Brain Anatomy-Guided MRI Analysis for Assessing Clinical Progression of Cognitive Impairment with Structural MRI

Lintao Zhang[1], Jinjian Wu[2], Lihong Wang[3], Li Wang[1], David C. Steffens[3], Shijun Qiu[2], Guy G. Potter[4(✉)], and Mingxia Liu[1(✉)]

[1] Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
mxliu@med.unc.edu

[2] The First School of Clinical Medicine, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China

[3] Department of Psychiatry, University of Connecticut School of Medicine, University of Connecticut, Farmington, CT, USA

[4] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA
guy.potter@duke.edu

Abstract. Brain structural MRI has been widely used for assessing future progression of cognitive impairment (CI) based on learning-based methods. Previous studies generally suffer from the limited number of labeled training data, while there exists a huge amount of MRIs in large-scale public databases. Even without task-specific label information, brain anatomical structures provided by these MRIs can be used to boost learning performance intuitively. Unfortunately, existing research seldom takes advantage of such brain anatomy prior. To this end, this paper proposes a brain anatomy-guided representation (BAR) learning framework for assessing the clinical progression of cognitive impairment with T1-weighted MRIs. The BAR consists of a *pretext model* and a *downstream model*, with a shared brain anatomy-guided encoder for MRI feature extraction. The pretext model also contains a decoder for brain tissue segmentation, while the downstream model relies on a predictor for classification. We first train the pretext model through a brain tissue segmentation task on 9,544 auxiliary T1-weighted MRIs, yielding a generalizable encoder. The downstream model with the learned encoder is further fine-tuned on target MRIs for prediction tasks. We validate the proposed BAR on two CI-related studies with a total of 391 subjects with T1-weighted MRIs. Experimental results suggest that the BAR outperforms several state-of-the-art (SOTA) methods. The source code and pre-trained models are available at https://github.com/goodaycoder/BAR.

Keywords: Structural MRI · Brain anatomy · Cognitive impairment

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43993-3_11.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14227, pp. 109–119, 2023.
https://doi.org/10.1007/978-3-031-43993-3_11

1 Introduction

Brain magnetic resonance imaging (MRI) has been increasingly used to assess future progression of cognitive impairment (CI) in various clinical and research fields by providing structural brain anatomy [1–6]. Many learning-based methods have been developed for automated MRI analysis and brain disorder prognosis, which usually heavily rely on labeled training data [7–10]. However, it is generally time-consuming and tedious to collect category labels for brain MRIs in practice, resulting in a limited number of labeled MRIs [11].

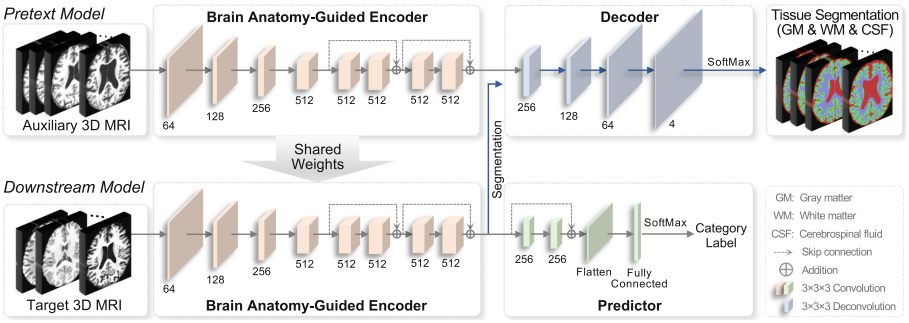

Fig. 1. Illustration of brain anatomy-guided representation (BAR) learning framework for assessing the clinical progression of cognitive impairment. The BAR consists of a *pretext model* and a *downstream model*, with a shared brain anatomy-guided encoder for MRI feature extraction. The pretext model also contains a decoder for brain tissue segmentation, while the downstream model relies on a predictor for prediction. The pretext model is trained on the large-scale ADNI [12] with 9,544 T1-weighted MRIs, yielding a generalizable encoder. With this learned encoder frozen, the downstream model is then fine-tuned on target MRIs for prediction tasks.

