

# Prediction of Alzheimer's Disease Based on Coordinate-Dense Attention Network

Yongmei TANG<sup>a,b</sup>, Xiangyun LIAO<sup>b,1</sup>, Weixin SI<sup>b</sup> and Zhigang NING<sup>a</sup>

<sup>a</sup>University of South China, Hengyang, China

<sup>b</sup>Shenzhen Institute of Advanced Technology Chinese Academy of Sciences, Shenzhen, China

**Abstract.** Alzheimer's disease (AD) is a degenerative disease of the nervous system. Mild cognitive impairment (MCI) is a condition between brain aging and dementia. The prediction will be divided into stable sMCI and progressive pMCI as a binary task. Structural magnetic resonance imaging (sMRI) can describe structural changes in the brain and provide a diagnostic method for the detection and early prevention of Alzheimer's disease. In this paper, an automatic disease prediction scheme based on MRI was designed. A dense convolutional network was used as the basic model. By adding a channel attention mechanism to the model, significant feature information in MRI images was extracted, and the unimportant features were ignored or suppressed. The proposed framework is compared with the most advanced methods, and better results are obtained.

**Keywords.** Alzheimer's disease, deep learning, convolutional neural network

## 1. Introduction

Alzheimer's disease is a common brain disorder characterized by the loss of connections between neurons in the brain and the shrinkage of the rest of the brain. Studies show that more than six million Americans may have Alzheimer's disease, second only to heart disease and cancer as the leading cause of death among the elderly [1]. Patients with mild cognitive impairment have memory disorder and are more likely to develop Alzheimer's disease. Therefore, MCI plays an important role in the early diagnosis and intervention of AD [2]. Accurate and efficient prediction of sMCI and pMCI can timely intervene and prevent the occurrence of Alzheimer's disease and reduce the potential patient population.

Due to the success of untreatable Alzheimer's disease, intervention and treatment in the early stages of Alzheimer's disease have become critical. Li *et al.* [3] use principal component analysis (PCA) to obtain features, and then use unsupervised deep learning training and finally use SVM classification. In addition to SVM, others also commonly use logistic regression (LR) [4] and random forest (RF) [5]. Many studies using MRI area (ROI) are interested in classification. Calvini *et al.* [6] around the medial temporal lobe and hippocampus were extracted to classify, most based on ROI measures are interested in artificial selection area, this inevitably can add human factors in the process of feature extraction. Some potential details may be ignored to affect the results.

---

<sup>1</sup> Corresponding Author, Xiangyun LIAO, Shenzhen Institute of Advanced Technology Chinese Academy of Sciences, Shenzhen, China; E-mail: xiangyun-l@163.com.

With the continuous development of deep learning, excellent results have been achieved in various fields, and it has also been applied to the diagnosis of Alzheimer's disease. Qiu *et al.* [7] connected Fully Convolutional Networks (FCN) and Muti-Layer Perception (MLP) to predict AD state by combining MRI data and non-imaging data, to achieve visualization of disease risk. Alejandro *et al.* [8] using transfer learning (TL) technology, ResNet feature extractor and SVM classifier were used to identify sagittal MRI images. In this paper, DenseNet was used as the basic model to predict AD disease by adding channel attention mechanism and extracting significant channel differences according to the subtle differences between sMCI and pMCI lesions, and found that adding attention mechanism has a good effect on the early prediction of AD.

## 2. Related Work

General Classification of Alzheimer's disease includes methods based on voxel morphological measurement and region of Interest measurement (ROI). Klöppel *et al.* [9] segmentation of gray matter images from MRI images, extraction of voxels, and use support vector machine to assign weight to voxels. Vemuri *et al.* [10] added demographic information, such as age and gender, to sMR scan information from the structural abnormality index (STAND) of tissues in sMR scan, and added genetic information, APOE genotype information. Finally, support vector machine SVM was used to optimize classification. However, the 3D image data processing method based on voxel is difficult to calculate due to its large amount of computation.

Regions of Interest (ROI) based classification methods generally rely on prior knowledge to divide regions, and the hippocampus, entorhinal cortex and medial temporal lobe are often used to classify Alzheimer's disease. Zhang *et al.* [11] combined three biomarkers, MRI, PET, and CSF, using an atlas based approach and automatic labeling by atlas distortion algorithm. Lauge *et al.* [12] using hippocampal structure as a classification feature, demonstrated that texture is more meaningful for early diagnosis. Feature extraction also focuses on the organizations and regions with high correlation with classification, but at the same time, some global information will be lost.

