Region Ensemble Network for MCI Conversion Prediction with a Relation Regularized Loss

Yuan-Xing Zhao[1,2], Yan-Ming Zhang[2], Ming Song[2,3],
and Cheng-Lin Liu[1,2,4(✉)]

[1] School of Artificial Intelligence, University of Chinese Academy of Sciences,
Beijing 100149, China
{yuanxing.zhao,liucl}@nlpr.ia.ac.cn

[2] NLPR, Institute of Automation, Chinese Academy of Sciences,
Beijing 100190, China
{ymzhang,msong}@nlpr.ia.ac.cn

[3] Brainnetome Center, Institute of Automation, Chinese Academy of Sciences,
Beijing 100190, China

[4] CAS Center for Excellence of Brain Science and Intelligence Technology,
Beijing 100190, China

Abstract. Despite many recent advances, computer-aided mild cognitive impairment (MCI) conversion prediction is still a very challenging task due to: 1) the abnormal areas are subtle compared to the size of the whole brain, 2) the features' dimension is much larger than the number of samples. To tackle these problems, we propose a region ensemble model using a divide and conquer strategy to capture the disease's finer representation. Specifically, the features are independently extracted from non-overlapping regions and then fused to describe the subject according to the attention scores. Moreover, we design a novel loss that models the relationship between different stages of the disease to regularize the training process explicitly. Experiments on public data sets for MCI conversion prediction demonstrate that our method has achieved state-of-the-art performance. Specifically, the area under the receiver operating characteristic curve (AUC) is improved from 79.3% to 85.4%. Beyond that, each region's contribution can be assessed quantitatively, using the proposed method.

Keywords: Alzheimer's disease · Mild cognitive impairment · Region ensemble network · Relation regularized loss

1 Introduction

Mild cognitive impairment (MCI) is the prodromal stage of Alzheimer's disease (AD). A systematic review found that 32% of individuals with MCI would convert to AD within five years' follow-up. Hence, identifying which individuals with MCI are more likely to develop AD is a primary goal of current

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12905, pp. 185–194, 2021.
https://doi.org/10.1007/978-3-030-87240-3_18

research [1]. Before some noticeable symptoms of the disease, several subtle structural changes have already happened in the brain. As an essential computer-aided diagnosis technique, structural magnetic resonance imaging (sMRI) can non-invasively capture such changes. Therefore many machine learning or deep learning-based methods have been applied to AD diagnosis and MCI conversion prediction based on sMRI [2,3] and have reported remarkable success.

In general, all of these methods can be grouped into detection-dependent approaches and detection-free approaches, depending on whether they need a separate model to detect regions of interest (ROI) or not. Detection-dependent methods [4–9] first locate ROI based on prior domain knowledge using an independent detection model. It then constructs a diagnosis model based on the ROI's feature. These methods reduce the feature's dimension using the whole brain's sub-areas but may miss some critical regions in practice. Thus, it is hard for them to achieve high performance. To tackle this limitation, current state-of-the-art methods adopt the detection-free approach. Taking the whole brain as the input, the methods in [10–12] locate abnormal areas and predict the result simultaneously. In this way, they can extract the critical areas and discard useless regions in a data-driven and target-consistent way. While these methods are more powerful and flexible in principle, they tend to suffer from severe over-fitting due to the limited number of training samples. One way to alleviate this problem is to use auxiliary data. For example, the works [10] and [12] both pre-train an AD diagnosis model first and then fine-tune the model for the MCI conversion prediction task.

This paper proposes a region ensemble model together with a relation regularized loss for the MCI conversion prediction task. The model's core idea is to divide the brain into non-overlapping regions and learn a region-based diagnosis sub-network for each region. In this way, sub-networks overcome the curse of dimensionality by focusing on small regions of the brain. Finally, we construct an ensemble model by weighted fusion of the regional features with attention scores. Additionally, we propose a relation regularized loss based on an assumption of the disease to regularize the training process. More importantly, this loss allows our method to incorporate auxiliary samples to improve the performance.

