# Deep Grading based on Collective Artificial Intelligence for AD Diagnosis and Prognosis

Huy-Dung Nguyen, Michaël Clément, Boris Mansencal, Pierrick Coupé

▶ **To cite this version:**

Huy-Dung Nguyen, Michaël Clément, Boris Mansencal, Pierrick Coupé. Deep Grading based on Collective Artificial Intelligence for AD Diagnosis and Prognosis. Workshop on Interpretability of Machine Intelligence in Medical Image Computing at MICCAI 2021, 2021, Strasbourg, France. 10.1007/978-3-030-87444-5_3 . hal-03370898

**HAL Id: hal-03370898**
**https://hal.science/hal-03370898**

Submitted on 8 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Deep Grading based on Collective Artificial Intelligence for AD Diagnosis and Prognosis

Huy-Dung Nguyen[1], Michaël Clément[1], Boris Mansencal[1], and Pierrick Coupé[1]

[1]Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, 33400 Talence, France

**Abstract.** Accurate diagnosis and prognosis of Alzheimer's disease are crucial to develop new therapies and reduce the associated costs. Recently, with the advances of convolutional neural networks, methods have been proposed to automate these two tasks using structural MRI. However, these methods often suffer from lack of interpretability, generalization, and can be limited in terms of performance. In this paper, we propose a novel deep framework designed to overcome these limitations. Our framework consists of two stages. In the first stage, we propose a deep grading model to extract meaningful features. To enhance the robustness of these features against domain shift, we introduce an innovative collective artificial intelligence strategy for training and evaluating steps. In the second stage, we use a graph convolutional neural network to better capture AD signatures. Our experiments based on 2074 subjects show the competitive performance of our deep framework compared to state-of-the-art methods on different datasets for both AD diagnosis and prognosis.

**Keywords:** Deep Grading · Collective Artificial Intelligence · Generalization · Alzheimer's disease classification · Mild Cognitive Impairment

## 1 Introduction

The first cognitive symptoms of Alzheimer's disease (AD) appear right after the morphological changes caused by brain atrophy [10]. Those changes can be identified with the help of structural magnetic resonance imaging (sMRI) [2]. Recently, with the advances of convolutional neural networks (CNN), methods have been proposed for automatic AD diagnosis using sMRI. Despite encouraging results, current deep learning methods suffer from several limitations. First, deep models lack transparency in their decision-making process [31, 38]. Therefore, this limits their use for computer-aided diagnosis tools in clinical practice. Second, for medical applications, the generalization capacity of classification models is essential. However, only a few works have proposed methods robust to domain shift [13, 34]. Third, current CNN models proposed for AD diagnosis and prognosis still perform poorly [35]. Indeed, when properly validated on external datasets, current CNN-based methods perform worse than traditional approaches (*i.e.*, standard linear SVM).

In this paper, to address these three major limitations, we propose a novel interpretable, generalizable and accurate deep framework. An overview of our

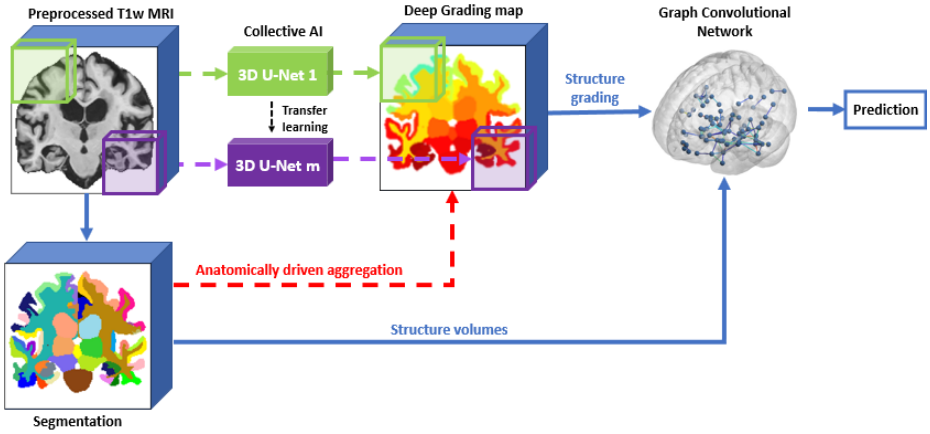


**Fig. 1.** Overview of our processing pipeline. The MRI image, its segmentation and the deep grading map illustrated are from an AD subject.

