IMAGE & SIGNAL PROCESSING

A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks

Farheen Ramzan[1] · Muhammad Usman Ghani Khan[1] · Asim Rehmat[1] · Sajid Iqbal[1,2] · Tanzila Saba[3] · Amjad Rehman[4] · Zahid Mehmood[5]

Abstract

Alzheimer's disease (AD) is an incurable neurodegenerative disorder accounting for 70%–80% dementia cases worldwide. Although, research on AD has increased in recent years, however, the complexity associated with brain structure and functions makes the early diagnosis of this disease a challenging task. Resting-state functional magnetic resonance imaging (rs-fMRI) is a neuroimaging technology that has been widely used to study the pathogenesis of neurodegenerative diseases. In literature, the computer-aided diagnosis of AD is limited to binary classification or diagnosis of AD and MCI stages. However, its applicability to diagnose multiple progressive stages of AD is relatively under-studied. This study explores the effectiveness of rs-fMRI for multi-class classification of AD and its associated stages including CN, SMC, EMCI, MCI, LMCI, and AD. A longitudinal cohort of resting-state fMRI of 138 subjects (25 CN, 25 SMC, 25 EMCI, 25 LMCI, 13 MCI, and 25 AD) from Alzheimer's Disease Neuroimaging Initiative (ADNI) is studied. To provide a better insight into deep learning approaches and their applications to AD classification, we investigate ResNet-18 architecture in detail. We consider the training of the network from scratch by using single-channel input as well as performed transfer learning with and without fine-tuning using an extended network architecture. We experimented with residual neural networks to perform AD classification task and compared it with former research in this domain. The performance of the models is evaluated using precision, recall, f1-measure, AUC and ROC curves. We found that our networks were able to significantly classify the subjects. We achieved improved results with our fine-tuned model for all the AD stages with an accuracy of 100%, 96.85%, 97.38%, 97.43%, 97.40% and 98.01% for CN, SMC, EMCI, LMCI, MCI, and AD respectively. However, in terms of overall performance, we achieved state-of-the-art results with an average accuracy of 97.92% and 97.88% for off-the-shelf and fine-tuned models respectively. The Analysis of results indicate that classification and prediction of neurodegenerative brain disorders such as AD using functional magnetic resonance imaging and advanced deep learning methods is promising for clinical decision making and have the potential to assist in early diagnosis of AD and its associated stages.

Keywords Alzheimer's disease · Functional magnetic resonance imaging (fMRI) · Diagnosis · Multi-class · Classification · Deep learning · Residual neural networks

---

✉ Sajid Iqbal
sajid.iqbal@bzu.edu.pk

[1] Department of Computer Science and Engineering, University of Engineering and Technology (UET), Lahore, Pakistan

[2] Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan

[3] College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia

[4] College of Computer and Information Sciences, Prince Sultan University , Riyadh 11586, Saudi Arabia

[5] Department of Computer Engineering, University of Engineering and Technology, Taxila 47050, Pakistan

Introduction

Alzheimer's disease (AD) is an incurable neurodegenerative disorder with an unrelenting progression, affecting memory and cognitive abilities of a person. AD pathogenesis is believed to be triggered due to the overproduction of amyloid-β (Aβ) [1, 2] and hyper phosphorylation of tau [3, 4] protein. This results in accumulation of Aβ plaques and tau neurofibrillary tangles, disrupting the nucleocytoplasmic transport between neurons leading to cell death. Initially, the hippocampus region [5] is affected by the disease. Since the hippocampus is associated with memory and learning, therefore, memory loss is one of the early symptoms of AD. The exact cause of AD is unknown and, in some cases, it is believed to be genetic.

Dementia is a general term used for memory-related neurological disorders; however, Alzheimer's disease is the most common type of dementia. According to the World Alzheimer's Report 2015 [6], approximately 50 million people are suffering from dementia where AD accounts for 70–80% of cases. It has been estimated that by 2050, 131.5 million people will be suffering from AD worldwide. The rate of prevalence of AD globally is alarming that in every 3 s a person falls prey to it [7]. Also, AD gets the 6th place among the leading causes of death in the aging population. The total estimated cost to combat the disease worldwide in 2015 was $818 billion [6]. The cost on AD is reaching up to trillion dollars by 2019 and this cost is estimated to reach up to 2 trillion dollars by 2030 [7]. The percentage of people with AD increases with age: 3% people of age 65–74, 17% people of age 75–84 and 32% people of age 85 or older have Alzheimer's disease [8].

AD is a progressive disorder that starts with mild symptoms and gets worse progressively. Researchers believe that Alzheimer's related brain changes may begin 20 years or more before any symptoms of AD appears [8]. There are various stages of the disease, that are termed as: cognitively normal (CN), significant memory concern (SMC), early mild cognitive impairment (EMCI), mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD) [9]. CN subjects show normal signs of aging with no signs of depression and dementia. In SMC, the subject has normal cognitive functions but show slight memory concerns. Subjects retain older memories by facing difficulties in forming and retaining new ones. EMCI, MCI, and LMCI are the stages during which disease has progressed and start affecting daily life activities. The patient shows symptoms including loss of motor functions, speech difficulties, memory concerns and ability to read and write. Levels of MCI are determined by a Wechsler Memory Scale (WMS) neuropsychological test [10]. AD is the advanced and final stage of the disease leading to death. There is no cure for AD but the right medication and proper care can help to manage symptoms. Although AD can't be cured however cognitive decline can be slowed down in the early stages of the disease. Therefore, early-stage detection of AD is highly desirable in order to increase the quality of patients' lives and to improve the developments in drug trials.

In recent years, the growth of neurodegenerative disorders such as AD has gained much interest from researchers worldwide to develop high performing methods for diagnosis, treatment, preventive therapies, and target drug discovery by studying the pathological processes associated with each stage of AD [11]. The rate of progression of AD varies from patient to patient and individuals may show different symptoms in a certain disease stage [12]. That makes classification of AD stages a challenging task for diagnosis and prognosis. New research developments have made it possible to diagnose AD using advanced diagnostic tools and biomarker tests. Various invasive and non-invasive neuroimaging technologies are used for AD diagnoses such as structural Magnetic Resonance Imaging (sMRI), functional MRI (fMRI), Positron Emission Tomography (PET), Computerized Tomography (CT), Electroencephalography (EEG), Magnetoencephalography (MEG) and Cerebrospinal fluid (CSF) biomarkers. The neuroimaging data acquired from these technologies are used for providing a computer-aided diagnosis to aid physicians and clinicians in order to improve health-care systems for AD.

Recently, resting-state functional magnetic resonance imaging (rs-fMRI) is being increasingly utilized to study the pathogenesis of AD and its stages. rs-fMRI is non-invasive and has shown great applicability to map how AD spreads in the living brain. Various studies have tested the accuracy of AD-related fMRI measurements and found positive predictability of disease related to cognitive decline [13, 14]. Resting-State fMRI captures the changes in blood oxygenation levels of subjects in the rest state. Therefore, brain regions affected by neurodegeneration show different patterns of blood oxygenation levels. Also, normal healthy subjects and AD patients show different patterns of blood oxygenation activities, which may directly be related to disease pathology and can be used to distinguish various stages of AD for diagnostic decision making. Various researchers have targeted the problem of computer-aided AD classification and diagnosis from rs-fMRI data. In this respect, one of the earlier methods used for AD diagnosis was based on using statistical techniques such as the General Linear Model (GLM) that have been applied for the analysis of fMRI [15, 16]. This method detects activated brain regions by performing a correlation between the template model and fMRI time sequences. GLM is a time-consuming algorithm that uses voxel as a parameter of measurement and is single variate [15].

Table 1 A literature review of Alzheimer's disease classification using deep learning techniques

Study	Modality	Dataset	Sample Size	Methods	Classification Accuracy (%)						
					CN/AD	CN/MCI	AD/MCI	MCI-C/MCI-NC	AD/CN/MCI	AD/CN/cMCI/MCI	sMCI/pMCI
Suk et al. (2013) [23]	PET, MRI, CSF	ADNI	51 AD, 99 MCI (43 MCI-C, 56 MCI-NC)	SAE, SVM	95.9	85	–	75.8	–	–	–
Suk et al. (2014) [13]	PET+MRI	ADNI	93 AD, 204 MCI, 101 CN	DBM	95.4	85.7	–	75.92	–	–	–
Zhu et al. (2014) [24]	PET, MRI, CSF	ADNI	51 AD, 52 CN, 99 MCI	MSLF, SVM	95.9	82	–	72.6	–	–	–
Li et al. (2015) [25]	PET, MRI, CSF	ADNI	51 AD, 99 MCI (43 MCI-C, 56 MCI-NC), 52 CN	PCA, MDLD, RBM, SVM	91.4	77.4	70.1	57.4	–	–	–
Liu et al. (2015) [26]	PET, MRI	ADNI	80 AD, 160 MCI-C, 214 MCI-NC, 204 CN	SAE, SVM, Zeromask	91.4	82.1	–	–	53.8	–	–
Payan et al. (2015) [27]	MRI	ADNI	755 AD, 755 MCI, 755 CN	SPAE, 3D CNN	95.4	92.1	86.8	–	89.4	–	–
Liu et al. (2016) [28]	MRI	ADNI	97 AD, 128 NC, 234 MCI	ISML, MVFE, SCFE, SVM	93.8	89.1	–	–	–	–	–
Zu et al. (2016) [29]	PET, MRI	ADNI	51 AD, 99 MCI, 52 CN	LAMTFS, SVM	96.0	80.3	–	69.78	–	–	–
Sarraf et al. (2016) [14]	fMRI	ADNI	28 AD, 15 CN	LeNet-5	96.85	–	–	–	–	–	–
S. Sarraf et al. (2016) [30]	MRI, fMRI	ADNI	fMRI (52 AD, 92 CN), MRI (211 AD, 91 CN)	LeNet, GoogleNet	fMRI: 99.9, MRI: 98.84	–	–	–	–	–	–
F. Li et al. (2017) [31]	MRI	ADNI	199 AD, 229 CN	3D CNN	88.31	–	–	–	–	–	–
N. Amoroso et al. (2018) [32]	MRI	ADNI	60 AD, 60 NC, 60 cMCI, 60 MCI	RF, DNN, Fuzzy logic	–	–	–	–	–	38.8	–
M. Liu et al.(2018) [33]	MRI, PET	ADNI	93 AD, 204 MCI (76 pMCI, 128 sMCI), 100 CN	2D CNN, 3D CNN	93.26	74.34	–	–	–	–	–
C. Yang et al. (2018) [34]	MRI	ADNI	47 AD, 56 CN	3D CNN, 3DVGGNet, 3DResNet	76.6	–	–	–	–	–	–
D. Lu et al. (2018) [11]	MRI, PET	ADNI	304 CN, 409 sMCI, 112 pMCI, 226 AD	Multiscale DNN	93.58	–	–	–	–	–	81.55
S.H. Wang et al. (2018) [35]	MRI	Local Hospitals, OASIS	98 AD, 98 CN	CNN	97.65	–	–	–	–	–	–
A. Khvostikov et al. (2018) [36]	MRI, DTI	ADNI	48 AD, 108 MCI, 58 CN	3D CNN	0.967	0.658	0.8	–	–	–	–
J. Shi et al. (2018) [37]	MRI, PET	ADNI	51 AD, 99 MCI (43 MCI-C, 56 MCI-NC), 52 CN	MM-SDPN, SVM	97.13	87.24	–	78.88	–	57.00	–

