

RNN-based longitudinal analysis for diagnosis of Alzheimer's disease

Ruoxuan Cui^a, Manhua Liu^{a,b,*}, the Alzheimer's Disease Neuroimaging Initiative

^a Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, 200240 China.

^b Shanghai Engineering Research Center for Intelligent Diagnosis and Treatment Instrument, Shanghai Jiao Tong University, China



ARTICLE INFO

Article history:

Received 8 July 2018

Received in revised form

30 September 2018

Accepted 21 January 2019

Keywords:

Alzheimer's disease diagnosis

Longitudinal analysis

Convolutional neural networks (CNNs)

Recurrent neural network

Magnetic resonance images

ABSTRACT

Alzheimer's disease (AD) is an irreversible neurodegenerative disorder with progressive impairment of memory and other mental functions. Magnetic resonance images (MRI) have been widely used as an important imaging modality of brain for AD diagnosis and monitoring the disease progression. The longitudinal analysis of sequential MRIs is important to model and measure the progression of the disease along the time axis for more accurate diagnosis. Most existing methods extracted the features capturing the morphological abnormalities of brain and their longitudinal changes using MRIs and then designed a classifier to discriminate different groups. However, these methods have several limitations. First, since the feature extraction and classifier model are independent, the extracted features may not capture the full characteristics of brain abnormalities related to AD. Second, longitudinal MR images may be missing at some time points for some subjects, which results in difficulties for extraction of consistent features for longitudinal analysis. In this paper, we present a classification framework based on combination of convolutional and recurrent neural networks for longitudinal analysis of structural MR images in AD diagnosis. First, Convolutional Neural Networks (CNN) is constructed to learn the spatial features of MR images for the classification task. After that, recurrent Neural Networks (RNN) with cascaded three bidirectional gated recurrent units (BGRU) layers is constructed on the outputs of CNN at multiple time points for extracting the longitudinal features for AD classification. Instead of independently performing feature extraction and classifier training, the proposed method jointly learns the spatial and longitudinal features and disease classifier, which can achieve optimal performance. In addition, the proposed method can model the longitudinal analysis using RNN from the imaging data at various time points. Our method is evaluated with the longitudinal T1-weighted MR images of 830 participants including 198 AD, 403 mild cognitive impairment (MCI), and 229 normal controls (NC) subjects from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results show that the proposed method achieves classification accuracy of 91.33% for AD vs. NC and 71.71% for pMCI vs. sMCI, demonstrating the promising performance for longitudinal MR image analysis.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Alzheimer's disease (AD) is an irreversible and chronic neurodegenerative disorder with progressive impairment of the memory and other important mental functions. According to the survey, AD often happens in people aged more than 65, and it has been officially listed as the sixth-leading cause of death in the United States (Ventura et al., 2008). The population of AD is around 90 million at present, which is expected to reach 300 million by 2050 (Zhu et al., 2015). Currently there is no effective cure for AD, but it is of great interest to develop treatments that can delay its progres-

sion, especially if AD can be diagnosed at an early stage when those treatments would have the most impact. Thus, accurate and early diagnosis of AD is important for patient care and treatment. Mild Cognitive Impairment (MCI) is known as a prodromal stage of AD with a high risk of progressing to AD. It can be categorized as progressive MCI (pMCI) and stable MCI (sMCI). pMCI means that MCI subjects will convert to AD after some time while sMCI subjects are stable and will not convert. Recently, machine learning methods have been widely investigated in brain imaging analysis for computer aided diagnosis of diseases (Jiao et al., 2017; Lei et al., 2016, 2017; Liu et al., 2013, 2014; Liu et al., 2015; Wang et al., 2016; Zhang et al., 2018).

Structural magnetic resonance imaging (MRI) is a non-invasive medical imaging modality to capture the internal body structures. It is the most sensitive imaging scan of brain in routine clinical

* Corresponding author.

E-mail address: mhliu@sjtu.edu.cn (M. Liu).

practice to visualize brain anatomical structures. MRI scans provide detailed information about the internal anatomical structures and the morphology of brain tissues such as white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). They have been recognized as an important image biomarker for AD progression and been widely studied to develop computer-aided diagnosis systems using pattern recognition methods (Hinrichs et al., 2009, 2011; Hosseini-Asl et al., 2016; Jiao et al., 2017; Kloppel et al., 2008, epub; Liu et al., 2016, 2018; Yu et al., 2017). The raw brain images are too huge and noisy to be directly used for diagnosis. To facilitate the morphological analysis, MR brain image is partitioned into multiple anatomical regions, i.e., regions of interest (ROIs), by grouping voxels through the warping of a labeled atlas and then the regional measurements are computed as the features for image classification (Liu et al., 2015; Zhang et al., 2011). The ROI-based feature extraction can significantly reduce the feature dimension and provide robust representations, but some subtle abnormal changes may be ignored. In addition, the ROIs are made by prior hypotheses and the abnormal brain regions relevant to AD might not fit well to the pre-defined ROIs, thus limiting the representation power of extracted features. To capture richer information, voxel-wise features were extracted after registering all brain images to associate each voxel with a vector of scalar measurements for AD diagnosis (Kloppel et al., 2008, epub). The brain volume is segmented to three types of tissues, i.e., GM, WM, and CSF, and the voxel-wise tissue density maps are computed for classification. The voxel-wise features can capture the subtle abnormal changes, but they are of huge dimensionality, far more features than training subjects.

Recently, deep learning methods have gained a good reputation especially to extract informative features for computer vision and medical image analysis (Shen et al., 2017). Instead of manually extracting features on the expert's knowledge about the target domain, deep learning can discover the latent and discriminant representations of image data by incorporating the feature extraction into the task learning process. In addition, deep learning can construct multi-layer neural networks to transform image data to task outputs (e.g., disease/normal) while learning hierarchical feature representations from data. Thus, complex patterns can be discovered with deep learning. Convolutional neural networks (CNNs) had been studied for AD diagnosis (Adrien, 2015; Hosseini-Asl et al., 2016). Hosseini-Asl et al. (Hosseini-Asl et al., 2016) have proposed a deep 3D-CNNs to learn generic features to predict AD using the structural MRI scans. In this method, the deep 3D-CNNs were built upon a 3D-CAES (convolutional Autoencoders) pre-trained with the rigidly registered training images to capture anatomical shape variations, followed by fully connected and softmax layers for classification. Adrien et al. (Adrien, 2015) proposed a deep learning classification algorithm for AD diagnosis using both the structural MRI and functional MRI. In this method, the CNN model was built with one convolutional layer trained with sparse Autoencoder, which were explored to extract the imaging features for AD classification.

