Early Diagnosis of Alzheimer's Disease Based on Selective Kernel Network with Spatial Attention

Huanhuan Ji[1], Zhenbing Liu[1], Wei Qi Yan[2(✉)], and Reinhard Klette[2]

[1] Guilin University of Electronic Technology, Guilin, China
[2] Auckland University of Technology, Auckland, New Zealand
wyan@aut.ac.nz

Abstract. Alzheimer's disease (AD) is a neurodegenerative disorder which leads to memory and behaviour impairment. Early discovery and diagnosis can delay the progress of this disease. In this paper, we propose a new deep learning method called selective kernel network with attention for early diagnosis of AD using magnetic resonance imaging. Generally, deep learning methods for high-accuracy recognition are based on structure of deep neural networks by stacking a myriad of convolutional layers in the model. In this paper, the structure of SKANet is constructed similarly to that of ResNeXt by repeating residual blocks with the same topology and group convolution for saving computational costs. Different from ResNeXt, the primary convolution is replaced by using selective kernel convolution to adaptively adjust the receptive field based on imported information. Then, attention mechanism is added to the bottom of the block to emphasize on important features and suppress unnecessary ones for more accurate representation of the network. The block is termed as selective kernel with attention block that consists of a sequence of operations followed by the order: a convolution with kernel size 1×1, a selective kernel convolution, a convolution with kernel size 1×1, and spatial attention mechanism. The effectiveness of this proposed model is verified based on the Alzheimer's Disease Neuroimaging Initiative dataset. Our experimental results show superiority of the proposed model for the early diagnosis of AD. The classification accuracy of AD and mild cognitive impairment reaches up to 98.82%.

Keywords: Alzheimer's disease diagnosis · MRI · Spatial attention mechanism · Selective kernel

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder characterised by aphasia, amnesia, visual impairment, and behavioural changes such as dementia. This disease can be divided into three stages including *normal control* (NC), *mild cognitive impairment* (MCI), and AD. There are distinctions of the brain structure

S. Palaiahnakote et al. (Eds.): ACPR 2019, LNCS 12047, pp. 503–515, 2020.
https://doi.org/10.1007/978-3-030-41299-9_39

between the three stages. MCI, a prodromal stage of AD, is an important phase for clinical trials. In MCI, individuals have a mild change in behaviour that can be observed easily by anybody who is close to the patients. AD is an incurable disease that often occurs on the elderly. But we can diagnose it early and take a timely solution to delay the deterioration of this disease. The early diagnosis of AD has become very important and attracted numerous researchers to investigate this problem. It is reported [3] that there will be 1 out of 85 people troubled with this disease by 2050.

Magnetic resonance imaging (MRI) provides a clear structure of the brain in a 3D view noninvasively. Thus, we capture the mild changes due to the atrophic process [9]. A plethora of *computer-aided diagnosis* (CAD) approaches [1,24,31] have been proposed for the early diagnosis of AD based on MRI. Meanwhile, *white matter* (WM) and *grey matter* (GM) of brain play a pivotal role on the diagnosis of AD. A series of methods have been proposed based on the GM and WM from MRI for the diagnosis of AD [7,8,30]. In this paper, GM and WM from MRI are preprocessed and classified using our proposed method.

Multiple machine learning methods have been proposed to aid the diagnosis of AD. The diagnosis through MRI using machine learning methods is split into three independent phases [31]: (1) Predetermination of *region-of-interest* (ROI), (2) extraction or selection of features from ROIs, and (3) construction of a classifier for classifications.

T-test [26] was proposed to extract texture features, classify the images, and diagnose AD by using collaborative representation. The multi-atlas-based method [25] exploits useful information in feature representations to classify the AD. A *support vector machine* (SVM) was used as a classifier to classify the features extracted from each atlas space; the results of multiple atlases were combined by using majority voting methods for making a final decision. A new multimodal data fusion and classification method was set forth [33] based on a kernel combination on the diagnosis of AD which provides a unified way to combine heterogeneous data, particularly for the cases that different types of data cannot be directly concatenated. Generally, the classifiers based on those data from ROIs often suffer overfitting due to its high dimensionality compared with a small number of samples for model training. Other research work also has been adopted for the diagnosis of AD with a sub-optimal performance because of the heterogeneous nature between the features extraction and classifier training. Thus, various methods [2,5,23,35] have been proposed to improve the diagnosis accuracy of AD.