*Abbreviations: ADNI = Alzheimer's Disease Neuroimaging Initiative, SAE = Stacked Autoencoder, SVM = Support Vector Machine, AD = Alzheimer's Disease, MCI = Mild Cognitive Impairment, MCI-C = MCI converter, MCI-NC = MCI Non-converter, DBM = Deep Boltzmann Machine, MSLF = Matrix-Similarity based Loss Function, PCA = Principal Component Analysis, MDLD = Multitask Deep Learning With Dropout, RBM = Restricted Boltzmann Machine, SPAE = Sparse Autoencoder, MVFE = MultiView Feature Extraction, SCFE = Subclass Clustering-based Feature Selection, pMCI = Progressive MCI, sMCI = Stable MCI, ISML = Influence of Subclass Number, LAMTFS = Label-aligned Multi-task Feature Selection, RF = Random Forest, DNN = deep Neural Network, OASIS = Open Access Series of Imaging Studies, MM-SDPN = Multimodal Stacked Deep Polynomial Network

Table 2   Alzheimer's disease classification using fMRI and deep learning techniques

Study	Modality	Dataset	Sample Size	Methods	Targeted Alzheimer's Disease Stages					
					CN	SMC	EMCI	MCI	LMCI	AD
Sarraf et al. (2016) [14]	fMRI	ADNI	28 AD, 15 CN	LeNet-5	✓	–	–	–	–	✓
S. Sarraf et al. (2016) [30]	MRI, fMRI	ADNI	fMRI (52 AD, 92 CN), MRI (211 AD, 91 CN)	LeNet, GoogleNet	✓	–	–	–	–	✓
Y. Kazemi et al. (2018) [39]	fMRI	ADNI	55 CN, 25 SMC, 46 EMCI, 39 LMCI, 29 AD	2D CNN, AlexNet	✓	✓	✓	–	✓	✓
Proposed Work	**fMRI**	**ADNI**	**25 CN, 25 SMC, 25 EMCI, 13 MCI, 25 LMCI, 25 AD**	**2D CNN, ResNet-18, Transfer learning**	✓	✓	✓	✓	✓	✓

Independent Component Analysis (ICA) is another statistical technique used for the analysis of neuroimaging data. Oghabian et al. [17] have applied ICA algorithm to distinguish between healthy, MCI and AD patients. They used fMRI data from 15 normal, 11 MCI and 14 AD subjects and applied seven steps pre-processing pipeline. Different pre-processing techniques have been applied in this study including MCFLIRT based head motion correction [18, 19], slice-timing correction, mean intensity normalization, spatial smoothing using FSLBET based brain extraction [20], high pass filtering, and Gaussian smoothing. After applying various pre-processing steps, the ICA algorithm has been applied to fMRI activation patterns. They obtained a difference of 0.0097, 0.0051, and 0.0168 between control and MCI, between control AD and between AD and MCI subjects respectively.

Another common method for neuroimaging data analysis is based on Multi-Voxel Pattern Analysis (MVPA) techniques [21, 22]. This method is based on supervised linear regression and determines specific functional activities of various brain regions by using their neural dynamics. Coutanche et al. [22] have applied MVPA to determine symptoms in patients. And it was found that MVPA methods can be used to classify various stages of a disease. In MVPA based approaches, multiple classifiers are used to obtain the best results. To classify fMRI data non-linear classifiers are used such as Support Vector Machine (SVM).

The traditional machine learning techniques require handcrafted feature extraction before classification. However, for automatic analysis of neuroimaging data, manual extraction of features is suboptimal. Approaches based on user-defined features have limitations. Improved performance can be obtained by learning features specific to the problem of interest. Recently, deep learning methods are being used in the domain of neuro-imaging for automated feature extraction and analysis of brain data by using improved processing power and graphical processing units. In deep learning techniques, feature extraction is automatic, thus, models based on this achieve improved performance.

In this respect, H.I Suk et al. [13] have applied deep learning to classify three disease stages including MCI, MCI converter, and AD. The dataset includes scans from 128 MCI, 76 MCI converters, 93 AD and 101 normal control (NC) subjects. These scans were pre-processed by applying methods of skull-striping, spatial normalization, and cerebellum removal. For feature extraction from images, an auto-encoder network has been applied. After feature extraction, SVM based classification has been performed and accuracies of 95.35%, 85.67%, and 75.92% have been achieved for AD vs. NC, MCI vs. NC and MCI-converter vs. MCI respectively. Siqi Liu et al. [26] presented a multi-modal method to extract neuro-imaging features for AD diagnosis. The zero-masking method was used to learn low-level features and stacked autoencoder
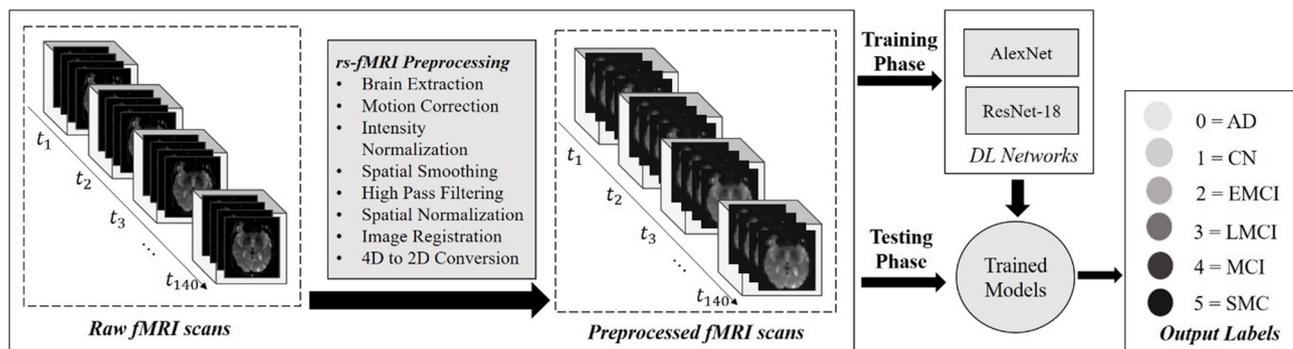

Fig. 1 Computational steps for multi-class AD classification

Table 3A   Overview of the study groups in the rs-fMRI dataset

Study Group	No. of Subjects	Mean Age
CN	25	75.091
SMC	25	72.512
EMCI	25	71.874
LMCI	25	72.277
MCI	13	75.000 s
AD	25	74.69
Total	138	–

network was used for learning high-level features. The extracted features have been classified by applying SVM classifier and an accuracy of 86.86% have been achieved.

Payan et al. [27] presented an algorithm for classification of three stages of AD including MCI, AD and normal control (NC). The algorithms were based on applying a 3D CNN with an autoencoder network and 2D CNN to classify brain scans. An accuracy of 89.47% and 85.53% have been achieved with 3D CNN and 2D CNN models respectively. Siqi Liu et al. [38] also achieved a classification accuracy of approximately 85.53% with the same network architecture for 2D CNNs. Sarraf et al. [14] performed research for classification of AD. The study was based on classifying AD patients from normal control subjects using MRI and fMRI scans. Two network architectures have been applied for binary classification. These CNN based architectures were based on LeNet-5 and GooleNet. They achieved an average accuracy of 99% with LeNet and 100% with GoogleNet using fMRI data.

Table 1 presents a review of the studies based on Alzheimer's disease classification using deep learning techniques. Most of the studies have used structural MRI or PET scans and are based on the classification of a few disease stages i.e. AD, CN, and MCI. There is a limited number of studies that have used fMRI data for multi-class AD diagnosis and classification. Some of the

Table 3B   Characteristics of the rs-fMRI dataset

Characteristic	Description
Acquisition Scanner	Philips Medical systems
EPI	140 scans/volume
Field strength	3.0
Flip angle	80 degree
TE = 30.001	30.0001
Width, Height	64,64
Number of slices	6720
Pixel spacing	3.3125
Slice thickness	3.313
Format	DICOM

studies that have used fMRI data for AD classification has been listed in Table 2 along with the stages of the target disease.