In recent years, deep learning technology has achieved good results in the classification of Alzheimer's disease. Kanghan *et al.* [13] based on the unsupervised learning of convolutional autoencoder (CAE) for the classification of AD and NC, proposed the end-to-end concept for classification. Lian *et al.* [14] proposed hierarchical full convolutional networks (H-FCN) to identify multi-scale discriminant locations, such as differentiated local plaques and regions in the brain. Compared with fixed interest area is extracted (ROI). Zhao *et al.* [15] used 3D multi-information generative adversarial network (mi-GAN) to generate predicted MRI images, and then used DenseNet to classify pMCI and sMCI.

## 3. Materials and Methods

### 3.1. Datasets and Preprocessing

This paper used data from a public data set of Alzheimer's Disease Neuroimaging (ADNI) (<http://adni.loni.usc.edu/>) download. In this study, follow-up data from ADNI were used to study the progression of patients with mild cognitive impairment (MCI)

over 6 months to 2 year period. Patients who developed and developed Alzheimer's disease during this period were considered progressive mild cognitive impairment (pMCI), and those who did not deteriorate were considered stable mild cognitive impairment (sMCI). T1-weighted MRI rapid gradient echo (MPRAGE) images of 969 subjects were downloaded from ADNI, including 210 AD, 274 NC, 183 pMCI and 302 sMCI. The demographic information of all subjects is shown in Table 1.

**Table 1.** Demographic Information of the Subjects Included in the Studied Datasets

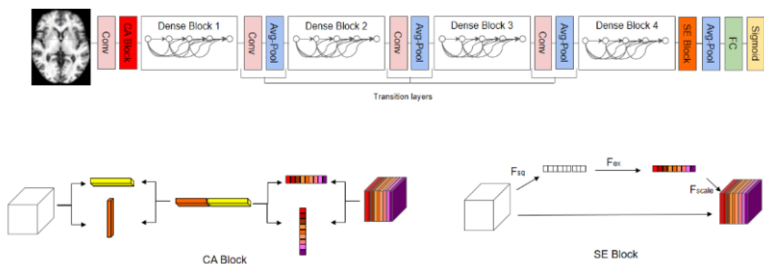
Group	Gender(male/ Female)	Age
AD	210(110/100)	75.68+7.67
pMCI	183(105/78)	75.07+7.67
sMCI	302(168/134)	72.90+7.12
NC	274(129/145)	74.98+6.31

Downloaded from ADNI open source data set of the original MRI image preprocessing, use spm8 (<https://www.fil.ion.ucl.ac.uk/spm/>) software kit pictures into the DICOM format NIFTI format, then the skull stripping, cerebellum resection, and the MRI images after skull removal were registered into the MNI152 standard space by affine matching to eliminate the global linear differences. Finally, all images were resampled and cropped to 121×145×121.

3.2. Methods

3.2.1. Dense Attention Network

This section will introduce a 3D dense channel attention network. The structure of the network is shown in Figure 1, which is mainly composed of a 3D dense Convolutional Network (DenseNet) [16], Coordinate Attention (CA) [17] and Squeeze-and-Excitation Attention (SE) [18] network. The problem of gradient explosion occurs when the network is too deep. The dense Connection solves this problem well. A densely connected convolution network has strong regularization effect and can reduce the overfitting on small training sets. From the perspective of feature channels, SE automatically obtains the importance of each channel through learning, which enhances the model's ability to extract important information. CA aggregates features from the two spatial directions and embed location information into channel attention, which can obtain not only channel information, but also direction and position information.



**Figure 1.** The architecture of dense attention connected convolution.

### 3.2.2. Dense Convolutional Network

The dense convolution network [16] does not directly add and fuse feature maps, but connects them on feature dimensions. Any input layer is connected with the subsequent layer, and the connected feature map is used as the input of the next layer:

