3D Global Fourier Network for Alzheimer's Disease Diagnosis Using Structural MRI

Shengjie Zhang[1,2], Xiang Chen[1,2], Bohan Ren[3], Haibo Yang[1,2], Ziqi Yu[1,2], Xiao-Yong Zhang[1,2(✉)], and Yuan Zhou[4(✉)]

[1] Institute of Science and Techonology for Brain-inspired Intelligence, Fudan University, Shanghai, China
[2] Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China
xiaoyong_zhang@fudan.edu.cn
[3] Department of Cyber Science and Technology, Beihang University, Beijing, China
[4] School of Data Science, Fudan University, Shanghai, China
yuanzhou@fudan.edu.cn

Abstract. Deep learning models, such as convolutional neural networks and self-attention mechanisms, have been shown to be effective in computer-aided diagnosis (CAD) of Alzheimer's disease (AD) using structural magnetic resonance imaging (sMRI). Most of them use spatial convolutional filters to learn local information from the images. In this paper, we propose a 3D Global Fourier Network (GF-Net) to utilize global frequency information that captures long-range dependency in the spatial domain. The GF-Net contains three primary components: a 3D discrete Fourier transform, an element-wise multiplication between frequency domain features and learnable global filters, and a 3D inverse Fourier transform. The GF-Net is trained by a multi-instance learning strategy to identify discriminative features. Extensive experiments on two independent datasets (ADNI and AIBL) demonstrate that our proposed GF-Net outperforms several state-of-the-art methods in terms of accuracy and other metrics, and can also identify pathological regions of AD. The code is released at https://github.com/qbmizsj/GFNet.

Keywords: Alzheimer's disease · Global fourier network · Multi-instance learning · MRI

1 Introduction

Alzheimer's disease (AD) has become one of the most prevalent neurological disorders due to the increasing population of AD [1]. In AD patients, their cognitive function is progressively impaired, accompanied by irreversible neurological

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16431-6_4.

L. Wang et al. (Eds.): MICCAI 2022, LNCS 13431, pp. 34–43, 2022.
https://doi.org/10.1007/978-3-031-16431-6_4

damage. An early diagnosis of AD could provide valuable information in the subsequent treatment, thereby delaying the onset of late stage symptoms, such as amnesia. Such a diagnosis can be achieved by structural magnetic resonance imaging (sMRI) which can capture the morphological changes induced by brain atrophy in a non-invasive way [2].

Existing methods for computer-aided diagnosis (CAD) of AD using sMRI could be categorized into traditional machine learning (ML) and deep learning (DL) [3]. Traditional ML methods follow the pipeline of regions-of-interest (ROIs) extraction and feature classification [4–10]. However, manual ROI extraction or feature construction is laborious and time-consuming. DL methods, especially convolutional neural networks (CNN), extract image features automatically, hence may greatly improve the performance of AD diagnosis [11–19].

Currently, several DL methods have been proposed for this task. Some methods replace the machine learning classifiers with neural networks while keeping the pre-defined ROIs [17,18]. Some other methods train CNN in a multi-instance learning strategy [14]. Typically, these DL methods use spatial convolution to extract local features [15]. However, spatial convolution means using local receptive fields in vision, which ignores global connections between disease-related brain regions.

To address this problem, we propose a *global Fourier network* (GF-Net). The GF-Net divides an input image into several sub-regions (patches) and consists of a patch embedding operation and a sequence of *global Fourier blocks* (GF block). Each block contains a 3D discrete Fourier transform, an element-wise multiplication, and a 3D inverse Fourier transform. The element-wise multiplication is introduced to learn filters in the frequency domain to capture global information. The GF-Net is trained by a *multi-instance learning* (MIL) strategy, which randomly drops some patches (filling with zero) to generate additional instances for training. Our model was applied on two AD datasets and showed promising performance. To illustrate the effectiveness of the extracted features, we present two interpretability analyses, *saliency map* and *Shapley value*, to highlight the pathological regions.