Classifying different stages of AD is a challenging task due to overlapping features of different stages. Most of the work in literature is directed towards the binary classification i.e. presence or absence of AD from neuro-scans. Little work is done to classify two or more stages of this disease. In this research, the objective is to perform a multi-class classification of 6 AD stages that include CN, SMC, EMCI, MCI, LMCI, and AD. Classifying data with similar features among different classes is a challenging task. Another challenge is the availability of large datasets with ground truth labels. In order to overcome this problem transfer learning approach, in addition to training the model from scratch, has been used in this study to improve performance. We have used resting-state fMRI to perform multi-class AD classification by applying image processing and deep learning methods. We used Resnet-18 as a base architecture and empirically performed analysis by using two approaches. First, by training ResNet-18 from scratch by randomly initializing the network parameters and reducing the number of input channels to one. Second, by initializing weights from pre-trained model and using two strategies for transfer learning: (i) by replacing the last dense layer of the original network with the new dense layer and, (ii) re-training all the convolution layers of the network with our dataset. Several experiments are performed by tuning hyperparameters of algorithms and classifiers, to get optimal accuracy.

This paper is organized into the following sections. Section 1 provides an introduction and literature review. Section 2 presents the methods and materials used to conduct this research. The experimental details and results are listed in Section 3. Finally, the conclusion has been presented in the last section.

Materials and Methods

The research methodology consists of multiple steps including data acquisition, pre-processing, deep learning-based feature extraction and classification followed by evaluation. Neuroimaging data are acquired from a well-known database on Alzheimer's disease. Pre-processing techniques are applied to remove noise and artifacts from data. Preprocessed data is then fed to CNN based neural networks for feature extraction and classification of multiple stages of AD. These computational steps are graphically presented in Fig. 1.

Neuroimaging Dataset

Neuroimaging data is acquired from Alzheimer's disease Neuroimaging Initiative (ADNI) database [40] that has been used in various studies [13, 23] for AD classification.

ADNI is an extensive multisite study aimed at developing genetic, biochemical, neuroimaging and clinical bio-markers for AD diagnosis, prognosis, and tracking. ADNI contains neuroimages in various modalities including MRI, fMRI, PET, and DTI. In this research, we have used rs-fMRI brain scans provided by ADNI. The dataset contains fMRI scans from 138 subjects including 25 CN, 25 SMC, 25 EMCI, 25 LMCI, 13 MCI, and 25 AD. The age of the subjects is greater than 71 and each of them has been diagnosed and labeled as one of the AD stages based on their scores in cognitive tests i.e. mini-mental state examination (MMSE) [41] and clinical dementia rating (CDR) [42]. The characteristics of the fMRI dataset used for experimental analysis are given in Tables 3A and 3B.

Preprocessing of Resting-State fMRI Data

Researchers have used various preprocessing steps on this dataset [14, 30]. Data preprocessing is applied to remove noise and artifacts from data that can improve the quality of images and leads to better feature extraction. For preprocessing rs-fMRI, the standard pipeline consisting of various steps is used. Firstly, the dataset is converted from DICOM to NIFTI format by using the conversion toolbox from Chris Rorden [43]. Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL) [44, 45] is used for preprocessing the data.

Brain extraction is performed on scans to remove non-brain tissues such as neck tissues and skull. For this purpose, FSL-BET toolbox [46] is used, which performs brain extraction by estimating the intensity histogram-based threshold, the center of the gravity and radius of the sphere of the brain's surface. Inside the brain, the tessellated surface is initiated, which slowly updates one vertex at a time until a complete surface is achieved. Then, *motion correction* is applied to remove and correct the effect of subjects' head motion during data acquisition sessions. We performed motion correction by using FSL-MCFLIRT toolbox [19, 47]. *Slice timing correction* is also performed by using FEAT [48] module of FSL library. The method of slice timing correction applies interpolation to transform the voxel time-series either forward

or backward in time to make the temporal adjustment. The interpolation method used for this study is sinc interpolation based on Hanning windowed method to adjust the voxel time-series by a fraction of scanner's TR (Repetition Time) with respect to the middle of TR.

Intensity normalization is applied on data to ensure that each volume has the same mean intensity. *Spatial smoothing* is applied to reduce the noise level while preserving the underlying signal. Its purpose is to improve the signal to noise gain. The extent of spatial smoothing is set according to the size of the underlying signal. We performed spatial smoothing by using a 5 mm FWHM Gaussian kernel. The kernel size selection corresponds to what has been recommended in the literature for this dataset. Then, temporal *high-pass filtering* is applied to remove the low-frequency noise signals as a result of some psychological artifacts such as breathing, heartbeat or scanner drift for the rs-fMRI time series. High pass filtering is performed by using a temporal filter of cut-off frequency 0.01 HZ. We also applied *spatial normalization* on images by first putting the images in T1 weighted space by using a linear-transformation with 7 Degree of Freedom (DOF). The images are then registered to a standard space of MNI152 template, which is a reference template of brain-derived from the average of 152 MRI scans. To register images to MNI152 space, a linear transformation with 12 DOF (such as translation, scaling, shear, and rotation) is applied. In this study, spatial normalization is performed by FSL-FLIRT [47, 49, 50] toolbox.

After applying the preprocessing methods on fMRI data, preprocessed $64 \times 64x48x140$ 4D fMRI scans are obtained in which each scan contains $64 \times 64x48$ 3D volumes per time course (140 s). These 4D scans are then converted to 2D images along with image height and time axis. This results in 6720 images of size 64x64 per fMRI scan. The first and last three slices are removed as they contain no functional information. Therefore, from each scan information from 44 slices is used. Hence, 6160 2D images are obtained from each fMRI scan and are saved in portable network graphics (PNG) format.

The data acquired from ADNI is processed and converted to 2D images by using the aforementioned pre-processing methods. In this way, we have created a dataset that was used for training deep learning networks. The characteristics of the preprocessed dataset are given in Table 4.

Deep Learning Methods for RS-fMRI Data

We performed our experiments on CNN based architectures to train and evaluate our dataset. Prior work on rs-fMRI based Alzheimer's disease classification is mainly based on LeNet [14], GoogleNet [30] and AlexNet architectures. Due to the outstanding performance of Residual neural network [51] in the computer vision domain, our focus in this study is the ResNet-18 architecture [52]. We used this architecture by
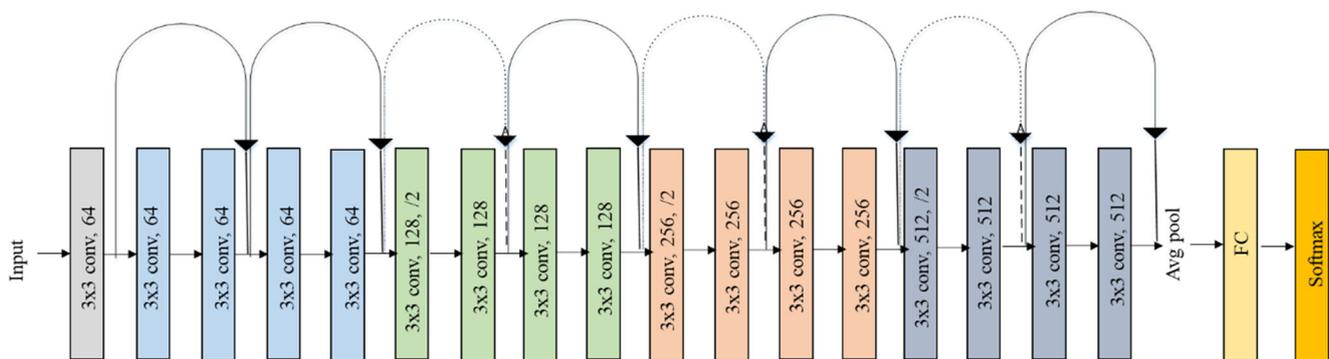
Table 4 Characteristics of the preprocessed dataset

Group	Rs-fMRI Scans	No. of Images
CN	25	154,000
SMC	25	154,000
EMCI	25	154,000
LMCI	25	154,000
MCI	13	80,080
AD	25	154,000
Total	*138*	*850,080*

Fig. 2 Original ResNet-18 Architecture

training from scratch as well as by adapting the pre-trained network for our task through transfer learning. The details of network architectures are given in this section.

Residual Neural Network for AD Classification

ResNet was developed by Kaiming He et al. [51] in 2016. A residual learning method was proposed to train deeper networks that are practically difficult to train. Network layers were reformulated to learn residual functions with reference to the layer inputs. The results indicate that the deeper networks based on residual learning can achieve better optimization and high accuracy. Experimental evidence [53, 54] revealed that network depth is crucial to achieving better performance. But deeper networks are difficult to train and the increased number of layers may not ensure better learning. Also when deep networks start convergence, accuracy gets saturated at a point and then starts to decrease rapidly. The use of residual learning in deeper networks solves the problem of accuracy degradation in deeper networks. In plain networks, several layers are stacked together to directly learn the desired mapping. However, in residual networks, the layers are stacked to learn a residual mapping. The mapping function, denoted by $H(x)$, is fitted by a few stacked layers. The idea of residual learning is hypothesized as, if several nonlinear layers can asymptotically estimate a complicated mapping function, then they can asymptotically estimate the residual function denoted as $F(x)$. The underlying mapping is given by:

$$H(x) = F(x) + x \qquad (1)$$

And the residual function is given by:

$$F(x) = H(x) - x \qquad (2)$$

The stacked layers explicitly learn the residual function $F(x)$ rather than learning the original function $H(x)$. This method assumes that residual mapping function is easier to optimize than the original function. For example, if an identity mapping is optimized than the residual can be pushed to zero easily rather than approximating the identity mapping from a few stacked non-linear layers. After approximating the residual function, original mapping function is calculated as $H(x) = F(x) + x$. This mapping function $F(x) + x$ is realized as residual shortcut connection in a feedforward neural network and performs element-wise addition. In a residual network, these connections approximate an identity mapping. Their output is then added back to the stacked layers. Addition of these connections in the networks does not introduce more complexity or parameters. Also, these residual networks can be easily trained with SGD based backpropagation.