Deep learning is a revolutionary methodology compared with traditional machine learning approaches. Instead of separating feature extraction and classifier training, deep learning can train a model and learn its parameters from end to end without an engagement of human specialists. *Convolutional neural networks* (CNNs) have significantly pushed computer vision forward based on the rich representation of digital images [13].

Deep learning also shows its powerful capability on the early diagnosis of AD recently [22]. A sparse autoencoder [10] was proposed to learn a set of bases from natural images and applied convolution to extract features from the MRI so as to classify the AD. LeNet-5 and convolutional neural networks were employed [32] to classify Alzheimer's disease with an accuracy of 96.85%. The 16-layer VGGNet was modified [4] for the 3-way classification between AD, MCI, and NC. A hierarchical and fully convolutional network [21] was put forward to identify multiscale discriminative location for the diagnosis of AD using MRI. Suk et al. [36] embarked on a CNN that applied multiple sparse regression models to various parameters of regularisation for AD diagnosis. Ensemble learning [16,29] applies voting methods to the final results with the base classifiers for the early diagnosis of AD.

From our observations, we see that the structure of deep learning models is prone to be complex. From LeNet [19] to ResNet [11], the networks are becoming increasingly much deeper and richer for feature representations. ResNet is constructed by stacking residual blocks with the same topology along with a skip connection. Inception [37–39] shows its depth as a neural network and also takes prominent effect on solving problems in computer vision. Xception [6] and ResNeXt [41] empirically reveal that the cardinality can boost the representation of networks and save the time of computations.

Generally, *receptive fields* (RFs) share the corresponding size in every layer for neural networks. Spillmann et al. [34] stated that the RF sizes can be adjusted by using stimuli. Li et al. [20] proposed selective kernel networks to adaptively adjust the RF size by using selective kernel (SK) convolution. In this paper, SK convolution is employed for our proposed network. SK convolution consists of three operations: Splitting, fusing, and selecting. During splitting, two kernels of the sizes 3×3 and 5×5 are adapted to extract the features from the input. A fusing operation is used to aggregate the information from the two branches so as to produce the global information and generate weights for the two branches. The two weights from the fusing operator are multiplied with the corresponding feature maps from the splitting operation during the selecting operation. Thus, the exports from two branches are operated by using element-wise summation to generate the final output of SK convolution.

The residual block of ResNeXt is used as the base for our model. In our proposed model, the main convolution is replaced by using SK convolution. The spatial attention mechanism is added to the end of this block so as to learn the attention in the spatial domain. Our work, reported in this paper, mainly has the following contributions:

- ResNeXt is set as the base architecture which is constructed by repeating the block with the same topology to share and save the hyperparameters (width and size of filters). Thus, a group convolution is introduced to reduce the computational cost.
- SK convolution is employed for our network to select the appropriate RF size and to increase the power of representations.
- Spatial attention is used to emphasise on those important features and suppress unnecessary ones to improve the representation of networks.

The rest of this paper is organised as follows. In Sect. 2, we introduce the progress of convolutional neural networks. Our method based on the proposed deep neural network is introduced in Sect. 3. In Sect. 4, experimental results are demonstrated along with our proposed model. Section 5 concludes.

2 Related Work

Convolutional neural networks with a deep structure by stacking convolutional layers have made a series of breakthroughs for image classification recently. YOLOv3 has been used on the human behaviour recognition with a good performance compared with traditional machine learning methods [27].

As the increase of layers, gradient vanishing and model degradation problems turned up during the training of neural networks. *Batch normalization* (BN) is introduced to reduce the problem of gradient vanishing. He et al. [11] used a shortcut to reduce gradient vanishing and model degradation. A residual learning framework is offered to train the network. ResNet has shown that residual learning is easily to be optimised and can gain high accuracy from considerably increased depth. With the stack of residual blocks, ResNet has shown a higher accuracy and superior performance on the classification.

