI volumes, each MRI scan is split into 2D slices, neglecting the connection among 2D image slices in an MRI volume. Instead, a CNN model could be followed by an RNN in a way that the model of 2D CNN + RNN can understand the connection among sequences of 2D image slices for an MRI. The issue is that the feature extraction step in the 2D CNN is independent of classification in the RNN. To tackle this, 3D CNNs can be employed instead of 2D CNNs to make voxel-based decisions. Our study's main contribution is to introduce transfer learning from a dataset of 2D images to 3D CNNs.

**Results:** The results on our MRI dataset indicate that sequence-based decisions improve the accuracy of slice-based decisions by 2% in classifying AD patients from healthy subjects. Also the 3D voxel-based method with transfer learning outperforms the other methods with 96.88% accuracy, 100% sensitivity, and 94.12% specificity.

**Conclusions:** Several implementations and experiments using CNNs on MRI scans for AD detection demonstrated that the voxel-based method with transfer learning from ImageNet to MRI datasets using 3D CNNs considerably improved the results compared with the others.

© 2021 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.8.2.024503](https://doi.org/10.1117/1.JMI.8.2.024503)]

**Keywords:** deep learning; convolutional neural network; Alzheimer's disease; magnetic resonance imaging; transfer learning.

Paper 20350R received Jan. 1, 2021; accepted for publication Apr. 13, 2021; published online Apr. 29, 2021.

## 1 Introduction

Alzheimer's disease (AD) is a fatal irreversible, progressive neurodegenerative disorder that causes brain cells to waste away and die. Typically, AD begins in middle/old age, with protein accumulation inside/around neurons. The most prevalent and one of the early symptoms of AD is problems remembering new things, since AD-related changes usually begin in the brain parts charged with learning. Symptoms include, but are not limited to, behavioral changes; deep confusion regarding time, events, and places; and doubts about family members and friends. They usually develop slowly and worsen over time, leading to a continual deterioration in memory and difficulty in swallowing, talking, and walking.<sup>1-3</sup>

---

\*Address all correspondence to Amir Ebrahimi, [Amir.Ebrahimi@uon.edu.au](mailto:Amir.Ebrahimi@uon.edu.au)

<sup>†</sup>Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, ADNI investigators contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

AD is the prevalent dementia type, which includes about 60% to 80% of the total cases of dementia.<sup>1</sup> Dementia refers to severe loss of memory and other cognitive capabilities that interfere with daily life. It is estimated to have affected about 50 million people in the world and 459,000 Australians in 2020.<sup>1,4</sup> The statistics are estimated to almost double every 20 years.<sup>5</sup> Dementia is the second leading cause of death among Australians, accounting for 15,016 deaths in 2019.<sup>6</sup>

Despite exploring various treatments for preventing or slowing AD, the success rate has been low, particularly in the latest phases of the disease.<sup>7</sup> Studies indicate that AD-related changes in the brain might begin about 20 years before symptoms emerge.<sup>1</sup> Therefore, there is a time gap that could be highly valuable to slowing AD's progression. Early detection of AD extends the independence of patients for a longer period. Recent research may enable greater comprehension of the disease and improvement of new treatments.<sup>2,8</sup>

In practice, AD detection is based on checking brain scans, a clinical assessment, and asking questions of the patient and their relatives.<sup>9,10</sup> This process is usually challenging because of the limited knowledge in identifying the parts of the brain affected by AD. Moreover, AD symptoms, like brain shrinkage, can also be observed in healthy, elderly normal control (NC) groups.<sup>11</sup> In the last two decades, a wide range of studies has been performed to detect AD using artificial intelligence. Common classification algorithms in machine learning, such as neural networks and support vector machines (SVMs), have been applied to brain scans such as positron emission tomography (PET) and magnetic resonance imaging (MRI). Detecting AD using these algorithms is difficult for scholars because of the low image quality, issues of brain segmentation and preprocessing steps, the absence of a database with a sufficient number of subjects, and the complexity of medical images. Successful classification necessitates a robust power to distinguish specific features in similar brain images.<sup>12</sup> In a systematic literature review,<sup>13</sup> 18 of 114 reviewed studies compared deep learning models with machine learning models. They all reported the superiority of the former. Therefore, in this paper, we only investigate and compare deep learning models.

The rise in the computation capacity of graphics processing units (GPUs) has supported the evolution of modern and innovative deep learning algorithms. As a subgroup of machine learning, deep learning models analyze data processing and pattern recognition in the human brain to solve complicated decision-making tasks. Deep learning approaches have enhanced intelligent systems in numerous areas.<sup>14</sup> Research on medical images has been encouraged by deep learning methods in applications using two-dimensional (2D) natural images.<sup>15,16</sup> Among deep learning models, convolutional neural networks (CNNs) have recently demonstrated revolutionized outcomes in disease detection and organ segmentation.<sup>17</sup> In contrast to traditional machine learning methods, CNNs can merge three main steps of classification: feature extraction, feature selection, and classification. It was recently stated that CNNs are the most frequently used method—about 70%—for AD detection.<sup>13</sup>

MRI is the most extensively utilized biomarker for the detection of AD using deep learning. It has been used in more than 80% of AD detection studies in single-modal approaches.<sup>13</sup> This paper plans to use MRI scans to classify AD patients from NCs using CNNs. We aim to use CNNs to uncover latent representations, discover relationships among slices of images, and recognize patterns related to AD in brain scans. Our research's main contribution is to expand the idea of transfer learning from 2D images to three-dimensional (3D) MRI scans. Hence, learnable parameters from 2D CNNs are transferred to 3D CNNs. To begin, the related work is first outlined to present the structure of CNNs and the background of employing 2D and 3D CNNs in AD detection using neuroimaging. Next, different types of CNNs in 2D and 3D approaches to manage MRI volumes with or without transfer learning are examined in detail. Subsequently, our proposed deep model for introducing the concept of transfer learning to 3D CNNs is explained. Finally, experimental results are discussed, followed by the conclusion.

## 2 Related Work

Given that the inputs are in the form of a vector, which is followed in many other algorithms using multi-layer perceptron, vectorization ruins the structure information of surrounding voxels

or pixels in images. CNNs are regarded as the most outstanding deep models for analysis of images. The brain's visual cortex is the inspirational source for CNNs. CNNs are designed to understand spatial information of images by stacking several convolutional layers to extract more abstract features.<sup>17</sup> Further, in contrast to multi-layer neural networks, there are significantly fewer parameters in CNNs because of shared weights in convolutional layers and the presence of pooling layers.<sup>17</sup>

Despite the initial success of CNNs,<sup>18</sup> their extensive application has not been realized until recent years—various novel approaches and computer systems have been developed to train them efficiently.<sup>17</sup> CNNs attracted great attention after they succeeded in the ImageNet competitions, for which they were successfully employed for a classification task on a database with about one million images with 1000 various classes.<sup>19</sup> This section aims to overview the basic concepts, architectures, and application of CNN models used in AD detection.

## 2.1 CNN Architecture

Typical CNNs consist of several layers, including but not limited to convolutional layers, activation layers, pooling layers, fully connected layers, and a Softmax layer. A CNN is trained as follows: a forward step computes the loss cost between ground-truth labels and predicted outputs and then a backward step applies penalization on learnable parameters. The performance of the CNN depends mainly on the architecture of the layer and the filters' settings, leading researchers to focus on developing different architectures to improve performance. Here the key elements in a CNN architecture are explained.