The architecture of the original ResNet-18 is shown in Fig. 2. There is a total of eighteen layers in the network (17 convolutional layers, a fully-connected layer and an additional softmax layer to perform classification task). The convolutional layers use 3×3 filters and the network is designed in such a way that if the output feature map is the same size then the layers have the same number of the filters. However, filters get doubled in the layers, if the output feature map is halved. The downsampling is performed by convolutional layers having a stride of 2. Lastly, there is an average-pooling followed by a fully-connected layer with a softmax layer. Throughout the network, residual shortcut connections are inserted between layers. There are two types of connections. The first type of connections, denoted by solid lines, are used when input and output have the same dimensions. Second types of connections, denoted by dotted lines, are used when dimensions increase. This type of connection still performs identity mapping but with zeros padding for increased dimensions with a stride of 2.

In order to benefit from the effects of a different design decision in deep learning, we designed several experiments by training modified ResNet-18 from scratch as well as performing transfer learning. Specifically, we used two strategies in our experiments for network training. First, we used a slightly modified version of ResNet-18 to perform training from scratch by randomly initializing the network parameters. We also reduced the number of input channels to one in order to perform training with the greyscale images.

Table 5 Adapted Architectures of ResNet-18 used for AD classification

Layer Type	1-Channel ResNet-18 (1CR)		Transfer Learning			
	Layer Parameters	*Output Size*	*Layer Parameters*	*Output size*	*Off the Shelf (OTS)*	*Fine-Tuned (FT)*
Convolution	7 × 7, 64-d, stride 2	112 × 112	7 × 7, 64-d, stride 2	112 × 112	Same	Fine tuned
Max Pooling	3 × 3, stride 2	56 × 56	3 × 3, stride 2	56 × 56	Same	Same
Convolution	1 × 1, 256-d, stride 1	56 × 56	1 × 1, 64-d, stride 1	56 × 56	Same	Fine tuned
Convolution	1 × 1, 64-d, stride 1	56 × 56	3 × 3, 64-d, stride 1	56 × 56	Same	Fine tuned
Convolution	3 × 3, 64-d, stride 1	56 × 56	3 × 3, 64-d, stride 1	56 × 56	Same	Fine tuned
Convolution	1 × 1, 256-d, stride 1	56 × 56	3 × 3, 64-d, stride 1	56 × 56	Same	Fine tuned
Convolution	1 × 1, 64-d, stride 1	56 × 56	3 × 3, 64-d, stride 1	56 × 56	Same	Fine tuned
Convolution	1 × 1, 64-d, stride 1	56 × 56	1 × 1, 128-d, stride 1	28 × 28	Same	Fine tuned
Convolution	1 × 1, 256-d, stride 1	56 × 56	3 × 3, 128-d, stride 2	28 × 28	Same	Fine tuned
Convolution	1 × 1, 64-d, stride 1	56 × 56	3 × 3, 128-d, stride 1	28 × 28	Same	Fine tuned
Convolution	3 × 3, 64-d, stride 1	56 × 56	3 × 3, 128-d, stride 1	28 × 28	Same	Fine tuned
Convolution	1 × 1, 256-d, stride 1	56 × 56	3 × 3, 128-d, stride 1	28 × 28	Same	Fine tuned
Convolution	1 × 1, 512-d, stride 2	28 × 28	1 × 1, 256-d, stride 2	14 × 14	Same	Fine tuned
Convolution	1 × 1, 128-d, stride 2	28 × 28	3 × 3, 256-d, stride 2	14 × 14	Same	Fine tuned
Convolution	3 × 3, 128-d, stride 1	28 × 28	3 × 3, 256-d, stride 1	14 × 14	Same	Fine tuned
Convolution	1 × 1, 512-d, stride 1	28 × 28	3 × 3, 256-d, stride 1	14 × 14	Same	Fine tuned
Convolution	1 × 1, 128-d, stride 1	28 × 28	3 × 3, 256-d, stride 1	14 × 14	Same	Fine tuned
Convolution	3 × 3, 128-d, stride 1	28 × 28	1 × 1, 512-d, stride 2	7 × 7	Same	Fine tuned
Convolution	1 × 1, 512-d, stride 1	28 × 28	3 × 3, 512-d, stride 2	7 × 7	Same	Fine tuned
Convolution	–	–	3 × 3, 512-d, stride 1	7 × 7	Same	Fine tuned
Convolution	–	–	3 × 3, 512-d, stride 1	7 × 7	Same	Fine tuned
Convolution	–	–	3 × 3, 512-d, stride 1	7 × 7	Same	Fine tuned
Average Pooling	7 × 7, stride 1	22 × 22	7 × 7, stride 1	1 × 1	Same	Same
Dense	6-d	1 × 1	6-d	1 × 1	Same	Same
Softmax	6-d	1 × 1	6-d	1 × 1	Same	Same

Secondly, we used a pre-trained network for weight initialization and transfer learning was performed. Since the trained model was for a different domain and task, we adapt the network to perform our task. In order to transfer the knowledge from a pre-trained network, we performed transfer learning in two ways. In the first approach, we performed *off-the-shelf (OTS)* [55, 56] transfer learning by replacing the last dense layer of the original network with the new dense layer to match the number of classes for our task. In the off-the-shelf approach, all the layers except the last layer (classifier) of the network are used for feature extraction and the weights of the last layer are adapted to the new task. The second approach is *fine-tuning (FT)*, in which more than one layers of the network are re-trained from the samples of the new task. For fine-tuning approach, we re-trained all the convolution layers of the network with our dataset. For both approaches of transfer learning, we used weights of ResNet-18 network trained on ImageNet as a starting point [57].

The details of the networks used for experiments are given in Table 5. We used the ResNet-18 architecture in our experiments and the table presents the difference between the original and our networks. There are three networks in the table, 1-Channel ResNet-18, Off the Shelf (OTS) and Fine-Tuned (FT). 1-Channel ResNet-18 represents the network that was trained from scratch by using greyscale images. For transfer learning, we used three additional convolution layers. The words "same" and "fine-tuned" are used to represent the difference between OTS and FT networks. The layer parameters of each network are given in the table. The ResNet with 18 layers has 2.37 million parameters and ResNet with 21 layers has 11.18 M parameters.

Evaluation Measures

A common approach of evaluating results of machine learning models is using precision, recall [58] f1-measure and area under the receiver operating characteristic (AROC) curve [54]. These measures have been originated form Information Retrieval. In this research, we have evaluated our multi-class AD classification models by using the aforementioned evaluation measures.

Table 6   Details of dataset split used for training

Type	Split Ratio	No. of Images
Training dataset	70%	595,056
Validation dataset	20%	170,016
Testing dataset	10%	85,005
Total	*100%*	*850,080*

Precision

Precision or Confidence [58] is the proportion of predicted positive cases that are actually real positives. It is also called Positive Predictive Value (PPV). Precision is defined as:

$$Precision = Confidence = \frac{TP}{TP + FP} \tag{3}$$

where TP denotes true positive and FP denotes false positive.

Recall

Recall (also named as sensitivity) [58] is the proportion of actual positive cases that are correctly predicted. This measures the coverage of actual positive cases and reflects correct predicted cases. It is termed as True Positive Rate (TPR) and is given as:

$$Recall = Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

where TP denotes true positive and FN denotes false negative.

F1-Measure

F- Measure is a combined measure [58] that captures the tradeoffs associated with precision and recall and is defined as:

$$F-Measure = \frac{1}{\alpha \frac{1}{P} + (1-\alpha)\frac{1}{R}} = \frac{(1 + \beta^2)PR}{\beta^2 P + R} \tag{5}$$

where P denotes precision and R denotes recall. The harmonic mean is considered a very conservative average, that's why a balanced measure is used called F1-measure with $\alpha = 1$, $\beta = 1/2$ and is defined as:

$$F1-Measure = \frac{2*PR}{P + R} \tag{6}$$

Area under Receiver Operating Characteristic (AROC) Curve

Receiver Operating Characteristics (ROC) analysis [58] has been borrowed from Signal Processing in medical sciences. It has become a standard tool for evaluation by comparing the true positive rate (TPR) and false-positive rate (FPR). In behavioral sciences, AROC curve represents the combination of sensitivity (TPR) and specificity (TNR). It allows to compare the performance of classifier models and takes values between 0 and 1. Best classifier model is the one which is closest to 1
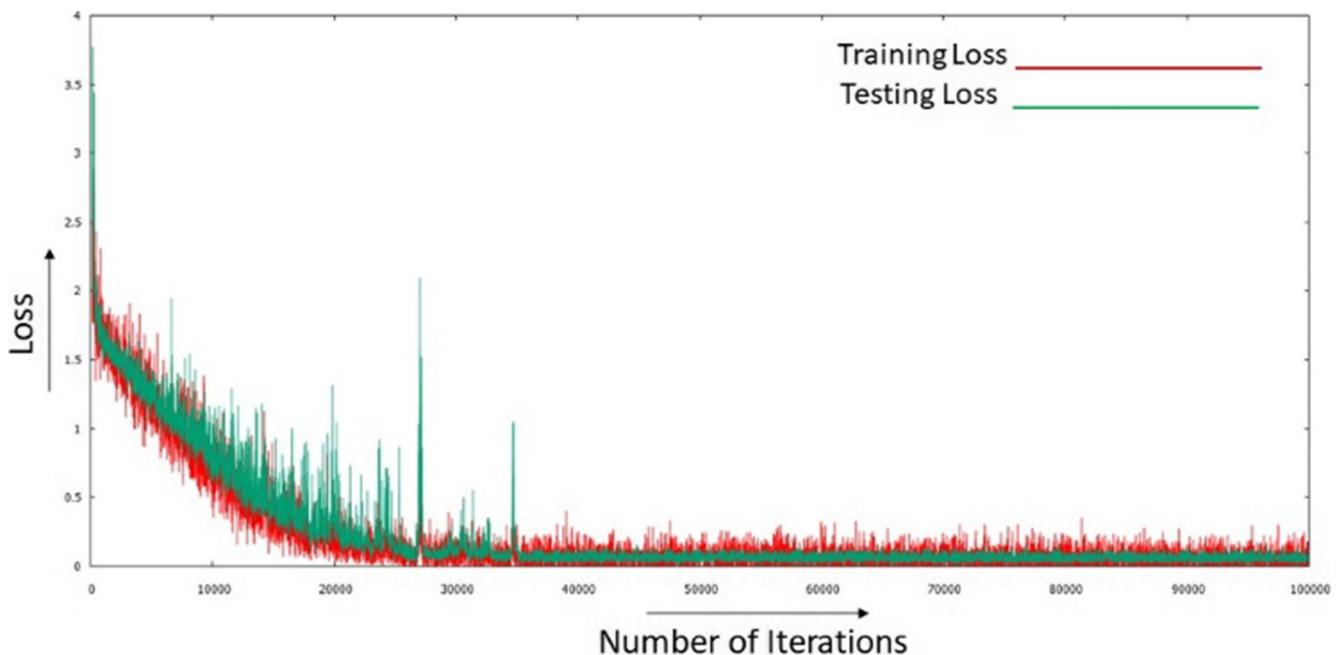

