Cost-Sensitive Meta-learning for Progress Prediction of Subjective Cognitive Decline with Brain Structural MRI

Hao Guan[1], Yunbi Liu[1,2], Shifu Xiao[3], Ling Yue[3(✉)], and Mingxia Liu[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mxliu@med.unc.edu

[2] School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

[3] Department of Geriatric Psychiatry, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200030, China

Abstract. Subjective cognitive decline (SCD) is a preclinical phase of Alzheimer's disease (AD) which occurs before the deficits could be detected by cognitive tests. It is highly desired to predict the progress of SCD for possible intervention of AD-related cognitive decline. Many neuroimaging-based methods have been developed for AD diagnosis, but there are few studies devoted to automated progress prediction of SCD due to the limited number of SCD subjects. Even though some studies proposed to transfer models (trained on AD/MCI) to SCD analysis, the significant domain shift between their data distributions may degrade the prediction performance. To this end, this paper tackles the problem of learning a model from the source data for which can directly generalize to an unseen target domain for SCD prediction. We propose a cost-sensitive meta-learning scheme to simultaneously improve the model generalization and its sensitivity in MRI-based SCD detection. During training, the source domain is divided into virtual meta-train and meta-test sets to explicitly simulate the scenario for early-stage detection of AD. Considering the importance of sensitivity for progressive status detection, we further introduce cost-sensitive learning to enhance the meta-optimization process by encouraging the model to gain higher sensitivity for SCD detection with simulated domain shift. Experiments conducted on the large-scale ADNI dataset and a small-scale SCD dataset have demonstrated the effectiveness of the proposed method.

Keywords: Subjective cognitive decline · Meta-learning · Brain MRI

1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease that usually starts slowly and worsens over time. As illustrated in Fig. 1, the spectrum of AD has been extended to an earlier stage even before its prodromal stage

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12905, pp. 248–258, 2021.
https://doi.org/10.1007/978-3-030-87240-3_24

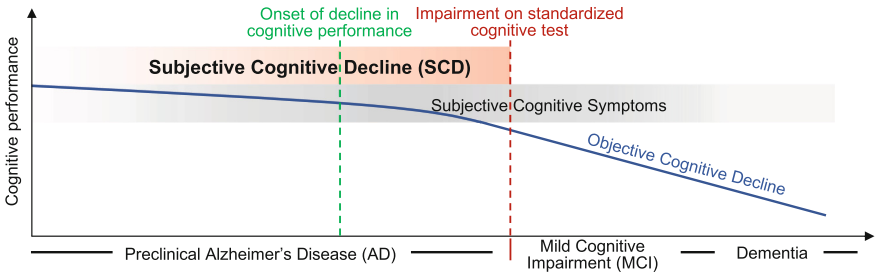


Fig. 1. Progress of disease pathology and clinical status of AD.

(*i.e.*, mild cognitive impairment, MCI), called subjective cognitive decline (SCD) or subjective memory complaint (SMC) that occurs before the deficits could be detected by cognitive tests [1,2]. In the literature, increasing evidence has shown that subjects with SCD have an increased risk of underlying AD pathology [3,4]. Therefore, predicting the future progress of subjects with SCD is fundamental for possible intervention of AD-related cognitive declines.

Neuroimaging-based methods have been widely used for AD/MCI diagnosis [5–8]. But only a few studies are devoted to SCD progress prediction, due to several challenges. (**i**) The number of SCD subjects is usually very limited (*e.g.*, tens), making it difficult to train a robust model with good generalization ability. Previous studies [9,10] proposed to augment data samples for model training, but usually failed to deal with domain shift between different domains/datasets. (**ii**) Even though some domain adaptation methods were designed to enhance the transferability of a learning model [11–16], they often require a part of labeled/unlabeled target samples to facilitate adaptation which cannot be satisfied in real applications. (**iii**) SCD appears at the preclinical stage of AD even without significant objective impairment in the brain. These challenges make it difficult to design a robust model for reliable SCD progress prediction.

