Anatomy-Aware Gating Network for Explainable Alzheimer's Disease Diagnosis

Hongchao Jiang[1] and Chunyan Miao[1,2]

[1] Alibaba-NTU Singapore Joint Research Institute, Nanyang Technological University, Singapore
[2] College of Computing and Data Science, Nanyang Technological University, Singapore

Abstract. Structural Magnetic Resonance Imaging (sMRI) is a non-invasive technique to get a snapshot of the brain for diagnosing Alzheimer's disease. Existing works have used 3D brain images to train deep learning models for automated diagnosis, but these models are prone to exploit shortcut patterns that might not have clinical relevance. We propose an Anatomy-Aware Gating Network (AAGN) which explicitly extracts features from various anatomical regions using an anatomy-aware squeeze-and-excite operation. By conditioning on the anatomy-aware features, AAGN dynamically selects the regions where atrophy is most discriminative. Once trained, we can interpret the regions selected by AAGN as explicit explanations for a given prediction. Our experiments show that AAGN selects regions well-aligned with medical literature and outperforms various convolutional and attention architectures. The code is available at `https://github.com/hongcha0/aagn`.

Keywords: Alzheimer's disease · Structural Magnetic Resonance Imaging (MRI) · Explainable AI (XAI) · Deep Learning.

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disease characterized by the accumulation of amyloid plaques and neurofibrillary tangles in the brain. Research has shown that changes in the brain occur years before any signs of cognitive decline [8]. Structural Magnetic Resonance Imaging (sMRI) is a non-invasive technique to get a snapshot of the brain structure for analysis.

Machine learning methods for classifying sMRI scans process the data at different levels of granularity. Voxel Based Morphometry [1] processes the data at the voxel level, while Region of Interest (ROI) methods extract high-level information like shape and volume from key brain regions [25]. There is a trade-off between global (general patterns) and local information (fine-grained differences). Patch-based methods try to achieve a balance between local and global features by processing the data in small patches. They typically involve training individual classifiers on the patches or employing an exemplar-based approach

that compares patch intensities [5,12]. Deep learning methods, like Convolutional Neural Networks (CNNs), manage the tradeoff between the global and local information through a data-driven approach.

While CNNs can effectively process whole sMRI scans, they may fail to capture essential medical knowledge and rely on trivial or shortcut patterns [10,38]. The translational invariance of CNNs makes them ideal for image recognition tasks where spatial location is not critical. In contrast, AD-related atrophy in sMRI images is usually region specific (e.g., medial temporal lobe). Atrophy in other regions may be due to natural aging and cannot provide a definitive diagnosis. One solution is attention models that can focus on specific parts of the sMRI. The seminal work on Squeeze-and-Excitation Network (SENet) [14] models the cross-channel relationship to assign importance to each channel of the feature map. Subsequently, SENet was adapted for medical imaging tasks by introducing concepts like spatial excitation [31] to encode positional importance. Recent works [21,40] use Vision Transformers (ViTs) for AD diagnosis by employing self-attention to learn long-range dependencies between image patches.

An inductive bias that enables the network to reason about features at an anatomy-specific level will improve explainability and performance. Prior works [24,23] have tried to include knowledge about brain anatomy into deep learning models. Liu et al. [24] used a multivariate statistical test to identify important voxels as anatomical landmarks and extracted patches from the landmarks to train landmark-specific CNNs. However, the feature selection step operates independently of model training and is not trainable end-to-end.

We propose an Anatomy-Aware Gating Network (AAGN) that integrates anatomy information in a fully differentiable manner. AAGN uses an anatomy-aware squeeze-and-excite operation to extract features from each ROI in the brain. These features are then used in a Multiple-instance learning (MIL) framework [4]. Similar to how clinicians target specific areas in sMRI scans, AAGN selectively focuses on relevant ROI instances. Ilse et al. [16] proposed a selection mechanism for MIL based on self-attention to assign importance scores to instances in the bag. A softmax function is used to normalize the scores, which explicitly forces the model to focus on a single instance. In contrast, our proposed AAGN models each ROI's importance independently as a Bernoulli distribution and has the flexibility to use more than one ROI.