Fig. 3 Graphical trends of training loss vs. testing loss with 1CR Network
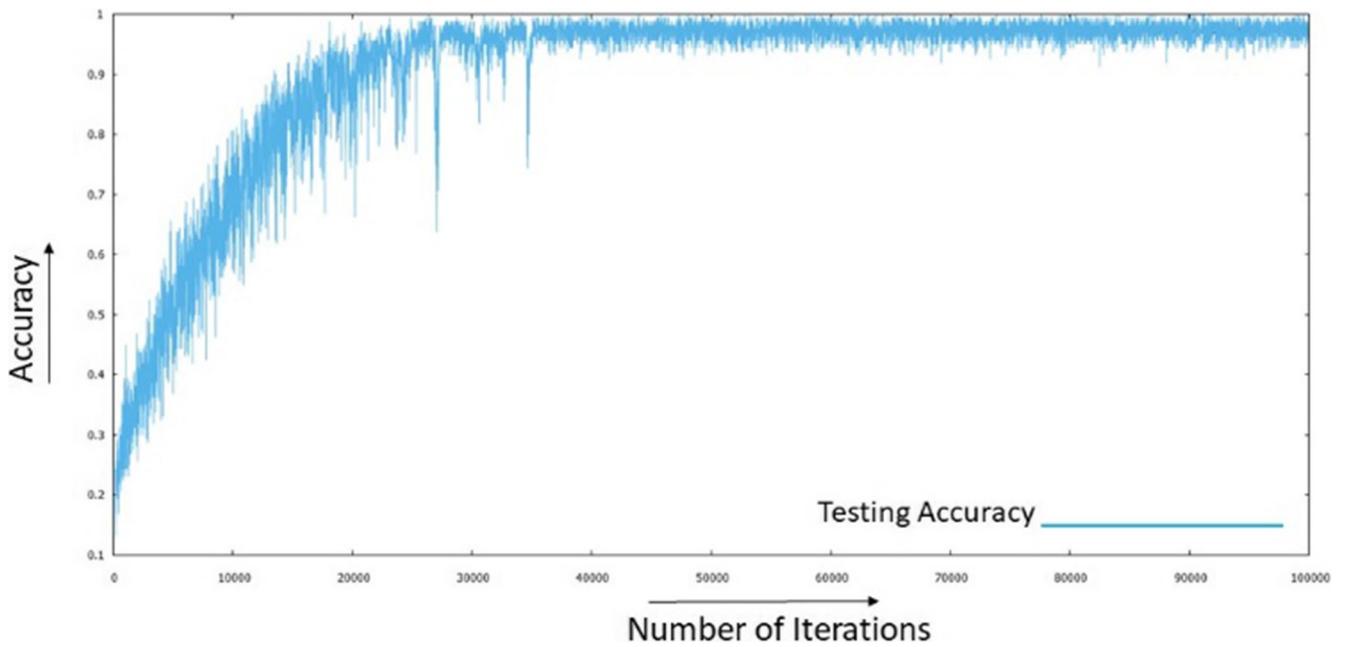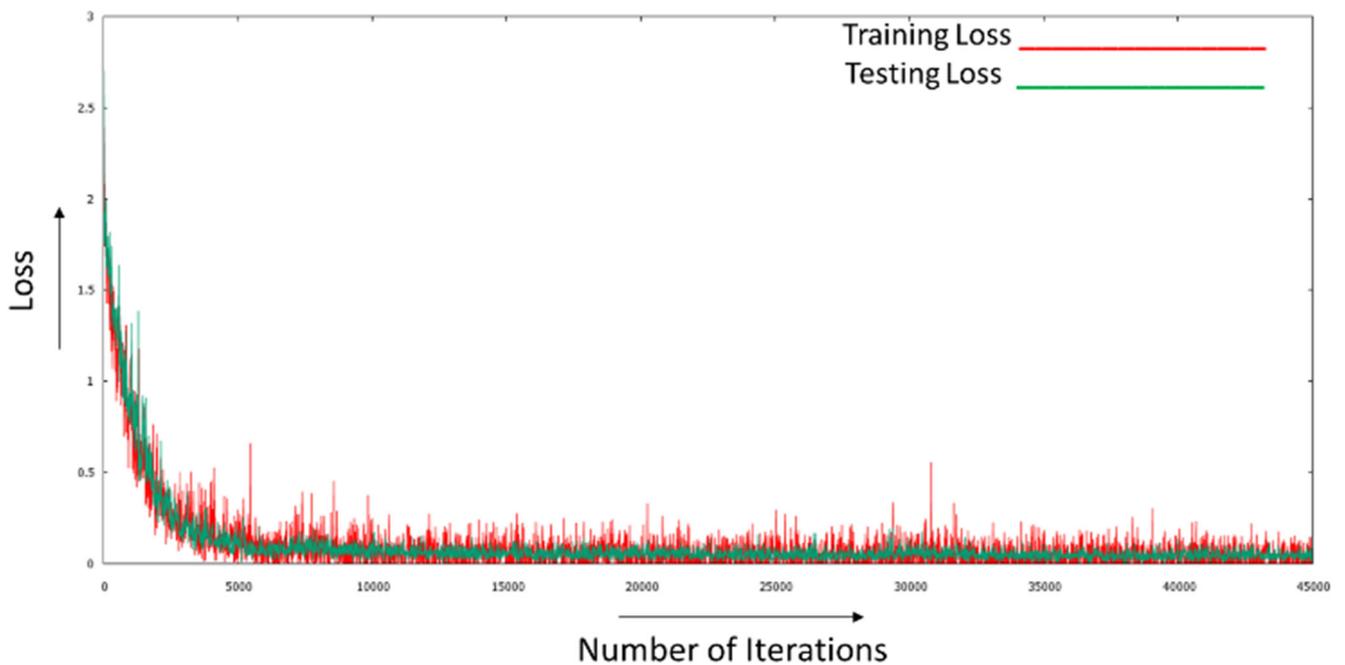
Fig. 4 Graphical trend of test accuracy with 1CR Network

and farthest from TPR = FPR. However, lower bound for classification is 0.5 in practical scenarios where classifier has no discrimination ability. Whereas classifier with must higher value than 0.5 has a much more discriminative ability. The approach to measuring the ROC curve is by calculating the area under the curve (AUC) and is given by:

$$AUC = \frac{TPR - FPR + 1}{2} = \frac{TPR - TNR}{2} = 1 - \frac{FPR + FNR}{2} \quad (7)$$

where *TPR* denotes true positive rate, *FPR* as false positive rate, *TNR* as true negative rate and *FNR* as false-negative rate.

Results and Discussions

This research is aimed at using rs-fMRI data to classify 6 stages of AD. We have applied different preprocessing methods before performing further analysis. The preprocess-
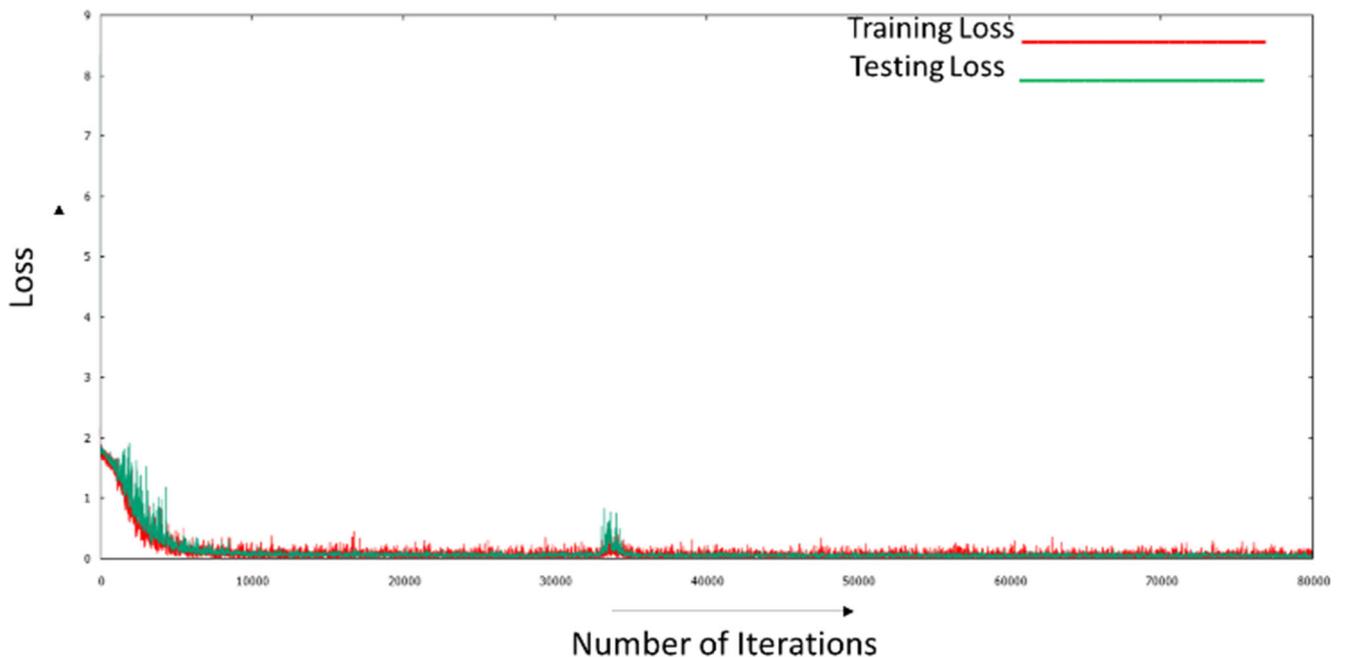