proposed pipeline is shown in Figure 1. First, we propose a novel Deep Grading (DG) biomarker to improve the interpretability of deep model outputs. Inspired by the patch-based grading frameworks [4,12,32], this new biomarker provides a grading map with a score between $-1$ and 1 at each voxel related to the alteration severity. This interpretable biomarker may help clinicians in their decision and to improve our knowledge on AD progression over the brain. Second, we propose an innovative collective artificial intelligence strategy to improve the generalization across domains and to unseen tasks. As recently shown for segmentation [6,18], the use of a large number of networks capable of communicating offers a better capacity for generalization. Based on a large number of CNNs (*i.e.*, 125 U-Nets), we propose a framework using collective artificial intelligence efficient on different datasets and able to provide accurate prognosis while trained for diagnosis task. Finally, we propose to use a graph-based modeling to better capture AD signature using both inter-subject similarity and intrasubject variability. As shown in [12], such strategy improves performance in AD diagnosis and prognosis.

In this paper, our main contributions are threefold:

- A novel deep grading biomarker providing interpretable grading maps.
- An innovative collective artificial intelligence strategy robust to unseen datasets and unknown tasks.
- A new graph convolutional network (GCN) model for classification offering state-of-the-art performance for both AD diagnosis and prognosis.

## 2 Materials and method

### 2.1 Datasets

The data used in this study, consisting of 2074 subjects, were obtained from multiple cohorts: the Alzheimer's Disease Neuroimaging Initiative (ADNI) [16], the

**Table 1.** Number of participants used in our study. Data used for training is in bold.

| Dataset | CN | AD | sMCI | pMCI |
|---|---|---|---|---|
| **ADNI1** | **170** | **170** | 129 | 171 |
| ADNI2 | 149 | 149 | - | - |
| AIBL | 233 | 47 | 12 | 20 |
| OASIS3 | 658 | 97 | - | - |
| MIRIAD | 23 | 46 | - | - |

Open Access Series of Imaging Studies (OASIS) [21], the Australian Imaging, Biomarkers and Lifestyle (AIBL) [7], the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) [27]. We used the baseline T1-weighted MRI available in each of these studies. Each dataset contains AD patients and cognitively normal (CN) subjects. For ADNI1 and AIBL, it also includes mild cognitive impairment (MCI), the early stage of AD composed of abnormal memory dysfunctions. Two groups of MCI are considered: progressive MCI (pMCI) and stable MCI (sMCI). The definition of these two groups is the same as in [35]. Table 1 summarizes the number of participants for each dataset used in this study. During experiments, AD and CN from ADNI1 are used as training set and the other subjects as testing set.

### 2.2 Preprocessing

All the T1w MRI are preprocessed using the following steps: (1) denoising [29], (2) inhomogeneity correction [33], (3) affine registration into MNI space ($181 \times 217 \times 181$ voxels at $1\text{mm} \times 1\text{mm} \times 1\text{mm}$) [1], (4) intensity standardization [28] and (5) intracranial cavity (ICC) extraction [30]. After that preprocessing, we use AssemblyNet [6] to segment 133 brain structures (see Figure 1). The list of structures is the same as in [14]. In this study, brain structure segmentation is used to determine the structure volume (*i.e.*, normalized volume in % of ICC) and aggregate information in the grading map (see Section 2.3 and Figure 1).

### 2.3 Deep Grading for disease visualization

In AD classification, most of deep learning models only use CNN as binary classification tool. In this study, we propose to use CNN to produce 3D maps indicating where specific anatomical patterns are present and the importance of structural changes caused by AD.