$$X_L = H_L([X_0, X_1, \dots, X_{L-1}]) \quad (1)$$

In Eq. (1),  $L$  represents the number of network layers, and  $H_L$  represents a group of convolution operations at the  $L$  layer. Each layer is composed of batch normalization (BN), correction linear unit (RELU), pooling and convolution (CONV).  $[X_0, X_1, \dots, X_{L-1}]$  represents the feature graph connection of the previous  $L-1$  layer, and  $X_L$  represents the feature graph obtained after the convolution operation of the  $L$  layer. The Dense connection structure in a densely connected convolution network needs to keep the size of the feature image constant, so the Dense Block and Transition Layer are stacked. After the last dense block, the global average pooling layer is used to reduce the number of parameters. Each group of dense blocks is composed of the batch normalization layer (BN) and the convolution kernel of the RELU activation function is  $1 \times 1 \times 1$  and  $3 \times 3 \times 3$ . The number of feature graphs output in each group is  $R = 32$ , which is used to control the number of network parameters. The transformation layer consists of a  $1 \times 1 \times 1$  convolution layer and a  $2 \times 2 \times 2$  pooling layer to reduce the resolution of feature maps.

### 3.2.3. Coordinate Attention

CA [17] proposed a novel attention network by embedding location information into channel attention, which decomposed attention into two features in different directions by global pooling, and then aggregated features along with these two directions respectively. Given the input feature  $x$  and compress them into  $h \times 1 \times d$  and  $1 \times w \times d$  in horizontal and vertical directions respectively, as shown in Eq. (2), Eq. (3),  $h$ ,  $w$ ,  $d$  represent the length, width and number of slices of the picture respectively. The horizontal and vertical pooling results were connected together, and the feature images were obtained through a  $1 \times 1 \times 1$  convolution. The feature images were then divided into two groups of feature images in horizontal and vertical directions. The weight data were obtained by  $1 \times 1 \times 1$  convolution and sigmoid respectively, and finally the obtained horizontal and vertical weights were multiplied to obtain the attention output feature.

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i, d) \quad (2)$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w, d) \quad (3)$$

### 3.2.4. Squeeze and Excitation Attention

SE [18] is an attention model based on channel dimensions.  $F_{sq}$ ,  $F_{ex}$  and  $F_{scale}$  are the three key operations of the compression activation module. Firstly, the input feature graph is compressed to  $1 \times 1 \times 1$  by  $F_{sq}$  operation to keep the number of feature images unchanged. Given the input feature  $x$ , the compression formula for channel  $C$  is Eq. (4). By activating  $F_{ex}$  operation, weight from 0 to 1 is generated for each feature graph. Redefine the relationship of each channel in the original feature graph, and  $F_{scale}$  adds weight to the picture. Channel attention networks often use global pooling to encode spatial

information and compress space into channel descriptors, ignoring location information. In this paper, SE is placed behind the last Dense Block.

$$Z_c = \frac{1}{H \times W \times D} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^D x_c(i, j, k) \quad (4)$$

## 4. Results and Discussion

### 4.1. Evaluation Metrics

We designed the classification of AD and CN and the predictive of sMCI and pMCI in two groups of controlled experiments. The performance of the classification is evaluated against six commonly used indicators. It includes accuracy, sensitivity, Specificity, Precision, Recall and F1 Score. These indicators are defined as Eq. (5) through Eq. (10):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$\text{Sensitive} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

TP indicates that the category is positive and the prediction is also positive, FP means the prediction is positive and the prediction is wrong, so it's a negative example. FN indicates that the prediction is negative and the prediction is wrong, so it is a positive example, TN indicates that the category is negative and the prediction is also negative.

### 4.2. Experimental Evaluation

In this part, we compared our experimental results with the most advanced convolutional neural networks, such as ResNet18 [19], DenseNet201, DenseNet264, We added attention to these networks to study the comparison between attentional mechanisms and those without. All experiments were implemented based on PyTorch with python3.8 programming language, executed on a server with an Intel Xeon Cascade Lake 8255C (2.5 GHz), Tesla V100-NVLINK-32G and Ubuntu18.04 operating system. The learning rate was attenuated from 0.0001 to 0.000001. In order to optimize the model, Adam was used as the optimization algorithm and the batch size was 8.

In this experiment, two groups of classification experiments will be carried out. Since the brain of AD and NC is quite different, it is easy to distinguish the two types. MCI is the early stage of AD, and there are only slight brain lesions. There is no significant difference in the changes of the lesion areas in the brain. The purpose of adding an attention mechanism is to make the model focus on the significant areas in the training process, excluding the influence of other organizational structures. sMCI and pMCI classification is a difficult task, and their classification accuracy is far lower than

that of AD and NC. The experimental results of these two groups of classification on different convolution models are shown in Table 2 and Table 3.