To address these issues, we propose a cost-sensitive meta-learning (CSML) framework for structural MRI-based progress prediction of SCD, as illustrated in Fig. 2. During training, the source domain is divided into virtual meta-train and meta-test sets to explicitly simulate the scenario for early-stage detection of AD (*e.g.*, AD+MCI→ SCD). Considering the importance of sensitivity for detecting progressive status, we further introduce a cost-sensitive learning technique to enhance the meta-optimization process, by encouraging the model to gain higher sensitivity for progressive SCD (pSCD) detection with simulated domain shift. To the best of our knowledge, this is among the first attempts to perform SCD detection by explicitly considering cross-site domain shift and requiring no target data for model training. The proposed CSML is expected to simultaneously improve the model generalization and sensitivity in SCD progress prediction. Experimental results on a small-scale SCD dataset and the large-scale ADNI dataset demonstrate the effectiveness of the proposed method.

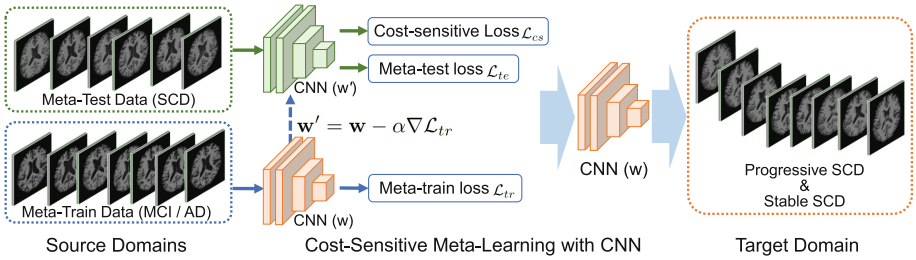


Fig. 2. Illustration of the cost-sensitive meta-learning (CSML) framework for SCD progress prediction with brain MRIs. CNN: Convolutional neural network.

2 Materials and Methodology

Datasets and MR Image Pre-processing. A total of 1,393 T1-weighted structural MRIs from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset[1] are used in this work to train a prediction model. Specifically, the ADNI contains 367 AD, 357 normal control (NC), 653 MCI subjects (253 pMCI, 400 sMCI), and 16 SCD subjects (11 pSCD and 5 sSCD). Among MCI subjects, 253 progressive MCI (pMCI) subjects would convert to AD within 36 months after baseline time, and the remaining 400 stable MCI (sMCI) would not. Also, 11 progressive SCD (pSCD) would convert to MCI within 36 months and the status of the 5 SCD subjects keep stable (called stable SCD, sSCD). We also collected 113 SCD subjects (with T1-weighted structural MRIs) from a local hospital as the target domain to evaluate the performance of the proposed model, including 40 pSCD subjects and 73 sSCD subjects. Note that we only used baseline MRIs in this work, without using longitudinal MRIs.

Following [17], all brain MRIs go through a standard pre-processing pipeline, including (i) skull stripping, (ii) intensity correction, (iii) re-sampling to the same resolution of $1 \times 1 \times 1$ mm^3 and (iv) spatial normalization to the Automated Anatomical Labeling (AAL) template. We employ the SPM software package[2] as the main tool to facilitate the MR image pre-processing.

Problem Formulation. In this work, we study the problem of learning a model based on N source domains, and apply it to precisely predict the future progress of SCD subjects in an unseen target domain. We treat different brain disorder categories (*e.g.*, AD, MCI, SCD) as different domains in this paper. Suppose there are N labeled source domains (*i.e.*, $\mathcal{D}_S = \{\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_N\}$), as well as a to-be-analyzed target domain (*i.e.*, $\mathcal{D}_T$). All source and target domains share the same input MR image space $\mathcal{X}$ (with possibly different marginal distributions), but have different category labels. Our aim is to *utilize samples from source domains to train a convolutional neural network (CNN) that can be well*

[1] https://ida.loni.usc.edu.
[2] https://www.fil.ion.ucl.ac.uk/spm.

generalized to an unseen target domain with high sensitivity to SCD progress prediction.

Meta-learning Scheme. Since no target samples (either labeled or unlabeled ones) are available for model training or fine-tuning, we develop a meta-learning scheme to simulate the real-world scenario for cross-domain classification [18–20]. Different from conventional meta-learning schemes that only simulate domain shift with the assumption that the source and target domains share the same class labels, we also simulate domain shift by using different but related subjects (*e.g.*, MCI) for SCD prediction in the proposed meta-learning strategy.