### 2.1.1 Convolutional layers

A convolutional layer is the first and underlying layer in the architecture, in which the convolution of an input image with a kernel is conducted, and feature maps are produced. Discriminative scale/shift-invariant features of local areas in the image are extracted by the first convolutional layers of deep CNNs. The last convolutional layers enable task-specific classification through the extracted features. The key benefit of convolutional layers is the weight-sharing idea in the same feature map, which decreases parameters and leads to model simplicity.<sup>20</sup>

### 2.1.2 Activation layers

A non-linear activation function such as a Sigmoid, Tanh, or ReLU typically follows convolutional layers to create a feature map corresponding to each filter. Introducing non-linearity allows models to learn complex representations. Old-style activation functions were in the form of a Sigmoid function, but this function's saturation property leads learning algorithms to operate poorly in the neural networks' training process. To solve this problem, ReLU turned out to be useful and popular. It was employed in most of the studies,<sup>13</sup> although Sigmoid or Tanh use is still widespread. ReLU applies an element-wise activation function  $\max(0, x)$ . The training time using ReLU was reported to be considerably faster than that of Tanh and Sigmoid.<sup>21</sup> However, to solve the issue of zero gradients in gradient-based learning algorithms, leaky ReLU was suggested with the output value of  $0.01x$  for the input value of  $x < 0$  instead of forcing negative values to 0.<sup>22</sup>

### 2.1.3 Pooling layers

A pooling layer is the third layer type that appears after convolutional layers. This layer performs the input feature map's down-sampling via the replacement of each area with its average or maximum. Pooling helps reduce the parameter number in a network and, simultaneously, retain the most influential features. Average pooling was often used historically,<sup>23</sup> but it has recently fallen out of favor compared with the max-pooling operation. Max pooling has been shown to have faster convergence and better classification performance; it can select superior invariant features and improve generalization.<sup>22,24</sup>

### 2.1.4 Fully connected layers

The fourth type of layer is the fully connected layer with a performance similar to that of traditional neural networks. It contains a large portion of learnable parameters in a CNN.<sup>20</sup> Following the previously mentioned chain of layers, feature maps are flattened to a vector, in which they are no longer spatially positioned. Then a fully connected layer is added. Fully connected layers are connected to the previous layer's feature elements. They help to understand non-linear associations among the local features extracted by convolutional layers.

### 2.1.5 Softmax layer

The Softmax layer categorizes subjects by selecting the highest anticipated probabilities of each category. The vector's largest value is underlined by the Softmax function, while the lower values are considerably suppressed.

## 2.2 CNN Models in the Literature

Deep learning approaches in AD detection using neuroimaging can be divided into two types: unsupervised and supervised.<sup>13,20</sup> Unsupervised models try to obtain an abstract representation from images. A typical unsupervised method comprises an auto-encoder (AE) for feature extraction and an SVM as the classifier. Nevertheless, supervised methods are more prevalent in the literature than are unsupervised methods. Feature extraction and classification are combined as a single entity in supervised approaches. As discussed previously, CNNs are the most used model for AD detection in deep intelligent systems.

The preference in some studies is to design the CNN structure.<sup>25,26</sup> Nevertheless, employing notable structures such as LeNet,<sup>18</sup> AlexNet,<sup>27</sup> CaffeNet,<sup>28</sup> VGGNet,<sup>29</sup> GoogLeNet,<sup>30</sup> ResNet,<sup>31</sup> DenseNet,<sup>32</sup> and Inception<sup>33</sup> can be helpful. According to previous studies and ImageNet competitions, these models have been successful in image classification.<sup>20</sup> This section explains the employed 2D/3D CNN models in the literature and the weight initialization process.

### 2.2.1 2D CNNs

CNNs were initially proposed to recognize patterns from 2D images. Although 3D CNNs can classify 3D brain scans, they need many parameters compared with 2D CNNs.<sup>34,35</sup> Therefore, using 2D CNNs is more common than using 3D for AD detection using 3D brain scans.<sup>13</sup> By dividing the MRI volumetric data into 2D image slices, 2D information can be extracted from 3D images. Assuming that certain features of interest in 3D MRIs are preserved in 2D images, this process reduces the number of parameters in CNNs. Here different deep models using 2D CNNs are reviewed.

Generally, 2D CNNs capture the middle part of brain scans as the input data and ignore the remainder. Some studies extract gray matter (GM) tissue as the input data. Research studies use standard planes of brain scans, such as the sagittal, coronal, or axial planes. The axial plane is the most widely used plane.<sup>13</sup> Farooq et al. employed axial slices of MRI (GM) so that slices were eliminated from the beginning and end, where there is no information. They implemented 2D CNNs based on GoogLeNet, ResNet-18, and ResNet-152.<sup>36</sup> Valliani and Soni<sup>37</sup> employed the median axial slice of subjects to train ResNet-18. Farooq et al. used 166 axial slices of MRI (GM) to train GoogLeNet, and ResNet-152.<sup>38</sup> Seven 2D CNNs on seven groups of 2D images slices (five mid-axial slices in each group), each consisting of three convolutional layers, was proposed by Luo et al.<sup>39</sup> In this study, a subject was classified as AD when at least one of the classifiers categorized it as AD.

After discarding the first and the last axial slices, assuming them to be without anatomical information, Wu et al.<sup>40</sup> combined every three neighboring slices into an RGB color image to train CaffeNet and GoogLeNet. By removing the last 10 axial slices from MRI (GM) and slices with 0 mean pixels, LeNet and GoogLeNet models were used by Sarraf and Tofghi for AD detection.<sup>41</sup> In two other studies, a sorting mechanism based on entropy was proposed to select the most informative slices from the axial plane of each MRI scan.<sup>42,43</sup> The highest entropy

computed from the histogram was associated with the most informative slices. The slices were then used to train VGGNet-16 and Inception-V4.

Gunawardena et al.<sup>44</sup> used several image slices from the coronal plane to train a 2D CNN with two convolutional layers. This work showed that a brain scan's coronal view covers the essential brain parts related to AD. The coronal view offers a discriminative advantage that was depicted by Wang et al. with DenseNet-121.<sup>45</sup> All planes were used, and the coronal plane was selected as the most accurate. In another study, 20 mid-coronal slices were employed for training a 2D CNN based on VGGNet-16.<sup>21</sup> Sagittal slices were also employed to train a 2D CNN with two convolutional layers<sup>46</sup> or six convolutional layers.<sup>22</sup>

The use of all three planes of 3D brain scans can offer complementary features useful for our classification process. Thus some research considered all image planes. Islam and Zhang<sup>47</sup> designed three 2D CNNs for three views. Each CNN comprised 4 dense blocks (12 convolutional layers in each) and 4 convolutional layers; the final classification was done by majority voting. In another multi-view study, different 2D CNNs (such as GoogLeNet, AlexNet, ResNet, and VGGNet) were employed with and without long short-term memory (LSTM).<sup>48</sup> LSTM is considered a type of recurrent neural network (RNN) with a more complicated structure. No significant differences have been reported among all views, and multi-view models were reported to have higher accuracy than were single-view models.<sup>48</sup> The most critical disadvantage of multi-view approaches is that they could lead to ambiguity in the final decision.