In summary, our key contributions are as follows:

- We propose AAGN, which consists of (1) an anatomy-aware squeeze-and-excite operation to extract enhanced ROI features for MIL, and (2) a gating network with a fully differentiable feature selection mechanism to identify relevant ROIs. AAGN improves diagnostic accuracy and effectively adapts to different task difficulties.
- AAGN provides interpretable diagnosis that is consistent with medical literature, enhancing its reliability and transparency for clinical use. Additionally, we show that AAGN's knowledge of brain anatomy is a strong inductive bias for better data-efficiency during training.

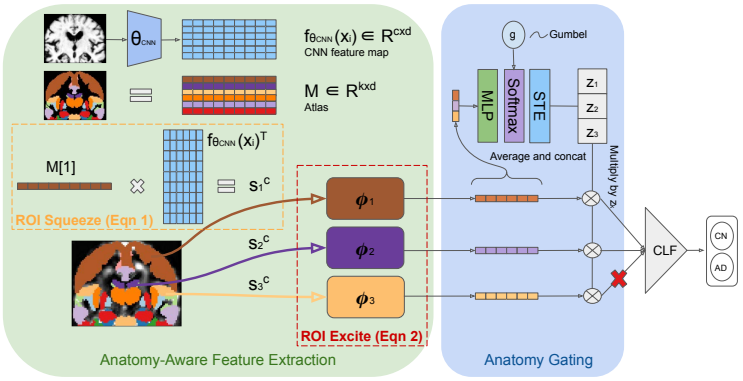


Fig. 1: Overview of AAGN. An anatomy-aware squeeze-and-excite operation extracts features from the CNN feature map for each ROI. The ROI feature embeddings are concatenated to form a context vector representing the overall brain state. The Anatomy Gating Network uses this context vector to dynamically determine which regions are relevant. A permutation-invariant classifier is trained on the selected regions and provides feedback on which ROIs to select.

2 Methodology

The classification pipeline is broken into a two-stage process: (1) finding a set of important ROIs and (2) using the ROIs to predict the disease state. More concretely, given a labeled dataset $D = \{x_i, y_i\}_{i=1}^{N}$ of sMRI images, we want to find for each x_i a set of discriminative regions $S_i = \{s_1, s_2, ..., s_m\}$ and $|S_i| \leq k$ where k is the number of brain regions. S_i can be obtained from expert annotation, but in our work we assume such information is unavailable. Instead, S_i is chosen to maximize the likelihood of D. For the disease state prediction, S_i is classified using a permutation invariant classifier (e.g., $\{s_1, s_2, s_3\}$ and $\{s_2, s_1, s_3\}$ produce the same output). An overview of the method can be found in Fig. 1.

2.1 Construction of ROI Instances for MIL

Atlas-guided segmentation [15] treats the segmentation of brain ROIs as an image registration task, where we try to find spatial correspondence between an image and brain atlas (single or multiple). For simplicity, we use a single atlas approach with anatomical parcellation obtained from the Harvard Oxford Atlas [6]. All sMRI are registered to the atlas using the deformable transformation method Symmetric Normalization (SyN) [2]. The ground truth mask of all ROIs is represented as a binary matrix $M \in R^{k \times d}$, where k is the number of ROIs and d is the number of voxels in the atlas.

2.2 Anatomy-Aware Feature Extraction

A shared CNN backbone, $f_{\theta_{CNN}}$, without downsampling layers, is used to extract a feature map spatially aligned with the brain atlas. Each voxel in the feature map has an increased receptive field, capturing information from a neighborhood to account for inaccuracies in registration. However, this may lead to some feature leakage across regions. In future work, we will explore other feature extraction techniques, such as slice-wise extraction [36].