Fig. 5 Graphical trends of training loss vs. testing loss with OTS Network
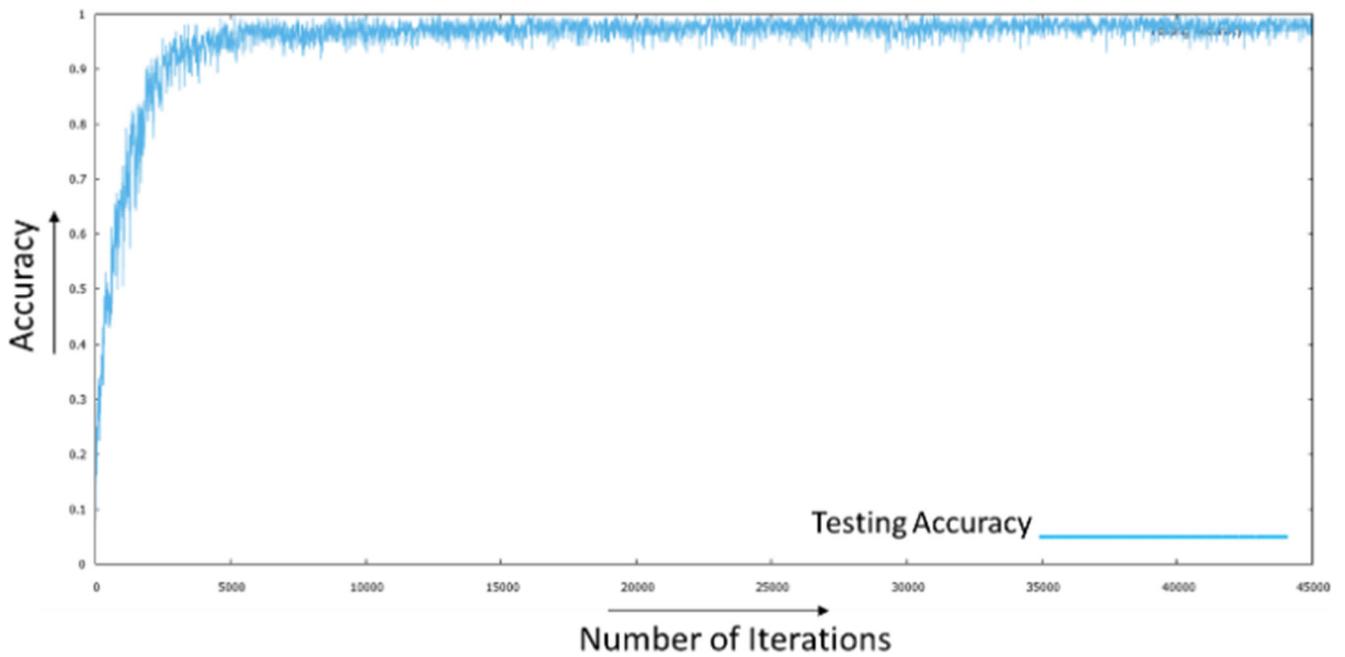
Fig. 6 Graphical trend of test accuracy with OTS Network

ing methods and algorithms used to analyze rs-fMRI data have been discussed in Section 2. This section provides details on the experiments conducted and discusses the results achieved. In the dataset, there are 138 4D scans and 850,080 2D images. For the evaluation, we split the dataset into a training dataset, validation dataset and testing dataset with 70%, 20%, and 10% split ratio, respectively as described in Table 6. The dataset was randomly shuffled before splitting. The validation set was used to determine the trend of learning during the training phase. We estimate the validation loss to

determine the best models. The testing set was used to perform inference on the learned model.

Experiments and Evaluation

We trained three ResNet-18 based networks (1CR, OTS, and FT) to classify different stages of AD. We used the same experimental setup in all the experiments. The input to the networks are images of size 224×224 that are resized to match the pre-trained network's input size. The learning rate

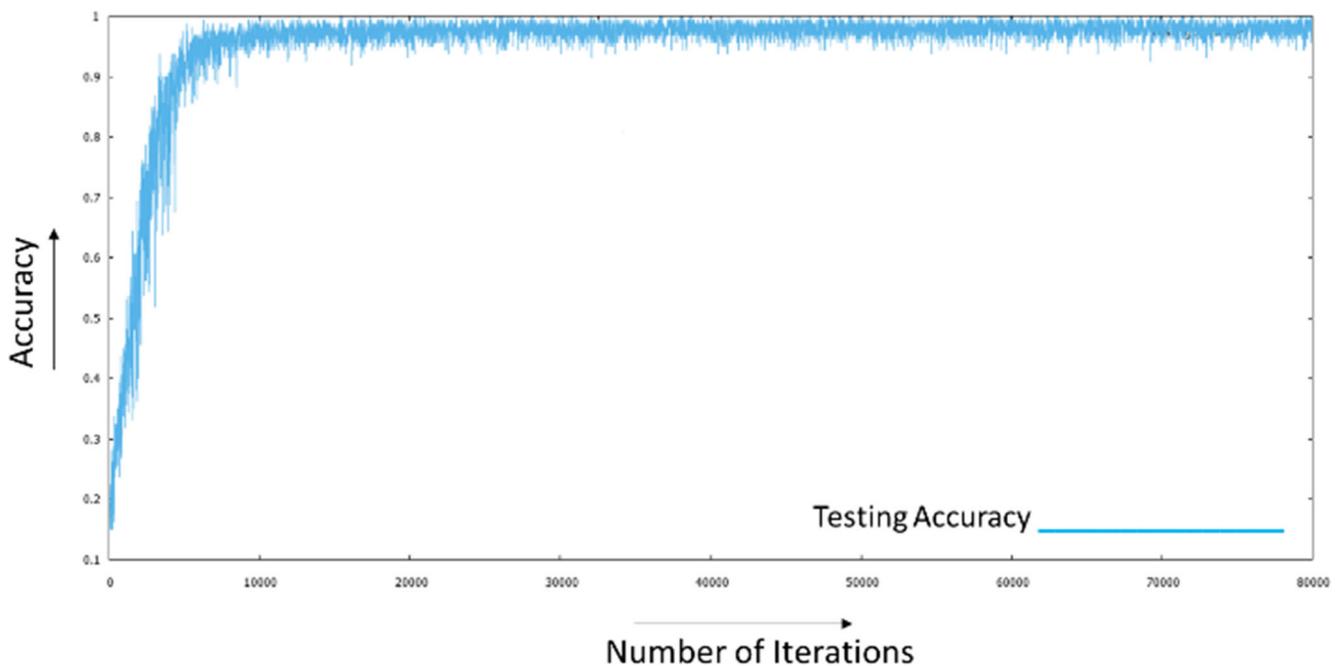**Fig. 7** Graphical trends of training loss vs. testing loss with FT Network

Fig. 8 Graphical trend of test accuracy with FT Network

is initialized to 0.001 which decreased by 10% every 25,000 iterations. Gamma is initialized as 0.1, momentum as 0.9 and weight decay factor is 0.0005. Stochastic gradient descent (SGD) based solver is used with a batch size of 32 images for training. The models are implemented on Caffe and trained on FloydHub cloud service with GPU Tesla K80.

The trends of training and testing loss and average test accuracy are shown in Figs. 3, 4, 5, 6, 7, 8.

Table 7 summarizes the testing accuracy and validation loss with the three networks for a multi-class AD classification task. The best average testing accuracy is achieved with the OTS network, however, the improvement is slight compared to the FT network. However, for "CN" stage, the FT network achieved approximately 3.12% improvement in accuracy than the OTS network. Table 8 summarizes the outcomes of our evaluation and Fig. 9 presents the ROC curves of three models. We evaluated three different experimental setups with varying weight initializations and network architectures. We

evaluated the classification models by using different measures such as precision, recall, f1-measure, and AROC curves analysis to evaluate all AD stages.

The results indicate variability in outcomes with respect to AD stages, especially for CN and SMC stages. Specifically, for the "CN" stage, we observe a standard deviation of 0.048 for precision, 0.017 for recall and 0.016 for f1-measure of three models. For the "SMC" stage, the standard deviation is 0.05 for precision, 0.014 for recall and 0.0197 for f1-measure of three models. For each AD stage, the results are better with either OTS or FT models. However, the average scores for all measures are higher for the OTS network. In particular, an average improvement of 0.069, 0.0055, 0.0055 and 0.0002 is observed for precision, recall, f1-measure, and AUC respectively with the OTS model. While, our models achieved high AUC for all AD stages, yet the applicability of such technology in a clinical assessment largely depends on the data available for model training and evaluation.