2 Methods

An overview of our architecture is given in Fig. 1. It consists of a patch embedding operation and a series of global Fourier blocks (GF block). Let $\mathbf{x} \in \mathbb{R}^{H\times W\times D\times C}$ denote the input image, where H, W, D denote the height, width, and depth respectively, C denotes the number of channels (1 for sMRI). We first extract several 3D patches $\{\mathbf{x}_i \in \mathbb{R}^{P\times P\times P\times C} : i = 1,\ldots,N\}$ from $\mathbf{x}$, where P is the patch size, $N = HWD/P^3$ is the number of patches. The patches are embedded in a E-dimensional space by linear transformation $\mathbf{z}_i = \mathbf{W}\mathrm{vec}(\mathbf{x}_i)$, which is implemented by E 3D convolutions with a stride equal to P. Denote the resulting tensor by $\mathbf{z} \in \mathbb{R}^{\tilde{H}\times\tilde{W}\times\tilde{D}\times E}$ ($\tilde{H} = \frac{H}{P}$, $\tilde{W} = \frac{W}{P}$, $\tilde{D} = \frac{D}{P}$). This tensor will go through a sequence of GF blocks for prediction.

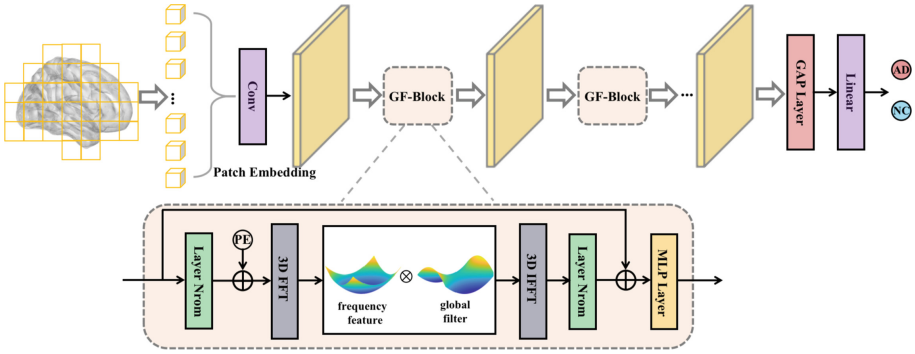

Fig. 1. The overall architecture of the GF-Net. It is a frequency attention model that consists of a sequence of blocks. Each block features a 3D discrete Fourier transform to convert features from the spatial domain to the frequency domain, an element-wise multiplication between frequency features and learnable global filters, and a 3D inverse Fourier transform to convert the features back to the spatial domain.

2.1 Global Fourier Block

A GF block adds $\mathbf{z}$ to its frequency filtered version before feeding it to a multi-layer perceptron (MLP) f_{MLP}:

$$f_{GF}(\mathbf{z}) = f_{MLP}(\mathbf{z} + f_{FF}(\mathbf{z})),$$

where $f_{FF} : \mathbb{R}^{\tilde{H}\times\tilde{W}\times\tilde{D}\times E} \to \mathbb{R}^{\tilde{H}\times\tilde{W}\times\tilde{D}\times E}$ denotes a channel-wise frequency filtering operation by Fourier transform. We first explain f_{FF} in details below.

Frequency Filtering by 3D Fourier Transform. Simply put, f_{FF} is composed of a layer normalization (LN), a positional embedding (PE), Fourier and inverse Fourier transform, and finally another LN (see Fig. 1). Given input $\mathbf{z}$, the LN operation normalizes it along the channels such that the values at each voxel has a zero mean and a standard deviation of 1. The PE operation adds a learnable vector to each voxel. After these two operations, denote the resulting tensor by $\tilde{\mathbf{z}}$ and its jth channel by $\tilde{\mathbf{z}}^j$. The 3D image $\tilde{\mathbf{z}}^j$ has a size $\tilde{H} \times \tilde{W} \times \tilde{D}$ and can be seen as a discretized version of a 3D function $\tilde{z}^j : \mathbb{R}^3 \to \mathbb{R}$. We use the 3D Fourier transform to capture the global information of this function:

$$\hat{z}^j(\xi) = \mathcal{F}(\tilde{z}^j)(\xi) = \int_{\mathbb{R}^3} \tilde{z}^j(\mathbf{y}) e^{-i2\pi\mathbf{y}\cdot\xi} d\mathbf{y},$$