To capture these anatomical alterations, we extend the idea of the patch-based grading (PBG) framework [4, 12, 32]. The PBG framework provides a 3D grading map with a score between $-1$ and 1 at each voxel related to the alteration severity. Contrary to previous PBG methods based on non-local mean strategy, here we propose a novel DG framework based on 3D U-Nets.

Concretely, each U-Net (similar to [6]) takes a 3D sMRI patch (*e.g.*, $32 \times 48 \times 32$) and outputs a grading map with values in range $[-1, 1]$ for each voxel. Voxels with a higher value are considered closer to AD, while voxels with a lower

value are considered closer to CN. For the ground-truth used during training, we assign the value 1 (resp. $-1$) to all voxels inside a patch extracted from an AD patient (resp. CN subject). All voxels outside of ICC are set to 0.

Once trained, the deep models are used to grade patches. These local outputs are gathered to reconstruct the final grading map (see Section 2.4). Using the structure segmentation, we represent each brain structure grading by its average grading score (see Figure 1). This anatomically driven aggregation allows better and meaningful visualization of the disease progression. In this way, during the classification step (see Section 2.5), each subject is encoded by an n-dimensional vector where n is the number of brain structures.

### 2.4 Collective AI for grading

As recently shown in [3, 35], current AD classification techniques suffer from a lack of generalization. In this work, we propose an innovative collective artificial intelligence strategy to improve the generalization across domains and to unseen tasks. As recently shown for segmentation [6, 18], the use of a large number of compact networks capable of communicating offers a better capacity for generalization. There are many advantages to using the collective AI strategy. First, it addresses the problem of GPU memory in 3D since each model processes only a sub-volume of the image. The use of a large number of compact networks is equivalent to a big neural network with more filters. Second, the voting system based on a large number of specialized and diversified models helps the final grading decision to be more robust against domain shift and different tasks.

Concretely, a preprocessed sMRI is decomposed into $k \times k \times k$ overlapping patches of the same size (*e.g.*, $32 \times 48 \times 32$). During training, for each patch localization in the MNI space, a specialized model is trained. Therefore, in our case ($k = 5$), we trained $m = k \times k \times k = 125$ U-Nets to cover the whole image (see Fig. 1). Moreover, each U-Net is initialized using transfer learning from its nearest neighbor U-Nets in the MNI space, except the first one trained from scratch as proposed in [6]. As adjacent patches have some common patterns, this communication allows grading models to share useful knowledge between them. For each patch, 80% of the training dataset (*i.e.*, ADNI1) is used for training and the remaining 20% for validation. The accuracy obtained on validation set is used to reconstruct the final grading map using a weighted average as follows:

$$G_i = \frac{\sum_{x_i \in P_j} \alpha_j * g_{ij}}{\sum_{x_i \in P_j} \alpha_j} \tag{1}$$

where $G_i$ is the grading score of the voxel $x_i$ in the final grading map, $g_{ij}$ is the grading score of the voxel $x_i$ in the local grading patch $P_j$, and $\alpha_j$ is the validation accuracy of the patch $j$. This weighted vote enables to give more weight to the decision of accurate models during the reconstruction.

### 2.5 Graph convolutional neural network for classification

The DG feature provides an inter-subject similarity biomarker which is helpful to detect AD signature. However, the structural alterations leading to cognitive decline could be different between subjects. Indeed, following the idea of [12], we model the intra-subject variabilities by a graph representation to capture the relationships between several regions related to the disease. We define an undirected graph $G = (N, E)$, where $N = \{n_1, \ldots, n_s\}$ is the set of nodes for the $s$ brain structures and $E = s \times s$ is the matrix of edge connections. In our approach, all nodes are connected with each other in a complete graph, where nodes embed brain features (*e.g.*, our proposed DG feature) and potentially other types of external features.