Comparative Analysis

We compared the performance of our ResNet-18 based networks with each other. We also compared our results with a previous study [39], that have worked on a similar problem but with 5 disease stages including AD, CN, EMCI, LMCI, and SMC. But in order to have a fair comparison (using same data samples, data split and a number of stages), we performed an additional evaluation by training an AlexNet used in [39] on our dataset. We evaluated the classification results in terms of average accuracy and classification accuracy of each AD stage. Table 9 presents the comparative analysis as well as

Table 7   Multi-class classification results with networks

Study Group	*1CR Network*	*OTS Network*	*FT Network*
CN	0.9593	0.9688	1.0000
SMC	0.9984	1.0000	0.9685
EMCI	0.9719	0.9738	0.9738
LMCI	0.9672	0.9738	0.9743
MCI	0.9713	0.9781	0.9740
AD	0.9729	0.9801	0.9801
Average Accuracy	**0.9737**	**0.9792**	**0.9788**
Validation loss	0.0739	0.0716	0.0393

Table 8 Evaluation of results for our models

Group	Network	CN	SMC	EMCI	LMCI	MCI	AD	Average
Precision	1CR	0.9979	0.8869	0.9849	0.9997	**1.0000**	0.9984	*0.9761*
	OTS	**0.9988**	0.8974	**1.0000**	**1.0000**	**1.0000**	0.9999	***0.9813***
	FT	0.8964	**1.0000**	0.9987	**1.0000**	**1.0000**	**1.0000**	*0.9810*
Recall	1CR	0.9593	0.9984	0.9719	0.9672	0.9713	0.9729	*0.9737*
	OTS	0.9689	**1.0000**	**0.9738**	0.9738	**0.9781**	**0.9801**	***0.9792***
	FT	**1.0000**	0.9685	**0.9738**	**0.9743**	0.9740	**0.9801**	*0.9789*
F1-Measure	1CR	0.9782	0.9393	0.9784	0.9832	0.9854	0.9855	*0.9741*
	OTS	**0.9836**	0.9459	**0.9867**	0.9867	**0.9889**	**0.9899**	***0.9796***
	FT	0.9454	**0.9840**	0.9861	**0.9869**	0.9868	**0.9899**	*0.9793*
AUC	1CR	0.9994	0.9995	0.9994	0.9995	0.9996	0.9995	*0.9995*
	OTS	**0.9996**	**0.9996**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	***0.9997***
	FT	**0.9996**	**0.9996**	0.9996	**0.9997**	**0.9997**	**0.9997**	***0.9997***

summarizes the classification results with other approaches. Figure 10 graphically presents the comparative analysis for 6 AD stages.

In our comparative analysis, we noticed that our models performed better than Y. Kazemi et al. [39] and our trained AlexNet model. When we compared our results with Y. Kazemi et al., we found that the FT model achieved higher classifying accuracy with each AD stage and OTS and 1CR achieved higher accuracy in all but "CN" stage. Particularly, the FT model achieved 1.66%, 2.3%, 1.74%, 1.54 and 3.04% improvement in accuracy with CN, SMC, EMCI, LMCI, and AD respectively. While, TS model achieved an improvement of 5.45%, 1.74%, 1.49% and 3.04 with SMC, EMCI, LMCI, and AD respectively. With 1CR model, we achieved an improvement of 5.29%, 1.55%, 0.83% and 2.32% with SMC, EMCI, LMCI, and AD respectively. To have a fair comparison, we compared our results with our trained AlexNet model and found that there is an approximately 4% improvement in accuracy with our models. Overall, when directly comparing our results to the former best results, we achieved state-of-the-art results with our FT model for all AD stages. While with OTS and 1CR models we achieved state-of-the-art results for all except "CN" stage. Comparing the average accuracy
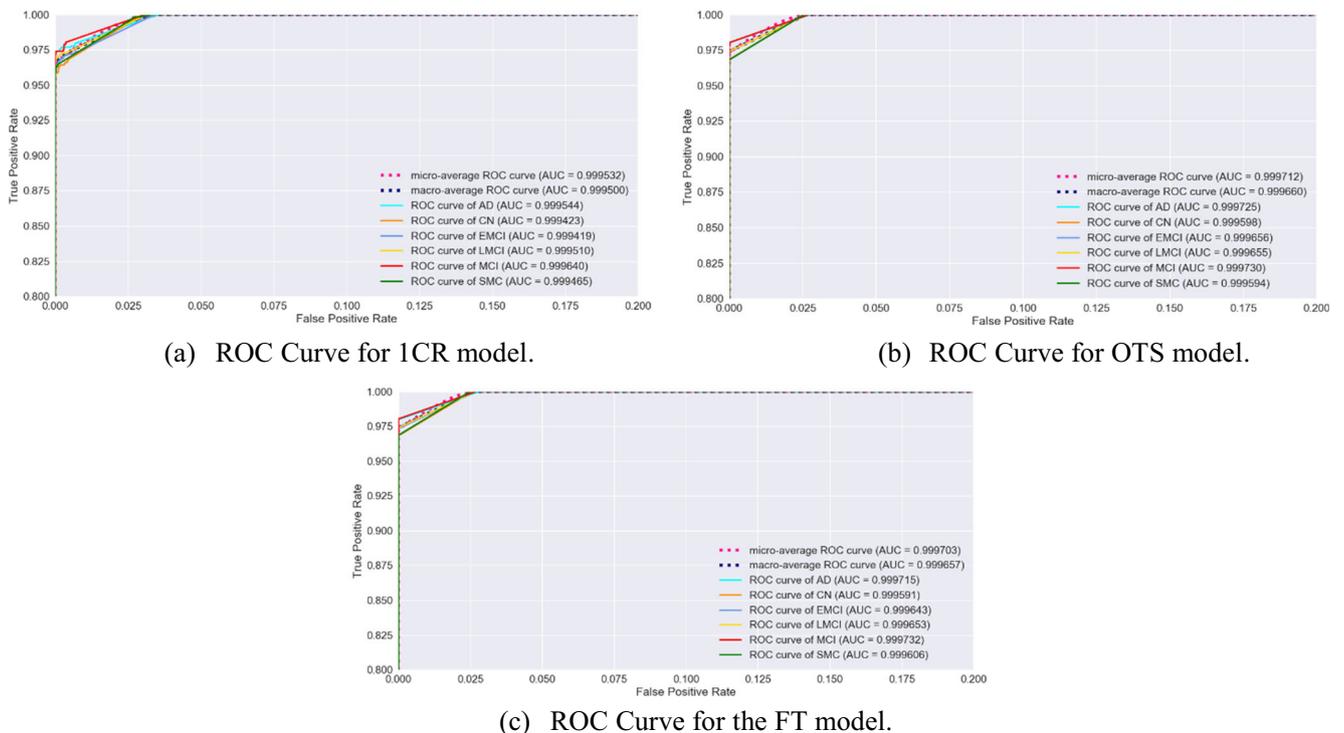

(a) ROC Curve for 1CR model.

(b) ROC Curve for OTS model.

(c) ROC Curve for the FT model.

Fig. 9 Receiver operating characteristic curves for multi-class AD classification

Table 9 Overview of classification accuracy (%)

Pathology	Y. Kazemi et al. [39]	AlexNet	1CR	OTS	FT
CN	98.34	91.73	95.93	96.88	**100.0**
SMC	94.55	**100.0**	99.84	**100.0**	96.85
EMCI	95.64	91.80	97.19	**97.38**	**97.38**
LMCI	95.89	90.50	96.72	97.38	**97.43**
MCI	–	94.80	97.13	**97.81**	97.40
AD	94.97	95.14	97.29	**98.01**	**98.01**
Average Accuracy	97.63	93.60	97.37	**97.92**	97.88

for all AD stages with the former approach i.e. AlexNet, also achieved improved performance. While the former study reported average accuracy of 97.63% with 5 disease stages and a larger dataset than ours, we still achieved better results for 6 AD stages with an average accuracy of 97.92% and 97.88% for OTS and FT models respectively.

Conclusion

Alzheimer's disease (AD) is an incurable neurological disorder affecting a significant world's population. The early diagnosis of AD is crucial to improve the quality of patients' lives and the development of improved treatment and targeted drugs. The present study was conducted to explore the effectiveness of the resting-state functional magnetic resonance imaging and advanced deep learning techniques to perform multi-class classification and diagnosis of AD and its progressive stages including CN, SMC, EMCI, MCI, LMCI, and AD. The study proposed to use deep residual neural networks combined with transfer learning approach for performing the classification of 6 AD stages. We incorporate different weight initialization schemes and network architectures to evaluate our dataset. We present a systematic evaluation of three networks including 1CR, OST and FT. 1CR network was exclusively trained on single-channel rs-fMRI images, while two networks were optimized on ImageNet dataset by retraining last dense layer in the OTS network and all the convolution layers in the FT network respectively. While state-of-the-art results are achieved with our models, however, in comparison

to a former study [39] FT network achieved higher accuracy for all AD stages with 1.66%, 2.3%, 1.74%, 1.54 and 3.04% improvement in accuracy for CN, SMC, EMCI, LMCI, and AD respectively. And Our OTS network achieved the best average accuracy of 97.92% for 6 disease stages. The use of residual learning, pre-training and transfer learning approach helped to achieve better performance. The results of this study indicate that integration of resting-state fMRI based neuroimaging and deep learning methods can assist the diagnostic decision making in early diagnosis of neurodegenerative disease, especially AD. The diagnosis of AD stages could aid drug discovery by providing better pathogenesis for measuring effects of target treatments that can slow down the disease progression. By combining clinical imaging with deep learning techniques can help to uncover patterns of functional changes in the brain, related to AD progression and could aid in the detection of risk factors and prognostic indicators.