Even without task-specific category label information, brain anatomical structures provided by auxiliary MRIs can be employed as a prior to boost disease progression prediction performance. Considering that there are a large number of unlabeled MRIs in existing large-scale datasets [12,13], several deep learning methods propose to extract brain anatomical features from MRI without requiring specific category labels. For instance, Song *et al.* [14] suggest that the anatomy prior can be utilized to segment brain tumors, while Yamanakkanavar *et al.* [15] discuss how brain MRI segmentation improves disease diagnosis. Unfortunately, there are few studies that try to utilize such brain anatomy prior for assessing the clinical progression of cognitive impairment with structural MRIs.

To this end, we propose a brain anatomy-guided representation (BAR) learning framework for cognitive impairment prognosis with T1-weighted MRIs, incorporated with brain anatomy prior provided by a brain tissue segmentation task. As shown in Fig. 1, the BAR consists of a *pretext model* and a *downstream model*, with a shared anatomy-guided encoder for MRI feature extraction. These

two models also use a decoder and a predictor for brain tissue segmentation and disease progression prediction, respectively. The pretext model is trained on 9,544 MRI scans from the public Alzheimer's Disease Neuroimaging Initiative (ADNI) [12] without any category label information, yielding a generalizable encoder. The downstream model is further fine-tuned on target MRIs for CI progression prediction. Experiments are performed on two CI-related studies with 391 subjects, with results suggesting the efficacy of BAR compared with state-of-the-art (SOTA) methods. The pretext model can also be used for brain tissue segmentation in other MRI-based studies. To the best of our knowledge, this is the first work that utilizes anatomy prior derived from large-scale T1-weighted MRIs for automated cognitive decline analysis. To promote reproducible research, the source code and trained models have been made publicly available to the research community (see https://github.com/goodaycoder/BAR).

2 Materials and Proposed Method

Data and Preprocessing. The *pretext model* is trained via a tissue segmentation task on auxiliary MRIs (without category label) from **ADNI**. A total of 9,544 T1-weighted MRIs from 2,370 ADNI subjects with multiple scans are used in this work. To provide accurate brain anatomy, we perform image preprocessing and brain tissue segmentation for these MRIs to generate ground-truth segmentation of three tissues, *i.e.*, white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF), using an in-house toolbox iBEAT [16] with manual verification.

The *downstream model* is trained on 1) a late-life depression (**LLD**) study with 309 subjects from two sites [17,18], and 2) a type 2 diabetes mellitus (**DM**) study with 82 subjects from the First Affiliated Hospital of Guangzhou University of Chinese Medicine. Subjects in LLD are categorized into three groups: 1) 89 non-depressed cognitively normal (CN), 2) 179 depressed but cognitively normal (CND), 3) 41 depressed subjects (called CI) who developed cognitive impairment or even dementia in the follow-up years. Category labels in the LLD study are determined based on subjects' 5-year follow-up diagnostic information, while MRIs are acquired at baseline time. The DM contains 1) 45 health control (HC) subjects and 2) 37 diabetes mellitus patients with mild CI (MCI). Detailed image acquisition protocols are given in Table SI of *Supplementary Materials*. All MRIs are preprocessed via the following pipeline: 1) bias field correction, 2) skull stripping, 3) affine registration to the MNI space, 4) resampling to $1 \times 1 \times 1$ mm^3, 5) deformable registration to AAL3 [19] with SyN [20], and 6) warping 166 regions-of-interest (ROIs) of AAL3 back to MRI volumes.

Proposed Method. While it is often challenging to annotate MRIs in practice, there are a large number of MRIs (without task-specific category labels) in existing large-scale datasets. Even without category labels, previous studies propose to extract anatomical features (*e.g.*, ROI volumes of GM segmentation maps) to characterize brain anatomy [21,22]. Such brain anatomy prior learned via

tissue segmentation can be employed to boost learning performance intuitively. Accordingly, we propose a brain anatomy-guided representation (BAR) learning framework for progression prediction of cognitive impairment, incorporated with brain anatomy prior provided by brain tissue segmentation. As shown in Fig. 1, the BAR consists of a *pretext model* for brain tissue segmentation and a *downstream model* for disease progression prediction, both equipped with brain anatomy-guided encoders (shared weights) for MRI feature learning.

