2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Fusion of Multi-view 2D and 3D Convolution Neural Network based MRI for Alzheimer's Disease Diagnosis

Hezhe Qiao, Lin Chen* and FanZhu*

Abstract— **Alzheimer's disease (AD) is a neurodegenerative disease leading to irreversible and progressive brain damage. Close monitoring is essential for slowing down the progression of AD. Magnetic Resonance Imaging (MRI) has been widely used for AD diagnosis and disease monitoring. Previous studies usually focused on extracting features from whole image or specific slices separately, but ignore the characteristics of each slice from multiple perspectives and the complementarity between features at different scales. In this study, we proposed a novel classification method based on the fusion of multi-view 2D and 3D convolutions for MRI-based AD diagnosis. Specifically, we first use multiple sub-networks to extract the local slice-level feature of each slice in different dimensions. Then a 3D convolution network was used to extract the global subject-level information of MRI. Finally, local and global information were fused to acquire more discriminative features. Experiments conducted on the ADNI-1 and ADNI-2 dataset demonstrated the superiority of this proposed model over other state-of-the-art methods for their ability to discriminate AD and Normal Controls (NC). Our model achieves 90.2% and 85.2% of accuracy on ADNI-2 and ADNI-1 respectively, thus it can be effective in AD diagnosis. The source code of our model is freely available at https://github.com/fengduqianhe/ADMultiView.**

I. INTRODUCTION

Alzheimer's disease (AD) is a kind of neurodegenerative disorder with the impairment of memory and brain damage. Early diagnosis of AD is important for slowing down the development of Alzheimer's disease. Magnetic resonance imaging (MRI) is a typical biomarker to determine whether the subject is an AD patient [1]. MRI-based AD diagnoses have received increasing attention in recent years as its promising results in quantifying the stage of disease of AD and normal controls (NC), making it possible for medical scientists to identify imaging biomarkers of AD from anatomical MRI [2]. Machine-learning methods, including traditional voxel-based morphology analysis and recent deep learning based models, have been proposed to assist MRI-based AD diagnosis and have achieved promising results.

A large number of voxel-based morphology (VBM) studies based on traditional statistical methods have explored the difference in the brain between AD and NC. Studies based on VBM calculate the inherent characteristics of certain biomarkers, such as the hippocampus volumes [3], cortex sickness [4], subcortical volumes [5]. Most of the VBM based methods require certain expert knowledge like brain atlas and a complex procedure for handcrafted feature extraction, which are time-consuming and difficult to implement.

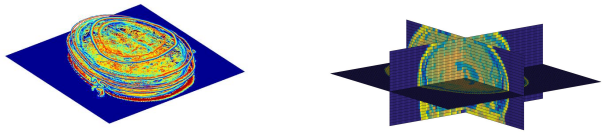


Fig. 1. The slice-level information of slice. The left is the slices at different positions in axial plan. The right is the slice in different dimensions including sagittal plane, coronal plane, and axial plan.

With the continuous development of deep learning, several attempts based on deep learning have been employed to analyze the MRI data by constructing an end to end model avoiding manually extracting features [6]. 3D convolution neural network (CNN) could directly perform feature extraction on the whole image at subject-level [9]. Since the much useless information in the full MRI and few MRIs are difficult for modeling at the subject level, methods based on 3D-Patch and 2D-slice have achieved good results [11]. Multiple classifiers are trained at different locations in MRI, and finally integrated for decision-making [10]. Compared with the modeling at subject level, the patch and slice carries more local features, but lose some global information. On the basis of purely using deep learning methods, some methods focus on the specific biomarker like hippocampus [12], amygdala, posterior temporal lobe combining VBM and deep learning methods [13].

Although deep learning based model have achieved great classification performance for AD diagnosis, it is still an undetermined since subjects' MRIs have relatively small differences in anatomic abnormalities, and it is necessary to dig out moderately subtle changes in disease progression from high denominational of MRI sequences data. As shown in the Figure 1, there is a certain amount of local information in each slice of the MRI at different position and directions. Considering global information in subject-level and various information in slices-level could be both valuable, in this work, we integrate the local slice-level and global subject-level features to complete information complementation between multiple scales and achieve a

This research is supported in part by the National Nature Science Foundation of China under grants 61902370 and 61802360.

Hezhe Qiao is with Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, 400714 Chongqing, China and University of Chinese Academy of Sciences, 100049 BeiJing, China.