Inception networks [37–39] are with the successes of all multi-branch networks where every branch is designed elaborately. The appearance of inception made the model much wider. Inception v1 (GoogleNet) [37] was designed by using a sufficient size of filters to fit various RFs. Meanwhile, 1×1 convolution is introduced so as to reduce the dimension of the inputs and computational cost. Thus, the 5×5 convolution kernel is replaced by using two 3×3 convolution operators to reduce the complexity; correspondingly, $n \times n$ convolution is replaced by using $n \times 1$ and $1 \times n$ kernels; the 1-dimensional convolution is employed to accelerate the computations. Our model follows the idea of the Inception networks with filters for multiple branches.

The significant contribution of group convolution is to reduce the computational cost. AlexNet [17] is the neural network that firstly took advantage of group convolution successfully. Then, the group convolution was employed on ResNeXt [41] to reduce computations and improve the outcomes of classifications. In this paper, group convolution is introduced based on our designed neural network to reduce the computational cost.

An attention mechanism is an important part of our proposed model to capture useful information. Attention mechanisms have been employed on various fields such as image captioning [42]. According to an attention mechanism, the informative features are strengthened; simultaneously, the less useful ones are suppressed [14,15,18,28]. Wang et al. [40] created a residual attention network by stacking attention modules to make the model working. Furthermore, SENet [12] was proposed to use a gating mechanism on channels to improve the representation and performance of this network. Beyond channel attention, CBAM [43] also introduced spatial attention in the network to capture the valid information of inputs in the same way.

Unlike existing networks, our proposed neural network introduces the SK convolution to adaptively adjust the RF sizes and spatial attention to improve the representation of networks.

3 Methods

In this section, the details of our model are delineated which encapsulate the architecture of this neural network. First, our model is constructed based on ResNeXt with SK convolution. The spatial attention mechanism is added to the block for strengthening representation capability of the network.

In our model, SK convolution is employed to adaptively adjust the RF sizes. SK convolution is implemented by introducing two branches with kernel sizes of 3×3 and 5×5. The SK convolution is constructed by using the three operations of splitting, fusing, and selection.

The image slices from MRI are sent to the model for training the classifier from end to end. Then, the features of these slices, extracted from a convolutional layer with kernel size 1×1, are sent to the SK convolution for adaptively adjusting RF size. SK convolution is conducted by following two paths with kernel sizes 3×3 and 5×5. The splitting outputs are grouped into two branches.

During the splitting operation, a convolution operation based on a 3×3 kernel is the standard; 5×5 convolution is a dilated one with dilation size 2. The outputs from two branches are fused by using element-wise summation. Global average pooling is employed on the fusing operation to conduct channel-wise statistics. Full connection is applied to the channel-wise statistics by using adaptive selection so as to squeeze the network. Soft attention is added into the squeezed network, mixed with selected spatial scales and different weights which are applied to the two branches.

The weights are multiplied with the features extracted from splitting operations with different kernel sizes. We fuse outputs from the two branches via element-wise summation; the final outputs of SK convolution are conducted with kernel size 1×1; the spatial attention module is embedded to the bottom of the block.

In the spatial attention, the outputs of 1×1 convolution are operated by using max pooling and average pooling to generate two 2D maps on each channel. Then, the two maps are concatenated and convoluted by using a standard convolution of size 7×7 to produce a 2D spatial attention map which determines how to emphasise or suppress the feature maps.

Similar to ResNeXt, our proposed model is constructed by using the stack of bottleneck blocks called *selective kernel attention* (SKA). Each SKA block consists of a sequence of operations, following the order: a standard convolution with 1×1 kernel, a SK convolution, a standard convolution with 1×1 kernel, and spatial attention as shown in Fig. 1.

Different to ResNeXt, the SK convolution is used to replace the large kernel convolution in the bottleneck block of ResNeXt so as to select appropriate RF sizes in an adaptive manner, the spatial attention is added to the bottom of the block. The architecture of our proposed model is similar to the ResNeXt for two reasons: (1) Compared with standard convolution, it has a lower computational cost by introducing the group convolution; (2) the network is one of the popular networks constructed by repeating a block with the same topology so as to reduce the choices of hyperparameters and the complexity of the design.