In summary, the main contributions of this paper are three-fold. First, we propose a novel region ensemble model that uses a divide-and-conquer strategy and attention mechanism to extract the discriminative features and locate abnormal areas. Second, we propose a relation regularized loss to regularize the model's training process through additional samples. Third, on public data sets (ADNI-1 and ADNI-2), our method outperforms competing methods with a large margin and achieves the state-of-the-art performance.

2 Method

2.1 Region Ensemble Network

As shown in Fig. 1, our diagnosis model consists of three sequential components: 1) feature extraction sub-network, 2) region-based diagnosis sub-network, and

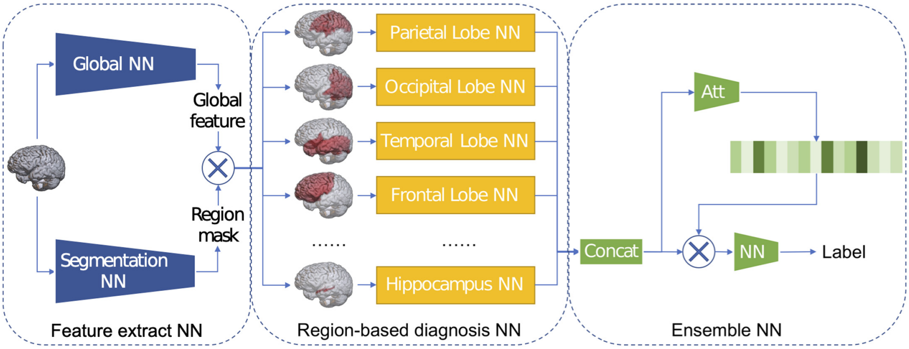

Fig. 1. The architecture of the region ensemble model. 15 region-based diagnosis sub-networks are adopted in our model.

3) ensemble sub-network. We first extract the global feature map from the whole brain through the feature extraction sub-network. This feature map then voxel-wisely multiplies with the region masks, produced by a segmentation model, to create a raw feature map for each brain region. After that, region-based diagnosis sub-networks generate discriminative features from the regional feature maps. Finally, the ensemble sub-network fuses the regional representations with an attention mechanism to produce the final representation and make a classification.

In this framework, the segmentation network is trained separately using the dataset in [13], while the other parts are trained in an end-to-end manner.

Feature Extract Sub-Network. In this stage, a sub-network is used to extract a raw feature map for each region-based diagnosis sub-network. Concretely, we first extract the whole brain's feature map $\boldsymbol{F}^g$ using three feature extraction blocks, the first and the third blocks followed by a max-pooling layer. A block is composed by stacking a convolution layer, a batch normalization (BN) [14], a parametric rectified linear units (PReLU) [15], and a convolutional block attention module (CBAM) [16]. Meanwhile, we perform 3D whole brain segmentation using an auxiliary segmentation network and get one region mask $\boldsymbol{M}^r$ for each region r. After that, the input feature map $\boldsymbol{F}^{r,in}$ for the r-th region diagnosis sub-network is calculated by $\boldsymbol{F}^{r,in} = \boldsymbol{M}^r \otimes \boldsymbol{F}^g$, where $\otimes$ denote an element-wise multiplication.

Region-Based Diagnosis Sub-Network. For each region, we use an independent diagnosis sub-network to extract the discriminative features from $F^{r,in}$. We first adopt a convolution layer followed by a max-pooling layer to reduce the feature map scale. Next, 14 feature extraction blocks, each composed by stacking a convolution layer, a BN, a PReLU, and a CBAM, are used to obtain the final regional feature map $\boldsymbol{F}^{r,out}$. Finally, a convolution layer with kernel size $1 \times 1 \times 1$

combining with a softmax layer is applied to classify voxels in the feature map. Note that the sizes of all regional feature maps are same: $\boldsymbol{F}^{r,out} \in \mathbb{R}^{d\times L\times W\times H}$ for all r. Here, d is the number of channels, and L, W, H is the length, width, height of $\boldsymbol{F}^{r,out}$.