As illustrated in Fig. 2, the source domains $\mathcal{D}_S$ are divided into a *virtual meta-train set* and a *virtual meta-test set*. To simulate SCD prediction on an independent target domain, the virtual test set consists of SCD samples whereas the virtual training set involves other types of samples (*e.g.*, MCI). Denote $\mathbf{w}$ as the parameters (weights) of a CNN model. In the meta-learning process, $\mathbf{w}$ goes through a two-step update and optimization. In the *first* step, a meta-train classification loss $\mathcal{L}_{tr} = \mathcal{F}(\mathcal{D}_{tr}; \mathbf{w})$ is computed using samples in the meta-train set, and the gradient with respect to $\mathbf{w}$ can be calculated and updated as:

$$\mathbf{w}' = \mathbf{w} - \lambda \nabla \mathcal{L}_{tr}(\mathcal{D}_{tr}; \mathbf{w}), \tag{1}$$

where λ is an update rate for the gradient of the CNN model. In the *second* step, the model with $\mathbf{w}'$ is applied to the meta-test set, yielding a meta-test loss $\mathcal{L}_{te} = \mathcal{G}(\mathcal{D}_{te}; \mathbf{w}')$. In Fig. 2, the CNNs in different colors merely indicate different update stages. By jointly optimizing the meta-test and meta-train losses, the CNN is expected to avoid bias towards the training set and learn some domain-invariant features that can be generalized to unseen target domains.

Cost-Sensitive Learning. Previous learning models for disease diagnosis or domain adaptation often treat misclassification errors equally for both positive or negative samples. Actually, incorrectly predicting a positive subject (unhealthy) as a negative case (healthy) is much more costly, because it will make subjects miss the best time for early intervention and treatment. To this end, we employ cost-sensitive learning [21–23] strategy to reduce the false negative rate and increase the sensitivity of our prediction model to possible patients. Specifically, we introduce a cost-sensitive loss $\mathcal{L}_{cs}$ into the meta-learning process of the proposed CNN. We impose this loss on the meta-test set to encourage the model to pursue high prediction accuracy of progressive SCD on unseen target domain.

Let $y \in \{0, 1\}$ denote the ground-truth label for a subject, where 0 represents sSCD and 1 is for pSCD, and the corresponding prediction by the model is $\hat{y}$. We design a cost matrix as follows:

$$\mathbf{M} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{ij} indicates the cost for a model when it predicts a subject belonging to the i-th category into the j-th category. For a specific prediction result $\hat{y}$, the

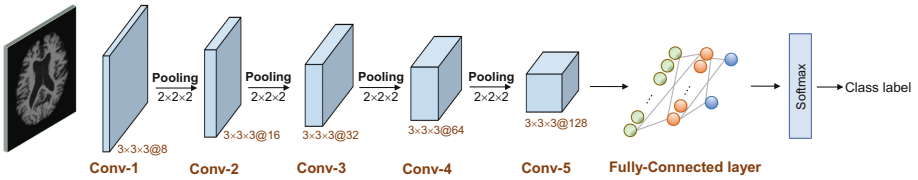


Fig. 3. Structure of the backbone CNN for MRI-based classification. The term $3 \times 3 \times 3@N$ indicates N convolution filters with the kernel size of $3 \times 3 \times 3$.

cost-sensitive loss can be calculated by the inner product of $\hat{y}$ and the row of $\mathbf{M}$ corresponding to its label y, *i.e.*, $\mathcal{L}_{cs} = \langle \hat{y}, M(y+1, \cdot) \rangle$. For example, for a pSCD subject ($y = 1$), the prediction of the learning model is: $\hat{y} = [1, 0]$. Then, the cost is $\langle [1, 0], [C_{21}, C_{22}] \rangle = C_{21}$. In this work, we empirically set C_{21} and C_{12} in $\mathbf{M}$ to η (η=2) and 1, respectively, whereas the other elements to 0.