In this section, we introduce our neural network having 50 layers, called SKA-50. Table 1 displays the structure of the 50-layer network with the number of stacked blocks on each stage. During each stage, path M is set as 2 which means that there are two different kernels that are selected and aggregated. The group number G is set as 32 to control the cardinality on each path; r is the reduction ratio to control the numbers of parameters in the fusing operation.

4 Experiments

The data for our experiments is taken from the public dataset *Alzheimer's Disease Neuroimaging Initiative* (ADNI) at adni.loni.usc.edu. ADNI was designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking AD. In this paper, 490 MRI images were grouped into 143 AD (108 train samples, 35 test samples), 203 MCI (153 train samples, 50 test samples), and 144 NC (108 train samples, 36 test samples) in `.nii` format which were used to validate the effect of our model. The proportion between male and female samples in each category is roughly equal.

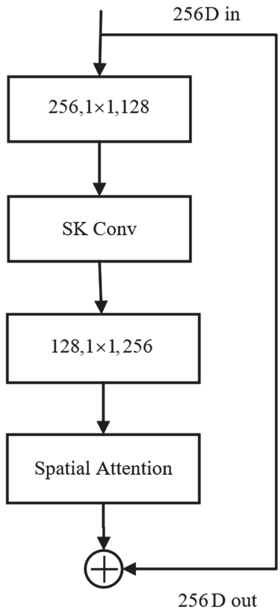

Fig. 1. A block of an SKANet.

There are minor head movements and noises in the MRI images downloaded directly from ADNI. We corrected the images to exclude the negative influence due to head movements by using *statistical parametric mapping* (SPM). Then, the images were resized into $192\times192\times160$. WM and GM are segmented from the resized images by using SPM. Then, the segmented GM and WM images were split into 192 slices in `.tif` format. According to prior knowledge, consecutive sets of 20 slices with significant brain structure from each MRI were selected from GM and WM. The slices from WM were added to the corresponding slices of GM and resized to 224×224 as the inputs of our model.

All experiments were implemented based on Python 3.5 and Pytorch framework. Our model was trained from scratch for 60 epochs on an Nvidia Tesla P100 GPU. The Adam optimiser was used for model training; the learning rate was set to be 10^{-3} at the beginning; the attenuation started from 10^{-5}; the size of mini batch was initialised as 20; the loss was measured by using the cross-entropy function. The data was preprocessed by using SPM based on MATLAB.

Our model was mainly validated on the early diagnosis of AD (i.e., AD *vs* MCI, and MCI *vs* NC). Classification *accuracy* (ACC), *specificity* (SPE),

Table 1. The 50-layer networks whose structures are similar to that of ResNeXt. The three columns show the structures of ResNeXt-50, SKNet-50, and SKANet-50 with 32×4D, respectively. The filter sizes and feature map dimensions are shown in brackets.d

Output	ResNeXt-50	SKNet-50	SKANet-50
112×112	7×7, 64, stride 2		
56×56	3×3 max pool, stride 2		
56×56	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128, \; G=32 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 128 \\ \mathrm{SK}[M=2, G=32, r=16], 128 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 128 \\ \mathrm{SK}[M=2, G=32, r=16], 128 \\ 1\times1, 256 \\ \text{spatial attention} \end{bmatrix}\times3$
28×28	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256, \; G=32 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 256 \\ \mathrm{SK}[M=2, G=32, r=16], 256 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 256 \\ \mathrm{SK}[M=2, G=32, r=16], 256 \\ 1\times1, 512 \\ \text{spatial attention} \end{bmatrix}\times4$
14×14	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512, \; G=32 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 512 \\ \mathrm{SK}[M=2, G=32, r=16], 512 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 512 \\ \mathrm{SK}[M=2, G=32, r=16], 512 \\ 1\times1, 1024 \\ \text{spatial attention} \end{bmatrix}\times6$
7×7	$\begin{bmatrix} 1\times1, 1024 \\ 3\times3, 1024, \; G=32 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 1024 \\ \mathrm{SK}[M=2, G=32, r=16], 1024 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 1024 \\ \mathrm{SK}[M=2, G=32, r=16], 1024 \\ 1\times1, 2048 \\ \text{spatial attention} \end{bmatrix}\times3$
1×1	7×7 global average pool, 2-D f_c, softmax		