(1) Pretext Model for Segmentation. To learn brain anatomical features from MRIs in a data-driven manner, we propose to employ a segmentation task for pretext model training. As shown in the top of Fig. 1, the pretext model consists of 1) a *brain anatomy-guided encoder* for MRI feature extraction and 2) a *decoder* for segmentation. The brain anatomy-guided encoder takes large-scale auxiliary 3D MRIs without category labels as input, and outputs 512 feature maps. It contains 8 convolution blocks, with each block containing a convolution layer (kernel size: $3 \times 3 \times 3$), followed by instance normalization and Parametric Rectified Linear Unit (PReLU) activation. The first 4 blocks downsample the input with a stride of $2 \times 2 \times 2$. The channel numbers of the eight blocks are [64, 128, 256, 512, 512, 512, 512, 512], respectively. A skip connection is applied to sum the input and output of every two of the last 4 blocks for residual learning.

The decoder takes the 512 feature maps as input and outputs segmentation maps of three tissues (*i.e.*, WM, GM, and CSF), thus guiding the encoder to learn brain anatomical features. The decoder contains four deconvolution blocks with 256, 128, 64 and 4 channels, respectively. Each deconvolution block shares the same architecture as the convolution block in the encoder. The output of the decoder is then fed into a SoftMax layer to get four probability maps that indicate the probability of a voxel belonging to a specific tissue (*i.e.*, background, WM, GM, and CSF). Besides, the reconstruction task can be used to train the pretext model instead of segmentation when lacking ground-truth segmentation maps. For *problems without ground-truth segmentation maps*, we can resort to an MRI reconstruction task to train the pretext model in an unsupervised manner.

(2) Downstream Model for Prediction. As shown in the bottom panel of Fig. 1, the downstream model takes target MRIs as input and outputs probabilities of category labels. It consists of 1) a *brain anatomy-guided encoder* and 2) a *predictor* for prognosis. This encoder shares the same architecture and parameters as that of the pre-trained pretext model. With the encoder frozen, predictor parameters will be updated when the downstream model is trained on target MRIs. The predictor has two convolution blocks (kernel size: $3 \times 3 \times 3$, stride: $2 \times 2 \times 2$, channel: 256) with a skip connection, followed by a flatten layer, an FC layer, and a SoftMax layer. The architecture of the predictor can be flexibly adjusted according to the requirements of different downstream tasks.

(3) Implementation. The proposed BAR is trained via two steps. 1) The pretext model is first trained on 9,544 MRIs from ADNI, with ground-truth segmentation as supervision. The Adam optimizer [23] with a learning rate of 10^{-4} and dice loss are used for training (batch size: 4, epoch: 30). 2) We then share

Table 1. Performance of fourteen methods in two MRI-based depression recognition tasks (*i.e.*, CND vs. CN and CI vs. CND classification) on the LLD study.