Fig. 10 Comparative analysis of classification results

Compliance with Ethical Standards

Ethical Approval   We declare that all human and animal studies have been approved by the Medical University of South Carolina Institutional Review Board and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Informed Consent   We declare that all patients gave informed consent prior to inclusion in this study.

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

References

1. Selkoe, D. J., and Hardy, J., The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol. Med. 8(6):595–608, Jun. 2016.
2. Qiang, W., Yau, W.-M., Lu, J.-X., Collinge, J., and Tycko, R., Structural variation in amyloid-β fibrils from Alzheimer's disease clinical subtypes. Nature 541(7636):217–221, Jan. 2017.
3. Eftekharzadeh, B., Daigle, J. G., Kapinos, L. E., Coyne, A., Schiantarelli, J., Carlomagno, Y., Cook, C., Miller, S. J., Dujardin, S., Amaral, A. S., Grima, J. C., Bennett, R. E., Tepper, K., DeTure, M., Vanderburg, C. R., Corjuc, B. T., DeVos, S. L., Gonzalez, J. A., Chew, J. et al., Tau Protein Disrupts Nucleocytoplasmic Transport in Alzheimer's Disease. Neuron 99(5):925–940.e7, Sep. 2018.
4. Khan, S. S., Bloom, G. S., Tau: The Center of a Signaling Nexus in Alzheimer's Disease, Front. Neurosci., vol. 10, Feb. 2016.
5. Bachstetter, A. D., Van Eldik, L. J., Schmitt, F. A., Neltner, J. H., Ighodaro, E. T., Webster, S. J., Patel, E., Abner, E. L., Kryscio, R. J., and Nelson, P. T., Disease-related microglia heterogeneity in the hippocampus of Alzheimer's disease, dementia with Lewy bodies, and hippocampal sclerosis of aging. Acta Neuropathol. Commun. 3(1):32, Dec. 2015.
6. Prince, M. J., World Alzheimer report 2015: The global impact of dementia: An analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International, 2015.
7. "2018 Alzheimer's disease facts and figures," Alzheimer's Dement., vol. 14, no. 3, pp. 367–429, Mar. 2018.
8. "2018 ALZHEIMER'S DISEASE FACTS AND FIGURES Includes a Special Report on the Financial and Personal Benefits of Early Diagnosis."
9. "ADNI | Background & Rationale." .
10. Montgomery, V., Stabler, A., Harris, K., and Lu, L., B-26 Effects of Delay Duration on the Wechsler Memory Scale Logical Memory Performance of Older Adults with Probable Alzheimer's Dementia, Probable Vascular Dementia, and Normal Cognition. Arch. Clin. Neuropsychol. 30(6):531, 2015.
11. Wyss-Coray, T., Ageing, neurodegeneration and brain rejuvenation. Nature 539(7628):180, 2016.
12. A. Association and others, "2017 Alzheimer's disease facts and figures," Alzheimer's Dement., vol. 13, no. 4, pp. 325–373, 2017.
13. Suk, H.-I., Lee, S.-W., Shen, D., and Initiative, A. D. N., othersHierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage 101:569–582, 2014.
14. Sarraf, S., Tofighi, G., Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data, in Future Technologies Conference (FTC), pp. 816–820, (2016).
15. Monti, M. M., Statistical analysis of fMRI time-series: A critical review of the GLM approach. Front. Hum. Neurosci. 5:28, 2011.
16. Pernet, C. R., The General Linear Model: Theory and Practicalities in Brain Morphometric Analyses, in Brain Morphometry, Springer, pp. 75–85, (2018).
17. Oghabian, M. A., Batouli, S. A. H., Norouzian, M., Ziaei, M., and Sikaroodi, H., Using functional magnetic resonance imaging to differentiate between healthy aging subjects, Mild Cognitive Impairment, and Alzheimer's patients. J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci. 15(2):84, 2010.
18. Marchitelli, R., Collignon, O., and Jovicich, J., Test–retest reproducibility of the intrinsic default mode network: Influence of functional magnetic resonance imaging slice-order acquisition and head-motion correction methods. Brain Connect. 7(2):69–83, 2017.
19. Goto, M., Abe, O., Miyati, T., Yamasue, H., Gomi, T., and Takeda, T., Head motion and correction methods in resting-state functional MRI. Magn. Reson. Med. Sci. 15(2):178–186, 2016.
20. Rajagopalan, V., and Pioro, E. P., Disparate voxel based morphometry (VBM) results between SPM and FSL softwares in ALS patients with frontotemporal dementia: which VBM results to consider? BMC Neurol. 15(1):32, 2015.
21. Carp, J., Park, J., Polk, T. A., and Park, D. C., Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. Neuroimage 56(2):736–743, 2011.
22. Coutanche, M. N., Thompson-Schill, S. L., and Schultz, R. T., Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. Neuroimage 57(1):113–123, 2011.
23. Suk, H.-I., Shen, D., Deep learning-based feature representation for AD/MCI classification, in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 583–590, (2013).
24. Zhu, X., Suk, H.-I., and Shen, D., A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. Neuroimage 100:91–105, Oct. 2014.
25. Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., and Li, J., A robust deep model for improved classification of AD/MCI patients. IEEE J. Biomed. Heal. Informatics 19(5):1610–1616, Sep. 2015.
26. S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham, and others, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," IEEE Trans. Biomed. Eng., vol. 62, no. 4, pp. 1132–1140, 2015.
27. Payan, A., Montana, G., Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, arXiv Prepr. arXiv1502.02506, (2015).
28. Liu, M., Zhang, D., Adeli, E., and Shen, D., Inherent structure-based Multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis. IEEE Trans. Biomed. Eng. 63(7):1473–1482, Jul. 2016.
29. Zu, C., Jie, B., Liu, M., Chen, S., Shen, D., Zhang, D., and the A. D. N. Initiative, Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. Brain Imaging Behav. 10(4):1148–1159, Dec. 2016.
30. S. Sarraf, G. Tofighi, and for the A. D. N. Initiative, "DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI," bioRxiv, p. 070441, Aug. 2016.
31. Li, F., Cheng, D., Liu, M., Alzheimer's disease classification based on combination of multi-model convolutional networks, in Imaging Systems and Techniques (IST), 2017 IEEE International Conference on, pp. 1–5, (2017).
32. Amoroso, N., Diacono, D., Fanizzi, A., La Rocca, M., Monaco, A., Lombardi, A., Guaragnella, C., Bellotti, R., and Tangaro, S., Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge. J. Neurosci. Methods 302:3–9, 2018.
33. Liu, M., Cheng, D., Wang, K., Wang, Y.A. D. N., Initiative, and others, "Multi-Modality Cascaded Convolutional Neural Networks

for Alzheimer's Disease Diagnosis," Neuroinformatics, pp. 1–14, (2018).

34. C. Yang, A. Rangarajan, and S. Ranka, "Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification," arXiv Prepr. arXiv1803.02544, 2018.

35. S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of Alzheimer's Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling," J. Med. Syst., vol. 42, no. 5, p. 85, 2018.

36. A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, "3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies," arXiv Prepr. arXiv1801.05968, 2018.

37. Shi, J., Zheng, X., Li, Y., Zhang, Q., and Ying, S., Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J. Biomed. Heal. informatics 22(1):173–183, 2018.

38. S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on, 2014, pp. 1015–1018.

39. Y. Kazemi and S. Houghten, "A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data," in 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2018, pp. 1–8.

40. "ADNI | Alzheimer's Disease Neuroimaging Initiative." .

41. S. T. Creavin, S. Wisniewski, A. H. Noel-Storr, C. M. Trevelyan, T. Hampton, D. Rayment, V. M. Thom, K. J. E. Nash, H. Elhamoui, R. Milligan, A. S. Patel, D. V Tsivos, T. Wing, E. Phillips, S. M. Kellman, H. L. Shackleton, G. F. Singleton, B. E. Neale, M. E. Watton, et al., "Mini-mental state examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations," Cochrane Database Syst. Rev., Jan. 2016.

42. Kim, J. W., Byun, M. S., Sohn, B. K., Yi, D., Seo, E. H., Choe, Y. M., Kim, S. G., Choi, H. J., Lee, J. H., Chee, I. S., Woo, J. I., and Lee, D. Y., Clinical dementia rating orientation score as an excellent predictor of the progression to Alzheimer's disease in mild cognitive impairment. Psychiatry Investig. 14(4):420–426, Jul. 2017.

43. C. Rorden, "dcm2nii DICOM to NIfTI conversion." 2012.

44. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M., FSL. Neuroimage 62(2):782–790, Aug. 2012.

45. Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. M., Bayesian analysis of neuroimaging data in FSL. Neuroimage 45(1): S173–S186, 2009.

46. Smith, S. M., Fast robust automated brain extraction. Hum. Brain Mapp. 17(3):143–155, 2002.

47. Jenkinson, M., Bannister, P., Brady, M., and Smith, S., Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17(2):825–841, 2002.

48. Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M., Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14(6):1370–1386, 2001.

49. Jenkinson, M., and Smith, S., A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5(2): 143–156, 2001.

50. Greve, D. N., and Fischl, B., Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48(1): 63–72, 2009.

51. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

52. He, K., Zhang, X., Ren, S., and Sun, J., Identity mappings in deep residual networks. Cham: Springer, 2016, 630–645.

53. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv Prepr. arXiv1409.1556, 2014.

54. C. Szegedy, W. Liu, and Y. Jia, "C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, arXiv: 1409.4842."

55. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" pp. 3320–3328, 2014.

56. A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition." pp. 806–813, 2014.

57. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L., ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115(3):211–252, Dec. 2015.

58. I. H. Witten, E. Frank, and M. a Hall, Data Mining: Practical Machine Learning Tools and Techniques (Google eBook). 2011.