sensitivity (SEN), and the *area under receiver operating characteristic curve* (AUC) are used to evaluate the results of classification. The AUC is calculated based on all possible pairs of SEN and SPE obtained by changing the thresholds of the classification scores which are yielded by using the trained networks.

Let TP, TN, FP, and FN be the rates of true positive, true negative, false positive, and false negative, respectively. The common definitions are as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$SEN = \frac{TP}{TP + FN} \tag{2}$$

$$SPE = \frac{TN}{TN + FP} \tag{3}$$

Our model is constructed based on ResNeXt with SK convolution and a spatial attention mechanism. In this section, we will assert the effects of SK convolution and spatial attention by controlling variables to execute the experiments on the 50-layer networks. The results from our model are compared with those of the ResNeXt-50 network with SK convolution called SK-50, ResNeXt network with the spatial attention, and ResNeXt-50 to verify the effectiveness of our idea.

Table 2. Accuracy, sensitivity, specificity, and AUC of various classification methods for MCI and NC.

Models	ACC (%)	SEN (%)	SPE (%)	AUC
ResNeXt-50	84.88	72.22	94.00	0.96
ResNeXt-50+attention	86.05	75.00	94.00	0.96
SK-50	87.21	80.56	92.00	0.94
SKA-50	89.53	86.11	92.00	0.97

Table 2 shows our results based on MCI and NC from those models. From Table 2, we see that the accuracy of SK-50 is higher than ResNeXt-50 that reflects the model has become better after introduced the SK convolution. Furthermore, both accuracy and sensitivity of SK-50 are higher than ResNeXt-50. Meanwhile, the performance of SKA-50 is superior to the ResNeXt with spatial attention that also illustrates the effectiveness of SK convolution.

Compared to ResNeXt-50, the accuracy and sensitivity of ResNeXt with the attention block have been greatly improved that show the effectiveness of spatial attention. Compared to the SK-50, the accuracy and sensitivity of SKA-50 have increased by 2.32% and 5.55%, respectively. The AUC of SKA-50 also was raised, which shows the superior impact of the spatial attention.

Table 3. Accuracy, sensitivity, specificity, and AUC of various classification methods for AD and MCI.

Models	ACC (%)	SEN (%)	SPE (%)	AUC
ResNeXt-50	95.29	100.00	88.57	1.00
ResNeXt-50+attention	97.65	98.00	97.14	0.99
SK-50	97.65	96.00	100.00	1.00
SKA-50	98.82	98.00	100.00	1.00

We furthermore validate our model for classification based on AD and MCI to verify the effectiveness of SK convolution and spatial attention. Table 3 shows the classification results based on AD and MCI of those models. From Table 3, we see that our proposed model with SK convolution and spatial attention has taken a great step that also illustrates the effectiveness of SK convolution and spatial attention. In general, our model has significant improvement after introduced the SK convolution and spatial attention. The receiver operating characteristic (ROC) curves based on MCI and NC are shown in Fig. 2.

From Fig. 2, we see that our model has the best AUC on the early diagnosis of AD compared with other models. In general, the performance of our model with SK convolution is superior to the model without SK convolution. The model becomes better after adding the spatial attention block. From Tables 2 and 3, we see that our model SKA-50 with SK convolution and spatial attention has the best performances for the early diagnosis of AD.

We compared our results with previous work regarding the early diagnosis of AD. Table 4 shows the results of previous studies and our results. From Table 4, we see that our results have a better performance on the classifications of MCI. Our model has a higher accuracy of classification based on AD and MCI.