where $\xi \in \mathbb{R}^3$. For a discretized image $\tilde{\mathbf{z}}^j$, a fast Fourier transform (FFT) implementation can be used to convert it to a tensor in complex values. Combining all these tensors, we have $\hat{\mathbf{z}} \in \mathbb{C}^{\tilde{H}\times\tilde{W}\times\tilde{D}\times E}$, which can be seen as features in the frequency domain. These features are multiplied element-wisely by a learnable filter $\mathbf{K} \in \mathbb{C}^{\tilde{H}\times\tilde{W}\times\tilde{D}\times E}$:

$$\hat{\mathbf{z}}' = \mathbf{K} \odot \hat{\mathbf{z}}$$

where $\odot$ is the element-wise multiplication (Hadamard product). Then, a 3D inverse Fourier transform is performed channel-wisely to convert the frequency domain features to the spatial domain. Denote the output by $\tilde{\mathbf{z}}' \in \mathbb{R}^{\tilde{H} \times \tilde{W} \times \tilde{D} \times E}$. Finally, another LN is applied to $\tilde{\mathbf{z}}'$ and the result is the output of f_{FF}.

Theoretically, multiplication in the frequency domain is equivalent to convolution in the spatial domain. We have some empirical results showing that as the filter size increases in spatial convolution, the performance improves in our framework, possibly due to better learned global information (see Fig. S1 in the supplementary material). However, the computational complexity of spatial convolution in the 3D domain increases rapidly as the filter size increases. Hence, filtering in the frequency domain becomes a more viable approach. Similar results have also been found in a Transformer-like framework [20].

Multi-layer Perceptron Layer. The MLP layer in the GF block reshapes $\mathbf{z} + f_{FF}(\mathbf{z})$ to a matrix of size $\tilde{H}\tilde{W}\tilde{D} \times E$ and linearly transforms its columns, followed by an activation function. The linear transformation is designed to keep the resulting size unchanged (i.e. $\tilde{H}\tilde{W}\tilde{D} \times E$) such that the resulting matrix can be reshaped back to the original size $\tilde{H} \times \tilde{W} \times \tilde{D} \times E$. Hence, the whole GF block preserves the tensor size.

The above GF block is repeated multiple times. The output of the last GF block is fed to a classifier layer for prediction.

2.2 Classifier Layer and Loss Function

The classifier layer contains a global average pooling (GAP) layer and a linear layer. The GAP layer calculates the average value along the channel dimension. Then the resulting 3D image is flattened and the linear layer is applied to the flattened vector to produce the logits. Finally, a standard cross entropy loss is used for training.

2.3 Multi-instance Learning Strategy

To avoid the over-fitting problem, we drop some patches randomly to generate multiple instances for training. Dropping patches is implemented by filling the patches with zero [21]. Hence, for each image, we obtain multiple instances that have the same size and share the same class label as the original image [22]. For example, given an image, we can randomly pick 30, 40, 50 patches for dropping, which results in 4 different instances that share the same label.

3 Experiments and Results

3.1 Dataset Description and Experimental Setup

The brain structural images (T1WI) acquired at 3.0T MRI systems, were provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL). We

pre-processed the sMRI images by spatial normalization, intensity normalization, and background removal. The spatial normalization registered the images to the MNI152 template by affine transformation as implemented by the FSL package in Python. The intensity normalization transforms the intensities linearly such that they have a zero mean and a standard deviation of 1. Any voxel with intensity lower than -1 was set to -1 and those with intensity higher than 2.5 were truncated to 2.5. The background removal step stripped the skull according to the MNI152 template and set the voxels outside the skull to -1. Quality check of these steps was performed by visual inspection. After quality check, 417 images were left from ADNI (229 for NC, 188 for AD), and 380 images from AIBL (320 for NC, 62 for AD). All of them were acquired within 6 months from the date of diagnosis.

In the experiment, 60% of the ADNI data were used for network training, 20% for validation and the remaining 20% for testing. The entire AIBL data were kept for testing. The experiment was repeated 10 times with random training/validation/test split and the mean and standard deviation of the accuracy/sensitivity/specificity/F1-score were calculated.