Method	CND vs. CN Classification					CI vs. CND Classification				
	AUC (%)	ACC (%)	SEN (%)	SPE (%)	F1s (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	F1s(%)
SVM-GM	51.6(0.8)	52.3(3.8)*	48.7(12.2)	56.0(8.6)	49.9(8.4)	54.0(1.5)	50.8(1.1)*	51.6(2.4)	50.0(0.0)	50.5(1.8)
SVM-WM	57.9(2.6)	53.3(4.7)*	59.3(6.4)	47.3(6.4)	55.9(4.9)	38.7(3.5)	48.7(3.1)*	42.1(9.8)	55.0(7.1)	44.0(6.8)
XGB-GM	43.9(2.8)	42.3(1.5)*	54.0(2.8)	30.7(4.3)	48.3(1.4)	49.3(7.0)	44.1(9.5)*	36.8(12.9)	51.0(8.2)	38.7(12.5)
XGB-WM	50.8(2.5)	53.0(2.7)*	64.0(3.6)	42.0(3.8)	57.6(2.6)	61.7(5.0)	57.4(3.9)*	50.5(7.1)	64.0(6.5)	53.5(4.9)
ResNet18	65.4(7.2)	58.7(8.0)*	52.7(10.1)	64.7(9.0)	61.0(7.7)	61.3(8.1)	56.9(7.8)	**59.0**(14.6)	55.0(26.7)	**56.8**(4.9)
ResNet34	58.9(7.3)	57.7(6.1)*	56.0(8.3)	59.3(7.6)	58.3(6.2)	58.5(4.6)	56.4(8.1)	47.4(5.3)	65.0(13.7)	51.6(6.9)
ResNet50	63.0(5.4)	57.0(3.6)	67.3(8.6)	46.7(8.2)	60.9(4.5)	55.3(2.6)	50.8(4.6)	41.0(10.1)	60.0(13.7)	44.3(6.7)
Med3D18	57.9(2.4)	54.7(1.8)*	53.3(9.7)	56.0(9.3)	55.0(4.3)	59.6(9.1)	57.4(10.3)*	50.5(16.9)	64.0(10.8)	52.8(15.4)
Med3D34	58.7(7.2)	56.7(4.9)*	50.0(7.8)	63.3(11.1)	59.1(6.6)	57.6(4.6)	55.9(2.1)*	56.8(6.9)	55.0(7.9)	55.5(3.2)
Med3D50	60.2(2.5)	59.0(4.3)*	50.0(16.8)	**68.0**(13.9)	54.0(8.5)	47.3(9.2)	46.7(6.4)	45.3(24.6)	48.0(23.9)	43.0(13.6)
SEResNet	66.4(2.1)	57.7(4.3)	60.0(12.7)	55.3(14.6)	58.2(6.1)	59.5(5.5)	53.9(2.6)*	55.8(28.5)	52.0(28.0)	51.4(13.6)
EfficientNet	59.3(9.9)	58.0(5.9)*	64.7(25.7)	51.3(17.1)	54.1(6.7)	56.4(8.9)	53.3(5.2)*	56.8(31.9)	50.0(29.8)	49.4(22.9)
MobileNet	55.3(8.6)	54.3(3.5)*	44.7(31.2)	64.0(31.7)	53.9(20.6)	58.5(10.1)	53.3(5.8)*	56.8(40.0)	50.0(35.9)	47.2(27.5)
BAR (Ours)	**70.5**(4.1)	**65.0**(1.7)	**77.3**(5.5)	52.7(3.6)	**68.8**(2.4)	**64.5**(5.8)	**59.0**(5.1)	50.5(8.0)	**67.0**(6.7)	54.4(6.4)

the encoder of the pre-trained pretext model with the downstream model and fine-tune the predictor via a cross-entropy loss (batch size: 2, learning rate: 10^{-4}, epoch: 90). The learning rate of fine-tuning decays by 0.1 every 30 epochs. The BAR is implemented on PyTorch with NVIDIA TITAN Xp (memory: 12GB).

3 Experiment

Experimental Setting. Three classification tasks are performed: (1) depression recognition (*i.e.*, CND vs. CN classification) on LLD, (2) CI identification (*i.e.*, CI vs. CND classification) on LLD, and (3) MCI detection (*i.e.*, MCI vs. HC classification) on DM. The partition of training/test set is given in Table SII of *Supplementary Materials*. Such partition is repeated five times independently to avoid any bias introduced by random partition, and the mean and standard deviation results are recorded. The training data is duplicated and augmented using random affine transform. Five evaluation metrics are used, including area under ROC curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1-Score (F1s). Besides, we perform tissue segmentation by directly applying the trained pretext model to target MRIs from LLD and DM studies and visually compare the results of our BAR with those of FSL [24].