Most of existing methods in recent studies have focused on using the imaging data from single-time point to detect the pathological changes for AD diagnosis. In practice, there are also increasing amount of longitudinal data collected at the follow-up time points which often provide useful information about the pathological progression of disease. With the availability of these longitudinal image data at multiple time points, it is possible to use them for improving AD prediction (Jiao et al., 2017; Li et al., 2012; Zhang and Shen, 2012). In (Li et al., 2012), the longitudinal cortical thickness changes are extracted for analyzing the development of pathology in AD. In addition, both baseline and longitudinal biomarkers are combined and utilized to more accurately predict the future clinical changes of MCI patients (Zhang and Shen, 2012). Although these methods can incorporate the information from multiple time-

points for improving the classification accuracy, they require that every subject should have longitudinal images at all time points, and a subject with incomplete data (which occurs very often in the longitudinal study) cannot be used. This will throw away a large amount of useful information and reduce the number of training samples. It is important to investigate how to use the incomplete longitudinal data.

The longitudinal analysis of sequential MR images is important to model and measure the disease progression along the time axis for more accurate diagnosis. In this work, we present a classification framework based on combination of convolutional and recurrent neural networks for longitudinal analysis of structural MR images at variable-length time series for AD diagnosis. First, a deep 3D CNN model is constructed to learn the spatial features from the structural MR images for the task of disease classification. After that, a deep RNN model with cascaded three bidirectional gated recurrent units (BGRU) layers is constructed on the outputs of CNN from the imaging data of multiple time points for extracting the longitudinal features, leading to a classification predicting score. BGRU can process the varying length image sequences to alleviate the problem of incomplete longitudinal data. The proposed method can automatically learn the spatial and longitudinal features from the imaging data at multiple time points of variable length for classification. Our method is evaluated using T1-weighted structural MR brain images on 830 subjects including 198 AD patients, 403 MCI (including 236 sMCI and 167 pMCI) subjects, and 229 normal controls (NC) from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results demonstrate the promising performance of the proposed method.

The rest of this paper is organized as follows. Section 2 presents the materials of longitudinal imaging data used in this paper and the proposed deep learning method. Experiments and results are provided in Section 3. Section 4 concludes this paper.

2. Materials and method

In this section, we will present the brain image sets used in this work and the proposed classification algorithm. The MR image is a powerful brain imaging modality often used as biomarkers to help physicians for AD diagnosis. It is still challenging to make use of the high-dimensional and longitudinal image data to improve AD diagnosis. In this work, we propose a deep learning framework based on combination of CNN and RNN to integrate the spatial and longitudinal features of MR images for AD diagnosis. There are three main advantages to apply the deep learning model for our task. First, the deep architecture of CNNs can explicitly make use of the spatial structure of brain images and gradually extract the spatial features from the low-, mid- to high-levels useful for classification task. Second, cascading RNN can model and measure the disease progression with the images at multiple time points. Finally, the spatial and longitudinal features are combined into a classification framework for more accurate diagnosis. Fig. 1 shows the flowchart of the proposed classification algorithm based on combination of CNN and RNN, which consists of three main steps: image processing, spatial feature learning by 3D CNN, RNN based longitudinal analysis and final classification, as detailed below.

2.1. Materials

In this study, the data used were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which can be freely downloaded from the website (www.loni.ucla.edu/ADNI). The ADNI database 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 (FDA),

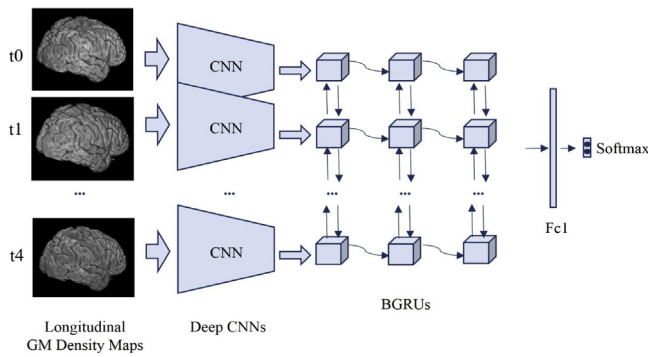


Fig. 1. The flowchart of our proposed classification framework, where a deep CNN is constructed to capture the spatial structural features of the images from each time point while the stacked BGRUs are trained to capture the longitudinal features of the multiple time points. On top of stacked BGRU layers, one fully connected layer and softmax layer are appended to make final classification prediction.

Table 1
Demographic characteristics of the studied subjects from ADNI database (The values are denoted as mean \pm standard deviation).

Diagnosis	Number	Age	Gender (M/F)	MMSE
AD	198	75.65 \pm 7.73	103/95	23.36 \pm 1.99
pMCI	167	74.89 \pm 6.83	102/65	26.53 \pm 1.70
sMCI	236	74.89 \pm 7.74	158/78	27.17 \pm 1.83
NC	229	75.99 \pm 5.01	119/110	29.09 \pm 0.98

private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public–private partnership. The primary goal of the ADNI was to test whether serial MRI, Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression was intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, M.D., VA Medical Center and University of California, San Francisco. ADNI was the result of efforts of many co-investigators from a broad range of academic institutions and private corporations. The study subjects were recruited from over 50 sites across the U.S. and Canada and gave written informed consent at the time of enrollment for imaging and genetic sample collection and completed questionnaires approved by each participating sites Institutional Review Board (IRB). The initial goal of ADNI was to recruit 800 adults, aged from 55 to 90, to participate in the research.

Although the proposed method makes no assumption on a specific neuroimaging modality, MR images is widely available, non-invasive and often used as the first biomarker in the diagnostics of AD. In ADNI, MR images were acquired according to the ADNI acquisition protocol in (Jack et al., 2008). The MR image sets included standard T1-weighted MR images acquired sagittally using volumetric 3D MPRAGE with $1.25 \times 1.25 \text{ mm}^2$ in-plane spatial resolution and 1.2 mm thick sagittal slices. Most of these images were obtained with 1.5T scanners, while a few were acquired using 3 T scanners. Detailed information about MR acquisition procedures is available at the ADNI web site. In this work, we use the T1-weighted MR brain image data acquired with 1.5 T scanners from 830 ADNI participants including 198 AD, 403 MCI (236 sMCI and 167 pMCI), 229 NC subjects to test our proposed method. Table 1 presents the demographic details of the studied subjects in this paper, where MMSE denotes the Mini-Mental State Examination (MMSE). These subjects have MR imaging data at different numbers of time-points. The maximum number of time points is 6 including baseline, 6-month, 12-month, 18-month, 24-month and

36-month, denoted as t_0 , t_1 , t_2 , t_3 , t_4 and t_5 , respectively. Table 2 shows the different numbers of the available studied subjects at different time points from ADNI database.