3.2 Implementation Details

Our model was implemented in PyTorch and accelerated by 2 NVIDIA A-6000 GPUs and 4 NVIDIA V-100 GPUs. The Gaussian error linear unit (GELU) was used as the activation function in the MLP of the GF blocks. The patch size was set to $10 \times 10 \times 10$. The batch size was set to 10. The learning rate was set to 0.0001 and the maximal number of training epochs was set to 1000 with a weight decay of 0.9 for each 100 epochs. The number of GF blocks was set to 8. For each image, we dropped 1/16, 1/8, 1/4, 1/2 patches according to the patch indices randomly 10 times, resulting in a training set whose size is 40 times the original size.

3.3 Experimental Results on ADNI and AIBL

To show the effectiveness of the proposed GF-Net, we compared it with VBM [23], CNN3D [24], ResNet3D [25], FCN [14], and ViT3D [26]. VBM calculates gray matter (GM) densities and uses a linear SVM classifier. The other competing methods were run according to their online code. The results are shown in Table 1.

Table 1. Results for AD classification (AD vs. NC) on ADNI and AIBL. The results (in %) were calculated based on 10 random training/validation/test splits. The rows in italic have their numbers directly copied from the papers. The "# of Images" column has a format of (number of images for AD, number of images for NC). The best result in each category is boldfaced.

Dataset	Method	# of Images	Accuracy	Sensitivity	Specificity	F1-score	AUC
ADNI	VBM	(188,229)	82.5 ± 4.4	75.5 ± 4.7	88.9 ± 3.2	82.8 ± 5.7	80.1 ± 4.2
	CNN3D		84.8 ± 3.8	86.2 ± 3.3	84.5 ± 4.8	85.3 ± 4.2	84.6 ± 4.4
	FCN		78.6 ± 5.7	82.2 ± 4.3	76.5 ± 6.4	79.2 ± 6.8	78.2 ± 5.3
	ViT3D		85.5 ± 2.9	87.9 ± 3.6	86.8 ± 3.7	87.3 ± 3.6	85.7 ± 3.5
	ResNet3D		87.7 ± 3.5	90.2 ± 2.8	89.7 ± 3.0	90.0 ± 3.5	86.2 ± 4.0
	Salvatore et al. [5]	*(137,162)*	*76.0*	–	–	–	–
	Cuingnet et al. [6]	*(137,162)*	*88.6*	*81.0*	***95.0***	*87.4*	–
	Eskildsen et al. [7]	*(194,226)*	*86.7*	*80.4*	*92.0*	*85.8*	–
	Cao et al. [8]	*(192,229)*	*88.6*	*85.7*	*90.4*	*88.0*	–
	Lin et al. [19]	*(188,229)*	*88.8*	–	–	–	–
	Tong et al. [10]	*(198,231)*	*90.0*	*86.0*	*93.0*	*89.4*	–
	Li et al. [18]	*(199,229)*	*89.5*	*87.9*	*90.8*	*89.3*	–
	Qiu et al. [16]	*(188,229)*	*83.4*	*76.7*	*88.9*	*82.4*	–
	H-FCN [14]	*(389,400)*	*90.5*	*89.7*	*91.3*	*90.5*	–
	DA-Net [15]	*(389,400)*	*92.4*	*91.0*	*93.8*	***92.4***	–
	GF-Net	(188,229)	$\mathbf{94.1 \pm 2.8}$	$\mathbf{93.2 \pm 2.4}$	90.6 ± 2.6	91.8 ± 2.6	$\mathbf{93.5 \pm 2.7}$
AIBL	VBM	(62,320)	81.7 ± 4.4	75.5 ± 4.7	86.3 ± 4.1	80.4 ± 6.1	81.5 ± 4.2
	CNN3D		86.2 ± 2.9	70.2 ± 5.7	88.7 ± 3.2	78.3 ± 4.8	80.3 ± 5.5
	FCN		77.2 ± 5.5	74.4 ± 4.6	78.8 ± 5.1	76.5 ± 4.8	76.8 ± 5.1
	ViT3D		87.5 ± 2.6	88.2 ± 3.4	91.8 ± 1.9	89.9 ± 4.0	87.9 ± 4.1
	ResNet3D		88.0 ± 3.6	91.1 ± 2.4	83.7 ± 4.6	87.2 ± 3.5	87.8 ± 3.7
	GF-Net		$\mathbf{93.2 \pm 2.4}$	$\mathbf{93.3 \pm 2.6}$	$\mathbf{94.6 \pm 3.3}$	$\mathbf{94.0 \pm 2.7}$	$\mathbf{93.8 \pm 2.9}$