Based on meta-learning and cost-sensitive learning schemes, the total loss for the CSML can be formulated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{tr} + \alpha \mathcal{L}_{te} + \beta \mathcal{L}_{cs}, \tag{2}$$

where α and β are parameters to control contributions of three losses. Our model is optimized via a two-step update strategy. First, the CNN is trained with the meta-train loss on meta-train set, and the weight $\mathbf{w}$ is updated according to Eq. (1). Then, the updated CNN is fed with the meta-test set, and the meta-test loss and cost-sensitive loss are utilized to further update the weight. The training on the meta-test set can be understood as a regularization term which enables the network to gain strong generalizability.

Network Architecture. We utilize an AlexNet-like 3D CNN as the backbone of our CSML model, with its architecture illustrated in Fig. 3. Specifically, this CNN contains 5 convolution layers and 3 fully-connected layers (with 128, 64 and 2 units, respectively). Each convolution layer consists of a 3D convolution filter (kernel size: $3 \times 3 \times 3$), batch normalization operation and ReLU activation function. We also add a $2 \times 2 \times 2$ max-pooling operation (stride: $2 \times 2 \times 2$) after each convolution layer. A softmax layer is used as the classification layer to outputs the probability of an input MRI belonging to each category.

Implementation Details. We implement the proposed CSML model using PyTorch[3]. The Adam algorithm is employed as the optimizer with a learning rate of 0.0001. The cross-entropy loss is used as the loss function for both $\mathcal{L}_{tr}$ and $\mathcal{L}_{te}$ for classification. Dropout with a probability of 0.5 is used during the training process to avoid overfitting. The gradient update rate λ in Eq. (1) is set to 0.0001. For simplicity, the parameters α and β in Eq. (2) are set to 1. The network is trained for 50 epochs with the batch size of 2.

[3] https://pytorch.org.

Table 1. Results of SCD progress prediction achieved by five methods.

Method	AUC (%)	ACC (%)	BAC (%)	SEN (%)	SPE (%)
Baseline-1	56.66 ± 0.87	56.63 ± 1.05	45.92 ± 1.52	21.50 ± 3.45	70.34 ± 1.27
Baseline-2	57.26 ± 1.03	59.12 ± 2.12	53.23 ± 1.66	35.00 ± 4.26	$\mathbf{71.45 \pm 2.81}$
ROI+SVM	58.18 ± 0.21	54.66 ± 0.01	52.07 ± 0.01	46.50 ± 0.02	61.64 ± 0.01
VoxCNN	56.30 ± 1.05	58.75 ± 2.81	53.42 ± 1.96	36.50 ± 4.58	69.67 ± 2.54
CSML (ours)	$\mathbf{65.36 \pm 0.63}$	$\mathbf{59.66 \pm 2.53}$	$\mathbf{60.51 \pm 1.65}$	$\mathbf{66.40 \pm 3.01}$	54.62 ± 1.96

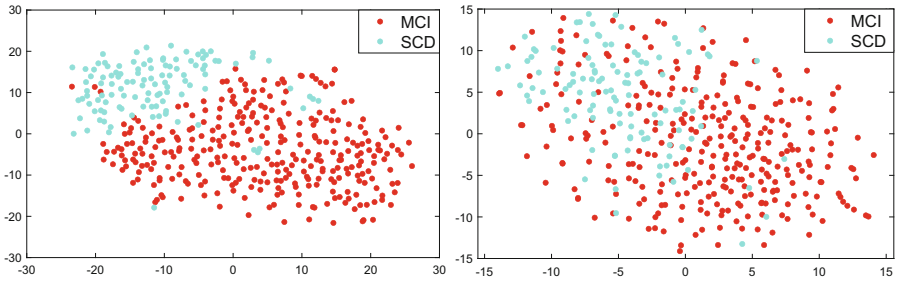
3 Experiment

Experimental Setup. Five metrics are used for performance evaluation, *i.e.*, area under receiver operating characteristic curve (AUC), classification accuracy (ACC), balanced accuracy (BAC), sensitivity (SEN) and specificity (SPE).

In the experiments, we use AD/NC, MCI, SCD samples in the ADNI as three source domains, denoted as $\mathcal{A}$ (with AD and NC), $\mathcal{M}$ (with pMCI and sMCI) and $\mathcal{S}$ (with pSCD and sSCD), respectively. The proposed CSML model is first trained on these three source domains, and then applied to the target domain (with 113 SCD subjects collected from a local hospital). In the training phase, we use the domain $\mathcal{S}$ as the meta-test set. And we treat the domain $\mathcal{M}$ as the meta-train set, considering the closer relationship between SCD and MCI when compared with SCD and AD (see Fig. 1).