The T1-weighted MR brain images of multiple time points were preprocessed to make the images from different systems more similar before performing classification. All MR images were preprocessed by applying the typical procedures of Anterior Commissure (AC)–Posterior Commissure (PC) correction, skull-stripping, and cerebellum removal (Wang et al., 2011a). Specifically, a correction of intensity inhomogeneity was first performed on the MR images using nonparametric nonuniform intensity normalization (N3) algorithm (Sled et al., 1998). Secondly, a robust and automated skull stripping method (Wang et al., 2011b) was performed for brain extraction and cerebellum removal followed by manually checking the skull-stripped images to ensure the clean and dura removal. Thirdly, each brain image is segmented into three kinds of tissue volumes, e.g., gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) volumes. The three tissue volumes of different time-point images of each subject will be spatially normalized together onto a standard space by a mass-preserving deformable warping algorithm called HAMMER (Shen and Davatzikos, 2003). During the image warping, the tissue volume within any size of region is preserved, i.e., it is increased if the region is compressed and vice versa. The warped mass-preserving tissue volumes reflect the spatial distribution of tissues in the original brain by taking into consideration the local tissue volumes prior to warping. The warped mass-preserving tissue volumes are also called as the tissue density maps. Since GM is more related to AD and MCI than WM and CSF, the GM density map is used as the input image for the following deep learning in this work.

2.2. Feature learning with 3D CNNs

Different from the conventional methods which directly use GM density map as the feature for classification, deep convolutional neural networks (CNN) are used to learn the spatial features from GM density map in this work. CNN is a special kind of multi-layer neural networks trained with a version of the back-propagation algorithm. It has been widely used in several domains such as image classification and object detection (He et al., 2015; Krizhevsky et al., 2012; LéCun et al., 1998). However, most of these mature CNN architectures are studied for 2D image recognition. In this work, CNN with the 3D convolutional kernel is employed to learn the rich features of 3D brain image. Typically, a deep CNN for feature extraction alternatively stacks several convolutional and sub-sampling layers followed by fully connected layers. In this work, we have made some changes on the traditional CNN network to learn the rich features better for image classification. First, to make use of the multi-level features, the features extracted from both low and high level convolutional layers are combined by shortcut connections, which concatenate the feature maps from different depths of CNN model. Second, in the pooling layer, we use the convolutional operation with $2 \times 2 \times 2$ stride instead of maximum operation to subsample the feature maps, which can learn the linear combination of the neurons for pooling. Fig. 2 shows the deep 3D CNN architecture in this work, which is composed of convolution block, full-connected block and Softmax layer. The convolution block includes the convolution, batch normalization, ReLU (Rectified Linear Unit) activation, dropout, sub-sampling layers while full-connected block includes three full-connected layers. Accordingly, Table 3 presents the structure details of deep 3D CNNs.

A typical convolutional layer convolves the input image with a set of kernel filters, followed by adding a bias term and applying a non-linear activation function, and finally produce a feature map

Table 2
Numbers of the available studied subjects at different time points from ADNI database.

Diagnosis	t0	t0+t1	t0+t1+t2	t0+t1+t2+t3	t0+t1+t2+t3+t4	t0+t1+t2+t3+t4+t5
AD	198	171	143	91	0	0
PMCI	167	164	154	140	97	14
SMCI	236	208	185	147	93	8
NC	229	215	204	161	16	0

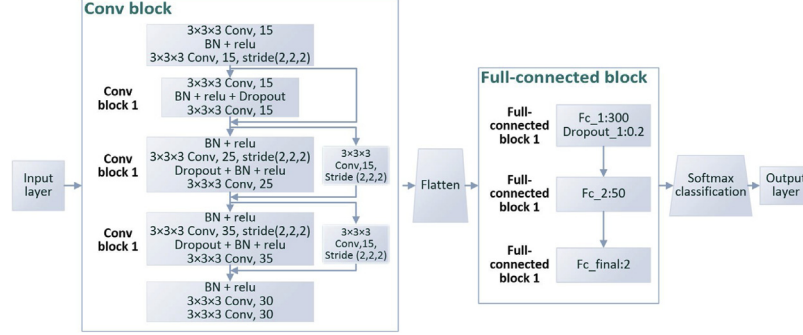


Fig. 2. The architecture of deep 3D CNN model consisting of the convolution block (denoted as “Conv block”), the full-connected block and Softmax layer. The convolution block includes the convolution, batch normalization, ReLU activation, dropout, sub-sampling layers while the full-connected block includes three full-connected layers.

Table 3
The structure details of deep 3D CNNs.

Layer ID	Layer name	Detail information	Output size	Parameter number
0	input		(1,50,42,42)	0
1	Conv_0.1	Padding: same	(15,50,42,42)	420
2	Bn_0		(15,50,42,42)	60
3	Conv_0.p	Stride: (2,2,2)	(15,24,20,20)	6090
4	Conv_1.1	Padding: same	(15,24,20,20)	6090
5	Bn_1		(15,24,20,20)	60
6	Conv_1.2	Padding: same	(15,24,20,20)	6090
7	Merge.1	(Conv_0.p, Conv_1.2)	(30,24,20,20)	0
8	Bn_2.1		(30,24,20,20)	120
9	Conv_2.1	Stride: (2,2,2)	(25,12,10,10)	20275
10	Bn_2.2		(25,12,10,10)	100
11	Conv_2.2	Padding: same	(25,12,10,10)	16900
12	Conv_2.cut	Stride: (2,2,2)	(15,12,10,10)	12165
13	Merge.2	(Conv_2.2, Conv_2.cut)	(40,12,10,10)	0
14	Bn_3.1		(40,12,10,10)	160
15	Conv_3.1	Stride: (2,2,2)	(35,6,5,5)	37835
16	Bn_3.2		(35,6,5,5)	140
17	Conv_3.2	Padding: same	(35,6,5,5)	33110
18	Conv_3.cut	Stride: (2,2,2)	(25,6,5,5)	27025
19	Merge.3	(Conv_3.2, Conv_3.cut)	(60,6,5,5)	0
20	Bn_4		(60,6,5,5)	240
21	Conv_4.1	Padding: valid	(30,4,3,3)	48630
22	Conv_4.2	Padding: valid	(30,2,1,1)	24330
23	Flatten		60	0
24	Full-connected 1		300	18300
25	Full-connected 2		50	15050
26	Full-connected 3		2	102
27	softmax		2	0

for each filter. The 3D convolution operation is defined as:

$$u_{kj}^l(x, y, z) = \sum_{\delta_x} \sum_{\delta_y} \sum_{\delta_z} F_k^{l-1}(x + \delta_x, y + \delta_y, z + \delta_z) \times W_{kj}^l(\delta_x, \delta_y, \delta_z) \quad (1)$$

where x, y and z denote the pixel positions for a given 3D image, $W_{kj}^l(\delta_x, \delta_y, \delta_z)$ is the j th 3D kernel weight connecting the k th feature map of the $l-1$ layer to the j th feature map of the l layer, F_k^{l-1} is the k th feature map of the previous $l-1$ layer, and $\delta_x, \delta_y, \delta_z$ are the kernel sizes corresponding to the x, y and z coordinates. The output $u_{kj}^l(x, y, z)$ is the convolutional response of kernel fil-

ter. Then, ReLU activation is applied after each convolution layer. The j th 3D feature map of l layer $F_j^l(x, y, z)$ is obtained by summation of the response maps of different 3D convolution kernels. By using 3D kernel to consider the spatial correlations, the 3D CNNs can take full advantages of the volumetric contextual information to learn the spatial features.