Several observations could be derived from Table 1: 1) The three DL methods—CNN3D, ViT3D, ResNet3D—yield better performance than the traditional VBM method, suggesting the superiority of DL algorithms on this task. 2) The proposed GF-Net outperforms the competing methods by a relatively large margin in all the metrics. We also calculated the p-values and the p-values were less than 0.05. 3) Our results are even better than almost all the results in the literature even though that our results are calculated based on 10 random splits.

Additional results on classifying progressive mild cognitive impairment (pMCI) and stable mild cognitive impairment (sMCI) are available in Table S1 in the supplementary material (AUC: 87.8 ± 3.6). Similar to classifying AD and NC, GF-Net achieved superior accuracy on classifying pMCI and sMCI.

3.4 Ablation Study

We performed two ablation studies. First, we investigated the influence of the number of GF blocks. Using the ADNI dataset, we selected the number of blocks from {4,6,8,10} and reported the results in Table 2. As shown in Table 2, most of the metrics increased when the number of blocks increased from 4 to 8. However,

when the number of blocks reached 10, the performance declined slightly, which could be attributed to the complexity of the network that results in over-fitting. Nevertheless, the proposed network in this case still outperformed other state-of-the-art methods with a large margin, demonstrating its robustness.

Table 2. Ablation study (results in %) on the number of blocks on ADNI.

# of blocks	Accuracy	Sensitivity	Specificity	F1-score	AUC
4	93.3 ± 3.1	92.1 ± 2.7	91.6 ± 2.4	91.8 ± 2.7	91.5 ± 3.6
6	93.5 ± 3.5	91.7 ± 2.3	91.3 ± 2.2	91.5 ± 3.1	93.1 ± 3.0
8	94.1 ± 2.8	93.2 ± 2.4	90.6 ± 2.6	91.8 ± 2.6	93.5 ± 2.7
10	93.4 ± 2.7	91.4 ± 2.6	91.4 ± 2.7	91.4 ± 3.2	92.5 ± 3.5

In the previous experiment, the patch size is fixed as $10 \times 10 \times 10$. Now, we show the effect of different patch sizes. Table 3 shows the classification results when the patch size ranges in $10 \times 10 \times 10$, $15 \times 15 \times 15$, $20 \times 20 \times 20$, $25 \times 25 \times 25$. From Table 3, we can see that our GF-Net is also not sensitive to the change of patch size. The patch size of $10 \times 10 \times 10$ shows the best performance, while the rest sizes also outperform the current state-of-the-art methods.

Table 3. Ablation study (results in %) on the patch size on ADNI.

Patch size	Accuracy	Sensitivity	Specificity	F1-score	AUC
$10 \times 10 \times 10$	94.1 ± 2.8	93.2 ± 2.4	90.6 ± 2.6	91.8 ± 2.6	93.5 ± 2.7
$15 \times 15 \times 15$	93.7 ± 3.1	91.2 ± 3.7	90.3 ± 3.4	90.8 ± 3.6	92.3 ± 2.9
$20 \times 20 \times 20$	93.5 ± 3.4	90.7 ± 3.9	91.1 ± 2.9	90.9 ± 3.1	92.0 ± 3.1
$25 \times 25 \times 25$	92.4 ± 2.5	89.9 ± 3.5	90.2 ± 3.1	90.0 ± 2.7	91.4 ± 3.6

3.5 Interpretation by Saliency Map and Shapley Value

To investigate the discriminative regions for classifying AD and NC, we employed two interpretation analyses, saliency map and Shapley value. Figure 2 shows the corresponding saliency map, which calculates the derivative of the logit with respect to the input image. Figure 2 shows that most of the discriminative voxels (in red) are located in the brain regions closely associated with clinical diagnosis, such as hippocampus, amygdala, and thalamus [27,28].