Competing Methods. We compare our BAR with two classic machine learning methods and five SOTA deep learning approaches, including (1) support vector machine (**SVM**) [25] with a radial basis function kernel (regularization: 1.0), (2) XGBoost (**XGB**) [26] (estimators: 300, tree depth: 4, learning rate: 0.2), (3) **ResNet**x with x convolution layers, (4) **Med3D**x [27] with x convolution layers, (5) **SEResNet** [28] that is an improved model by adding squeeze and excitation blocks to ResNet, (6) **EfficientNet** [29], and (7) **MobileNet** [30] that is an efficient lightweight CNN model. For SVM and XGB, we extract ROI-based WM and GM volumes of each MRI as input. All competing deep learning methods (with default architectures) take whole 3D MRIs as input and share the same training strategy as that used in the downstream model of the BAR. An early-stop training strategy (epoch: 90) is used in all deep learning models.

Table 2. Performance of fourteen methods in the MRI-based MCI detection task (*i.e.*, MCI vs. HC classification) on the DM study.

Method	AUC (%)	ACC (%)	SEN (%)	SPE (%)	F1s (%)
SVM-GM	52.5(1.7)	50.5(2.1)*	42.0(10.4)	59.0(9.6)	45.3(7.3)
SVM-WM	41.4(1.9)	47.5(1.8)*	43.0(13.0)	52.0(14.0)	44.0(9.2)
XGB-GM	39.1(5.2)	46.5(3.4)*	42.0(4.5)	51.0(4.2)	43.9(4.0)
XGB-WM	61.5(3.2)	55.5(2.1)*	48.0(2.7)	**63.0**(2.7)	51.9(2.5)
ResNet18	58.8(1.4)	55.5(2.7)*	60.0(7.1)	51.0(2.2)	57.3(4.5)
ResNet34	60.9(2.9)	55.5(1.1)*	59.0(5.5)	52.0(4.5)	56.9(2.7)
ResNet50	60.8(2.5)	58.0(2.1)	65.0(7.9)	51.0(4.2)	60.6(4.1)
Med3D18	56.0(3.1)	53.5(4.9)	58.0(22.5)	49.0(12.9)	53.9(11.7)
Med3D34	55.1(6.9)	51.5(5.8)	59.0(14.8)	44.0(20.7)	54.4(6.5)
Med3D50	61.3(5.4)	57.5(5.3)	63.0(6.7)	52.0(8.4)	59.7(5.0)
SEResNet	62.8(1.9)	59.0(2.9)	63.0(16.4)	55.0(17.0)	59.8(7.0)
EfficientNet	55.6(3.6)	53.5(8.0)*	64.0(19.8)	43.0(4.5)	56.9(12.5)
MobileNet	58.5(2.1)	53.0(4.1)*	59.0(26.1)	47.0(24.4)	53.3(13.5)
BAR (Ours)	**67.6**(1.7)	**60.5**(1.1)	**66.0**(10.8)	55.0(9.4)	**62.2**(4.5)

Results of Depression and CI Identification. In this task, we aim to recognize cognitively normal subjects with depression with a higher risk of progressing to CI than healthy subjects. The results of fourteen methods on the LLD study are reported in Table 1, where '*' denotes that the results of BAR and a competing method are statistically significantly different ($p < 0.05$ via paired t-test).

From the left of Table 1, we have the following observations on *CND vs. CN classification*. *First*, our BAR model generally outperforms thirteen competing methods in most cases. For instance, the BAR yields the results of AUC = 70.5% and SEN = 77.3%, which are 4.1% and 10.0% higher than those of the second-best methods (*i.e.*, SEResNet and ResNet50), respectively. This implies that the brain anatomical MRI features learned by our pretext model on large-scale datasets would be more discriminative, compared with those used in the competing methods. *Second*, among 10 deep models, our BAR produces the lowest standard deviation in most cases (especially on SEN and SPE), suggesting its robustness to bias introduced by random data partition in the downstream task. This could be due to the strong generalizability of the feature encoder guided by brain anatomy prior (derived from the auxiliary tissue segmentation task). *In addition*, our BAR significantly outperforms four machine learning methods and two lightweight deep models (*i.e.*, EfficientNet and MobileNet) with $p < 0.05$.