The convolutional layer is often followed by pooling layer. Taking small block of feature maps, the pooling layer produces a single output for the block. There are several ways for the pooling, such as taking the average value or the maximum, or a linear combination of the neurons in the block. To learn the features, we apply the convolution operation with $2 \times 2 \times 2$ stride to learn a linear combination of the neurons for pooling. Through pooling, the features become

more compact and efficient from low to high layers. The third type of layer is the fully connected layers. After several convolutional and pooling layers, the high-level reasoning of neural network is done by fully connected layers. All 3D feature maps are flattened and concatenated into a 1D vector as the input of fully connected layer. Each neuron of fully connected layer outputs the linear combination of all neurons from the previous layer and passed through a nonlinearity. The inputs and outputs of fully connected layers are concatenated into one-dimensional vector and are not spatially located.

Finally, a softmax layer is appended to the last fully connected layer and is fine-tuned back-propagation with negative log-likelihood to predict class probability. The softmax function is a derivation of logistic function that highlights the largest values in a vector while suppressing those significantly below the maximum. The softmax layer has been often used as the last layer of CNN architecture to generate a prediction score for each class label. The output of each node ranges from 0 and 1, and the sum of all the nodes is always 1.

In our implementation, the 3D deep CNN is built by stacking input block, Conv block, full-connected block and Softmax layer as shown in Fig. 3. Batch normalization and ReLU activation are added after each convolutional layer to improve the performance of the whole 3D CNN model. Feature subsampling is conducted by $2 \times 2 \times 2$ stride with convolution layers. The 3D convolutional kernels are randomly initialized from the Gaussian distribution and other trainable parameters are tuned using the standard back-propagation with stochastic gradient descent by minimizing the cross-entropy loss. In addition, the dropout strategy is used to reduce the co-adaption of intermediate features and overfitting problem and improve the generalization capability. For each time point, a deep CNN is constructed on the GM density maps to learn the features for the task of disease classification.

2.3. RNN based longitudinal analysis

With the increasing availability of longitudinal image data, the dynamic features that directly describe the temporal changes of GM tissue volumes can also be extracted to provide more information about the pathological development. The deep 3D CNN is constructed to learn the features of single point. To combine the features of multiple time points, one direct method is to concatenate the learned features by deep CNNs from multiple time points and design a classifier to make the final classification. However, this method ignores the correlation information between the sequential features. Recurrent neural networks (RNN) are effective to process and model the sequential data. To fully utilize the longitudinal image information, we propose the stacked RNNs to model and measure the temporal variations about the disease progression as shown in Fig. 3. These longitudinal measures will be used as the supplementary features in this study.

A RNN is a class of artificial neural network in which connections between nodes form a directed graph along a sequence. Different from traditional neural networks which assume that all inputs (and outputs) are independent of each other, RNN can model the dynamic temporal behavior for a time sequence with the output being dependent on the previous computations. The time sequences of inputs can be processed with the internal state (memory) of RNNs, which captures information about what has been computed so far. In theory, RNNs can make use of information in arbitrarily long sequences, but in practice they are limited to only a few steps. For longitudinal analysis of structural MR images, RNN is built to make use of the sequential MR images at multiple time points, which can model the progression of the disease. In the past decades, the performance of RNN was severely restricted due to the difficulty of training. Gradient mass and explosion are

the common unsolved problems until the emergence of LSTM—a special kind of RNN structure. In the traditional chain structure, a simple *tanh* activation unit is repeated in each layer. LSTM modifies it to a carefully designed structure, which contains three gate units and a memory cell unit. The three gates are *forget gate*, *input gate* and *output gate*, respectively. By updating the state of memory cell through three gates, LSTM can discard irrelevant information and effectively capture the helpful information in sequence. Gate recurrent unit (GRU) was proposed as a special kind of variants for LSTM (Cho et al., 2014). Through removing the memory cell from the original LSTM, GRU makes RNN simpler without degrading performance. It was concluded that GRU has a slightly better performance than LSTM (Chung et al., 2014). Thus GRU is used to build the RNN model.

Compared to LSTM, GRU has only two gates: *update gate* z and *reset gate* r , thus it has the advantages of less parameters and easier training. The *forget* and *output* gates of LSTM are merged into a single *update gate* z , which is used to get the current state of the output via linear interpolation in GRU, as shown in Fig. 3. When the input (x_i, h_{i-1}) denotes the features of the i^{th} time point and the previous hidden state, the *update gate* z and *reset gate* r are computed as:

$$z_i = \sigma(W^{xz}x_i + W^{hz}h_{i-1}) \quad (2)$$

$$r_i = \sigma(W^{xr}x_i + W^{hr}h_{i-1}) \quad (3)$$

where W^{xz} , W^{hz} , W^{xr} and W^{hr} are the corresponding weight matrices; σ is a logistic sigmoid function. The candidate state of the hidden unit is computed by:

$$\tilde{h}_i = \tan(W^{xh}x_i + W^{hh}(h_{i-1} \odot r_i)) \quad (4)$$

where \odot is an element wise multiplication. When r_i is close to 0 (off), the reset gate r effectively makes the unit act as if it is reading the first symbol of an input sequence, allowing it to forget the previously computed state. The i^{th} hidden activation state h_i of GRU is a linear interpolation between the previous state h_{i-1} and the candidate state \tilde{h}_i :

$$h_i = \tan(W^{xh}x_i + W^{hh}(h_{i-1} \odot r_i)) \quad (5)$$

The MR image of each time point is correlated with those of both its preceding and following time point in capturing the temporal variation. Thus, we apply the Bidirectional-GRU (BGRU), which consists of a forward GRU (i is from 1 to n) and a backward GRU (i is from n to 1), to capture correlation features of longitudinal image data. The inputs of each BGRU include the features of multiple time points generated by 3D CNNs, i.e., the outputs of fully connection layer. The outputs of both forward and backward GRUs are concatenated together to form the outputs of BGRU at the same time. In addition, to further enhance the longitudinal information flow in the network, three layers of BGRUs are stacked into a deep structure by taking the outputs of one BGRU as the input of another BGRU. Fig. 3 shows the flow chart of deep RNN network architecture based on cascaded BGRUs. In our deep learning method, RNN integrates the features of multiple time points to learn the longitudinal features. The first BGRU layer takes the feature vectors of multiple time points produced by CNN. The following 2 BGRU layers are cascaded to learn the high-level features of the time series features. The first two BGRU layers produce the outputs of multiple sequences, while the third BGRU layer generates single output. At last, the learned features from the outputs of BGRUs are concatenated to a fully connected layer and softmax layer for the classification task.