### 2.2.2 3D CNNs

Given that MRI scans are 3D images, and a spatial relationship exists among 2D image slices, using 3D CNNs is the trend. The most direct method for AD detection is to take the entire MRI scan as the input. However, a large number of parameters are involved in training on a small dataset, which could lead to overfitting.<sup>49</sup> In straightforward methods, 3D CNNs with 5<sup>50</sup> and 12<sup>51</sup> convolutional layers were proposed; 3D CNNs pretrained with AEs with 3 convolutional layers<sup>52</sup> were also suggested. Others, based on VGGNet and ResNet,<sup>53</sup> ResNet-18,<sup>54</sup> and ResNet-37,<sup>55</sup> were recently implemented. In another study, the features were combined from multi-scale 3D convolutional AEs with three hidden layers and a 3D CNN with six convolutional layers.<sup>56</sup>

A VGGNet-based 3D CNN was proposed by Tang et al. to reduce the gradient-vanishing impact through an additional shortcut to merge high- and low-level information.<sup>57</sup> A 3D CNN with seven convolutional layers was described by Wegmayr et al.<sup>58</sup> To capture input features on different scales in this work, three different sizes for filters were selected in the first convolutional layer. Dense connections were introduced to 3D CNN for AD detection by Wang et al.<sup>59</sup> It was reported that dense connections could enhance the gradients' propagation throughout the network with insufficient training data.

## 2.3 Transfer Learning

Successful deep learning methods in the classification of natural images have benefited studies of deep learning in the domain of medical images. However, it is still a challenge for researchers because of the low medical image acquisition quality and errors in preprocessing and segmentation; unavailability of a comprehensive dataset (including a vast number of subjects and biomarkers); low between-class variance in different stages of diseases; lack of expert knowledge, especially in identifying regions of interest (ROIs) in the medical images; and complexity of medical images compared with the usual natural images.

Overfitting is a challenging topic in deep learning that may occur because of issues such as a low number of subjects and a large number of learnable parameters. Some techniques are embedded into CNNs to avoid overfitting, such as max pooling and drop-out layers. Max pooling reduces the number of parameters and, subsequently, the dimension of extracted features to control overfitting and reach the invariance to scale, shift, and rotation.<sup>24</sup> Drop-out layers randomly drop neurons at each update of the training phase and force neurons to act independently.<sup>21</sup> Another idea is to discard fully connected layers applied to networks such as SqueezeNet. Removing them reduces the number of learnable parameters compared with VGGNet and reduces overfitting.



In addition to these CNN-related structures, L1 and L2 regularization have proven to prevent overfitting in the literature for training CNN models. A comprehensive overview of regularization methods, their working time, and object was recently studied together with advanced optimization methods on loss function during the training process.<sup>60,61</sup> The overfitting issue is worst when applying 3D CNNs, which require training a large number of parameters.<sup>49</sup> Transfer learning is another workaround to avoid overfitting.

Some studies trained a deep model from scratch; however, it is often not optimal. It is time consuming to train from scratch, and a dataset with adequate size is sometimes not available.<sup>21,49</sup> Datasets for conventional classifications contain millions of images, while MRI datasets contain only a few hundred images. It causes overfitting in the training procedure. Using weights from a pretrained CNN for a specific task and retraining it for a new task through simply fine-tuning the learnable parameters is common. It is possible because general features are extracted by convolutional layers of CNNs, which can be utilized for many tasks. Hence, it is possible to transfer weights of convolutional layers from one application to another. This practice is called “transfer learning.”

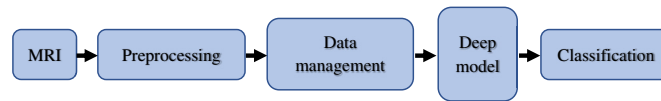
Considering transfer learning from the ImageNet database, 2D CNNs like VGGNet-16,<sup>21</sup> GoogLeNet and ResNet-152,<sup>38</sup> Inception-V4,<sup>62</sup> ResNet-18,<sup>37</sup> VGGNet-16 and Inception-V4,<sup>42</sup> VGGNet-16,<sup>43</sup> CaffeNet and GoogLeNet,<sup>40</sup> and DenseNet-121<sup>45</sup> were implemented for AD detection using MRI scans. In contrast, CNNs such as LeNet and GoogLeNet,<sup>41</sup> GoogLeNet, ResNet-18, and ResNet-152,<sup>36</sup> and 3D CNNs based on ResNet,<sup>55</sup> VGGNet,<sup>57</sup> and VGGNet and ResNet<sup>53</sup> were trained for AD detection from scratch. Transfer learning was reported to be quicker and outperformed training from scratch in 2D CNNs.<sup>37,42</sup> Competition was noted among famous 2D CNNs, like ResNet-152 and GoogLeNet.<sup>36,38</sup> Nevertheless, the performance of Inception-V4, ResNet, and CaffeNet was better than that of VGGNet-16, AlexNet, and GoogLeNet in some studies.<sup>40,42,62</sup> DenseNet outperformed ResNet and LeNet in another study.<sup>63</sup>

Generally, 2D CNNs with transfer learning performed well when a limited number of subjects were available. For more subjects, however, 3D CNNs requires many more learnable parameters to solve such a complex problem. Wang et al.<sup>59</sup> examined the effect of depth in CNNs. Results of this study indicated that shallow and very deep networks do not always yield proper outcomes. One crucial factor is the size of the training set, which significantly influences the performance of classifiers.<sup>59</sup> There are a limited number of AD patients in MRI datasets, which is particularly insufficient for deep models. Therefore, combining datasets, using several scans of the same subject in longitudinal datasets, and data augmentation were followed in some studies.

### 3 Materials and Methods

Training 2D CNNs is easy, but they are not optimized in encoding the spatial information of the 3D MRIs because of the non-existence of the third dimension in convolving kernels.<sup>49</sup> Conversely, 3D CNNs can obtain 3D information from the 3D MRIs. Regardless of the training complexity, they have shown a higher accuracy than 2D CNNs.<sup>57</sup> In Sec. 2.3, the success of transfer learning was explained. Training from scratch has been followed in many classification studies. However, according to the literature review, parameters such as weight and bias can be initialized by training deep models on other related or unrelated tasks. These parameters can extract general features from images, which can assist in our AD detection problem. The dataset from the source domain can be anything with millions of images, such as ImageNet or MNIST. ImageNet is a visual dataset that includes 1000 object categories, such as plants, tools, and many animals.<sup>19</sup> MNIST is a dataset of handwritten digits with 70,000 centered, fixed-size, and gray-scale images.<sup>64</sup> To train models from scratch, the Xavier initialization method is utilized.<sup>65</sup>

The block diagram of an AD detection system is presented in Fig. 1. Preprocessing steps, such as intensity normalization, registration, and tissue segmentation, prepare MRI scans to be used as input data for intelligent systems. After preprocessing, MRI scans are prepared for training the implemented deep models according to each model's structure. Slicing 3D MRI scans to 2D images, identifying ROIs, 2D/3D patch extraction, and resizing are possible data



**Fig. 1** The block diagram of an AD detection system.

management methods. Three approaches for AD detection were implemented in this paper, using CNNs. The slice-based approach extracts 2D image slices from 3D MRIs for data management, using a 2D CNN as the deep model. The sequence-based approach takes the extracted features from the previous approach and feeds them to a RNN. The voxel-based approach takes the entire 3D MRI for data management, using a 3D CNN as the deep model.

As discussed previously, various 2D CNN models have been proposed for AD detection using the ImageNet dataset. In Tables 1 and 2, all employed CNNs in this paper are listed together with their depth, number of layers, number of connections, number of convolution layers, number of fully connected layers, size on disk, number of learnable parameters, and image input size. The network depth is defined as the largest number of sequential convolutional layers and fully connected layers on a path from the input layer to the output layer. CNN structures are the same for 2D and 3D CNNs (i.e., number, type, and order of layers). The only differences are the dimensions of filters and input images, which can be 2D and 3D. Except for LeNet-5, which was trained on grayscale MNIST images, all networks' inputs are RGB images according to the ImageNet dataset's images. To adjust the models to our AD detection problem, the last layer of each 2D CNN model was changed to have two outputs with a weight/bias learning rate factor 10 times more than the previous layers. Applying this learning rate leads to faster training of the new layer than the other layers, which were already trained on the MNIST/ImageNet dataset.