Competing Methods. We compare our CSML with four methods, including (i) **Baseline-1** that trains the backbone CNN with AD and NC samples and directly transfer it to the target domain; (ii) **Baseline-2** that trains the backbone CNN with pMCI and sMCI samples and directly transfers it to the target domain; (iii) **ROI+SVM** that uses pSCD+MCI subjects as positive samples and sSCD+NC as negative samples. The normalized gray matter volumes of 90 ROIs in AAL are used as MRI features to train a linear support vector machine (SVM) which is transferred to the target domain; and (iv) **VoxCNN** that is a state-of-the-art deep learning method deliberately designed for dementia classification [24] and contains 10 convolution layers and two fully-connected layers.

Prediction Results. The results of SCD progress prediction of different methods are shown in Table 1. From this table, we can derive the following observations. (**i**) Compared with four competing methods, our CSML can produce the overall best performance, especially with the highest sensitivity. This implies that the proposed meta-learning and cost-sensitive learning strategies in CSML help boost the prediction performance. (**ii**) Two baseline CNNs (i.e., Baseline-1 and Baseline-2) are weak in SCD progress prediction with a bias towards negative samples (i.e., stable SCD). This indicates that simply using AD or MCI samples to train a deep network and applying it to SCD analysis cannot achieve satisfactory performance. (**iii**) Baseline-2 that uses MCI samples for training can



(a) Features from Baseline-2 (b) Features from the CSML model

Fig. 4. Visualization of feature distributions of MCI (source domain) and SCD subjects (target domain) from the baseline CNN and the proposed CSML model.

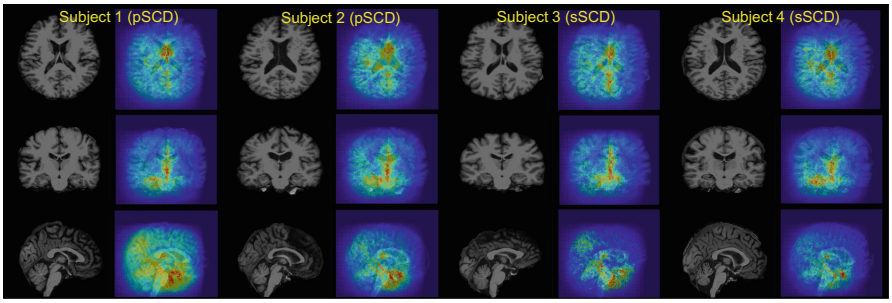


Fig. 5. Saliency maps generated by the proposed CSML model.

achieve better performance than Baseline-1 trained on AD/NC samples. This implies that a network trained on MCI conversion prediction can be reliably applied to SCD progress prediction, compared with that trained on AD/NC subjects. This may be because SCD is semantically closer to MCI in the AD spectrum.

Visualization Analysis. In Fig. 4, we visualize the feature distributions of MCI subjects in the source domain and SCD subjects in the target domain via t-SNE [25], where MRI features are generated from Baseline-2 and CSML, respectively. From Fig. 4(a), we can see that the feature distributions of MCI and SCD subjects generated from Baseline-2 (without meta-learning) have significant difference. As shown in Fig. 4(b), after the meta-training through our CSML model, the distribution gap between these two domains has been greatly reduced. This may indicate that our CSML can learn some domain invariant features, which can partly explain its superiority in cross-domain prediction.

We further visualize the saliency maps [26] generated by the proposed CSML for four SCD subjects in the target domain, as shown in Fig. 5. We compute the gradients of the output of the network w.r.t. each input MRI. Figure 5 shows

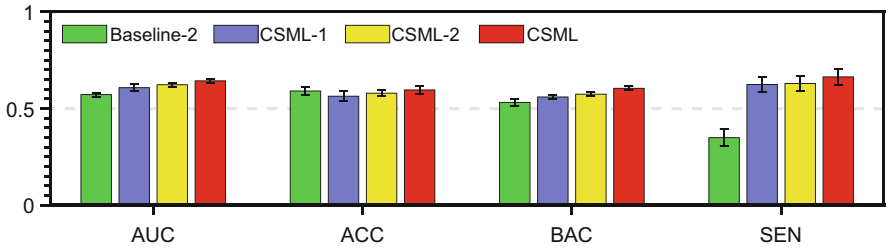


Fig. 6. Comparison between CSML and its variants in SCD progress prediction.