Indeed, besides the grading map, the volume of structures obtained from the segmentation could be helpful to distinguish AD patients from CN [12, 32]. It is due to the evidence that AD leads to structure atrophy. Age is also an important factor as, within sMRI, patterns in the brain of young AD patients could be similar to elder CN. Indeed, the combination of those features is expected to improve our classification performance. In our method, each node represents a brain structure and embeds a feature vector (DG, V, A) where V and A are respectively the volume of structures and subject's age. Finally, we use the graph convolutional neural network (GCN) [20] as the way to pass messages between nodes and to perform final classification.

### 2.6 Implementation details

First, we downsample the sMRI from $181 \times 217 \times 181$ voxels (at 1mm) to $91 \times 109 \times 91$ voxels to reduce the computational cost, then decompose them into $5 \times 5 \times 5$ overlapping patches of size $32 \times 48 \times 32$ voxels equally spaced along the three axis. For each patch, an U-Net is trained using mean absolute error loss, Adam optimizer with a learning rate of 0.001. The training process is stopped after 20 epochs without improvement in validation loss. We employed several data augmentation and sampling strategies to alleviate the overfitting issue during training. A small perturbation is first created in training samples by randomly translating by $t \in \{-1, 0, 1\}$ voxel in 3 dimensions of the image. We then apply the mixup [37] data augmentation scheme that was shown to improve the generalization capacity of CNN in image classification.

Once the DG feature is obtained, we represent each subject by a graph of 133 nodes. Each node represents a brain structure and embeds DG, volume and age features. Our classifier is composed of 3 layers of GCN with 32 channels, followed by a global mean average pooling layer and a fully connected layer with an output size of 1. The model is trained using the binary cross-entropy loss, Adam optimizer with a learning rate of 0.0003. The training process is stopped after 20 epochs without improvement in validation loss. At inference time, we randomly add noise $X \sim \mathcal{N}(0, 0.01)$ to the node features and compute the average of 3 predictions to get the global decision. Experiments have shown that it helps our GCN to be more stable.

For training and evaluating steps, we use a standard GPU (*i.e.*, NVIDIA TITAN X) with 12Gb of memory.

## 3 Experimental results

In this study, the grading models and classifiers are trained using ADNI1 dataset within AD and CN subjects. Then, we assess their generalization capacity in domain shift using AD, CN subjects from ADNI2, AIBL, OASIS, MIRIAD. The generalization capacity in derived tasks is performed using pMCI, sMCI subjects from ADNI1 (same domain) and AIBL (out of domain).

**Influence of collective AI strategy.** In this part, the DG feature is denoted as $DG_C$(resp. $DG_I$) when obtained with the collective (resp. individual) AI strategy. The individual AI strategy refers to the use of a single U-Net to learn patterns from all patches of sMRI. We compare the efficiency of $DG_C$ and $DG_I$ feature when using the same classifier (*i.e.*, SVM or GCN) (see Table 2). These experiments show that using $DG_C$ achieves better results in most configurations. When using SVM classifier, we observe a gain of 3.6% (resp. 0.8%) on average in AD/CN (resp. pMCI/sMCI) classification. The efficiency of $G_C$ feature is even better with GCN classifier, where a gain of 4.0% (resp. 3.5%) is observed.

**Influence of GCN classifier.** Besides the DG feature, the intra-subject variabilities are also integrated into our graph representation. Hence, it should be beneficial to use GCN to exploit all this information. In our experiments, GCN outperforms SVM in all the tests using either $DG_I$ or $DG_C$ feature (see Table 2). Concretely, using $DG_I$ feature, we observe a gain of 5.0% (resp. 7.6%) on average for AD/CN (resp. pMCI/sMCI) classification. These improvements are 5.4% and 10.6% when using $DG_C$ feature.

**Influence of using additional non-image features.** Moreover, we analyze the model performance using $DG_C$ with the structural volume $V$ and age $A$ as additional node features in our graph representation. By using the combined features, the performance on average in AD/CN and pMCI/sMCI is both improved by 0.3% and 1.4% compared to $DG_C$ feature (see Table 2). In the rest of this paper, these results are used to compare with current methods.