To extract features for each region s_k, we adapt Squeeze-and-Excitation [14] to be anatomy aware. The squeeze operation aggregates channel features by using global pooling, which treats all pixels equally. To extract anatomy-specific information, we use region masks during pooling. The squeezed channel features s_k^c also serve as the feature map for each region:

$$s_k^c = M[k] \times f_{\theta_{CNN}}(x_i) / \sum_{i=1}^{d} M[k,i]\,, \tag{1}$$

For the excitation operation, we apply an MLP with parameters θ_{EX}^k followed by a Sigmoid activation to obtain a relevance score for recalibrating each channel. Lastly, we use an MLP with parameters θ_{MLP}^k to learn region specific features and obtain the enhanced feature embedding for each region:

$$\phi_k(s_k) = f_{\theta_{MLP}^k}(f_{\theta_{EX}^k}(s_k^c) \cdot s_k^c)\,. \tag{2}$$

2.3 Anatomy Gating Network

The selection of an ROI can be modeled as a Bernoulli random variable. We use an MLP with parameters θ_{AGN} to model this distribution. The input to the MLP is a context vector representing the overall brain state, obtained by averaging and concatenating the ROI embeddings. The output α_k is the unnormalized score for ROI importance. The discrete decision z_k of selecting a region can be seen as taking a sample from the Bernoulli distribution. Since the sampling process is non-differentiable, previous works have used reinforcement learning or pathwise derivative estimator to obtain estimates of $\nabla_{\theta_{AGN}} \mathbb{E}_{z_k \sim p_{\theta_{AGN}}(z_k)}[f(z_k; \theta_{AGN})]$. In our work, we use a pathwise derivative estimator approach known as the Gumbel-Softmax trick [18,26]. Specifically, we use the binary version of Gumbel-Softmax:

$$z_k = sigmoid(\frac{\alpha_k + g_1 - g_2}{\tau}) = \frac{\exp\left((\log \alpha_k + g_1)/\tau\right)}{\exp\left((\log \alpha_k + g_1)/\tau\right) + \exp\left((g_2)/\tau\right)}\,, \tag{3}$$

where g_1 and g_2 are i.i.d samples drawn from the Gumbel distribution and τ is a hyperparameter. Lower τ values more closely approximate the categorical distribution but increases the variance of the gradients. Given that z_k is a real value approximation of a discrete random variable, it cannot be directly used for selecting regions. To binarize z_k while allowing gradient flow, we use the straight through estimator (STE) [3]:

$$STE(z_k) = \begin{cases} 1 & \text{if } z_k > 0.5, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \nabla_{STE} z_k = 1\,. \tag{4}$$

By setting the threshold value for z_k at 0.5, we do not encode any prior knowledge on which regions to select. Note that we can also set different threshold values for each region to bias the model to select medically relevant regions.

2.4 Set Prediction

For the final disease state prediction, we adopt the approach proposed in [39], which uses commutative summation to achieve permutation invariance. The embedding of the selected regions S_i are summed and fed into a classifier with parameters θ_{CLF} for prediction:

$$\hat{y} = f_{\theta_{CLF}} \left(\sum_{k=1}^{K} z_k \cdot \phi_k(s_k) \right) . \tag{5}$$

A Cross-entropy loss is used to maximize the likelihood of the prediction $\hat{y}$ with respect to the ground truth labels. The parameters of AAGN, $\theta_{AAGN} = (\phi_1, \phi_2, ..., \phi_k, \theta_{AGN})$ are updated jointly with the classifier:

$$L(\theta_{CLF}, \theta_{AAGN}) = -\mathbb{E}_{x,y\sim\hat{p}_D}\ y\ log\ f_{\theta_{CLF}}(S_i)\,. \tag{6}$$

3 Experiments

3.1 Dataset and Pre-processing

We use the 1.5T T1-weighted scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [17]. Each scan was acquired during the baseline visit and has a clinical diagnosis of AD, progressive mild cognitive impairment (pMCI), stable mild cognitive impairment (sMCI), or cognitively normal (CN). pMCI subjects are those who transition to AD within the subsequent 36 months. The subject demographic and data pre-processing pipeline can be found in supplementary.