Lin Chen and Fan Zhu are with the Chongqing Key Laboratory of Big Data and Intelligent Computing, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China 400714

Correspondence should be addressed to Lin Chen and Fan Zhu (chenlin@cigit.ac.cn, zhufan@cigit.ac.cn)

better prediction accuracy.

The contributions of this study are summarized as follows:

(1) The local feature is extracted from multi-view slices by using multiple 2D sub-networks and the fusion of slice-level feature were performed according to the different directions.

(2) A fusion of multi-view slice-level feature and global subject-level feature extracted by 3D CNN was further used to obtain more discriminative feature.

(3) For AD *vs.* NC classification, the accuracy of this method achieved to 90.2% and 85.2% on ADNI-1 and ADNI-2 respectively, which supports the efficiency of our method when help clinical diagnosis of Alzheimer's disease.

II. Methodology

Figure 2 shows the overall workflow of proposed model. This model consists of four modules: the input module, slice-level sub-network, subject-level 3D CNN, and the classification module. More details of each module are available in the following section.

A. Sub-networks for Slice-level Feature Extraction

The inputs of sub-networks are slices in the three directions including the sagittal plane, coronal plane, and axial plane, as the Figure 3 shown. Since there are various kinds of information on different slices in different directions, each slice is corresponding to a sub-network. The sagittal plane, coronal plane, and axial plane section of MRI are referred to as x, y, and z, respectively. Three transformations F_x, F_y, F_z mapping an input $I \in R^{D\times H\times W}$ to feature map in three dimensions. We take $S_x = [\mathrm{s}_x^1, \mathrm{s}_x^2, ..., s_x^N]$ as the slice cluster in the x direction. The 2D convolution layers were used to separately extract the features of each slice in the corresponding slice cluster.

$$i_x^i = F_x^i(s_x^i, w_x^i) \tag{1}$$

where F_x is the function of extracting feature which consists of multiple blocks. Each blocks consists of a 2D convolution layer, 2D barch normalize (BN) layer, rectified linear unit (RELU) layer, and 2D Maxpooling layer. w_x^i is the weight of convolutional layers crossponding the i_x^i the i th slice in x direction. Finally the 2D avgpooling layer was used to mapping the feature map into a vector, After the slice-level feature extracting module, we first cascade the slices at different locations in a same direction. $I_x = [i_x^1, i_x^2, ..., i_x^N]$ denotes the slice cluster embedding feature in direction x. Then we cascade the features in three direction. $I = [I_x, I_y, I_z]$ is the cascade output of the local embedding in three directions representing the local information of MRI.

B. 3D Neural Network for Subject-level Feature Extraction

In order to obtain global characteristics, we extend 2D convolution to 3D convolution for 3D MRI data. The 3D convolution operation is defined as

$$u_j^l(x,y,z) = \sum_{\delta_x}\sum_{\delta_y}\sum_{\delta_z} F_k^{l-1}(x+\delta_x, y+\delta_y, z+\delta_z) \times W_{kj}^l(\delta_x, \delta_y, \delta_z) \tag{2}$$

where (x,y,z) is the coordinates of pixel in 3D image, F_k^{l-1} is the k feature map of the l layer and $W_{kj}^l(\delta_x, \delta_y, \delta_z)$ is a three-dimensional convolution kernel connecting the kth feature map of the $l-1$ layer to the jth feature map of the l layer. $u_j^l(x,y,z)$ is the output of the convolutional layer, the new jth feature map of the l layer. There are four blocks in the subject-level 3D CNN, and each block consists of a 3D convolution layer, 3D BN layer, ReLu layer and 3D maxpooling layer. After multiple blocks, a 3D average pooling layer maps multiple-channel feature maps into a vector representing the global information.

C. Fully Connected Layer for classification

The local and global features extracted from the two branches are cascaded in a fully connected layer. The output of the final layer is the probability that the subject belongs to a certain category. We use the cross entropy loss L as our optimization goal which is defined as

$$L = -\frac{1}{C}\sum_{c=1}^{C}\frac{1}{N}\sum_{\mathbf{X}_n\in X} \mathrm{I}\{y_n^c = c\}\log(\mathrm{P}(y_n^c = c|\mathbf{X}_n; \mathrm{W})) \tag{3}$$

where L is the cross-entropy loss for classification and the $\mathrm{I}\{\cdot\}$ is an indicator function. When $\{\cdot\}$ is true, $\mathrm{I}\{\cdot\} = 1$, otherwise $\mathrm{I}\{\cdot\} = 0$. $\mathrm{P}(y_n^c = c|\mathbf{X}_n; \mathrm{W})$ denotes the probability of the subject X_n been correctly classifies as the group y_n^c with weights W.