**Comparison with state-of-the-art methods.** Table 3 summarizes the current performance of state-of-the-art methods proposed for AD diagnosis and prognosis classification that have been validated on external datasets. In this comparison we considered five categories of deep methods: patch-based strategy based on a single model (Patch-based CNN [35]), patch-based strategy based on multiple models (Landmark-based CNN [24], Hierarchical FCN [23]), ROI-based strategy based on a single model focused on hippocampus (ROI-based CNN [35]), subject-based considering the whole image based on a single model (subject-based CNN [35], 3D Inception-ResNet-v2 [26], Efficient 3D [36] and AD$^2$A [11]) and a classical voxel-based model using a SVM (Voxel-based SVM [35]).

For AD diagnosis (*i.e.*, AD/CN), all the methods show good balanced accuracy, although some of them failed to generalize on OASIS. In this scenario

**Table 2.** Validation of the collective AI strategy, GCN classifier, the combination of DG feature with other image and non-image features using GCN classifier. Red: best result, Blue: second best result. The balanced accuracy (BACC) is used to assess the model performance. The results are the average accuracy of 10 repetitions and presented in percentage. All the methods are trained on the AD/CN subjects of the ADNI1 dataset.

| Classifier | Features | AD/CN | | | | pMCI/sMCI | | Average | |
|---|---|---|---|---|---|---|---|---|---|
| | | ADNI2 | OASIS | MIRIAD | AIBL | ADNI1 | AIBL | AD/CN | p/sMCI |
| SVM | $DG_I$ | 83 | 83 | 88 | 79 | 65 | 66 | 83.3 | 65.5 |
| SVM | $DG_C$ | 83 | 84 | 91 | 87 | 68 | 64 | 86.3 | 66.0 |
| GCN | $DG_I$ | 84 | 88 | 96 | 82 | 68 | 73 | 87.5 | 70.5 |
| GCN | $DG_C$ | 87 | 89 | 100 | 88 | 70 | 76 | 91.0 | 73.0 |
| GCN | $DG_C, V, A$ | 87 | 88 | 98 | 92 | 74 | 74 | 91.3 | 74.0 |

**Table 3.** Comparison of our method with current methods in AD diagnosis and prognosis. Red: best result, Blue: second best result. The balanced accuracy (BACC) is used to assess the model performance. All the methods are trained on the AD/CN subject of the ADNI1 dataset (except [23] that is fined-tuned on MCI subjects for sMCI/pMCI task).

| Methods | AD/CN | | | | pMCI/sMCI | |
|---|---|---|---|---|---|---|
| | ADNI2 | OASIS | MIRIAD | AIBL | ADNI1 | AIBL |
| Landmark-based CNN [24] | 91 | - | 92 | - | - | - |
| Hierachical FCN [23] | 89 | - | - | - | 69 | - |
| Patch-based CNN [35] | - | 64 | - | 81 | 70 | 64 |
| ROI-based CNN [35] | - | 69 | - | 84 | 70 | 60 |
| Subject-based CNN [35] | - | 67 | - | 83 | 69 | 52 |
| Voxel-based SVM [35] | - | 70 | - | 88 | 75 | 62 |
| $AD^2A$ [11] | 88 | - | - | 88 | - | - |
| Efficient 3D [36] | - | 92 | 96 | 91 | 70 | 65 |
| 3D Inception-ResNet-v2 [26] | - | 85 | - | 91 | 42 | - |
| Our method | 87 | 88 | 98 | 92 | 74 | 74 |

(unseen datasets), our method obtained high accuracy for all the datasets. This confirms the generalization capacity of our approach against domain shift.