3.2 Baselines

Convolutional models: State of the art convolutional neural networks like 3D CNN [7,37] and ResNet [20] have been used for AD diagnosis.

Attention models: Squeeze-and-Excitation Network (SENet) [14] models channel-wise attention. SENet was later adapted [31] to incorporate: (1) spatial attention (sSENet) (2) both spatial and channel attention (scSENet). Recent works use Vision Transformers (ViTs), which employ self-attention to learn inter-patch dependencies. Trans-ResNet [21] augments ViT with patch features extracted by spatial convolution. GF-Net [40] improves ViT's ability to extract global features by learning filters in the frequency domain.

Weakly-supervised Attention models: Attention maps in models can be guided with weak supervision. Attention Gated Network (Attn Gated Network) [33] uses features from the last convolutional layer to provide global context to earlier layers. MWAN [22] uses a multitask approach by adding a classification head on the attention map to predict the coarse-grained subject-level label.

Deep MIL: Attention-based DMIL (Attn-based DMIL) [16] uses self-attention to weigh the importance of different patches extracted from the image.

3.3 Implementation details

For a fair comparison, we standardize the backbone architecture for all baselines as the 3D CNN (excl. ResNet and DenseNet). The 3D CNN has 6 convolutional blocks. Each block contains a $3 \times 3 \times 3$ convolution kernel with 64 channels, Instance Normalization, ReLU activation, and $2 \times 2 \times 2$ Max pooling. Baseline specific modifications are then added to the backbone. For example, SENet has a Squeeze-and-Excitation MLP for modeling channel relationships. For AAGN, we set $\tau = 1$ and carry out experiments using soft (i.e., z_k is continuous) and hard (i.e., z_k is binary) variants. All models are trained with batch size 8 and learning rate 1e-4 for 50 epochs using the Adam optimizer. Data augmentation in the form of random flipping and Gaussian smoothing is applied.

3.4 Evaluation Metrics

We conduct experiments for 2 classification tasks: (1) AD vs CN (2) pMCI vs sMCI. Following [37], the evaluation metric used is Balanced accuracy, which is the unbiased mean of sensitivity and specificity (i.e., $\frac{\text{sensitivity} + \text{specificity}}{2}$). For each experiment, we carry out 5-fold cross-validation and report the average balanced accuracy and AUC on the test set across all folds.

4 Results and discussion

4.1 Comparison of performance

Table 1 shows that AAGN outperforms all baselines in both the AD vs CN and the more challenging pMCI vs sMCI task. Notably, AAGN outperforms both attention and weakly-supervised attention models. This shows the effectiveness of our anatomy-aware squeeze-and-excite operation over self-attention ViTs (e.g., Trans-ResNet, GF-Net) and channel attention CNNs (e.g., scSENet) in extracting disease-specific patterns. Attention-based DMIL, despite being a MIL approach like AAGN, performs poorly, possibly due to the use of arbitrary patches instead of brain regions as instances (further analysis in supplementary).

Table 1: Comparison of performance for baselines and prior works. The best result is in **bold** while the second best is underlined.