Let $(\boldsymbol{X}, y)$ be a training sample. Here, $\boldsymbol{X}$ is the sMRI image, $y \in \{1, ..., C\}$ is the ground-truth label of $\boldsymbol{X}$, and C is the number of the categories. Then, the diagnosis loss is defined as

$$L^{voxel}\left(\boldsymbol{F}^{r,out}, y\right) = \frac{1}{L\times W\times H}\sum_{i,j,k=1}^{L,W,H}\sum_{c=1}^{C} y_c \log\left(P\left(\hat{y} = c|F_{i,j,k}^{r,out}\right)\right). \tag{1}$$

y_c is the binary indicator of the ground-truth label, which equals to 1 if $\boldsymbol{X}$ belong to class c and 0 otherwise. $P\left(\hat{y} = c|F_{i,j,k}^{r,out}\right)$ is the predicted probability for class c of voxel (i, j, k). It is noted that different from the common-used strategy, which first performs a global pooling to reduce $\boldsymbol{F}^{r,out}$ to a feature vector and then optimizes the loss defined on that vector, we optimize the loss defined on each voxel to prevent losing critical details.

Ensemble Sub-Network. We design an ensemble sub-network to automatically identify discriminative regions in the whole brain and perform classification. The structure of the ensemble sub-network is shown in Fig. 1. It includes two parts: an attention module and a classifier.

Because voxels have different discriminative abilities, we assign an attention score to each region's voxel independently. For simplicity, we only introduce the computation of attention score for one voxel. Given voxel (i, j, k)'s feature $\boldsymbol{f}^r = \boldsymbol{F}_{i,j,k}^{r,out} \in \mathbb{R}^d$ generated by the r-th region-based diagnosis sub-network, the attention module first transforms $\boldsymbol{f}^r$ into a scalar $f^r \in \mathbb{R}$ by

$$f^r = \delta\left(\boldsymbol{W}_2^r \delta\left(\boldsymbol{W}_1^r \boldsymbol{f}^r\right)\right), \tag{2}$$

where δ refers to the PReLU function, $\boldsymbol{W}_1^r \in \mathbb{R}^{3d\times d}$ and $\boldsymbol{W}_2^r \in \mathbb{R}^{1\times 3d}$ are learnable parameters. To consider the relationship among regions, we combine the values at the same location (i, j, k) in different regional feature maps into $\boldsymbol{f} = [f^1, f^2, ..., f^R]$ and get the final attention score by

$$\boldsymbol{a} = \sigma(\boldsymbol{W}_2^a \delta(\boldsymbol{W}_1^a \boldsymbol{f})) \ \in \mathbb{R}^R, \tag{3}$$

where σ refers to the sigmoid function, $\boldsymbol{W}_1^a \in \mathbb{R}^{\frac{R}{3}\times R}$, and $\boldsymbol{W}_2^a \in \mathbb{R}^{R\times \frac{R}{3}}$. R is the number of regions. After computing the scores for each voxel, we can get the attention score $\boldsymbol{A}_r \in \mathbb{R}^{L\times W\times H}$ for each regional feature map $\boldsymbol{F}^{r,out}$ and then the ensemble feature map is computed as

$$\boldsymbol{F}^e = \sum_{r=1}^{R} \boldsymbol{F}^{r,out} \otimes \boldsymbol{A}_r. \tag{4}$$

Finally, a convolution layer with kernel size 1 × 1 × 1 combining with a softmax layer is applied to $\boldsymbol{F}^e$ to classify all voxels. The probability of a given sMRI $\boldsymbol{X}$ to be predicted as class c is calculated by

$$P(\hat{y} = c|\boldsymbol{X}) = \frac{1}{L \times W \times H} \sum_{i,j,k=1}^{L,W,H} P\left(\hat{y} = c|F^e_{i,j,k}\right). \tag{5}$$

2.2 Relation Regularized Loss

The main challenge in MCI conversion prediction is the lack of training samples. To alleviate the problem, works [10] and [12] use AD and normal controls (NC) samples to pre-train a model and then fine-tune it on stable MCI (sMCI) and progressive MCI (pMCI) samples. In this work, we utilize AD/NC samples more sophisticatedly by introducing a novel ranking loss.