Table 4. Comparisons of the results of previous studies and our method (%)

References	AD *vs* NC			AD *vs* MCI			MCI *vs* NC		
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
Sarraf et al. [32]	96.85	-	-	-	-	-	-	-	-
Ortiz et al. [29]	90.00	86.00	94.00	84.00	79.00	89.00	83.00	67.00	95.00
Lian et al. [21]	90.00	82.00	97.00	-	-	-	-	-	-
Suk et al. [36]	91.02	92.72	89.94	-	-	-	73.02	77.60	68.22
Ji et al. [16]	98.59	97.22	100.00	97.65	96.00	100.00	88.37	80.56	94.00
Our method	98.59	97.22	100.00	**98.82**	**98.00**	100.00	**89.53**	**86.11**	92.00

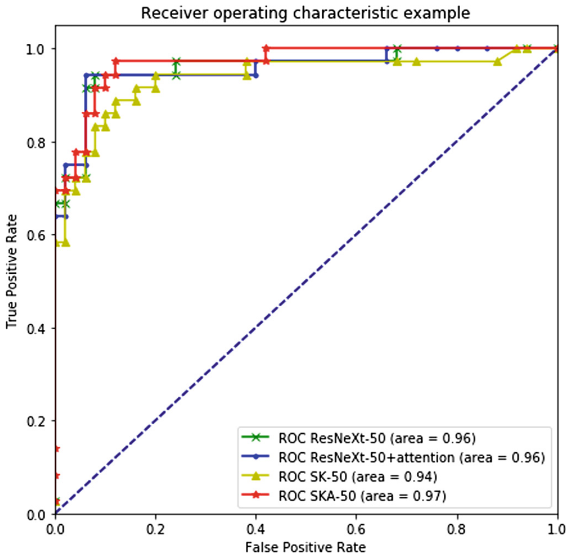

Fig. 2. The ROC curves for the comparisons between MCI and NC using our own methods.

5 Conclusion

In this paper, we constructed a model based on the ResNeXt with SK convolution and spatial attention to boost the representation of the network. Firstly, our model is constructed by stacking blocks with the same topology to reduce the hyperparameters and complexity of design. Then, selective kernel convolution was offered to adaptively adjust the RF sizes to capture more useful information through using dynamic selection mechanism. The spatial attention mechanism was attached to the module so as to improve the representation power of the network. In order to verify the effectiveness, we conducted the experiments for early diagnosis of AD. Our results show that the SKANet demonstrates superior performances for the early diagnosis of AD.

Acknowledgments. This project is supported by the study abroad program for graduate student of the Guilin University of Electronic Technology China and the National Natural Science Foundation of China under grants (61866009). The data used in this paper was downloaded from Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu). We are grateful to everyone who provided their support for this research project.

References

1. Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D.: Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. NeuroImage **145**, 137–165 (2017)
2. Arribas, J., Calhoun, V., Adali, T.: A automatic Bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from fMRI data. IEEE Trans. Bio-med. Eng. **57**(12), 2850–2860 (2010)
3. Brookmeyer, R., Johnson, E., Ziegler-Graham, K.: Forecasting the global burden of Alzheimer's disease. J. Alzheimers Assoc. **3**(3), 186–191 (2007)
4. Billones, D., Demetria, D., Hostallero, D.: DemNet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment. In: TENCON. IEEE, Singapore (2016)
5. Cheng, B., Zhang, D., Chen, S., Shen, D.: Predicting clinical scores using semi-supervised multimodal relevance vector regression. In: Suzuki, K., Wang, F., Shen, D., Yan, P. (eds.) MLMI 2011. LNCS, vol. 7009, pp. 241–248. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24319-6_30
6. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: CVPR. IEEE, Piscataway (2016)
7. Chyzhykand, D., Grana, M., Savio, A., Maiora, J.: Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI. Neurocomputing **75**(1), 72–77 (2012)
8. Cuingnet, R., Gerardin, E., Tessieras, J.: 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)
9. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. **6**(2), 67–77 (2010)
10. Gupta, A., Ayhan, M., Maida, A.: Natural image bases to represent neuroimaging data. In: ICML 2013, USA, pp. 987–994 (2013)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778. IEEE, Piscataway (2016)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR, pp. 7132–7141. IEEE, Piscataway (2018)
13. Huang, G., Liu, Z., Laurens, M.: Densely connected convolutional networks. In: CVPR, pp. 4700–4708. IEEE, Piscataway (2017)
14. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE TPAMI **20**(11), 1254–1259 (1998)
15. Itti, L., Koch, C.: Computational modelling of visual attention. Nat. Rev. Neurosci. **2**(3), 194–203 (2001)
16. Ji, H., Liu, Z., Yan, W., Klette, R.: Early diagnosis of Alzheimer's disease using deep learning. In: ICCCV, Korea (2019)
17. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: NIPS (2012)
18. Larochelle, H., Hinton, G.: Learning to combine foveal glimpses with a third-order Boltzmann machine. In: NIPS (2010)
19. Lecun, Y., Bottou, L., Bengio, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
20. Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: CVPR, pp. 510–519. IEEE, Piscataway (2019)