Table 4 shows the deep RNN network structures based on cascaded BGRUs, which consists of one masking layer, three BGRU layers, one full-connected layer and a softmax layer, generating a longitudinal classification prediction. For the longitudinal analysis with multiple time points, the MR images of some subjects

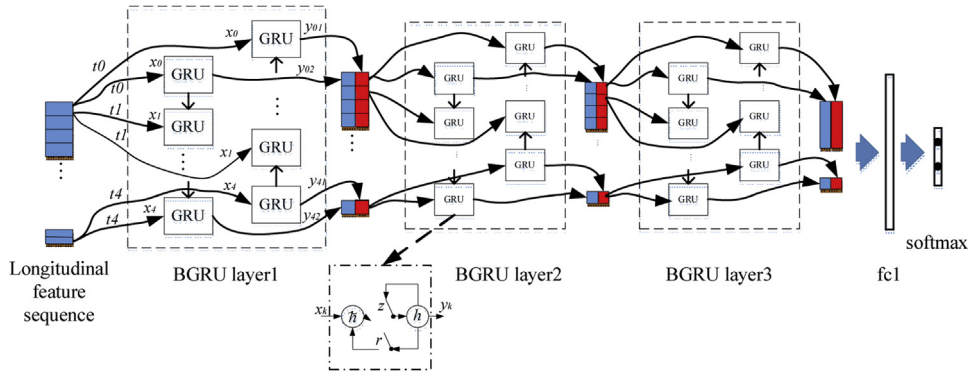


Fig. 3. Our deep recurrent neural network (RNN) based on three stacked layers of BGRU, where x_i is the feature vector generated from image of time point i as the inputs of BGRU while y_{i1} and y_{i2} are the output feature vectors of GRU from two directions (The internal structure of GRU is shown in the bottom, where r_i , z_i , h_i and h_i are the reset gate, update gate, candidate hidden layer and output layer, respectively). On top of stacked BGRU layers, one fully connected and one softmax layers are appended for classification prediction.

Table 4
The deep RNN network structures based on cascaded BGRUs.

Layer ID	Layer name	Detail information	Output size	Parameter number
0	Input		(5, 300)	0
1	Masking	Mask value = -1	(5, 300)	0
2	BGRU_1	300	(5, 600)	1081800
3	BGRU_2	200	(5, 400)	961200
4	Dropout_1	0.2	(5, 400)	0
5	BGRU_3	50	100	135300
6	Full_connected	100	50	5050
7	Dropout_2	0.2	50	0
8	Softmax		2	102

may be missing in the later time points. To address the missing data problem in longitudinal analysis, we take the advantage of the RNN's capability on analyzing sequence inputs of varying lengths with missing time points. The masking layer after the input layer is added to deal with the missing input data. In our implementation, the missing time points for every subject are padded with negative meaningless values (-1 in our experiments) in the masking layer so that the input sequences are all the same length for the RNN modelling. The RNN model can learn the negative values carry no information so that the sequences are not of the same length in terms of content, but the inputs of same length are required to perform the computation. Finally, the RNN model generate the features of same length.

2.4. Final classification

Construction of the whole deep network includes training of 3D CNN and RNN models. In our implementation, the deep 3D CNN models are pre-trained with the GM density data at all time points and then fine-tuned with the data at each time point, which can not only capture the changes of longitudinal data at different time points but also maintain the consistency of the learned features input to the BGRU network. The softmax function is used to directly map the outputs of the fully connected layer to the prediction scores of all class labels. Then, the trained parameters of 3D CNNs are used to fix all convolution and pooling layers of the 3D CNNs, while the parameters of RNN are fine-tuned jointly with the upper fully connected and softmax prediction layers. Finally, in the classification learning process, the initial-trained parameters of 3D CNNs and RNN are fixed, while the parameters of the upper fully connected layers and softmax prediction layer are fine-tuned jointly to combine the spatial and longitudinal features for the task-specific classification. The jointly fine training is included to adjust

the parameters to be able to handle the heterogeneity and produce a reliable estimate from the longitudinal images.

However, training a deep CNN is challenged by the risk of over-fitting as the current datasets for AD diagnosis are relatively small compared to other computer vision tasks such as face recognition. A common practice is to initialize the weights with the pre-trained models on some large dataset. To mitigate the problem, we adopted dropout layer for regularization (Srivastava et al., 2014). Fig. 4(a) and (b) show the curves of loss convergence for the CNN network training and validation on the classifications of AD vs. NC and pMCI vs. sMCI, respectively, while Fig. 4(c) and (d) show the curves of loss convergence for the RNN network training and validation on the classifications of AD vs. NC and pMCI vs. sMCI, respectively. From the results, we can see that the loss convergence for AD vs. NC classification is faster than that of pMCI vs. sMCI classification. This is because the classification task of pMCI vs. sMCI is more challenging than that of AD vs. NC. The over-fitting problem has been alleviated by using some dropout techniques and data augmentation in our proposed network. For training the deep CNN model, the batch size is set to 128, and the model begins to converge after iteration of 150 epochs. When training the BGRU model, the batch size is set as 60 and the model begins to converge after iteration of 500 epochs for classification of AD vs. NC.

3. Experimental results

3.1. Experiments

In this section, experiments are conducted to test the proposed classification algorithm based on combination of convolutional and recurrent neural networks for longitudinal analysis of structural MR images. The image sets used in this study and the corresponding image processing steps are illustrated in Section 2. There are MR images of 830 subjects including 198 AD, 167 pMCI, 236 sMCI and 229 NC subjects from ADNI. The GM density map size after above image processing is of $256 \times 256 \times 256$ voxels. We remove those pixels with zero intensity values to obtain the image size $200 \times 168 \times 168$ voxels. These images are subsampled by 4 as the inputs of proposed classification framework. Data augmentation is conducted by subsampling the image by shifting just ± 1 voxels along the three coordinates to generate additional images for training. Augmentation is not used on the validation and test sets.