The classification accuracy of the reference dataset (MNIST or ImageNet) is the most generic means to evaluate network accuracy. Networks that perform well on the reference dataset are likewise usually accurate when applied to other image datasets using transfer learning. This assumption is feasible because networks have learned to extract general informative features from images. However, high accuracy on the reference dataset does not guarantee the model's performance on other tasks; hence, trying multiple networks is required. Our LeNet-5 implementation achieved 98.48% accuracy on the MNIST test set after training on its training set. Other CNN models are implemented in MATLAB with the listed accuracies in the final column

**Table 1** Implemented 2D CNNs in our study.

Network	Depth	#Layers	#Connections	#Convolutional layers	#FCs	Size (MB)	#Parameters (millions)	Image input size	Reference accuracy (%)
LeNet-5	5	16	15	3	2	0.23	0.062	32 × 32	98.48
AlexNet	8	25	24	5	3	227	61.0	227 × 227 × 3	54.10
VGG-16	16	41	40	13	3	515	138	224 × 224 × 3	70.29
SqueezeNet	18	68	75	27	0	4.5	1.24	227 × 227 × 3	55.16
ResNet-18	18	71	78	20	1	44	11.7	224 × 224 × 3	69.49
VGG-19	19	47	46	16	3	535	144	224 × 224 × 3	70.42
GoogLeNet	22	144	170	58	1	27	7.0	224 × 224 × 3	66.25
Inceptionv3	48	315	349	94	1	87	23.9	299 × 299 × 3	77.07
ResNet-50	50	177	192	53	1	96	25.6	224 × 224 × 3	74.46
ResNet-101	101	347	379	104	1	167	44.6	224 × 224 × 3	75.96

**Table 2** Implemented 3D CNNs in our study.

Network	Size (MB)	#Parameters (millions)	Image input size
LeNet-5	0.26	0.26	$32 \times 32 \times 32$
ResNet-18	46	34	$224 \times 224 \times 224$
ResNet-50	132	48	$224 \times 224 \times 224$
ResNet-101	204	87	$224 \times 224 \times 224$

of Table 1. Standard ImageNet validation data were utilized to achieve these numbers according to the MATLAB documentation page.<sup>66</sup>

The knowledge from MNIST and ImageNet datasets can be transferred to our AD detection case. However, the main issue here is that these datasets have 2D images, and deep models pretrained on 2D images cannot be useful in a 3D workflow. This paper extends the concept of transfer learning from the 2D natural-images domain to the 3D MRI-images domain. To this end, we consider several approaches, explained in the following section.

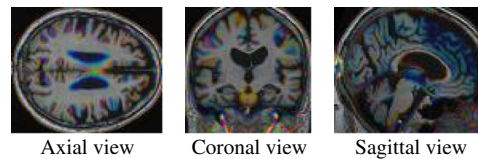
### 3.1 Dataset Preparation

The ADNI (AD Neuroimaging Initiative) study provided the dataset used here. ADNI is the most extensively utilized dataset in studies in this field.<sup>13,67</sup> Its main goal is to determine if it is possible to combine PET, MRI, and other biological markers, neuropsychological and clinical assessments to measure AD progress. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians in developing new treatments, monitoring their effectiveness, and reducing the time and cost of clinical trials. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration, private pharmaceutical companies, and non-profit organizations, as a \$60 million 5-year public-private partnership. Acquisition of these data was performed according to the ADNI acquisition protocol.<sup>67</sup> ADNI subjects aged 55 to 90 years were recruited from more than 50 sites across the United States and Canada. MRI scans of 132 subjects for each class (AD and NC) were employed in this study (only for baseline or screening); the dataset is available online upon request.

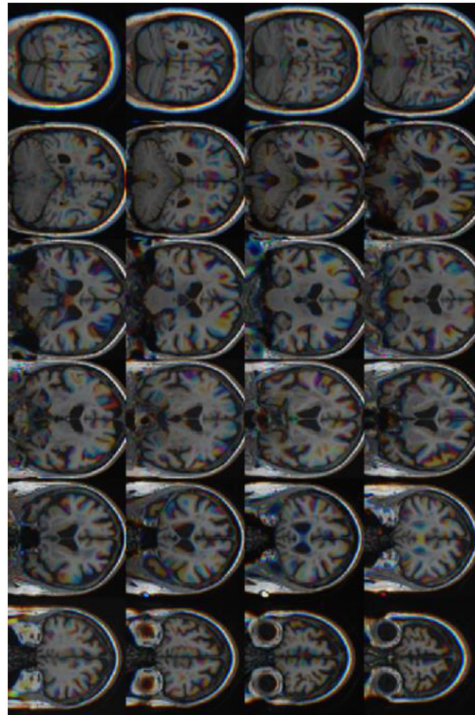
Classification performance is affected by image preprocessing. Among different preprocessing techniques for AD detection, registration and intensity normalization are the techniques most commonly used.<sup>13</sup> Normalization is mapping voxel/pixel intensities of all MRI scans to a common range. Registration is defined as a type of spatial adjustment of MRI scans to a reference anatomical space. Because of the differences among the brains of unique subjects, it is a necessity. Image registration is helpful to standardize MRI patterns. Intensity normalization (zero-center) and registration to Montreal Neurological Institute space<sup>68</sup> were utilized in this study using the SPM12 toolbox.<sup>69</sup>

After preprocessing, the dimension of each MRI was  $79 \times 95 \times 79$ . For 3D CNNs, the entire preprocessed MRI was used for training. For 2D CNNs, 2D slices of every view of MRI (axial, coronal, and sagittal) were extracted by disposing of the first and last image slices. By stacking three adjacent slices as RGB color channels, the remaining MRI slices were formed into 24 RGB coronal images, 19 RGB sagittal images, and 16 RGB axial images. Except for LeNet-5, since all 2D CNN models pretrained on the ImageNet dataset take RGB images as the input, RGB images are required. Some sample images of the ADNI dataset are shown in Figs. 2 and 3. Next, we resized each MRI to fit each 2D/3D CNN's input layer in Tables 1 and 2. Data augmentation was used because the number of patients was insufficient to train a deep structure to improve classification performance.<sup>37,70</sup> Data augmentation methods such as random translation, scaling, reflection, noise injection, rotation, blurring, cropping, and gamma correction are a process that increases data diversity to train models without gathering new data. In 3D approaches,  $\pm 5\%$  scaling and  $\pm 5$  pixel translation were performed on the training set only. In 2D approaches and