We also use Shapley value explanation to quantitatively evaluate the importance of the extracted features, i.e. the tensor before the last linear layer. Specifically, the features from all the instances are reduced to 3 dimensions by PCA, called PC_1, PC_1, PC_3. They are combined with non-imaging features: mini-mental state examination (MMSE), age, APOE, and gender. The combined features are fed into XGboosting for classification. Finally, the Shapley values are calculated for all the features and shown in Fig. 3.

It is not surprising that MMSE has the widest spread since MMSE measures the mental state of the patients (the lower, the more severe), which can be considered as another ground truth to classify AD and NC. Furthermore, our imaging

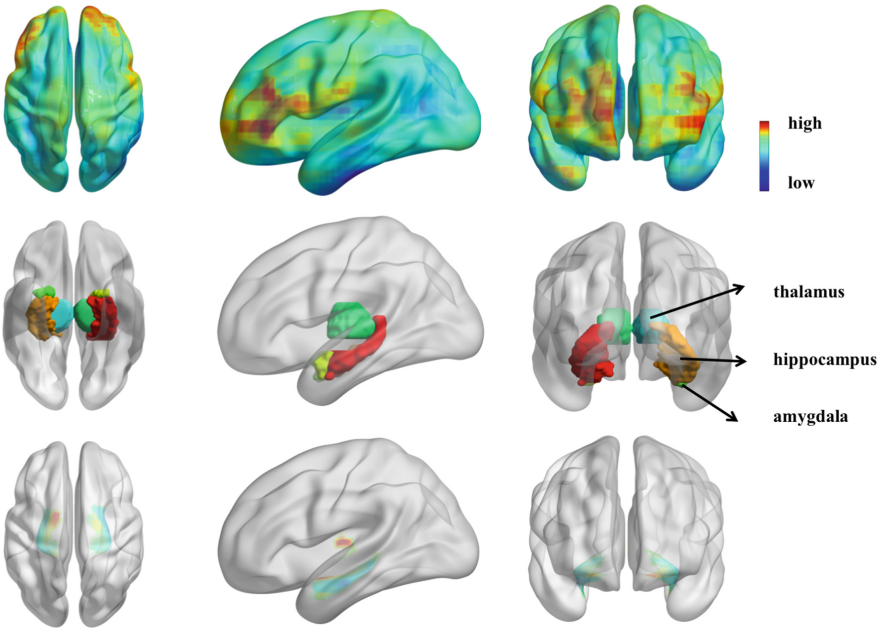

Fig. 2. Saliency map of the trained network on an example image. The three rows show the saliency map, clinically related reference regions, saliency map truncated to the reference regions respectively.

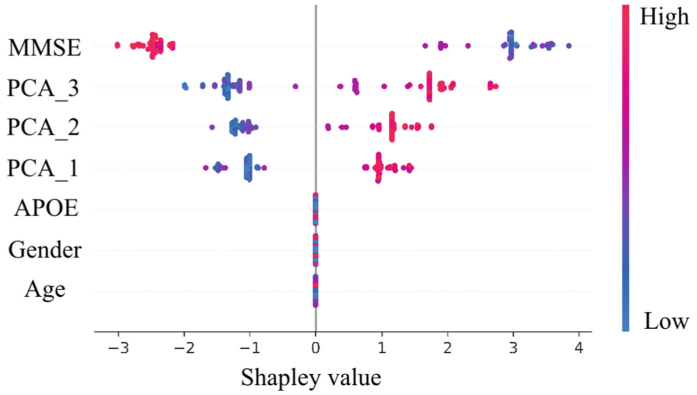

Fig. 3. Shapley values for analyzing importance scores of the image features extracted by GF-Net, when the image features are combined with the non-imaging features for classification. The spread of the Shapley values of a feature reflects its impact on the model output.

features have a spread close to MMSE, outperforming the other non-imaging features by a large margin. This quantitative interpretation demonstrates the effectiveness of our method again.