In this work, the proposed classification algorithm is implemented with the Keras library based on Tensorflow in Python. There are 5 deep CNNs independently trained to extract the spatial features for the individual time points with the output of the prediction scores for disease classification. The initial weights for the

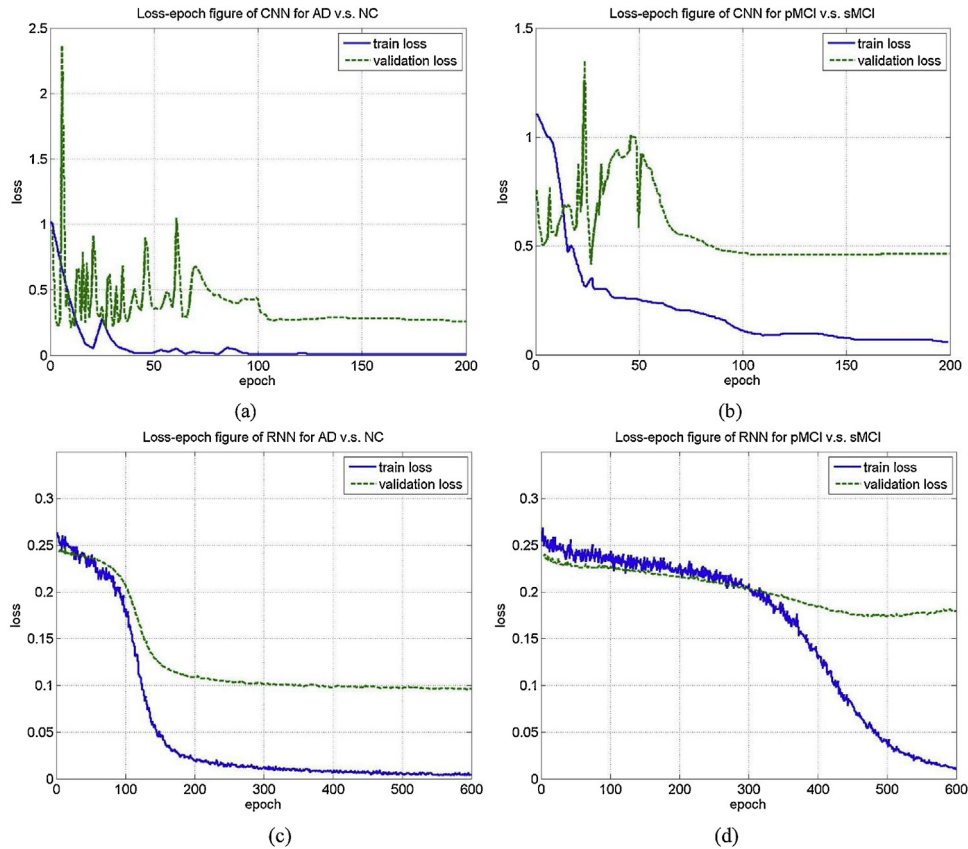


Fig. 4. The curves of loss convergence for CNN & RNN training and validation on the classification tasks: (a) CNN on AD vs. NC; (b) CNN on pMCI vs. sMCI; (c) RNN on AD vs. NC; and (d) RNN on pMCI vs. sMCI.

whole network is Xavier uniform. The Adadelta gradient descent algorithm (Zeiler, 2012) is used to train the deep CNNs with 0.8 learning rate. The parameter ρ is set to 0.7, ϵ is set to 1×10^{-8} with weight decay being 0. We use batch normalization layers, each followed with a *ReLU* activation layer, and the dropout layer parameter is set 0.2. As for training the BGRUs, Adam optimizer method (Kingma and Ba, 2014) is adopted for stochastic optimization. The first fully connected layer has 50 hidden neurons, followed with an ELU activation layer, and the final binary output layer is softmax classification function. To avoid overfitting problem, dropout is adopted in our network (Srivastava et al., 2014).

In our experiments, the proposed algorithm is tested on classifications of AD vs. NC and pMCI vs. sMCI. 5-fold cross-validation is used to avoid random factors affecting the results. Each time, one fold of the image set is used for testing, another one fold is used for validation while the left 3 folds are used for training. The validation part is used for stopping the training process to obtain the model weights with the optimized performance. To evaluate the classification performance, we compute the classification accuracy (ACC), the sensitivity (SEN), the specificity (SPE), and the area under receiver operating characteristic curve (AUC) as the performance measures. The receiver operating characteristic (ROC) curves are also shown for comparison.

3.2. The effectiveness of CNN and BGRU

In the first experiment, we test the effectiveness of CNN and RNN models on the classification performance. First, we only perform the CNN classification model on the baseline GM data without the following RNN model. Second, instead of using the CNN to extract the high-level features of GM density map, we use the traditional

principal component analysis (PCA) to extract the features as inputs to train the RNN model based on cascaded BGRUs for longitudinal analysis of the image data. For fair comparison, the number of components reduced by PCA is same as the number of features extracted by CNN, which is 300 in our experiments. Finally, we test our proposed classification algorithm based on combination of CNN and RNN models to extract the spatial and temporal features for classification. Fig. 5 shows the comparison of their corresponding ROC curves. Table 5 demonstrates the performance comparison of CNN, RNN, and their combination for classifications of AD vs. NC and pMCI vs. sMCI. These results show that the proposed algorithm based on combination of CNN and RNN can improve the classification performance compared to using CNN and RNN individually.

The second experiment is to test the performance improvement by using RNN for the longitudinal analysis with the MRI data from different number of multiple time points on classifications of AD vs. NC and pMCI vs. sMCI, as shown in Table 6. In the experiment, we test the proposed method by gradually adding more sequential MRIs from different number of time points for the RNN based longitudinal analysis. In Table 6, t_0 , t_1 , t_2 , t_3 , t_4 and t_5 denote the baseline, 6-month, 12-month, 18-month, 24-month and 36-month MRI brain images, respectively. From these results, we can see that the classification performances are gradually improved by adding more sequential data for longitudinal analysis.

3.3. Comparison with other methods

In this section, we compare the proposed method with other existing methods to show the performance. First, we compare our proposed method to two state-of-the-art longitudinal analysis methods. The same number of training and test subjects and the

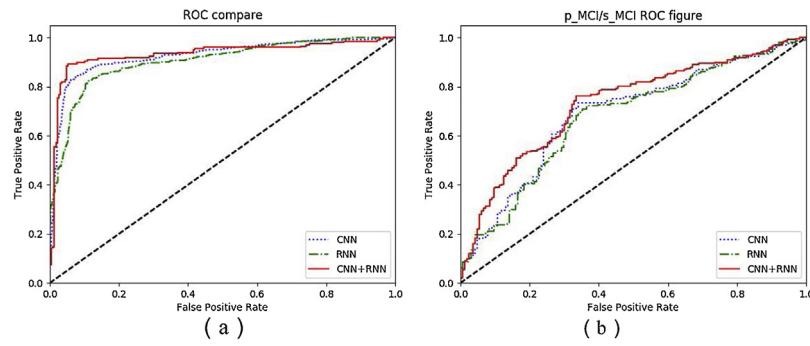


Fig. 5. Comparisons of ROC curves on classifications of (a) AD vs. NC and (b) pMCI vs. sMCI with CNN only, RNN, and the combination of CNN and RNN (denoted as “CNN + RNN”).