**Fig. 2** Sample images of the ADNI dataset.



**Fig. 3** 24 MRI coronal slices of one subject.

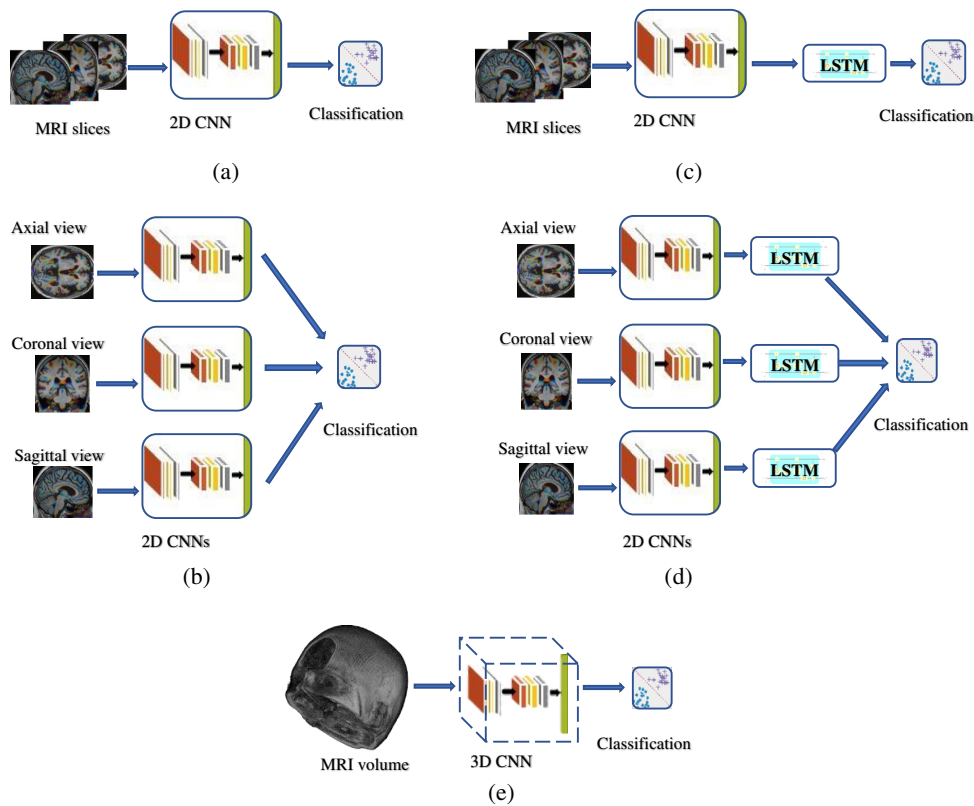
because of brain scans' symmetry, reflecting coronal and axial views was conducted along with the stated data augmentation transformations.

### 3.2 Approach 1: 2D CNNs

For approach 1, 2D CNNs have the ability to extract AD-related discriminative features, such as brain shrinkage, from each image slice, and classification of each subject is conducted based on image slices of that subject. Figure 4(a) indicates our single-view 2D CNN structure. Single-view (coronal, sagittal, and axial) image slices of all subjects on the training set are used to train a 2D CNN. For testing, all image slices of an MRI scan are then classified by the CNN model. The final classification is determined by a majority voting strategy on all image slices. The multi-view configuration follows a similar process, as depicted in Fig. 4(b), except another majority voting strategy is used on all three views to make the final decision.

### 3.3 Approach 2: 2D CNNs + LSTM

This approach aims to understand spatial connections in 2D image slices of MRIs by the combination of 2D CNNs with LSTM. After a preliminary phase of feature extraction with a CNN model on the ADNI database, an LSTM structure is employed to acquire informative features to detect AD on a sequence of image slices. In the single-view mode, after feature extraction by a CNN structure on each MRI slice, an LSTM model is trained on the extracted features from



**Fig. 4** Architectures of the applied approaches: (a) single-view 2D CNN; (b) multi-view CNNs; (c) single-view 2D CNN + LSTM; (d) multi-view CNNs + LSTM; and (e) 3D CNN.

MRI slices of a specific subject on each view, as shown in Fig. 4(c). The features extracted after the second to last layer of the CNN (before the final layer) feed the LSTM model. It is expected that LSTM can comprehend the connections between sequences of images in each subject and make a sequence-based decision based on all input image slices, not individually. The multi-view configuration, as can be observed in Fig. 4(d), also follows this process, with the difference being that a majority voting approach was employed over all three views for the final classification. Two LSTM layers (100 hidden units for each of them) were designed for the LSTM network by trial and error.

### 3.4 Approach 3: 3D CNNs

The whole preprocessed MRIs were used to train 3D CNNs to make voxel-based decisions. According to Table 2, 3D CNNs have many learnable parameters, and training them is computationally expensive. However, we utilized this type of deep model to be able to compare them with 2D CNNs. To build 3D CNNs, we extended the 2D filters of some CNN models in Table 1 in the third dimension to have 3D filters. Any other layers explained in Sec. 2.1 were adjusted according to the new filters. Training 3D CNNs has many learnable parameters compared with 2D CNNs. This issue makes the backpropagation learning process difficult to converge, especially when training from scratch. To transfer the learnable parameters from pretrained 2D CNNs (on MNIST or ImageNet) to 3D CNNs, we duplicated 2D filters (copying them repeatedly) through the third dimension. This is possible because an MRI scan can be converted into a sequence of image slices. To the best of our knowledge, 3D CNNs with transfer learning have not been employed for any classification task. In the training process, we expect that 3D models learn AD-related features in each slice of MRI and understand AD-related patterns across image slices. The 3D models are listed in Table 2, along with the number of learnable parameters and the networks' size on a memory. The number of learnable parameters increases with the extension of the filters' dimensions. A diagram of our 3D structure can be observed in Fig. 4(e).

## 4 Results and Discussion

As previously discussed, the ADNI dataset was utilized in this paper. For our experiments, we utilized 100, 16, and 16 subjects per class to train, validate, and test, respectively. For each subject, one MRI scan was used. Therefore, each set contains the same number of MRI scans for each class. The reason for selecting the same number of samples for each class was to avoid the prediction bias because of an imbalanced dataset. It is possible to access many MRI scans from healthy people to include in our dataset and increase the dataset size. However, the number of subjects with AD is limited in medical datasets. Therefore, to avoid class imbalance and subsequent prediction bias to one of the classes—NC in this case—the same number of MRI scans was selected for each class.

The same training parameters were used for all 2D CNNs that contained mini-batch size = 64, initial learning rate = 0.0003, and L2 regularization = 0.0005. Stochastic gradient descent (SGD) with momentum = 0.9 was the optimizer algorithm with early stopping according to the validation set. However, mini-batch size = 8 was used for 3D CNNs because of the available computational resources. Moreover, we utilized the same training parameters for all LSTM models that contained mini-batch size = 64, initial learning rate = 0.01, and L2 regularization = 0.0001. SGD was the optimizer, with momentum = 0.9, and the maximum number of epochs was 50, with early stopping according to the validation set.

The training set was shuffled in every epoch of the training process, and simultaneously, data augmentation was implemented for CNN models. Consequently, a CNN model observed altered input data in a different arrangement in each epoch. MATLAB deep learning toolbox was used in this paper to implement and train our models on an NVIDIA DGX station. This system has 4 GPUs with 32 GB memory for each and 256 GB RAM. However, we utilized only one GPU to reduce the chance of parallel computational issues in multi-core systems. The same training, validation, and test sets were employed for all models to ensure objective comparisons. In the following figures and tables, “TL” and “Sc” denote transfer learning and training from scratch, respectively.

The performance of classification by utilized CNNs is shown in Table 3. In this table, accuracy refers to the percentage of correctly classified test subjects. Sensitivity refers to the percentage of evaluated test subjects suffering from AD who were correctly classified as such, while specificity is the percentage of evaluated healthy test subjects correctly classified as healthy. Sensitivity is often more important for screening tests in medical tasks.