From the right of Table 1, we can see that the overall results of fourteen methods in *CI vs. CND classification* are usually worse than *CND vs. CN classification*. This suggests that the task of CI vs. CND classification is more challenging, which could be due to the more imbalanced training data in this task (as shown in Table SII of *Supplementary Materials*). On the other hand, the proposed BAR

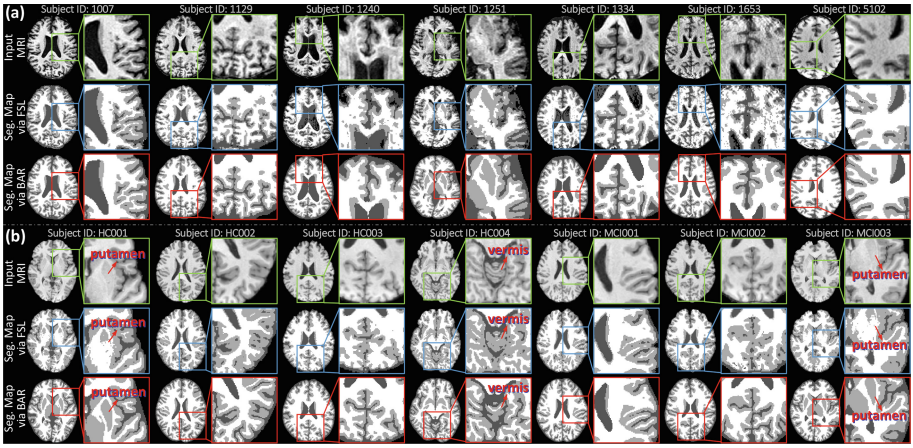

Fig. 2. Tissue segmentation (seg.) maps of white matter, gray matter and cerebrospinal fluid produced by FSL and our BAR on (a) LLD study and (b) DM study.

still performs best in terms of AUC=64.5% and SPE=67.0%, which are 2.8% and 2.0% higher than those of the second-best competing methods (*i.e.*, XGB-WM and ResNet34), respectively. These results further demonstrate the superiority of the BAR in MRI-based depression recognition.

Results of MCI Detection. The results of different methods in MCI detection (*i.e.*, MCI vs. HC classification) on the DM study are reported in Table 2. There are a total of 42 subjects (*i.e.*, 17 MCI and 25 HC) used for training in this task, which are fewer but more balanced than the two tasks in the LLD study (see Table SII). It can be observed from Tables 1 and 2 that the proposed BAR yields relatively lower standard deviations in terms of AUC and ACC in MCI vs. HC classification, compared with the two tasks on the LLD study. These results imply that data imbalance may be an important issue affecting the performance of deep learning models when the number of training samples is limited.

Segmentation Results. The pre-trained pretext model can also be used for brain tissue segmentation in downstream studies. Thus, we visualize brain segmentation maps generated by FSL and our BAR for target MRIs in both LLD and DM studies in Fig. 2. Note that T1-weighted MRIs in the LLD study are collected from 2 sites and have more inconsistent image quality when compared to those from DM. From Fig. 2, we have several interesting observations.

First, the segmentation results generated by the proposed BAR are generally better than those of FSL in most cases, especially for those *cortical surface areas* on the two studies. For instance, the WM region in segmentation maps generated by our BAR is much cleaner than that of FSL, indicating that our model is not sensitive to noise in MRI. Even for the LLD study with significant

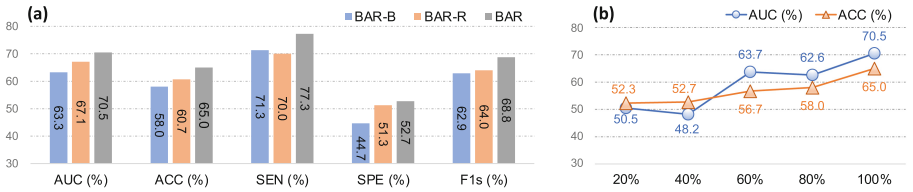

Fig. 3. (a) Results achieved by BAR with different pretext tasks in CND vs. CN classification on LLD. (b) Results of BAR with different numbers of MRIs from the target domain for downstream model training in CND vs. CN classification on LLD.