D. Complexity Analysis

We analysis the complexity of our proposed framework by considering the two branches, respectively. For the global subject-level module, the time complexity of each layer is $O(xyzK_{global}^3)$, where K_{global} is the size of 3D convolution kernel. x,y,z is the size of feature map which vary in different layers. For the slice-level module, since we combine feature information in three dimension, the time complexity is $O(N(xyK_{slice}^2 + yzK_{slice}^2 + xzK_{slice}^2))$ for each layer, where N is the number of slice in each direction and K_{slice} is the size of convolution kernel for slice.

E. Dataset and preprocessing

The dataset used in this study was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database which is publicly available on the website http://adni.loni.usc.edu/ [14]. Detailed information about MR acquisition procedures is available at the ADNI website. For this study, we extract ADNI-1 and ADNI-2 datasets from ADNI with corresponding T1-weight MR brain images. Table I shows the demographic details of the studied subjects. In order to verify the effectiveness of our model, following study [7][8], we first use the ADNI-1 as training dataset and evaluate on ADNI-2. In the second group, the ADNI-2 was used as the training dataset and ADNI-1 was used as the test set.

To learn valuable information for the training model, multiple processing operations have been performed on T1 weighted MRI, first we perform anterior commissure (AC)-posterior commissure (PC) correction on all MRIs

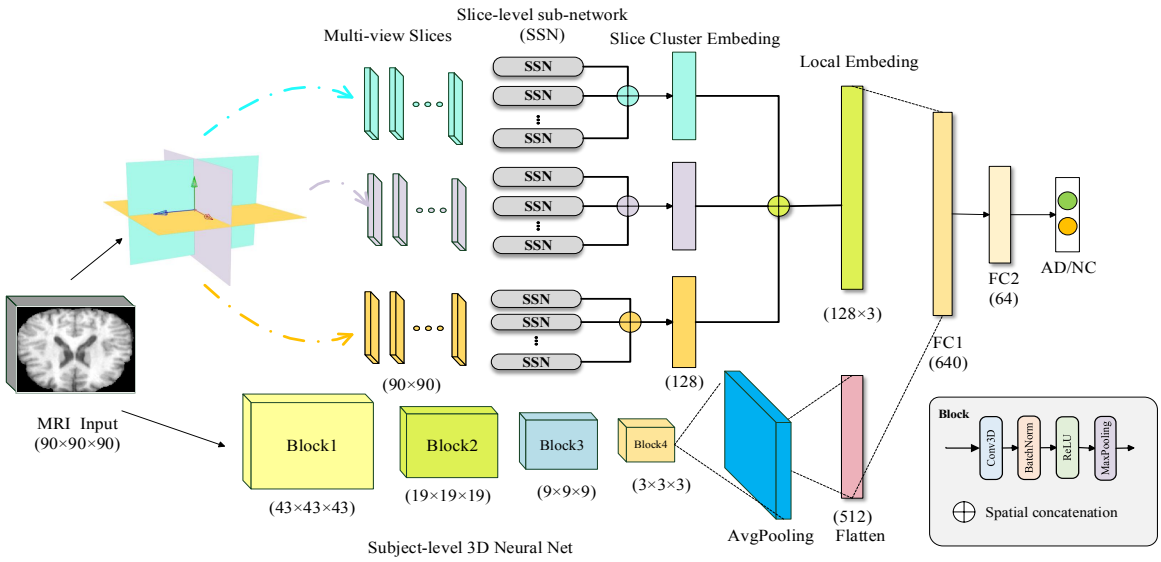


Fig. 2. The overall framework of our proposed model

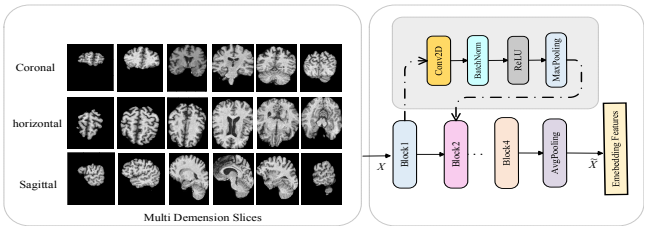


Fig. 3. The slice-level sub-network for multi-view slices feature extracting

and removed invalid areas of each sMRI, leaving only the brain locations. All the images were pre-processed by skull stripping and intensity normalize. Finally, as the input of the neural network, all MRIs were resized to the same size. In the experiment, we uniformly set the size of the model's input to $90 \times 90 \times 90$.