Table 5

The performance comparison of CNN, RNN, and the combination of CNN and RNN.

Perf. (%)	AD vs. NC (%)				pMCI vs. sMCI (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
CNN	88.99	84.85	92.58	92.53	70.22	62.87	75.42	68.88
RNN	85.01	82.32	87.34	90.30	68.49	64.67	71.19	67.52
CNN + RNN	91.33	86.87	95.20	93.22	71.71	65.27	76.27	73.03

Table 6

The performance comparison with the MRI data from different numbers of time points.

Perf. (%)	AD vs. NC (%)				pMCI vs. sMCI (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
t0	88.52	84.85	91.70	90.35	69.48	62.87	74.15	68.78
t0+t1	89.93	85.35	93.89	91.43	69.73	62.28	75.00	69.56
t0+t1+t2	90.16	85.35	94.32	91.89	70.22	63.47	75.00	71.03
t0+t1+t2+t3	90.40	86.87	93.45	92.78	70.97	64.07	75.85	71.57
t0+t1+t2+t3+t4	91.33	86.87	95.20	93.22	71.22	64.67	75.85	72.89
t0+t1+t2+t3+t4+t5	–	–	–	–	71.71	65.27	76.27	73.03

Table 7

The performance comparison of longitudinal analysis with different number of time points for classification of AD vs. NC.

Method	Evaluation	t0	t0+t1	t0+t1+t2	t0+t1+t2+t3	t0+t1+t2+t3+t4
SVM averaging	ACC	85.48	85.71	86.18	86.65	86.65
	SEN	80.81	80.81	81.31	81.82	81.31
	SPE	89.52	89.96	90.39	90.83	91.27
CNN averaging	ACC	87.35	88.06	88.52	88.76	88.99
	SEN	82.32	83.33	84.34	84.34	84.85
	SPE	91.70	92.14	92.14	92.58	92.58
BGRU	ACC	88.52	89.93	90.16	90.40	91.33
	SEN	84.85	85.35	85.35	86.87	86.87
	SPE	91.70	93.89	94.32	93.45	95.20

Table 8

The performance comparison of longitudinal analysis with different number of time points for classification of pMCI vs. sMCI.

Method	Evaluation	t0	t0+t1	t0+t1+t2	t0+t1+t2+t3	t0+t1+t2+t3+t4	t0+t1+t2+t3+t4+t5
SVM averaging	ACC	64.52	65.01	65.26	65.26	65.51	66.00
	SEN	53.89	54.49	53.89	55.09	53.89	55.09
	SPE	72.03	72.46	73.31	72.46	73.73	73.73
CNN averaging	ACC	68.24	68.49	68.49	69.23	69.73	70.22
	SEN	61.68	62.87	61.08	62.28	62.87	62.87
	SPE	72.88	72.46	73.73	74.15	74.58	75.42
BGRU	ACC	69.48	69.73	70.22	70.97	71.22	71.71
	SEN	62.87	62.28	63.47	64.07	64.67	65.27
	SPE	74.15	75.00	75.00	75.85	75.85	76.27

same original features are used for fair comparison. We implement these methods with our best efforts. One direct method for longitudinal analysis is to cascade the extracted original features at each time point into a long vector, and then employ principal component analysis (PCA) for feature reduction on the cascaded features followed by a support vector machine (SVM) classifier. The classifier

outputs at multiple time points are ensemble by averaging for final classification. This method is denoted as “SVM Averaging”. In addition, we also compare our proposed method with another related method, which directly combine the results of 3D CNN from multiple time points by using averaging (denoted as “CNN averaging”), instead of modelling the longitudinal features with cascaded BGRU.

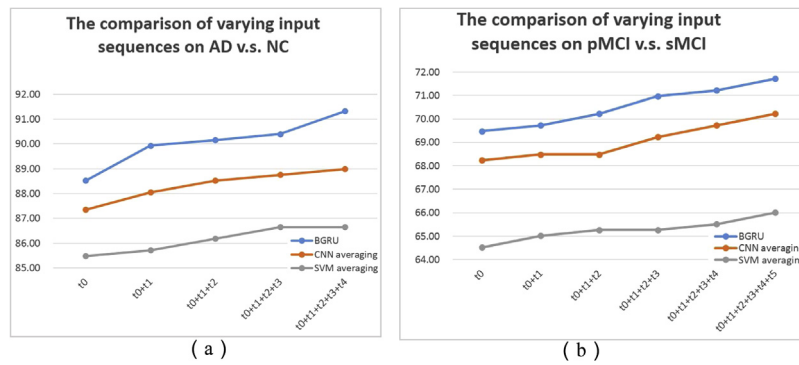


Fig. 6. The comparison of classification performances for longitudinal analysis with the input sequences of varying length for classifications of (a) AD vs. NC and (b) pMCI vs. sMCI.

These methods are tested with different numbers of time points by gradually adding one time point from the baseline, 6-month, 12-month, 18-month, and 24-month brain images. Table 7 and 8 demonstrates the performance comparison of longitudinal analysis with different numbers of time points for classification of AD vs. NC and pMCI vs. sMCI, respectively. Accordingly, Fig. 6(a) and (b) show the comparison of classification accuracy improvement when increasing the number of input sequences for longitudinal analysis in classifications of AD vs. NC and pMCI vs. sMCI, respectively. From these results, we can observe that the classification accuracy is improved by increasing the number of time points and our method performs better than the CNN averaging method for different lengths of input sequence, especially with 5 time points. It shows that BGRU can learn the longitudinal features from the time sequential data to improve the classification.

3.4. Discussion

Although there are many methods proposed to extract the spatial and longitudinal feature representations in longitudinal analysis of sequential MRIs for computer-aided AD diagnosis, most of them are based on hand-crafted measures that require time-consuming pre-processing procedures such as segmentation of ROIs. To this end, we develop a classification framework based on combination of CNN and RNNs for longitudinal analysis of structural MR images at variable-length time series for AD diagnosis. The deep 3D CNN is constructed to learn the spatial features of the structural MR images for the task of AD classification, while the RNN with cascaded three BGRU layers is constructed on the outputs of CNN at multiple time points for extracting the longitudinal features for AD classification. Instead of independently performing feature extraction and classifier training, the proposed method is a data-driven method to jointly learn the spatial and longitudinal features and disease classifier model for optimal classification performance. In addition, different from existing methods which cannot deal with the missing data at some time points, the proposed method can take advantages of RNN to model the longitudinal image analysis with variable length of input sequences.

As for the computation cost, the proposed deep network model includes both the training and testing stages. In the training stage, it takes about 25 min to train the proposed deep network model. In the test stage, it takes 8.5 ms to test for a given image. All experiments are conducted on PC with GPU NVIDIA GeForce GTX1080Ti of 12GB memory.