21. Lian, C., Liu, M., Zhang, J., Shen, D.: Hierarchical fully convolution network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Trans. PAMI **12**, 1–14 (2018)
22. Litjens, G., Kooi, T., Bejnordi, B., Setio, A., Ciompi, F., Ghafoorian, M.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
23. Liu, F., Wee, C., Chen, H., Shen, D.: Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification. NeuroImage **84**, 466–475 (2014)
24. Liu, M., Zhang, D., Chen, S., Xue, H.: Joint binary classifier learning for ECOC-based multi-class classification. IEEE Trans. Pattern Anal. Mach. Intell. **38**(11), 2335–2341 (2016)
25. Liu, M., Zhang, D., Shen, D.: View-centralized multi-atlas classification for Alzheimer's disease diagnosis. Hum. Brain Mapp. **36**(5), 1847–1865 (2015)
26. Liu, Z., Xu, T., Ma, C., Yang, H.: T-test based Alzheimer's disease diagnosis with multi-feature in MRIs. Multimedia Tools Appl. **77**(22), 29687–29703 (2018)
27. Lu, J., Yan, W., Nguyen, M.: Human behaviour recognition using deep learning. In: AVSS (2018)
28. Mnih, V., Heess, N., Graves, A.: Recurrent models of visual attention. In: NIPS (2014)
29. Ortiz, A., Munilla, J., Gorriz, M.: Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int. J. Neural Syst. **26**(7), 1650025 (2016)
30. Ortiz, A., Munilla, J., Martínez-Murcia, F.J., Górriz, J.M., Ramírez, J.: Learning longitudinal MRI patterns by SICE and deep learning: assessing the Alzheimer's disease progression. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 413–424. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_36
31. 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**, 530–548 (2017)
32. Sarraf, S., Tofighi, G.: Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In: Future Technologies Conference, pp. 816–820. IEEE, San Francisco (2016)
33. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging **21**(11), 1421–1439 (2002)
34. Spillmann, L., Dresp-Langley, B., Tseng, C.: Beyond the classical receptive field: the effect of contextual stimuli. J. Vis. **15**(9), 7 (2015)
35. Suk, H., Lee, S., Shen, D.: Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. Brain Struct. Funct. **221**(15), 2569–2587 (2016)
36. Suk, H., Lee, S., Shen, D.: Deep ensemble learning of sparse regression models for brain disease diagnosis. Med. Image Anal. **37**, 101–113 (2017)
37. Szegedy, C., et al.: Going deeper with convolutions. In: CVPR, pp. 1–9. IEEE, Piscataway (2015)
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR, pp. 2818–2826. IEEE, Piscataway (2016)
39. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: AAAI, San Francisco (2017)
40. Wang, F., et al.: Residual attention network for image classification. In: CVPR, pp. 3156–3164. IEEE, Piscataway (2017)

41. Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: CVPR, pp. 1492–1500. IEEE, Piscataway (2017)
42. You, Q., Jin, H., Wang, Z., Fang, C., Luo, J.: Image captioning with semantic attention. In: CVPR, pp. 4651–4659. IEEE, USA (2016)
43. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1