Table 2. Results of Classification for AD vs. NC.

Model	Accuracy	Sensitivity	Specificity	Precision	Recall	F1 Score
ResNet18	0.869	0.825	0.903	0.866	0.845	0.825
CSE_ResNet18	0.883	0.825	0.927	0.896	0.859	0.825
DenseNet201	0.876	0.730	0.987	0.978	0.836	0.730
CSE_DenseNet201	0.890	0.809	0.951	0.927	0.864	0.809
DenseNet264	0.897	0.825	0.951	0.928	0.873	0.825
CSE_DenseNet264	0.904	0.920	0.891	0.965	0.892	0.920

Table 3. Results of Classification for sMCI vs. pMCI.

Model	Accuracy	Sensitivity	Specificity	Precision	Recall	F1 Score
ResNet18	0.770	0.740	0.788	0.677	0.707	0.740
CSE_ResNet18	0.791	0.722	0.833	0.722	0.722	0.722
DenseNet201	0.791	0.685	0.855	0.740	0.711	0.685
CSE_DenseNet201	0.812	0.611	0.933	0.846	0.709	0.611
DenseNet264	0.819	0.740	0.866	0.769	0.740	0.754
CSE_DenseNet264	0.826	0.796	0.844	0.754	0.774	0.796

As can be seen from the table, the classification accuracy of sMCI and pMCI in our experiment is 0.826, the sensitivity is 0.796, and the specificity is 0.844, which is superior to other models in overall classification. By adding attention mechanism to other models, the accuracy of sMCI and pMCI was improved, indicating that attention mechanism has a good effect on the prediction of MCI to AD transformation.

At the same time, we compare our method with the most advanced method in the present paper [20], [21], [14], [22] as shown in Table 4 for MCI to AD conversion of prediction of our method is better than the existing methods, and different papers with convolution neural network contrast experiment, it shows our network for locating sMCI and pMCI lesion area has obvious effect.

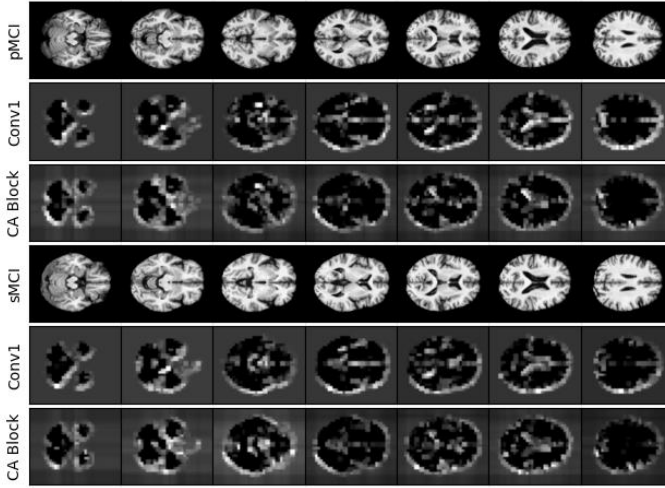
Table 4. Compare with the latest method on ADNI dataset.

References	AD vs NC			sMCI vs pMCI		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Cao et al. 2017	0.89	0.86	0.90	0.70	0.67	0.72
Liu et al. 2018	0.91	0.88	0.93	0.76	0.42	0.82
Lian et al. 2020	0.90	0.82	0.96	0.81	0.53	0.85
Zhu et al. 2021	0.92	0.91	0.93	0.80	0.77	0.82
Ours	0.90	0.92	0.89	0.83	0.80	0.84

4.3. Discussion

In order to understand how the network improves the classification effect, we conducted two groups of comparison experiments with and without attention in Densenet264. Figure 2 is the visual image of the experiment's features. pMCI and sMCI are the input of our model, and Conv1 represents the output of Conv1 without CA. CA Block represents the output of CA Block after adding the CA model. It was obvious from the images that compared with Conv1 images, CA Block images had more prominent

features in the hippocampus and cerebral cortex, while other useless features were ignored, which happened to be important diagnostic regions of Alzheimer's disease, proving that our model was superior to other models in feature extraction.