NC/sMCI/pMCI/AD labels are intrinsically ordered because MCI is a prodromal stage of AD, and its structural changes are between AD and NC [17]. Hence, we make an assumption as follows. We defined $P^c = P(\hat{y} = c|\boldsymbol{X})$ as the predicted probability of $\boldsymbol{X}$ belonging to class c, For a training sample $(\boldsymbol{X}, y)$,

1) if $y = \text{NC}$, we have $P^{\text{NC}} > P^{\text{sMCI}} > P^{\text{pMCI}} > P^{\text{AD}}$.
2) if $y = \text{sMCI}$, we have $P^{\text{sMCI}} > P^{\text{NC}}$ and $P^{\text{sMCI}} > P^{\text{pMCI}} > P^{\text{AD}}$.
3) if $y = \text{pMCI}$, we have $P^{\text{pMCI}} > P^{\text{AD}}$ and $P^{\text{pMCI}} > P^{\text{sMCI}} > P^{\text{NC}}$.
4) if $y = \text{AD}$, we have $P^{\text{AD}} > P^{\text{pMCI}} > P^{\text{sMCI}} > P^{\text{NC}}$.

In order to enforce such relationship, we define a ranking loss as

$$L^{rank}\left(P^{c_1}, P^{c_2}, z\right) = z \exp\left(-z\left(P^{c_1} - P^{c_2}\right)\right), \tag{6}$$

where $z \in \{0, 1\}$. For $z = 0$, L^{rank} equals to 0 and plays no role in learning. For $z = 1$, minimizing L^{rank} constrains the model to obey the relation $P^{c_1} > P^{c_2}$. It is noted that we only optimize the relation of $P^{c_1} > P^{c_2}$, because $P^{c_1} > P^{c_2}$ and $P^{c_1} < P^{c_2}$ are equivalent. To represent the pairwise relation between the predicted probabilities, a difference matrix $\boldsymbol{D} \in \mathbb{R}^{4\times4}$ is defined as $D_{ij} = P^{c_i} - P^{c_j}$. For each label c, a relation matrix $\boldsymbol{Z}^c \in \{0, 1\}^{4\times4}$ is defined as $Z^c_{ij} = 1$ for class c, $P^{c_i} > P^{c_j}$, according to relations explained above. Specifically, The rows and columns of the matrix are set in order of NC, sMCI, pMCI, and AD, (e.g. $Z^{NC}[2, 3]$ denotes the relation between the p^{sMCI} and p^{pMCI}) we have the following $\boldsymbol{Z}^c$

$$\boldsymbol{Z}^{\text{NC}} = \begin{bmatrix} 0&1&1&1 \\ 0&0&1&1 \\ 0&0&0&1 \\ 0&0&0&0 \end{bmatrix}, \boldsymbol{Z}^{\text{sMCI}} = \begin{bmatrix} 0&0&0&0 \\ 1&0&1&1 \\ 0&0&0&1 \\ 0&0&0&0 \end{bmatrix}, \boldsymbol{Z}^{\text{pMCI}} = \begin{bmatrix} 0&0&0&0 \\ 1&0&0&0 \\ 1&1&0&1 \\ 0&0&0&0 \end{bmatrix}, \boldsymbol{Z}^{\text{AD}} = \begin{bmatrix} 0&0&0&0 \\ 1&0&0&0 \\ 1&1&0&0 \\ 1&1&1&0 \end{bmatrix}.$$

Finally, the relation regularized loss can be defined as

$$L^{rank}(\boldsymbol{D}, \boldsymbol{Z}^c) = \frac{1}{nonzero(\boldsymbol{Z}^c)} \sum_{i,j=1}^{4} Z^c_{ij} \exp\left(-Z^c_{ij} D_{ij}\right), \tag{7}$$

where $nonzero(Y)$ is the number of non-zero elements in Y.