2D CNN made 24, 16, and 19 decisions on coronal, axial, or sagittal views in the single-view mode of the 2D approach. The majority voting on the whole pile of images for each MRI view of a particular subject informed the final decision. In the multi-view mode, CNNs determined all image slices of the related view separately. A majority voting approach on all three views of a particular subject determined the final decision. Both transfer learning and training from scratch were employed in this approach. Transfer learning yielded about 2% accuracy enhancement on average. Generally, transfer learning helps deep models avoid overfitting.

SqueezeNet and ResNet-18 performed best on our dataset in this approach. The core idea behind ResNet is presenting a “shortcut connection” that skips some layers to avoid the vanishing gradient problem. The main notion of SqueezeNet is to use  $1 \times 1$  (point-wise) filters to replace  $3 \times 3$  filters to have fewer computations and down-sample late in the model to retain a large feature map. Multi-view ResNet-18 using transfer learning had 84.38% accuracy, 87.5% sensitivity, and 81.25% specificity. Multi-view SqueezeNet using transfer learning had 90.62% accuracy, 81.25% of sensitivity, and 100% specificity. The same performance was achieved by SqueezeNet + LSTM.

In the 2D CNNs + LSTM approach, a CNN was assigned to extract features from an image slice, and an LSTM model was assigned to locate the connection among sequences of image slices for a subject. After extracting features via a CNN structure on each MRI slice, the LSTM model made a decision based on the extracted features from MRI slices of a specific subject on each view in the single-view mode. In other words, the LSTM model made a single decision on either the sagittal, coronal, or axial view. In the multi-view mode, an LSTM model independently decided on all slices of a view. The final decision was reached by a majority voting approach on all three LSTM models allocated to each view.

**Table 3** Classification accuracy of implemented deep models (AD versus NC).

CNN model	2D CNNs (Sc)					2D CNNs (TL)					2D CNNs (TL) + LSTM					3D CNN		
	Axial	Coronal	Sagittal	Multi-view	Axial	Coronal	Sagittal	Multi-view	Axial	Coronal	Sagittal	Multi-view	Axial	Coronal	Sagittal	Multi-view	Sc	TL
LeNet-5	78.12	78.12	84.38	84.38	81.25	81.25	81.25	84.38	84.38	81.25	81.25	84.38	84.38	84.38	84.38	87.50	81.25	81.25
AlexNet	68.75	78.12	84.32	78.12	75.00	81.25	84.18	84.38	84.38	81.25	81.25	84.38	90.62	90.62	81.25	84.38	—	—
VGGNet-16	71.88	84.38	84.38	81.25	84.38	81.25	84.38	84.38	81.25	81.25	84.38	84.38	81.25	84.38	81.25	84.38	—	—
SqueezeNet	78.12	78.12	78.12	81.25	84.38	81.25	87.50	90.62	87.50	81.25	90.62	90.62	87.50	90.62	90.62	90.62	—	—
ResNet-18	81.25	87.50	84.38	84.38	87.50	81.25	81.25	84.38	87.50	81.25	81.25	84.38	87.50	81.25	78.12	81.25	68.75	96.88
VGGNet-19	71.88	81.25	87.50	81.25	84.38	81.25	84.38	78.12	81.25	84.38	78.12	81.25	81.25	84.38	87.50	84.38	—	—
GoogleNet	75.00	78.12	75.00	78.12	71.88	81.25	84.38	81.25	87.50	81.25	84.38	81.25	87.50	87.50	75.00	87.50	—	—
Inceptionv3	78.12	78.12	81.25	78.12	81.25	78.12	78.12	78.12	81.25	78.12	78.12	78.12	75.00	84.38	78.12	81.25	—	—
ResNet-50	78.12	81.25	81.25	78.12	81.25	81.25	78.12	87.50	87.50	81.25	81.25	87.50	87.50	87.50	81.25	87.50	62.50	90.62
ResNet-101	78.12	78.12	78.12	81.25	81.25	78.12	75.00	78.12	81.25	78.12	78.12	78.12	81.25	81.25	81.25	78.12	—	93.75

Because of the better performance compared with training from scratch in the 2D CNNs approach, only transfer learning was utilized in this approach. The performance of models was compared, with and without LSTM. Since LSTM models incorporated temporal dependencies for the classification, almost 2% accuracy enhancement on average was observed. The results in Table 3 indicate there were not any significant differences among all views. However, multi-view models were more robust to some extent and showed greater accuracy compared with single-view models.

In the 3D CNNs approach, each CNN made a decision based on the whole MRI volume. A 3D CNN was responsible for feature extraction from each image slice and finding the relation between sequences of images for each subject simultaneously. Training some models was impossible using available hardware. We implemented and trained LeNet-5, ResNet-18, ResNet-50, and ResNet-101 in the 3D workflow. Both transfer learning and training from scratch were employed in this approach. Because of the computational requirements, ResNet-101 was trained using 4 GPUs simultaneously with parallel processing. In this situation, the 3D ResNet-101 did not converge after 2000 epochs while training from scratch. According to the results in Table 3, transferring knowledge from ImageNet to 3D CNNs improved the results significantly (96.88% accuracy on 3D ResNet-18) compared with other approaches. However, training 3D CNNs from scratch had a poor performance on our MRI dataset. This is because 3D CNNs have many learnable parameters, which makes the training process challenging. The situation was the worst for deeper models like ResNet-101. As shown in Table 3, transfer learning and training from scratch yielded the same performance for LeNet-5. However, transfer learning overtook training from scratch in accuracy in deeper models like ResNet-18 and ResNet-50. For ResNet-101, the training process could not converge for our available hardware resources when training from scratch.

The training time of different CNNs and different approaches in both transfer learning and training from scratch modes is shown in Fig. 5. Further, the number of required epochs is shown in Fig. 6. Only the results of multi-view approaches are given in both figures. All models have the

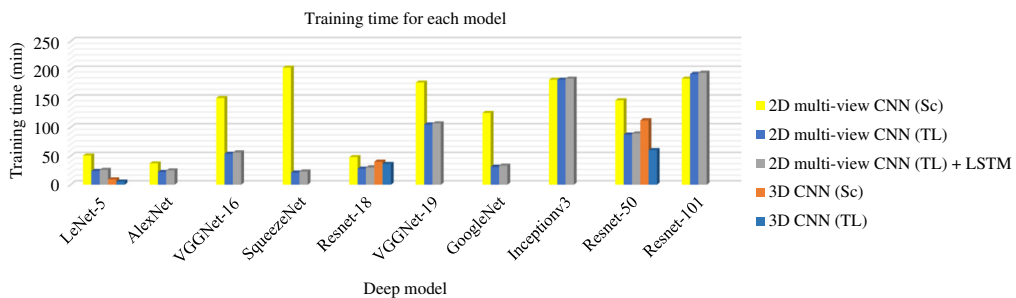


Fig. 5 Training time for implemented deep models in different approaches.