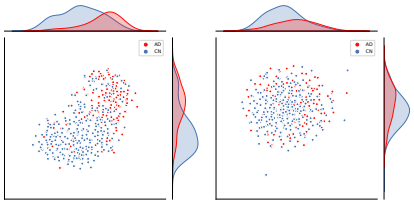
Method	AD vs CN		pMCI vs sMCI	
	Bal. Acc (%)	**AUC (%)**	**Bal. Acc (%)**	**AUC (%)**
Coupe, 2012 [5]	89.5	–	71.0	–
Tong, 2014 [35]	88.8	–	69.8	–
Moradi, 2015 [27]	–	–	70.2	76.6
Liu, 2018 [24]	**90.8**	**95.9**	62.3	77.6
Shmulev, 2018 [34]	–	–	67.5	73.0
Lian, 2020 [23]	89.5	95.1	69.0	78.1
Nguyen, 2022 [29]	90.3	–	–	–
Guan, 2023 [11]	–	–	73.0	75.7
3D CNN	88.4 ± 1.7	94.7 ± 2.0	69.3 ± 3.2	76.3 ± 3.8
3D ResNet-18	88.3 ± 1.8	94.8 ± 1.6	65.7 ± 3.4	74.5 ± 4.1
3D DenseNet-121	85.3 ± 3.2	93.6 ± 1.6	70.4 ± 3.6	77.2 ± 3.8
3D SENet	89.2 ± 3.0	95.0 ± 1.5	68.9 ± 4.4	75.9 ± 4.7
3D sSENet	88.9 ± 2.3	94.2 ± 2.2	66.7 ± 3.6	73.8 ± 2.0
3D scSENet	89.1 ± 2.2	94.9 ± 2.0	69.6 ± 4.6	75.3 ± 3.1
Attn Gated Network	87.9 ± 1.7	94.1 ± 1.5	67.8 ± 4.8	74.8 ± 4.9
Trans-ResNet	87.3 ± 3.7	93.1 ± 2.0	66.3 ± 4.6	72.5 ± 5.9
GF-Net	88.4 ± 1.4	93.9 ± 1.5	67.8 ± 3.0	72.9 ± 5.4
MWAN	88.8 ± 2.9	94.2 ± 2.3	67.9 ± 3.2	75.5 ± 4.3
Attn-based DMIL	88.0 ± 3.1	94.0 ± 1.9	65.8 ± 2.5	74.5 ± 4.6
AAGN (Soft)	90.1 ± 2.0	94.7 ± 1.5	$\mathbf{74.8 \pm 4.8}$	$\mathbf{81.3 \pm 3.2}$
AAGN (Hard)	$\underline{90.3 \pm 2.8}$	$\underline{95.1 \pm 1.4}$	$\underline{73.5 \pm 4.4}$	$\underline{81.3 \pm 3.7}$

4.2 Model analysis

To test the inductive bias of using anatomy information, we compare the model features at random initialization [9]. AAGN is able to achieve distinct separation of the AD and CN samples even without training (Fig. 2).

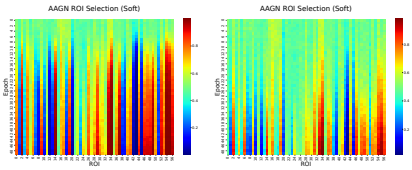
The ROI selection behaviour of AAGN during training is shown in Fig. 3. The x-axis is the index for each ROI (refer supplementary) and the y-axis is the training epoch. We can see that the training process is stable and converges to a fixed set of ROIs (i.e., column gets warmer in color along the y-axis).

AAGN is able to adapt to different tasks complexity. It selects more regions for AD vs CN compared to pMCI vs sMCI (i.e., more columns with warm color). This is because AD brains are significantly more atrophied than healthy brains, which means a model can look at almost any region of the brain to discriminate the two. In contrast, for pMCI vs sMCI, atrophy differences are less obvious and AAGN needs to be more careful in its selection. AAGN thus takes a longer time to explore and converge to a set of useful ROIs for pMCI vs sMCI (i.e., warm colors only appear around epoch 25). Such behaviour is ideal, as we want to avoid premature selection of suboptimal ROIs.



(a) AAGN (b) 3D CNN

Fig. 2: t-SNE plots at random initialization without training. Best viewed in color and zoomed in.



(a) AD vs CN (b) pMCI vs sMCI

Fig. 3: ROI selection during training. Best viewed in color and zoomed in.