The overall loss optimized in our method is defined as

$$Loss(\boldsymbol{X}, y) = L^{voxel}(\boldsymbol{F}^{e}, y) + \frac{\lambda_1}{R}\sum_{r=1}^{R} L^{voxel}(\boldsymbol{F}^{r,out}, y) + \lambda_2 L^{rank}(\boldsymbol{D}, \boldsymbol{Z}^{y}), \quad (8)$$

where λ_1, λ_2 are hyperparameters to control the influences of L^{rank} and L^{voxel}.

3 Experiments

3.1 Dataset and Evaluation Metrics

We perform experiments on the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [18]. Following [9–12], in all experiments, we treat ADNI-1 as the training set and leave ADNI-2 for testing to make an easier comparison. The training and testing set contains 226 sMCI vs. 167 pMCI and 239 sMCI vs. 38 pMCI, respectively. We also collect 199 AD and 229 NC samples in ADNI-1 as the additional samples to optimize the proposed relation regularized loss. Diagnostic performance is assessed using four metrics: classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and AUC.

3.2 Implementation Details

Since our method needs voxel-level annotation to extract brain regions, we use the dataset in [13] and the method in [19] to train a segmentation model which segments the whole brain into 134 regions. Then, we apply this model to sMRI images in ADNI and obtain the initial region annotations. Since small regions may result in unstable results, we merge the 134 regions into 15 according to anatomy knowledge. The resulting regions are shown in Fig. 3.

sMRI images are processed following a standard pipeline. Specifically, we use the segmentation model mentioned above to simultaneously strip the skull and split the brain region. All subjects were aligned by affine registration to Colin27 template [20] to remove the global linear difference. After that, voxels were resampled to an identical spatial resolution ($1 \times 1 \times 1\,\mathrm{mm}^3$), using SimpleITK [21]. To handle sMRI with different sizes, we crop or pad them (for large or small sMRI) into $160 \times 196 \times 152$ for both the training and testing phases. In the feature extract sub-network, the channel in the first building block is 16, and is increased by 16 after each block. In the region-based diagnosis sub-network, the channel in the first block is 24, and is increased by 14 after each block.

Stochastic gradient descent with momentum is used as the optimizer, and the learning rate is set to 0.05 initially and is decreased during the training process. The dropout is set to 0 initially and is increased to 0.1 until the 25th epochs. The batch size is 4 for each GPU. The method is implemented by PyTorch [22], and all experiments are conducted with two TITAN GPUs with 12 GB RAM.

3.3 Ablation Studies

To better understand our method, we conduct ablation experiments to examine how each proposed component affects performance. 1) The baseline model adopts neither region partition nor relation regularized loss and is trained with traditional cross-entropy loss. For a fair comparison, we expand the baseline model's channels so that the model's size is nearly the same as the proposed model. 2) The region ensemble model has 15 region diagnosis sub-networks based on the non-overlapped brain regions. 3) The baseline model is trained by the proposed relation regularized loss. 4) The region ensemble model is trained by the proposed relation regularized loss.

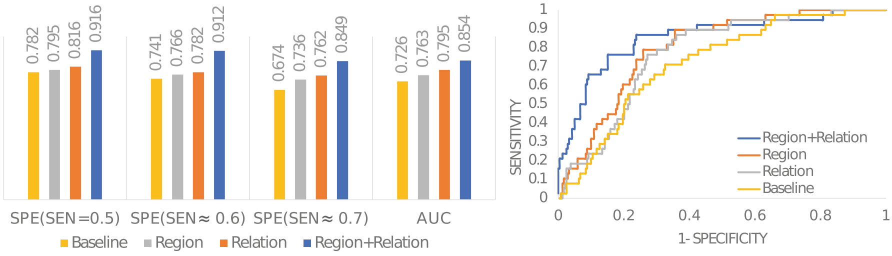

Fig. 2. Contribution of the proposed components in MCI conversion prediction

The results are shown in Fig. 2. The four models are denoted as Baseline, Region, Relation, and Region+Relation, separately. We have the following observations. First, our method consistently improves with each component's addition. Second, the region ensemble model outperforms the whole brain classification model. It implies that the region ensemble model can capture more helpful information. Third, the model trained with the relation regularized loss is more accurate than the baseline model and the region ensemble model. It indicates that using more training samples and exploiting the task's intrinsic structure are the critical factors for obtaining high performance.