For AD prognosis (*i.e.*, pMCI/sMCI), we observe a significant drop for all the methods. This drop is expected since pMCI/sMCI classification is more challenging and since models are trained on a different task (*i.e.*, AD/CN). For this task, our method is generally robust, especially on AIBL. Moreover, our approach is the only deep learning method that performs competitively with the SVM model [35] on ADNI1, while significantly better on AIBL. In this scenario (unknown task), our method obtains the highest accuracy on average. These results highlight the potential performance of our method on unseen tasks.

**Interpretation of collective deep grading.** To highlight the interpretability capabilities offered by our DG feature, we first compute the average DG map for each group: AD, pMCI, sMCI and CN (see Figure 2). First, we can note that the average grading maps increase between each stage of the disease. Second, we estimated the top 10 structures with highest absolute value of grading score over

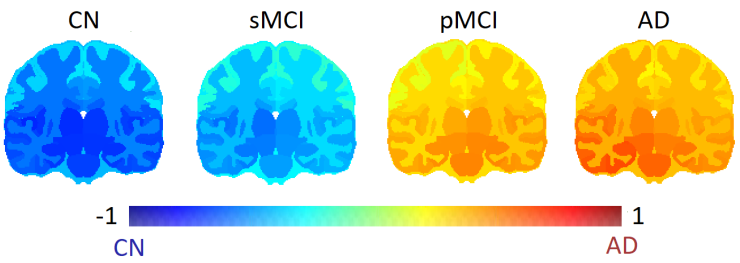


**Fig. 2.** Average grading map per group of subjects.

all the testing subjects. The found structures are known to be specifically and early impacted by AD. These structures are: *bilateral hippocampus* [9], *left amygdala* and *left inferior lateral ventricle* [5], *left parahippocampal gyrus* [19], *left posterior insula* [8], *left thalamus proper* [17], *left transverse temporal gyrus* [25], *left ventral diencephalon* [22]. While other attention-based deep methods failed to find structures related to AD [3], our DG framework shows high correlation with current physiopathological knowledge on AD [15].

## 4 Conclusion

In this paper, we addressed three major limitations of CNN-based methods by introducing a novel interpretable, generalizable and accurate deep grading framework. First, deep grading offers a meaningful visualization of the disease progression. Second, we proposed a collective artificial intelligence strategy to improve the generalization of our DG strategy. Experimental results showed a gain for both SVM and GCN in all tasks using this strategy. Finally, we proposed to use a graph-based modeling to better capture AD signature using both inter-subject similarity and intra-subject variability. Based on that, our DG method showed state-of-the-art performance in both AD diagnosis and prognosis.

**Acknowledgments** This work benefited from the support of the project DeepvolBrain of the French National Research Agency (ANR-18-CE45-0013). This study was achieved within the context of the Laboratory of Excellence TRAIL ANR-10-LABX-57 for the BigDataBrain project. Moreover, we thank the Investments for the future Program IdEx Bordeaux (ANR-10-IDEX-03-02), the French Ministry of Education and Research, and the CNRS for DeepMultiBrain project.