TABLE I

DEMOGRAPHIC CHARACTERISTICS OF THE STUDIED SUBJECTS FROM TWO DIFFERENT DATABASES INCLUDING ADNI-1 AND ADNI-2. (THE VALUES ARE DENOTED AS MEAN ± STANDARD DEVIATION)

Data	Diagnosis	Number	Age(Years)	Sex(M/F)	MMSE
ADNI-1	NC	229	76.2±5.1	104/95	29.2±1.0
	AD	183	75.6±7.6	74/67	23.1±2.5
ADNI-2	NC	184	77.3±6.7	76/70	28.8±1.7
	AD	143	75.6±7.8	67/44	21.9±3.8

F. Experimental setting

The proposed model is implemented on the Pytorch library and trained on 1 NVIDIA GeForce GTX 1080Ti GPU with 11G memory. To avoid the influence of models with different parameters, the batch size of all models is set to 12 during the training process. The optimization method is stochastic gradient descent with an attenuation factor. We adjusted the learning rate to make each model converge to an optimal value. To avoid over-fitting, we added an early stopping mechanism during the training process. To evaluate the performance of our proposed model, we calculate the following four measures: classification accuracy (ACC), sensitivity (SEN), specificity (SPE), receiver operating characteristic sensitivity (ROC) curve, and under ROC curve (AUC).

III. RESULTS AND DISCUSSION

A. Comparison with Different Methods

We first compare the proposed method with four methods including conventional approaches and deep learning based methods.

(1) Voxel+SVM [15]: All MRIs are first normalized to the AAL template using a non linear image registeration technique. Then the MRI was segmented into three types, gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). The GM tissue density was mapped to a vector and used the input of support vector machine (SVM) for classification.

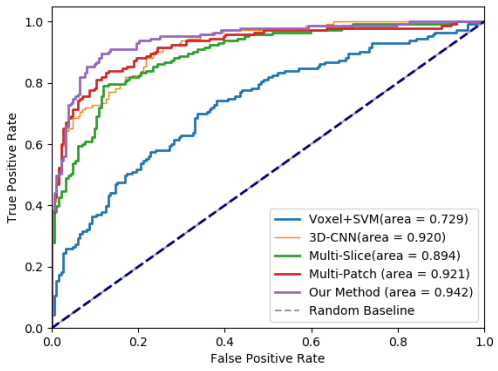
(2) 3D-CNN [16]: In this model, we only use a 3D CNN to extract global subject-level features of MRI. It is worthy noting that this model is a part of our proposed model.

(3) Multi-Slice: As another part of our model, this method only focusing on the slice feature extracting by cascading the feature of all slices in three direction.

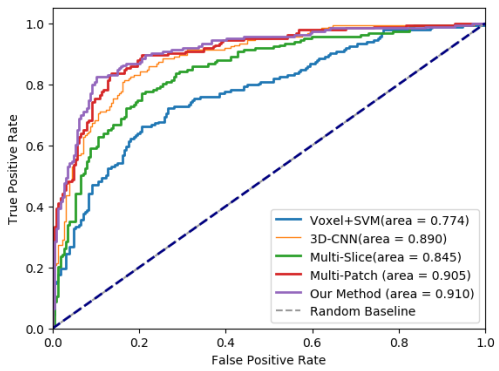
(4) Multi-Patch [16]: In this method, we spilt the MRI into multiple patches and train a feature extractor for each patch. Finally, the features were cascade to construct the embedding feature of the entire MRI.

B. Results on ADNI-2

In this section, we use ADNI-1 as the training dataset and evaluate on the ADNI-2 dataset. Table III illustrates the comparison results and Figure 4 (a) shows the ROC curves of



(a)



(b)

Fig. 4. Comparisons of ROC curves on AD *vs.* NC task. (a) is the model was trained on ADNI-1 and evaluated on ADNI-2. (b) is the model was trained on ADNI-2 and evaluated on ADNI-1.