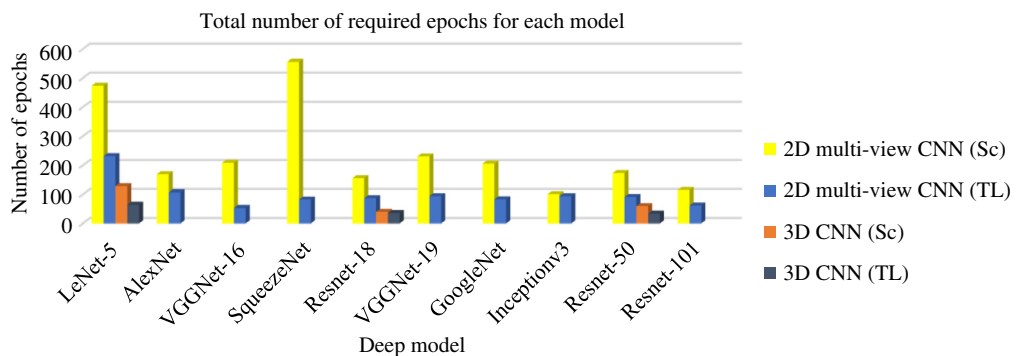
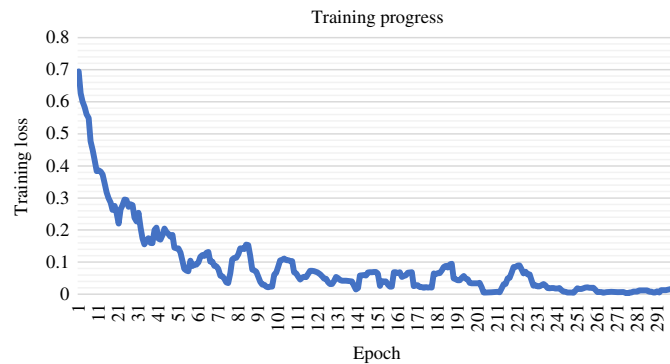


Fig. 6 The number of required epochs for implemented deep models in different approaches.





**Fig. 7** SGD optimizer's training loss for 3D ResNet-18 with transfer learning with moving average filter of 10.

same training parameters, except mini-batch size, 64 for 2D CNNs and 8 for 3D CNNs. Also the same hardware/software resources were used to train the models. Therefore, we omitted 3D ResNet-101 with transfer learning because it did not converge on a single GPU. As expected, deeper models or models with more learnable parameters took longer to train. For example, the training progress—SGD optimizer's training loss versus epochs—of 3D ResNet-18 with transfer learning is shown in Fig. 7. Training from scratch was more time consuming and required more epochs compared with transfer learning. Comparing transfer learning in 2D and 3D CNNs, approximately the same training time was obtained in our experiments, though 3D CNNs required fewer epochs. Deeper models need fewer epochs in their training process; however, each epoch was more time consuming.

## 5 Conclusions

This paper has presented the design, implementation, and experiments of several MRI-based AD detection approaches using CNNs. In the first approach, a 2D CNN was trained on MRI slices in the single-view mode. Then for classification, the CNN model made decisions on all slices of one patient on the particular view. A majority voting mechanism was applied in multi-view mode to make the final decision on three views of an MRI volume. In the second approach, an LSTM model was used to classify a sequence of MRI slices in multi-view and single-view modes. In these approaches, there were not any significant differences among all views. However, the multi-view models were slightly more robust and accurate compared with single-view models. In the third approach, a 3D CNN was employed to classify MRI volumes, each in a single decision.

In general, SqueezeNet and ResNet-18 have the best performance on our dataset. Transfer learning was used in the 2D CNN approach, yielding  $\sim 2\%$  accuracy enhancement on average compared with training from scratch. Further, transferring knowledge from ImageNet to the MRI dataset from ADNI using 3D CNNs considerably improved the results compared with training from scratch. To the best of our knowledge, 3D CNNs with transfer learning have not been employed for any classification task. These 3D models can be employed for classifying 3D images and videos. Comparing different approaches, 2D CNN models with LSTM achieved  $\sim 2\%$  accuracy enhancement on average compared with 2D CNN models alone. Also 3D CNNs with transfer learning considerably improved the results to 96.88% accuracy, 100% sensitivity, and 94.12% specificity. However, training 3D CNNs from scratch performed poorly on our MRI dataset. In conclusion, deep learning methods can be utilized for accurate detection of AD. However, the necessity of a large dataset is a weakness of these approaches.

## Disclosures

The authors have no conflicts of interest to declare, financial or otherwise, relevant to this article's content.

## Acknowledgments

Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI), National Institutes of Health (Grant No. U01 AG024904), and DOD ADNI (Department of Defense Award No. W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; Johnson & Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provide funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (<https://www.fnih.org>). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory of Neuro Imaging (LONI) at the University of Southern California.

## Code, Data, and Materials Availability

Data used in the preparation of this article were obtained from the ADNI study. The data are available online upon request (<http://adni.loni.usc.edu/data-samples/access-data/>). The models implemented in this study are available in MathWorks file exchange (<https://au.mathworks.com/matlabcentral/profile/authors/12814523?utf8=%E2%9C%93&detail2=&detail=fileexchange>).

## References

1. Alzheimer's Association, "2020 Alzheimer's disease facts and figures," *Alzheimer's Dementia* **16**(3), 1–94 (2020).
2. M. Prince, R. Bryce, and C. Ferri, *World Alzheimer Report 2011: The Benefits of Early Diagnosis and Intervention*, Alzheimer's Disease International (2011).
3. S. Klöppel et al., "Diagnostic neuroimaging across diseases," *Neuroimage* **61**(2), 457–463 (2012).
4. Dementia Australia, *The Dementia Guide for People Living with Dementia, their Families and Carers*, Dementia Australia, <https://www.dementia.org.au/sites/default/files/resources/The-Dementia-Guide-2019-v2.pdf> (2020).
5. M. Prince et al., "The global prevalence of dementia: a systematic review and metaanalysis," *Alzheimer's Dementia* **9**(1), 63–75.e2 (2013).
6. Australian Bureau of Statistics, "Causes of Death, Australia, 2019," Australian Bureau of Statistics, <https://www.abs.gov.au/statistics/health/causes-death/causes-death-australia/latest-release> (2020).
7. F. Mangialasche et al., "Alzheimer's disease: clinical trials and drug development," *Lancet Neurol.* **9**(7), 702–716 (2010).
8. S. Paquerault, "Battle against Alzheimer's disease: the scope and potential value of magnetic resonance imaging biomarkers," *Acad. Radiol.* **19**(5), 509–511 (2012).
9. J. P. Lerch et al., "Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls," *Neurobiol. Aging* **29**(1), 23–30 (2008).
10. E. Gerardin et al., "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," *Neuroimage* **47**(4), 1476–1486 (2009).

11. Y. Kazemi and S. K. Houghten, "A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Computat. Biol.*, pp. 1–8 (2018).
12. M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: overview, challenges and the future," in *Classification in BioApps*, N. Dey, A. Ashour, and S. Borra, Eds., pp. 323–350, Springer (2018).
13. A. Ebrahimihafez, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review," *Comput. Methods Prog. Biomed.* **187**, 105242 (2020).
14. L. Deng and D. Yu, "Deep learning: methods and applications," *Found. Trends Signal Process.* **7**(3–4), 197–387 (2014).
15. J. Ker et al., "Deep learning applications in medical image analysis," *IEEE Access* **6**, 9375–9389 (2018).
16. D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).
17. G. Litjens et al., "A survey on deep learning in medical image analysis" *Med. Image Anal.* **42**, 60–88 (2017).
18. Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).
19. O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vision* **115**(3), 211–252 (2015).
20. Y. Guo et al., "Deep learning for visual understanding: a review," *Neurocomputing* **187**, 27–48 (2016).
21. C. D. Billones et al., "DemNet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment," in *Proc. IEEE Region 10 Conf.*, pp. 3724–3727 (2016).
22. S.-H. Wang et al., "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling," *J. Med. Syst.* **42**(5), 85 (2018).
23. A. Karwath et al., "Convolutional neural networks for the identification of regions of interest in PET scans: a study of representation learning for diagnosing Alzheimer's disease," in *Proc. Conf. Artif. Intell. Med. Europe*, pp. 316–321 (2017).
24. D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Int. Conf. Artif. Neural Networks (ICANN)*, Springer, p. 92–101 (2010).
25. S. Bringas et al., "Alzheimer's disease stage identification using deep learning models," *J. Biomed. Inf.* **109**, 103514 (2020).
26. M. Raju et al., "Multi-class diagnosis of Alzheimer's disease using cascaded three dimensional-convolutional neural network," *Phys. Eng. Sci. Med.* **43**, 1219–1228 (2020).
27. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, p. 1097–1105 (2012).
28. Y. Jia et al., "Caffe: convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, pp. 675–678 (2014).
29. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition" arXiv:1409.1556 (2014).
30. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–9 (2015).
31. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).
32. G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, Vol. 1(2), pp. 4700–4708 (2017).
33. C. Szegedy et al., "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st Assoc. for the Adv. Artif. Intell. Conf. Artif. Intell.*, Vol. 4, p. 12 (2017).
34. D. Cheng and M. Liu, "Combining convolutional and recurrent neural networks for Alzheimer's disease diagnosis using PET images," in *Proc. IEEE Int. Conf. Imaging Syst. Tech.* pp. 1–5 (2017).