Table 2. Comparison with state-of-the-art methods in SCD progress prediction.

Method	Model	AUC (%)	ACC (%)	SEN (%)	SPE (%)
Yue *et al.* [27]	Cost-Sensitive SVM	59.8	56.6	55.0	**57.5**
Felpete *et al.* [28]	Random Forest	58.2	52.2	50.0	53.4
Liu *et al.* [9]	GAN	60.2	57.5	**70.0**	50.7
CSML (Ours)	Cost-Sensitive CNN	**65.0**	**59.3**	67.5	54.8

that even if it is not very prominent in the saliency maps, SCD may be related to several brain regions, such as the ventricle and hippocampus.

Ablation Study. Compared with baseline CNNs that are trained on the source domains and directly transferred to the target domain, our CSML model consists of meta-learning and cost-sensitive learning on virtual meta-train and meta-test data. To study their effects, we compare CSML with its two variants, including (i) **CSML-1** that uses $\mathcal{A}$ as the meta-train data without cost-sensitive learning, and (ii) **CSML-2** that uses $\mathcal{M}$ as the meta-train data without cost-sensitive learning. These two variants are compared with CSML and the Baseline-2 (trained on $\mathcal{M}$ without meta-learning), with results shown in Fig. 6. It can be seen from Fig. 6 that three meta-learning based methods (i.e., CSML, CSML-1 and CSML-2) generally outperform the Baseline-2 without meta-learning. Besides, our CSML achieves better SEN performance, compared with the three competing methods without cost-sensitive learning. We evaluate the results of CSML and its 2 variants (i.e., CSML-1 and CSML-2) through pair-wise t-test. The p-value for results of CSML vs. CSML-1 is 0.029, while that for CSML vs. CSML-2 is 0.033. This suggests that there is significant difference ($p < 0.05$) between CSML and CSML-1/CSML-2. These results verify the efficacy of the proposed meta-learning and cost-sensitive learning strategies. Also, CSML-2 (with MCI as meta-train data) usually outperforms CSML-1 (with AD/NC as meta-train data), which is consistent with the results of two baseline CNNs in Table 1.

Comparison with State-of-the-Arts. We further compare the proposed method with several state-of-the-art (SOTA) methods for MRI-based SCD

progress prediction, including 1) Cost-Sensitive SVM (CSVM) [27], 2) Random Forest (RF) [28] and 3) Generative Adversarial Network (GAN) [9]. We reproduced these algorithms and performed classification experiments using the same data set as this work for a fair comparison. The CSVM is implemented by an SVM with linear kernel trained with the same cost matrix adopted in our CSML. The RF is implemented by an ensemble of 100 bagged classification trees. The GAN is built with a generator (3 convolution layers, 3 residual blocks and 2 deconvolution layers) and a discriminator (5 convolution layers). The prediction results of four different methods are reported in Table 2. It can be seen from Table 2 that, compared with the three SOTA methods, our CSML can produce better AUC and ACC results and comparable performance in terms of SEN and SPE. This may be attributed to the meta-learning scheme that can enhance the generalization ability of the prediction model and the cost-sensitive learning scheme that can produce competing prediction accuracy for both positive and negative samples.

4 Conclusion

In this paper, we present a cost-sensitive meta-learning (CSML) framework for SCD progress prediction based on brain structural MRIs. We train the CSML model on a relatively large-scale source domain and apply it to an unseen target domain with small-scale data for cross-domain SCD progress prediction. Specifically, the source domain (training set) is divided into a virtual meta-train set and a virtual meta-test set to simulate the domain shift among different domains. We also introduce cost-sensitive learning into the meta-training process which can further enhance the sensitivity for identifying progressive SCD subjects. Experimental results on the public ADNI database (with 1,393 subjects) and a private SCD dataset (with 113 subjects) demonstrate the effectiveness of the proposed method over previous state-of-the-arts. As a future work, we intend to leverage multi-modality data (such as MRI, PET and demographic information) for model training to further improve the performance of SCD progress prediction.