TABLE II

RESULTS FOR AD *vs.* CN CLASSIFICATION TASKS WITH MODELS TRAINED ON ADNI-1 AND EVALUATED ON ADNI-2

Method(%)	ACC	SEN	SPE	AUC
Voxel+SVM	75.9	67.7	81.0	72.9
3D-CNN	86.5	81.8	86.6	92.0
Multi-Slice	84.0	78.3	84.2	89.4
Multi-Patch	83.4	82.5	80.2	92.1
Our Method	**90.2**	**88.1**	**89.3**	**94.2**

TABLE III

RESULTS FOR AD *vs.* CN CLASSIFICATION TASKS WITH MODELS TRAINED ON ADNI-2 AND EVALUATED ON ADNI-1

Method(%)	ACC	SEN	SPE	AUC
Voxel+SVM	75.4	72.8	71.6	77.4
3D-CNN	83.0	73.7	**85.9**	89.0
Multi-Slice	77.6	61.2	70.8	84.5
Multi-Patch	80.8	**77.6**	78.2	90.5
Our Method	**85.2**	77.5	82.3	**91.0**

different models. Results indicate that the proposed method is able to deliver better performance than other methods. Specifically, our model effectively improves the classification effect of the model. Compared with the method using whole image alone, after adding slice feature from multi-view, AUC was improved to 94.2%. The results indicating these two branches could cooperate with each other well. In addition, our method shows a significantly improvement on SEN compared to other methods in comparison. This is especially important, as positive subjects are of special interest in clinical diagnosis. The results suggest that the local and global feature through cascading the high-level semantic features can better obtain positive discriminant information and perform well in SEN than the method based on just slice or patch.

C. Results on ADNI1

In order to further demonstrate the effectiness of our proposed method, we also use the ADNI-2 as the training dataset and evaluate on ADNI-1. We report the comparison result in Table III and the ROC curves of the different models was shown in Figure 1 (b). The SEN of our method could achieve 77.5, although it is slight lower than Multi-Patch, it surpasses the other methods. Similar to the results on ADNI-2, our model also yields the best ACC and AUC when compared with the other methods. The results further indicate that the fusion of local features of slices in different directions and global feature of whole image will be more helpful for characterizing the characteristics of MRI.

It is noted that the performance of model trained with ADNI-2 is lower than the one trained with ADNI-1. The major reason is that ADNI-1 and ADNI-2 were acquired by 1.5T and 3.0T scanners, which leads to the different imaging quality that directly affects prediction accuracy.

D. Effect of Slices in Different Dimensions

To better understand the effectiveness of our model, we compared the effects of fusing local slice-level feature in different dimensions. As shown in Table IV, compared with the sagittal plane and the coronal plane, weighting in the axial plane performs better on SPE. When all three directions were considered, the model performs best in terms of ACC, SEN, and AUC, which shows that the integration of slices in multiple directions is conducive to the model extracting more refined features.

TABLE IV

THE EFFECTS OF FUSING LOCAL SLICE-LEVEL FEATURE IN DIFFERENT DIMENSIONS

Sagittal	Coronal	Axial	ACC	SEN	SPE	AUC
			86.5	81.8	86.7	93.8
√			88.1	83.2	88.8	94.1
	√		89.0	84.6	89.6	94.0
		√	89.3	85.3	**89.7**	94.1
√	√	√	**90.2**	**88.1**	89.4	**94.2**

TABLE V

THE PERFORMANCE COMPARISON OF OTHER MODEL REPORT IN THE LITERATURES AND OUR PROPOSED MODEL

Method(%)	ACC	SEN	SPE	AUC
SVM+Landmark [18]	82.2	77.4	86.1	88.1
Multi-Kernel [19]	88.6	85.7	90.4	89.8
Multi-Modal[20]	87.3	**88.4**	86.2	93.0
CNN +Landmark[21]	88.3	79.6	94.6	94.0
3D DenseNet [22]	88.9	86.6	90.8	92.5
MLP + BGRU [23]	89.7	86.8	**92.5**	92.1
Our Method	**90.2**	88.1	89.4	**94.2**

E. Comparison with Other Methods in Literature

We have compared our classification method with those reported in the literatures. Methods used in the comparison include traditional statistical methods and deep learning methods. The classical statistical methods compared here includes: SVM based on ROI or landmark [17][18], the multi-kernel learning combining feature selections, manifold learning and over-sampling with ROI [19], the multi-modal data of MRI and positron emission computed tomography (PET) were used to train a linear regression for the classification of subject [20]. The deep learning methods including Landmark-based deep multi-instance learning method [21], 3D DenseNet used to learn features of 3D patches based on the hippocampal segmentation results [22], RNN based longitudinal analysis on the MRI feature extracted by CNN [23]. As shown in Table V, we can see deep learning methods have achieve well performance with the statistical methods based on handcrafted features in terms of AUC. Meanwhile, our proposed model is able to generate more discriminative features by combining the local slice-level feature and global subject-level feature, thus outperforms other models in comparison in terms of ACC and AUC.