inter-site data heterogeneity, the boundary of WM and GM produced by BAR is more continuous and smoother, which is in line with the brain anatomy prior. *Second*, for MRIs with severe motion artifacts in the LLD study (IDs: 1240, 1334, and 1653), our method can produce high-quality segmentation maps, and the results are even comparable to those of MRIs without motion artifacts. This demonstrates that our model is robust to motion artifacts. The underlying reason could be that the pretext model is trained on large-scale MRIs, and thus, has good generalization ability when applied to MRIs with different image quality. *In addition*, both BAR and FSL often achieve better results in the DM study, since DM has relatively higher image quality than LLD. Still, the proposed BAR can achieve better segmentation results in many fine-grained brain regions, such as the *putamen region* (see HC001 and MCI003) and the *vermis region* (see HC004). These results demonstrate that our method has good adaptability when applied to classification and segmentation tasks in MRI-based studies.

Ablation Study. To validate the effectiveness of the learned brain anatomical MRI features, we further compare the BAR with its two variants (called **BAR-B** and **BAR-R**) that use anatomy prior derived from different pretext tasks in CND vs. CN classification on LLD. Specifically, the BAR-B is *trained from scratch* as a baseline on target data without any pre-trained encoder. The BAR-R trains the pretext model through an *MRI reconstruction task* in an unsupervised learning manner. As shown in Fig. 3(a), the BAR consistently performs better than its variants in terms of all five metrics. This implies that the learned MRI features guided by the segmentation task help promote prediction performance. Also, BAR and BAR-R outperform BAR-B in most cases, implying that brain anatomy prior derived from tissue segmentation or MRI reconstruction can help improve discriminative ability of MRI features and boost prediction performance.

Influence of Training Data Size. We also study the influence of training data size on BAR in CND vs. CN classification on LLD. With fixed test data, we randomly select a part of MRIs (*i.e.*, $[20\%, 40\%, \cdots, 100\%]$) from target training data to fine-tune the downstream prediction model. It can be observed from Fig. 3(b) that the overall performance in terms of AUC and ACC of our BAR

increases with the increase of training data, and it produces the best results when using all training data for model fine-tuning. This suggests that using more data for downstream model fine-tuning helps promote learning performance.

4 Conclusion and Future Work

In this paper, we develop a brain anatomy-guided representation (BAR) learning framework for MRI-based progression prediction of cognitive impairment, incorporated by brain anatomy prior (derived from an auxiliary tissue segmentation task). We validate the proposed BAR on two CI-related studies with T1-weighted MRIs, and the experimental results demonstrate its effectiveness compared with SOTA methods. Besides, the pretext model trained on 9,544 MRIs from ADNI can be well adapted to tissue segmentation in the two CI-related studies. There is significant intra- and inter-site data heterogeneity in LLD with two sites. It is interesting to reduce such heterogeneity using domain adaptation [31,32], which will be our future work. Aside from tissue segmentation, one can use other auxiliary tasks to model brain anatomy, such as brain parcellation and brain MRI to CT translation. Besides, it is meaningful to compare our method with other model pre-training strategies [33,34], which will also be our future work.

Acknowledgment. L. Zhang, G.G. Potter, and M. Liu were supported by NIH grants RF1AG073297 and R01MH108560.

References

1. Ashtari-Majlan, M., Seifi, A., Dehshibi, M.M.: A multi-stream convolutional neural network for classification of progressive MCI in Alzheimer's disease using structural MCI images. IEEE J. Biomed. Health Inform. **26**(8), 3918–3926 (2022)
2. El-Gamal, F.E.Z.A., et al.: A personalized computer-aided diagnosis system for mild cognitive impairment (MCI) using structural MRI (sMRI). Sensors **21**(16), 5416 (2021)
3. Gonuguntla, V., Yang, E., Guan, Y., Koo, B.B., Kim, J.H.: Brain signatures based on structural MRI: classification for MCI, PMCI, and AD. Hum. Brain Mapp. **43**(9), 2845–2860 (2022)
4. Guo, M., et al.: A novel conversion prediction method of MCI to AD based on longitudinal dynamic morphological features using ADNI structural MRIs. J. Neurol. **267**(10), 2983–2997 (2020). https://doi.org/10.1007/s00415-020-09890-5
5. Lombardi, G., et al.: Structural magnetic resonance imaging for the early diagnosis of dementia due to Alzheimer's disease in people with mild cognitive impairment. Cochrane Database Syst. Rev. (3) (2020)
6. Yin, C., et al.: Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment. Proc. Natl. Acad. Sci. **120**(2), e2214634120 (2023)
7. Noor, M.B.T., Zenia, N.Z., Kaiser, M.S., Mahmud, M., Al Mamun, S.: Detecting neurodegenerative disease from MRI: A brief review on a deep learning perspective. In: Liang, P., Goel, V., Shan, C. (eds.) BI 2019. LNCS, vol. 11976, pp. 115–125. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37078-7_12