We conduct analysis on the top regions selected by AAGN as shown in Tables 2 and 3. The regions picked like the Amygdala [30], Hippocampus [28], Parahippocampal [13], and Inferior Temporal Gyrus [32] are of medical relevance for AD. An interesting region selected is the Thalamus, which is seen as a potential research area for early-stage AD [19].

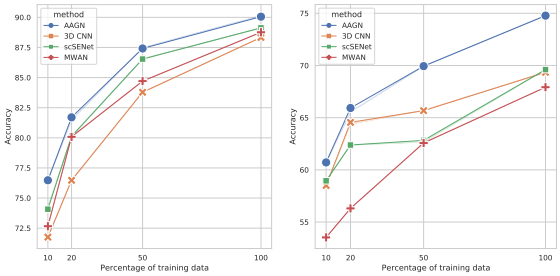
To assess data efficiency, we vary the amount of training data available and observe AAGN performs the best across all settings (Fig. 4).

Table 2: Top 5 ROI (AD vs CN)

Name	Probability (z_K)
Amygdala	0.998
Hippocampus	0.998
Parahippocampal Gyrus (post.)	0.998
Inferior Temporal Gyrus (post.)	0.992
Thalamus	0.985

Table 3: Top 5 ROI (pMCI vs sMCI)

Name	Probability (z_K)
Parahippocampal Gyrus (post.)	0.975
Amygdala	0.954
Parahippocampal Gyrus (ant.)	0.927
Hippocampus	0.902
Inferior Temporal Gyrus (post.)	0.867



(a) AD vs CN (b) pMCI vs sMCI

Fig. 4: Comparison of model's data-efficiency. Best viewed zoomed in.

5 Conclusion

AAGN is an explainable model that outperforms prior works on the clinically significant task of identifying prodromal AD subjects. It provides explicit explanations of which brain regions are important, enhancing transparency and aiding clinicians in identifying regions for further examination. AAGN's feature selection mechanism adapts to task difficulty and converges to ROIs well-aligned with medical literature. Additionally, AAGN's understanding of brain anatomy serves as a strong inductive bias, making it more data-efficient than other methods.

Acknowledgments. This research is supported by Alibaba Group and Alibaba-NTU Singapore Joint Research Institute (JRI), Nanyang Technological University, Singapore.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry—the methods. NeuroImage **11**(6), 805–821 (2000)
2. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. MedIA **12**(1), 26–41 (2008)
3. Bengio, Y., Léonard, N., Courville, A.: Estimating or propagating gradients through stochastic neurons for conditional computation. arXiv preprint arXiv:1308.3432 (2013)
4. Carbonneau, M.A., et al.: Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognition **77**, 329–353 (2018)
5. Coupé, P., et al.: Scoring by nonlocal image patch estimator for early detection of alzheimer's disease. NeuroImage: clinical **1**(1), 141–152 (2012)
6. Desikan, R.S., et al.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. NeuroImage **31**(3), 968–980 (2006)
7. Esmaeilzadeh, S., Belivanis, D.I., Pohl, K.M., Adeli, E.: End-to-end alzheimer's disease diagnosis and biomarker identification. In: International Workshop on Machine Learning in Medical Imaging. pp. 337–345. Springer (2018)
8. Frisoni, G.B., et al.: The clinical use of structural mri in alzheimer disease. Nature Reviews Neurology **6**(2), 67–77 (2010)
9. Gaier, A., Ha, D.: Weight agnostic neural networks. NeurIPS **32** (2019)
10. Geirhos, R., et al.: Shortcut learning in deep neural networks. Nature Machine Intelligence **2**(11), 665–673 (2020)
11. Guan, H., et al.: Attention-guided autoencoder for automated progression prediction of subjective cognitive decline with structural mri. IEEE journal of biomedical and health informatics **27**(6), 2980–2989 (2023)
12. Hett, K., et al.: Multi-scale graph-based grading for alzheimer's disease prediction. MedIA **67**, 101850 (2021)