IV. CONCLUSION

In this study, we proposed a novel method based on local slice-level feature and global 3D image feature for AD *vs.* NC classification, to address the problems of anatomical abnormal feature extraction and high-dimensional features optimization for MRI based AD diagnosis. Multi-scale information from local slice-level and subject-level can further improve the AD *vs.* NC classification ability of the model. Experiments conducted on the ADNI-1 and ADNI-2 dataset have proven the model's effectiveness in prediction accuracy . In future work, we will add prior knowledge and explore more deep learning algorithms for more discriminative feature extraction.

REFERENCES

[1] Jack Jr C R, Knopman D S, Jagust W J, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology, 2013, 12(2): 207-216.

[2] Jain R, Jain N, Aggarwal A, et al. Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cognitive Systems Research, 2019, 57: 147-159.

[3] Fuse H, Oishi K, Maikusa N, et al. Detection of Alzheimer's Disease with Shape Analysis of MRI Images. 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS). IEEE, 2018: 1031-1034.

[4] Luk C C, Ishaque A, Khan M, et al. Alzheimer's disease: 3-Dimensional MRI texture for prediction of conversion from mild cognitive impairment. Alzheimer's Dementia: Diagnosis, Assessment Disease Monitoring, 2018, 10(1): 755-763.

[5] Vu T D, Ho N H, Yang H J, et al. Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection. Soft Computing, 2018, 22(20): 6825-6833.

[6] Altinkaya E, Polat K, Barakli B. Detection of Alzheimer's Disease and Dementia States Based on Deep Learning from MRI Images: A Comprehensive Review. Journal of the Institute of Electronics and Computer, 2020, 1(1): 39-53.

[7] Liu M, Zhang J, Adeli E, et al. Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering, 2018, 66(5): 1195-1206.

[8] Lian C, Liu M, Zhang J, et al. Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE transactions on pattern analysis and machine intelligence, 2018, 42(4): 880-893.

[9] Jin D, Xu J, Zhao K, et al. Attention-based 3D convolutional network for Alzheimer's disease diagnosis and biomarkers exploration. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019: 1047-1051.

[10] Lian C, Liu M, Pan Y, et al. Attention-Guided Hybrid Network for Dementia Diagnosis With Structural MR Images. IEEE Transactions on Cybernetics, 2020.

[11] Ebrahimighahnavieh M A, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. Computer methods and programs in biomedicine, 2020, 187: 105242.

[12] Cui R, Liu M. Hippocampus Analysis by Combination of 3-D DenseNet and Shapes for Alzheimer's Disease Diagnosis. IEEE journal of biomedical and health informatics, 2018, 23(5): 2099-2107.

[13] Chen Y, Xia Y. Iterative sparse and deep learning for accurate diagnosis of Alzheimer's disease. Pattern Recognition, 2021, 116: 107944.

[14] Jack Jr C R, Bernstein M A, Fox N C, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2008, 27(4): 685-691.

[15] Ashburner J, Friston K J. Voxel-based morphometry the methods. Neuroimage, 2000, 11(6): 805-821.

[16] Wen J, Thibeau-Sutre E, Diaz-Melo M, et al. Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical image analysis, 2020, 63: 101694.

[17] Zhang J, Gao Y, Gao Y, et al. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. IEEE transactions on medical imaging, 2016, 35(12): 2524-2533.

[18] Zhang J, Liu M, An L, et al. Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. IEEE journal of biomedical and health informatics, 2017, 21(6): 1607-1616.

[19] Cao P, Liu X, Yang J, et al. Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures. Computers in biology and medicine, 2017, 91: 21-37.

[20] Ye T, Zu C, Jie B, et al. Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease. Brain imaging and behavior, 2016, 10(3): 739-749.

[21] Liu M, Zhang J, Adeli E, et al. Landmark-based deep multi-instance learning for brain disease diagnosis. Medical image analysis, 2018, 43: 157-168.

[22] Liu M, Li F, Yan H, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. NeuroImage, 2020, 208: 116459.

[23] Cui R, Liu M, Li G. Longitudinal analysis for Alzheimer's disease diagnosis using RNN. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 1398-1401.