Since some pathological patterns are affected by brain disease, only a subset of image regions may be closely related to the diagnosis of AD. Thus, it is also important to identify these regions for diagnosis and interpretations of the diseases. To achieve this, we adopt the visualization technique proposed in (Simonyan et al.,

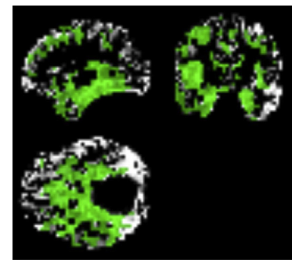


Fig. 7. The visualization of the regions most related to AD diagnosis by using the proposed classification method.

2013) to generate the image-specific class saliency map, highlighting the discriminative areas of the given image with respect to a class. For each test image, we generate the saliency map of the input GM density map with respect to AD. The saliency maps of all test images are averaged to get a mean saliency map. For better interpretation, thresholding is applied on the mean saliency map to obtain the areas most related to the physiological abnormality of AD as shown in Fig. 7, where the highlighted parts denote the most relevant regions to AD diagnosis. From these results, we can see that the highlighted parts cover the regions that are mostly affected by AD, such as hippocampus, amygdala, posterior temporal lobe etc. (Liu et al., 2015; Zhang et al., 2011).

4. Conclusion

In this paper, we have proposed a new classification framework based on combination of CNN and RNN to perform the longitudinal analysis of structural MR images for AD diagnosis. CNN model was proposed to extract the spatial features of each time point and generates single-time classification result, while RNN based on cascaded BGRU was used to model the temporal variations and extract the longitudinal features for improving disease classification. Experimental results on the ADNI dataset demonstrate the effectiveness of the proposed classification algorithm. In the future works, we will include other imaging features such as structural and functional connection networks of brain for RNN based longitudinal analysis. In addition, our work can be extended to other modalities such as fMRI and Positrons Emission Tomography (PET) images for AD diagnosis in the future, even other medical disease diagnosis fields.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (NSFC) under grants (No. 6181101049, 61375112, 61773263), The National Key Research and Develop-

ment Program of China (No.2016YFC0100903), the National Key Basic Research Program of China (973 Project, No.2015CB931802) and SMC Excellent Young Faculty program of SJTU.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, InnoGenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc., F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

References

Adrien, Pa.G.M., 2015. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv, 1502.02506 [cs.CV].

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv 1406.1078.

Chung, J., Gulcehre, C., Cho, K.H., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. Eprint Arxiv.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition., pp. 770–778.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., 2009. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48, 138–149.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., 2011. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55, 574–589.

Hosseini-Asl, E., Keynton, R., El-Baz, A., 2016. Alzheimer's disease diagnostics by adaptation of 3d convolutional network. In: 2016 IEEE International Conference on Image Processing (IcIp), pp. 126–130.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., War, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Jiao, Y., Zhang, Y., Wang, Y., Wang, B., Jin, J., Wang, X., 2017. A novel multilayer correlation maximization model for improving CCA-Based frequency recognition in SSVEP brain-computer interface. Int. J. Neural Syst., 1750039.

Kingma, D., Ba, J., 2014. Adam: a method for stochastic optimization. Comput. Sci. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox Jr., N.C., C.R. J., Ashburner, J., Frackowiak, R.S.J., epub 2008. Automatic classification of MR scans in Alzheimer's disease. Brain.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems., pp. 1097–1105.

LéCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE vol. 86, 2278–2324.

Lei, B., Chen, S., Ni, D., Wang, T., 2016. Discriminative learning for alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion. Front. Aging Neurosci. 8.

Lei, B., Yang, P., Wang, T., Chen, S., Ni, D., 2017. Relational-regularized discriminative sparse learning for alzheimer's disease diagnosis. IEEE Trans. Cybern., 1–12.

Li, Y., Wang, Y., G. W., S. F., Z. L., W. L., S. D., 2012. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. Neurobiol. Aging 33, 427, e415.

Liu, F., Suk, H.I., Wee, C.Y., Chen, H., Shen, D., 2013. High-Order Graph Matching Based Feature Selection for Alzheimer's Disease Identification., pp. 311–318.

Liu, F., Wee, C.Y., Chen, H., Shen, D., 2014. Inter-modality relationship constrained multi-modality multi-task feature selection for alzheimer's disease and mild cognitive impairment identification. Neuroimage 84, 466–475.

Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans. Biomed. Eng. 62, 1132–1140.

Liu, M., Zhang, D., Adeli, E., Shen, D., 2016. Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis. IEEE Trans. Biomed. Eng. 63, 1473–1482.

Liu, M., Zhang, J., Nie, D., Yap, P.T., Shen, D., 2018. Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. IEEE J. Biomed. Health Inform., pp. 1–1.

Shen, D., Davatzikos, C., 2003. Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. NeuroImage 18, 28–41.

Shen, D., Wu, G., Suk, H.I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221.

Simonyan, K., Vedaldi, A., Zisserman, A., 2013. Deep inside convolutional networks: visualising image classification models and saliency maps. Computer Science.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from Overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Ventura, S.J., Joyce, M.A., Martin, A., Sally, M.P.H., Curtin, C., 2008. National vital statistics reports from the CENTERS FOR DISEASE CONTROL AND PREVENTION national center for health statistics national vital statistics system births: final data for 1997. Idahoperinatal Org 58, 1–24.

Wang, Y., Nie, J., Yap, P.T., Shi, F., Guo, L., Shen, D., 2011a. Robust Deformable-surface-based Skull-stripping for large-scale Studies., pp. 635–642.

Wang, Y., Nie, J., Yap, P.T., Shi, F., Guo, L., Shen, D., 2011b. Robust Deformable-Surface-Based Skull-Stripping For Large-Scale Studies. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011, pp. 635–642.

Wang, H., Zhang, Y., Waytowich, N.R., Krusienski, D.J., Zhou, G., Jin, J., Wang, X., Cichocki, A., 2016. Discriminative feature extraction via multivariate linear regression for SSVEP-Based BCI. IEEE Trans. Neural Syst. Rehabil. Eng. 24, 532–541.

Yu, Z., Han, Z., Chen, X., Lee, S.W., Shen, D., 2017. Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis. Sci. Rep. 7, 6530.

Zeiler, M.D., 2012. ADADELTA: an adaptive learning rate method. Comput. Sci.

Zhang, D., Shen, D., 2012. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS One 7, e33182.

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55, 856.

Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., Cichocki, A., 2018. Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. Expert Syst. Appl., 96.

Zhu, X., Suk, H.I., Zhu, Y., Thung, K.H., Wu, G., Shen, D., 2015. Multi-view classification for identification of Alzheimer's disease. In: International Workshop on Machine Learning in Medical Imaging., pp. 255–262.