4 Conclusion

We propose a 3D GF-Net for AD diagnosis using sMRI. The GF-Net uses frequency filtering to capture disease-related global information. The network is trained by an MIL strategy to avoid the over-fitting problem. The classification results on ADNI and AIBL demonstrate that our method can significantly outperform other state-of-the-art methods for classifying AD and NC (also for classifying pMCI and sMCI). Two interpretability analyses, saliency map and Shapley value, show that our method could identify clinically meaningful regions.

Acknowledgments. This work was supported in part by Science and Technology Commission of Shanghai Municipality (20ZR1407800), Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), and Fudan Univerisity startup fund (JIH2305006Y).

References

1. Knopman, D.S., et al.: Alzheimer disease. In: Nature reviews Disease Primers, vol. 7.1, pp. 1–21 (2021)
2. Damulina, A., et al.: Cross-sectional and longitudinal assessment of brain iron level in Alzheimer disease using 3-T MRI. Radiology **296**(3), 619–626 (2020)
3. Wen, J., et al.: Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Med. Image Anal. **63**, 101694 (2020)
4. Zeng, N., et al.: A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. Neurocomputing **320**, 195–202 (2018)
5. Salvatore, C., et al.: Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front. Neurosci. **9**, 307 (2015)
6. Cuingnet, R., et al.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage **56**(2), 766–781 (2011)
7. Eskildsen, S.F., et al.: Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage **65**, 511–521 (2013)
8. Cao, P., et al.: Nonlinearity-aware based dimensionality reduction and oversampling for AD/MCI classification from MRI measures. Comput. Biol. Med. **91**, 21–37 (2017)
9. Sørensen, L., Nielsen, M., Initiative, A.D.N., et al.: Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination. J. Neurosci. Methods **302**, 66–74 (2018)
10. Tong, T., et al.: Multiple instance learning for classification of dementia in brain MRI. Med. Image Anal. **18**(5), 808–818 (2014)
11. Khvostikov, A., et al.: 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies. arXiv preprint arXiv:1801.05968 (2018)

12. Liu, M., et al.: Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018)
13. Zhou, B., et al.: Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929 (2016)
14. Lian, C., et al.: Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Trans. Pattern Anal. Mach. Intell. **42**(4), 880–893 (2018)
15. Zhu, W., et al.: Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI. IEEE Trans. Med. Imaging **40**(9), 2354–2366 (2021)
16. Qiu, S., et al.: Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain **143**(6), 1920–1933 (2020)
17. Li, H., Habes, M., Fan, Y.: Deep ordinal ranking for multi-category diagnosis of Alzheimer's disease using hippocampal MRI data. arXiv preprint arXiv:1709.01599 (2017)
18. Li, F., Liu, M., Initiative, A.D.N., et al.: Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Comput. Med. Imaging Graph. **70**, 101–110 (2018)
19. Lin, W., et al.: Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front. Neurosci. **12**, 777 (2018)
20. Rao, Y., et al.: Global filter networks for image classification. Adv. Neural. Inf. Process. Syst. **34**, 980–993 (2021)
21. He, K., et al.: Masked autoencoders are scalable vision learners (2021). arXiv: 2111.06377 [cs.CV]
22. Couture, H.D., Marron, J.S., Perou, C.M., Troester, M.A., Niethammer, M.: Multiple instance learning for heterogeneous images: training a CNN for histopathology. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 254–262. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_29
23. Ashburner, J., Friston, K.J.: Voxel-based morphometry-the methods. Neuroimage **11**(6), 805–821 (2000)
24. Kruthika, K.R. HD Maheshappa, Alzheimer's disease neuroimaging initiative. CBIR System using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis. Inf. Med. Unlocked **14**, 59–68 (2019)
25. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
26. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
27. Gerischer, L.M., et al.: Combining viscoelasticity, diffusivity and volume of the hippocampus for the diagnosis of Alzheimer's disease based on magnetic resonance imaging. NeuroImage Clin. **18**, 485–493 (2018)
28. Shao, W., et al.: Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. Comput. Med. Imaging Graph. **80**, 101663 (2020)