## References


1. Avants, B.B., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011)
2. Bron, E., et al.: Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. NeuroImage **111**, 562–579 (2015)
3. Bron, E., et al.: Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. NeuroImage: Clinical **31**, 102712 (2021)
4. Coupé, P., et al.: Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. NeuroImage: Clinical **1**(1), 141–152 (2012)
5. Coupé, P., et al.: Lifespan changes of the human brain in Alzheimer's disease. Nature Scientific Reports **9**(3998) (2019)
6. Coupé, P., et al.: AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation. NeuroImage **219**, 117026 (2020)
7. Ellis, K., et al.: The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. International psychogeriatrics / IPA **21**, 672–87 (2009)
8. Foundas, A., et al.: Atrophy of the hippocampus, parietal cortex, and insula in alzheimer's disease: a volumetric magnetic resonance imaging study. Neuropsychiatry, neuropsychology, and behavioral neurology **10**(2), 81—89 (1997)
9. Frisoni, G., et al.: The clinical use of structural MRI in Alzheimer's disease. Nature reviews. Neurology **6**, 67–77 (2010)
10. Gordon, B., et al.: Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: A longitudinal study. The Lancet Neurology **17** (2018)
11. Guan, H., et al.: Attention-guided deep domain adaptation for brain dementia identification with multi-site neuroimaging data. In: MICCAI Workshop on Domain Adaptation and Representation Transfer (2020)
12. Hett, K., et al.: Graph of brain structures grading for early detection of Alzheimer's disease. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2018)
13. Hosseini-Asl, E., et al.: Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: IEEE International Conference on Image Processing (ICIP) (2016)
14. Huo, Y., et al.: 3D whole brain segmentation using spatially localized atlas network tiles. NeuroImage **194**, 105–119 (2019)
15. Jack, C., et al.: Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. The Lancet Neurology **27**, 685–691 (2010)
16. Jack Jr., C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging **27**(4), 685–691 (2008)
17. de Jong, L.W., et al.: Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: An MRI study. Brain **131**(12), 3277–3285 (2008)
18. Kamraoui, R.A., et al.: Towards broader generalization of deep learning methods for multiple sclerosis lesion segmentation. arXiv 2012.07950 (2020)
19. Kesslak, J.P., et al.: Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in alzheimer's disease. Neurology **41**(1), 51–51 (1991)

20. Kipf, T., et al.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
21. LaMontagne, P.J., et al.: OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. medRxiv (2019)
22. Lebedeva, A.K., et al.: MRI-based classification models in prediction of mild cognitive impairment and dementia in late-Life depression. Frontiers in Aging Neuroscience **9**, 13 (2017)
23. Lian, C., et al.: Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(4), 880–893 (2020)
24. Liu, M., et al.: Landmark-based Deep Multi-Instance Learning for Brain Disease Diagnosis. Medical Image Analysis **43** (2017)
25. Liu, Y., et al.: Education increases reserve against Alzheimer's disease-Evidence from structural MRI analysis. Neuroradiology **54**, 929–38 (2012)
26. Lu, B., et al.: A practical Alzheimer disease classifier via brain imaging-based deep learning on 85,721 samples. bioRxiv (2021)
27. Malone, I.B., et al.: MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset. NeuroImage **70**, 33–36 (2013)
28. Manjón, J.V., et al.: Robust MRI brain tissue parameter estimation by multistage outlier rejection. Magnetic Resonance in Medicine **59**(4), 866–873 (2008)
29. Manjón, J.V., et al.: Adaptive non-local means denoising of MR images with spatially varying noise levels. Journal of Magnetic Resonance Imaging **31**(1), 192–203 (2010)
30. Manjón, J.V., et al.: NICE: Non-local Intracranial Cavity Extraction. International Journal of Biomedical Imaging (2014)
31. Nigri, E., et al.: Explainable deep CNNs for MRI-based diagnosis of Alzheimer's disease. In: International Joint Conference on Neural Networks (IJCNN) (2020)
32. Tong, T., et al.: A novel grading biomarker for the prediction of conversion from mild cognitive impairment to alzheimer's disease. IEEE Transactions on Biomedical Engineering (2016)
33. Tustison, N.J., et al.: N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging **29**(6), 1310–1320 (2010)
34. Wachinger, C., et al.: Domain adaptation for Alzheimer's disease diagnostics. NeuroImage **139** (2016)
35. Wen, J., et al.: Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical Image Analysis **63**, 101694 (2020)
36. Yee, E., et al.: Construction of MRI-based Alzheimer's disease score based on efficient 3D convolutional neural network: Comprehensive validation on 7,902 images from a multi-center dataset. Journal of Alzheimer's Disease **79**, 1–12 (2020)
37. Zhang, H., et al.: mixup: Beyond Empirical Risk Minimization. In: International Conference on Learning Representations (ICLR) (2018)
38. Zhang, X., et al.: An explainable 3D residual self-attention deep neural network for joint atrophy localization and Alzheimer's Disease diagnosis using structural MRI. IEEE Journal of Biomedical and Health Informatics (2021)