3.4 Comparing with SOTA Methods

We compare our method with several approaches for MCI conversion prediction. We trained the model on ADNI-1 (using 10% of subjects for validation) and then used it to diagnose the subjects from ADNI-2. The classification results are summarized in Tab. 1. The results of compared methods are referred from [9–12]. All methods have used the same testing data. Furthermore, works in [10] and [12] also use the same AD and NC sample as ours to pre-train the model. For a more thorough and comprehensive evaluation, we compared the model's performance at different SEN values. From Table 1, we can see that our approach yields better results, demonstrating the advantage of our proposed strategies, i.e., the region ensemble network and relation regularized loss. Due to the limitations of sample quantity, we cannot get the SEN's value at 0.6 and 0.7 exactly, hence, we set the SEN near 0.6 and 0.7. Additionally, we also trained and tested on the ADNI-1, using 5-folder cross-validation, obtaining a AUC of 0.82.

Table 1. Results for MCI conversion prediction on ADNI-2.

Methods	ACC	SEN	SPE	AUC
ROI [4]	0.661	0.474	0.690	0.638
VBM [5]	0.643	0.368	0.686	0.593
DMIL [9]	0.769	0.421	0.824	0.776
H-FCN [10]	0.809	0.526	0.854	0.781
IAF [11]	0.816	0.605	0.849	0.787
HybNet [12]	0.827	0.579	0.866	0.793
Ours(SEN $= 0.5$)	**0.859**	0.500	**0.916**	**0.854**
Ours(SEN ≈ 0.6)	**0.870**	**0.605**	**0.912**	**0.854**
Ours(SEN ≈ 0.7)	**0.830**	**0.711**	0.849	**0.854**

3.5 Interpreting the Model's Prediction

We try to provide insights into the MCI conversion prediction problem by analyzing our model's intermediate result. First, we examine how different brain

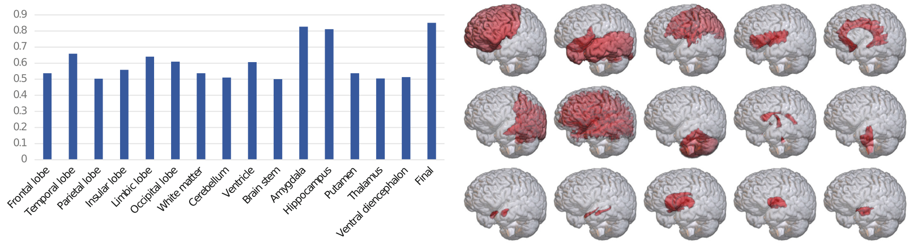

Fig. 3. The AUC of each brain region on ADNI-2 and the brain regions used to train the region diagnosis sub-networks, with the same order as the bins in the histogram, from left to right, top to bottom.

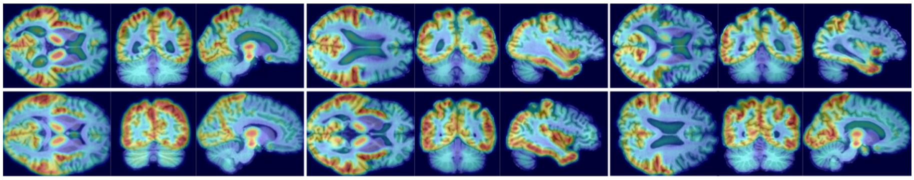

Fig. 4. Attention maps of sMCI subjects (top) and pMCI subjects (bottom).

regions are related to sMCI vs. pMCI classification. The AUC values for 15 region-based diagnosis sub-networks are shown in Fig. 3. We can see that Hippocampus and Amygdala are much more informative than other regions and can get similar AUC results as the ensemble model. On the other hand, Cerebellum, Parietal lobe, Thalamus, Ventral diencephalon, and Brain stem seemed valueless for the prediction.

In Fig. 4 we multiply each region mask $\boldsymbol{M}^r$ with the ensemble model's attention score $\boldsymbol{A}_r$ as each region's weight and visualize the weights at the individual level. As shown in the figure, our model can localize different subjects' abnormalities, which is valuable in clinical diagnosis.