**Figure 2.** Visualization results of Densenet264 and CSE\_Densenet264 feature maps.

## 5. Conclusion

In this paper, by adding attention mechanism to the model, significant information of MRI images can be extracted and combined with local and global information, so as to improve the accuracy of network classification. This method combines CA and SE attention and uses a dense attention network to automatically identify MRI lesions, for the early diagnosis and intervention treatment of AD, we put forward the method compared with the most advanced several methods has a better classification performance, especially in predicting the difficult MCI to AD transition task. MRI images can provide relatively simple information. In the future, we may add multimodal information, such as fluorodeoxyglucose positron emission tomography (FDG-PET), apolipoprotein-E (APOE) genotype, and age information as auxiliary information to further predict MCI to AD transformation.

## Acknowledgment

This work is supported by multiple grants, including: The National Key Research and Development Program of China (2020YFB1313900), National Natural Science Foundation of China (61902386), Shenzhen Science and Technology Program (JCYJ20180507182415428).

## References

- [1] Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement*. 2007 Jul;3(3):186-91.
- [2] Ben Ahmed O, Mizotin M, Benois-Pineau J, Allard M, Catheline G, Ben Amar C. Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. *Comput Med Imaging Graph*. 2015 Sep; 44: 13-25.
- [3] Li F, Tran L, Thung KH, Ji S, Shen D, Li J. A robust deep model for improved classification of AD/MCI patients. *IEEE J Biomed Health Inform*. 2015 Sep;19(5):1610-1616.
- [4] Liu Z, Maiti T, Bender AR. for the Alzheimer's Disease Neuroimaging Initiative. A role for prior knowledge in statistical classification of the transition from mild cognitive impairment to alzheimer's disease. *J Alzheimers Dis*. 2021 Aug 25.
- [5] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: a systematic review. *Front Aging Neurosci*. 2017 Oct 6; 9: 329.
- [6] Calvini P, Chincarini A, Gemme G, Penco MA, Squarcia S, Nobili F, Rodriguez G, Bellotti R, Catanzariti E, Cerello P, De Mitri I, Fantacci ME; MAGIC-5 Collaboration; Alzheimer's Disease Neuroimaging Initiative. Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer disease. *Med Phys*. 2009 Aug;36(8):3737-47.
- [7] Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, Chang GH, Joshi AS, Dwyer B, Zhu S, Kaku M, Zhou Y, Alderazi YJ, Swaminathan A, Kedar S, Saint-Hilaire MH, Auerbach SH, Yuan J, Sartor EA, Au R, Kolachalama VB. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*. 2020 Jun 1;143(6):1920-1933.
- [8] Puente-Castro A, Fernandez-Blanco E, Pazos A, Munteanu CR. Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. *Comput Biol Med*. 2020 May; 120:103764.
- [9] Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RS. Automatic classification of MR scans in Alzheimer's disease. *Brain*. 2008 Mar; 131(Pt 3):681-9.
- [10] Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage*. 2008 Feb 1;39(3):1186-97.
- [11] Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Alzheimer's disease neuroimaging initiative. multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011 Apr 1; 55(3): 856-67.
- [12] Sørensen L, Igel C, Liv Hansen N, Osler M, Lauritzen M, Rostrup E, Nielsen M. Early detection of Alzheimer's disease using MRI hippocampal texture. *Hum Brain Mapp*. 2016 Mar;37(3):1148-61.
- [13] Oh K, Chung YC, Kim KW, Kim WS, Oh IS. Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. *Sci Rep*. 2019 Dec 3;9(1):18150.
- [14] 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*. 2020 Apr; 42(4):880-893.
- [15] Zhao Y, Ma B, Jiang P, Zeng D, Wang X, Li S. Prediction of alzheimer's disease progression with multi-information generative adversarial network. *IEEE J Biomed Health Inform*. 2021 Mar;25(3):711-719.
- [16] Huang G, Liu Z, Laurens VDM, Weinberger KQ. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 2261-2269.
- [17] Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. *CVPR*. 2021.
- [18] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*. 2020 Aug; 42(8):2011-2023.
- [19] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. pp. 770-778.
- [20] Cao P, Liu X, Yang J, Zhao D, Huang M, Zhang J, Zaiane O. Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures. *Comput Biol Med*. 2017 Dec 1; 91: 21-37.
- [21] Liu M, Zhang J, Adeli E, Shen D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Med Image Anal*. 2018 Jan; 43:157-168.
- [22] Zhu W, Sun L, Huang J, Han L, Zhang D. Dual attention multi-instance deep learning for alzheimer's disease diagnosis with structural MRI. *IEEE Trans Med Imaging*. 2021 Sep;40(9):2354-2366.