8. Chen, X., Tang, M., Liu, A., Wei, X.: Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI. Ann. Transl. Med. **10**(14) (2022)
9. Garg, N., Choudhry, M.S., Bodade, R.M.: A review on Alzheimer's disease classification from normal controls and mild cognitive impairment using structural MR images. J. Neurosci. Methods **384**, 109745 (2022)
10. Scarpazza, C., et al.: Translating research findings into clinical practice: a systematic and critical review of neuroimaging-based clinical tools for brain disorders. Transl. Psychiatry **10**(1), 107 (2020)
11. Nanni, L., et al.: Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of Alzheimer's disease. Front. Neurol. **11**, 576194 (2020)
12. Jack Jr, C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med. **27**(4), 685–691 (2008)
13. LaMontagne, P.J., et al.: OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. MedRxiv 2019-12 (2019)
14. Song, B., Chou, C.R., Chen, X., Huang, A., Liu, M.C.: Anatomy-guided brain tumor segmentation and classification. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., Handels, H. (eds.) BrainLes 2016. LNCS, vol. 10154, pp. 162–170. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-55524-9_16
15. Yamanakkanavar, N., Choi, J.Y., Lee, B.: MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. Sensors **20**(11), 3243 (2020)
16. Wang, L., Wu, Z., Chen, L., Sun, Y., Lin, W., Li, G.: iBEAT V2.0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. Nat. Protocols **18**(5), 1488–1509 (2023)
17. Steffens, D.C., et al.: Methodology and preliminary results from the neurocognitive outcomes of depression in the elderly study. J. Geriatr. Psychiatry Neurol. **17**(4), 202–211 (2004)
18. Steffens, D.C., Wang, L., Manning, K.J., Pearlson, G.D.: Negative affectivity, aging, and depression: results from the neurobiology of late-life depression (NBOLD) study. Am. J. Geriatr. Psychiatry **25**(10), 1135–1149 (2017)
19. Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J., Joliot, M.: Automated anatomical labelling atlas 3. Neuroimage **206**, 116189 (2020)
20. 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. Med. Image Anal. **12**(1), 26–41 (2008)
21. Elsayed, A.S.A.: Region of interest based image classification: a study in MRI brain scan categorization. The University of Liverpool (United Kingdom) (2011)
22. Magnin, B., et al.: Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology **51**, 73–83 (2009)
23. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
24. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. NeuroImage **62**(2), 782–790 (2012)
25. Pisner, D.A., Schnyer, D.M.: Support vector machine. In: Machine Learning, pp. 101–121. Elsevier (2020)
26. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)

27. Chen, S., Ma, K., Zheng, Y.: Med3D: transfer learning for 3D medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
28. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
29. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114. PMLR (2019)
30. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
31. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Trans. Biomed. Eng. **69**(3), 1173–1185 (2021)
32. Guan, H., Liu, M.: DomainATM: domain adaptation toolbox for medical data analysis. NeuroImage **268**, 119863 (2023)
33. Zhou, Z., Sodha, V., Pang, J., Gotway, M.B., Liang, J.: Models genesis. Med. Image Anal. **67**, 101840 (2021)
34. Zhou, H.Y., Lu, C., Yang, S., Han, X., Yu, Y.: Preservational learning improves self-supervised medical image models by reconstructing diverse contexts. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3499–3509 (2021)