13. van Hoesen, G.W., et al.: The parahippocampal gyrus in alzheimer's disease: clinical and preclinical neuroanatomical correlates. Annals of the New York Academy of Sciences **911**(1), 254–274 (2000)
14. Hu, J., et al.: Squeeze-and-excitation networks. In: CVPR. pp. 7132–7141 (2018)
15. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. MedIA **24**(1), 205–219 (2015)
16. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
17. 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)
18. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)
19. 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)
20. Korolev, S., et al.: Residual and plain convolutional neural networks for 3d brain mri classification. In: IEEE ISBI. pp. 835–838. IEEE (2017)
21. Li, C., et al.: Trans-resnet: Integrating transformers and cnns for alzheimer's disease classification. In: IEEE ISBI. pp. 1–5. IEEE (2022)
22. Lian, C., Liu, M., Wang, L., Shen, D.: Multi-task weakly-supervised attention network for dementia status estimation with structural mri. IEEE TNNLS (2021)
23. Lian, C., et al.: Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri. IEEE TPAMI **42**(4), 880–893 (2020)
24. Liu, M., Zhang, J., Adeli, E., Shen, D.: Landmark-based deep multi-instance learning for brain disease diagnosis. MedIA **43**, 157–168 (2018)
25. Long, X., et al.: Prediction and classification of alzheimer disease based on quantification of mri deformation. PloS one **12**(3), e0173372 (2017)
26. Maddison, C.J., Mnih, A., Teh, Y.W.: The concrete distribution: A continuous relaxation of discrete random variables. arXiv preprint arXiv:1611.00712 (2016)
27. Moradi, E., et al.: Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. NeuroImage **104**, 398–412 (2015)
28. Mu, Y., Gage, F.H.: Adult hippocampal neurogenesis and its role in alzheimer's disease. Molecular neurodegeneration **6**(1), 1–9 (2011)
29. Nguyen, H.D., et al.: Interpretable differential diagnosis for alzheimer's disease and frontotemporal dementia. In: Wang, L., et al. (eds.) MICCAI 2022. pp. 55–65. Springer, Cham (2022)
30. Poulin, S.P., et al.: Amygdala atrophy is prominent in early alzheimer's disease and relates to symptom severity. Psychiatry Research: Neuroimaging **194**(1), 7–13
31. Roy, A.G., et al.: Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In: Frangi, A.F., et al. (eds.) MICCAI 2018. pp. 421–429. Springer, Cham (2018)
32. Scheff, S.W., Price, D.A., Schmitt, F.A., Scheff, M.A., Mufson, E.J.: Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and alzheimer's disease. Journal of Alzheimer's Disease **24**(3), 547–557 (2011)
33. Schlemper, J., et al.: Attention-gated networks for improving ultrasound scan plane detection. Medical Imaging with Deep Learning (2018)
34. Shmulev, Y., et al.: Predicting conversion of mild cognitive impairments to alzheimer's disease and exploring impact of neuroimaging. In: Graphs in biomedical image analysis and integrating medical imaging and non-imaging modalities, pp. 83–91. Springer (2018)

35. Tong, T., et al.: Multiple instance learning for classification of dementia in brain mri. MedIA **18**(5), 808–818 (2014)
36. Wang, C., et al.: Joint learning framework of cross-modal synthesis and diagnosis for alzheimer's disease by mining underlying shared modality information. MedIA **91**, 103032 (2024)
37. Wen, J., et al.: Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. MedIA p. 101694 (2020)
38. Winkler, J.K., et al.: Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA dermatology **155**(10), 1135–1141 (2019)
39. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. In: NeurIPS. vol. 30 (2017)
40. Zhang, S., et al.: 3d global fourier network for alzheimer's disease diagnosis using structural mri. In: Wang, L., et al. (eds.) MICCAI 2022. pp. 34–43. Springer, Cham (2022)