35. M. Liu, D. Cheng, and W. Yan, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Front. Neuroinf.* **12**, 35 (2018).
36. A. Farooq et al., "A deep CNN based multi-class classification of Alzheimer's disease using MRI," in *Proc. IEEE Int. Conf. Imaging Syst. Tech.*, pp. 1–6 (2017).
37. A. Valliani and A. Soni, "Deep residual nets for improved Alzheimer's diagnosis," in *Proc. 8th ACM Int. Conf. Bioinf., Comput. Biol., and Health Inf.*, pp. 615–615 (2017).
38. A. Farooq et al., "Artificial intelligence based smart diagnosis of Alzheimer's disease and mild cognitive impairment," in *Proc. Smart Int. Cities Conf. (ISC2)*, pp. 1–4 (2017).
39. S. Luo, X. Li, and J. Li, "Automatic Alzheimer's disease recognition from MRI data using deep learning method," *J. Appl. Math. Phys.* **5**(09), 1892–1898 (2017).
40. C. Wu et al., "Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks," *Quant. Imaging Med. Surg.* **8**(10), 992–1003 (2018).
41. S. Sarraf and G. Tofghi, "Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks," arXiv:1607.06583 (2016).
42. M. Hon and N. Khan, "Towards Alzheimer's disease classification through transfer learning," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, pp. 1166–1169 (2017).
43. R. Jain et al., "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognit. Syst. Res.* **57**, 147–159 (2019).
44. K. Gunawardena, R. Rajapakse, and N. Kodikara, "Applying convolutional neural networks for pre-detection of Alzheimer's disease from structural MRI data," in *Proc. 24th Int. Conf. Mechatron. Mach. Vision in Pract.*, pp. 1–7 (2017).
45. J. Islam and Y. Zhang, "Deep convolutional neural networks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using 3D brain MRI," in *Proc. Int. Conf. Brain Inf.*, pp. 359–369 (2018).
46. J. M. Ortiz-Suárez, R. Ramos-Pollán, and E. Romero, "Exploring Alzheimer's anatomical patterns through convolutional networks," *Proc. SPIE* **10160**, 10160Z (2017).
47. J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Inf.* **5**(2), 1–14 (2018).
48. A. Ebrahimi-Ghahnavieh, S. Luo, and R. Chiong, "Transfer learning for Alzheimer's disease detection on MRI images," in *IEEE Int. Conf. Industry 4.0, Artif. Intell., and Commun. Technol.*, pp. 133–138 (2019).
49. M. Liu et al., "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics* **16**(3–4), 295–308 (2018).
50. K. Bäckström et al., "An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images," in *Proc. IEEE 15th Int. Symp. Biomed. Imaging*, pp. 149–153 (2018).
51. S. Basaia et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clin.* **21**, 101645 (2019).
52. E. Hosseini-Asl, R. Keynto, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," in *Proc. IEEE Int. Conf. Image Process.*, pp. 126–130 (2016).
53. S. Korolev et al., "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imaging*, pp. 835–838 (2017).
54. A. Ebrahimi, S. Luo, and R. Chiong, "Introducing transfer learning to 3D ResNet-18 for Alzheimer's disease detection on MRI images," in *35th Int. Conf. Image and Vision Comput.*, New Zealand, p. 1–6 (2020).
55. H. Karasawa, C.-L. Liu, and H. Ohwada, "Deep 3D convolutional neural network architectures for Alzheimer's disease diagnosis," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, pp. 287–296 (2018).
56. F. Li, D. Cheng, and M. Liu, "Alzheimer's disease classification based on combination of multi-model convolutional networks," in *Proc. IEEE Int. Conf. Imaging Syst. Tech.*, p. 1–5 (2017).
57. H. Tang et al., "A fast and accurate 3D fine-tuning convolutional neural network for Alzheimer's disease diagnosis," in *Proc. Int. CCF Conf. Artif. Intell.*, pp. 115–126 (2018).

58. V. Wegmayr, S. Aitharaju, and J. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," *Proc. SPIE* **10575**, 10575S (2018).
59. H. Wang et al., "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," *Neurocomputing*, **333**, 145–156 (2019).
60. Q. Zheng et al., "Layer-wise learning based stochastic gradient descent method for the optimization of deep convolutional neural network," *J. Intell. Fuzzy Syst.* **37**(4), 5641–5654 (2019).
61. Q. Zheng et al., "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process," *IEEE Access* **6**, 15844–15869 (2018).
62. J. Islam and Y. Zhang, "A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data," in *Proc. Int. Conf. Brain Inf.*, pp. 213–222 (2017).
63. F. Li and M. Liu, for the Alzheimer's Disease Neuroimaging Initiative, "Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks," *Comput. Med. Imaging Graphics* **70**, 101–110 (2018).
64. Y. LeCun, C. Cortes, and C. J. Burges, "The MNIST database of handwritten digits, 1998" 10, 34, <http://yann.lecun.com/exdb/mnist> (1998).
65. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Stat.*, pp. 249–256 (2010).
66. MathWorks, "Pretrained deep neural networks," 2020, <https://mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html>.
67. C. R. Jack, Jr. et al., "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *J. Magn. Reson. Imaging* **27**(4), 685–691 (2008).
68. V. Fonov et al., "Unbiased average age-appropriate atlases for pediatric studies," *Neuroimage* **54**(1), 313–327 (2011).
69. W. D. Penny et al., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Elsevier (2011).
70. S. Esmaeilzadeh et al., "End-to-end Alzheimer's disease diagnosis and biomarker identification," in *Proc. Int. Workshop Mach. Learn. Med. Imaging*, pp. 337–345 (2018).

**Amir Ebrahimi** received his master's degree in digital electronics from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2012. Currently, he has been a PhD student in information technology since 2018 at the University of Newcastle, Callaghan, New South Wales, Australia. He has industrial experience in software/hardware programming for several companies. His current research interests are image processing, computer vision, machine learning, and deep learning.

**Suhui Luo** received his bachelor's and master's degrees in electrical engineering from Nanjing University of Posts and Telecommunications and his PhD from the University of Sydney. He is currently an associate professor in information technology at the University of Newcastle. He has conducted studies in areas ranging from medical imaging for computer-aided diagnoses to computer vision for intelligent driving systems and machine learning for enhancing cybersecurity. His main research interests include machine learning and cyber security.