4 Conclusion

MCI conversion prediction is a fundamental problem in the computer-aided diagnosis of Alzheimer's disease. This paper introduces a region ensemble model to predict the disease and identify the disease's critical brain regions. Additionally, we propose a relation regularized loss using the disease's intrinsic structure and AD/NC samples. Extensive experiments on public datasets show the superiority of our method. However, the critical brain region assessed by our method is relatively coarse because of the limitation of the GPU memory. In the future, we will investigate methods for evaluating the more delicate brain regions.

Acknowledgments. This work has been supported by the National Key Research and Development Program Grant 2018AAA0100400, the National Natural Science Foundation of China (NSFC) grants 61773376, 61836014, 61721004 and 31870984.

References

1. Association, A., et al.: 2020 Alzheimer's disease facts and figures. Alzheimer's Dement. **16**, 391–460 (2020)
2. Rathore, S., Habes, M., Iftikhar, M.A., Shacklett, A., Davatzikos, C.: A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. Neuroimage **155**, 530 (2017)
3. Leandrou, S., Petroudi, S., Reyes-Aldasoro, C.C., Kyriacou, P.A., Pattichis, C.S.: Quantitative MRI brain studies in mild cognitive impairment and alzheimer's disease: a methodological review. IEEE Rev. Biomed. Eng. **11**, 97–111 (2018)
4. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage **55**, 856–867 (2011)
5. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. Neuroimage **11**(6), 805–821 (2000)
6. Zhang, J., Gao, Y., Gao, Y., Munsell, B.C., Shen, D.: Detecting anatomical landmarks for fast alzheimer's disease diagnosis. IEEE Trans. Med. Imaging **35**, 2524–2533 (2016)
7. Lei, B., Yang, P., Wang, T., Chen, S., Ni, D.: Relational-regularized discriminative sparse learning for alzheimer's disease diagnosis. IEEE Trans. Cybern. **47**, 1102–1113 (2017)

8. Cheng, B., Liu, M., Zhang, D., Shen, D., Initiative, A.D.N., et al.: Robust multi-label transfer feature learning for early diagnosis of alzheimer's disease. Brain Imaging Behav. **13**, 138–153 (2019)
9. Liu, M., Zhang, J., Adeli, E., Shen, D.: Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018)
10. Lian, C., Liu, M., Zhang, J., Shen, D.: Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural MRI. IEEE Trans. Pattern Anal. Mach. Intell. **42**, 880–893 (2018)
11. Li, Q., et al.: Novel iterative attention focusing strategy for joint pathology localization and prediction of mci progression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 307–315 (2019)
12. Lian, C., Liu, M., Pan, Y., Shen, D.: Attention-guided hybrid network for dementia diagnosis with structural mr images. IEEE Trans. Cybern. 1–12 (2020, early access)
13. Landman, B., Warfield, S.: Miccai 2012 workshop on multi-atlas labeling. In: Medical Image Computing and Computer Assisted Intervention Conference (2012)
14. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456 (2015)
15. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
16. Woo, S., Park, J., Lee, J.-Y., So Kweon, I.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)
17. Coupé, P., Manjón, J.V., Lanuza, E., Catheline, G.: Lifespan changes of the human brain in alzheimer's disease. Sci. Rep. **9**(1), 1–12 (2019)
18. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Weiner, M.W.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. J. Magn. Reson. Imaging **27**, 685–691 (2010). http://adni.loni.usc.edu
19. Zhao, Y.-X., Zhang, Y.-M., Song, M., Liu, C.-L.: Multi-view semi-supervised 3d whole brain segmentation with a self-ensemble network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 256–265 (2019)
20. Holmes, C.J., Hoge, R., Collins, L., Woods, R., Evans, A.C.: Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. **3**, 324–333 (1998)
21. Lowekamp, B.C., Chen, D.T., Ibáez, L., Blezek, D.: The design of simpleitk. Front. Neuroinformatics **7**, 45 (2013)
22. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. **32**, 8026–8037 (2019)