Acknowledgements. H. Guan and M. Liu were partly supported by NIH grant (No. AG041721).

References

1. Jessen, F., et al.: A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimer's Dementia **10**(6), 844–852 (2014)
2. Rabin, L.A., Smart, C.M., Amariglio, R.E.: Subjective cognitive decline in preclinical Alzheimer's disease. Annu. Rev. Clin. Psychol. **13**, 369–396 (2017)
3. Amariglio, R.E., et al.: Subjective cognitive complaints and amyloid burden in cognitively normal older individuals. Neuropsychologia **50**(12), 2880–2886 (2012)
4. Jessen, F., et al.: AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. Alzheimer's Dementia **10**(1), 76–83 (2014)

5. Jack, C.R., Jr., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Resonan. Imaging Official J. Int. Soc. Magn. Resonan. Med. **27**(4), 685–691 (2008)
6. Liu, M., Zhang, J., Adeli, E., Shen, D.: Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. **43**, 157–168 (2018)
7. Wang, M., Zhang, D., Huang, J., Yap, P.T., Shen, D., Liu, M.: Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation. IEEE Trans. Med. Imaging **39**(3), 644–655 (2019)
8. Yao, D., Calhoun, V.D., Fu, Z., Du, Y., Sui, J.: An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. J. Neurosci. Methods **302**, 75–81 (2018)
9. Liu, Y., et al.: Joint neuroimage synthesis and representation learning for conversion prediction of subjective cognitive decline. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 583–592. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_57
10. Cheng, N., et al.: Self-weighted multi-task learning for subjective cognitive decline diagnosis. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12267, pp. 104–113. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59728-3_11
11. Wachinger, C., Reuter, M.: Domain adaptation for Alzheimer's disease diagnostics. NeuroImage **139**, 470–479 (2016)
12. Cheng, B., Liu, M., Zhang, D., Munsell, B.C., Shen, D.: Domain transfer learning for MCI conversion prediction. IEEE Trans. Biomed. Eng. **62**(7), 1805–1817 (2015)
13. Hosseini-Asl, E., Keynton, R., El-Baz, A.: Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 126–130. IEEE (2016)
14. Ghafoorian, M., et al.: Transfer learning for domain adaptation in MRI: application in brain lesion segmentation. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 516–524. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_59
15. Orbes-Arteaga, M., et al.: Multi-domain adaptation in brain MRI through paired consistency and adversarial learning. In: Wang, Q., et al. (eds.) DART/MIL3ID -2019. LNCS, vol. 11795, pp. 54–62. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33391-1_7
16. Li, W., Zhao, Y., Chen, X., Xiao, Y., Qin, Y.: Detecting Alzheimer's disease on small dataset: a knowledge transfer perspective. IEEE J. Biomed. Health Inform. **23**(3), 1234–1242 (2018)
17. Guan, H., Liu, Y., Yang, E., Yap, P.T., Shen, D., Liu, M.: Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification. Med. Image Anal. **71**, 102076 (2021)
18. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning, pp. 1126–1135. PMLR (2017)
19. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Learning to generalize: Meta-learning for domain generalization. In: Proceedings of the AAAI Conference on Artificial Intelligence (2018)
20. Andrychowicz, M., et al.: Learning to learn by gradient descent by gradient descent. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 3981–3989 (2016)
21. Ling, C.X., Sheng, V.S.: Cost-sensitive learning and the class imbalance problem. Encyclopedia Mach. Learn. **2008**, 231–235 (2011)

22. Kuo, W., Häne, C., Yuh, E., Mukherjee, P., Malik, J.: Cost-sensitive active learning for intracranial hemorrhage detection. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 715–723. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_82
23. Galdran, A., Dolz, J., Chakor, H., Lombaert, H., Ben Ayed, I.: Cost-sensitive regularization for diabetic retinopathy grading from eye fundus images. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 665–674. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_64
24. Korolev, S., Safiullin, A., Belyaev, M., Dodonova, Y.: Residual and plain convolutional neural networks for 3D brain MRI classification. In: ISBI, pp. 835–838 (2017)
25. Maaten, L.V.D., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11), 2579–2605 (2008)
26. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: visualising image classification models and saliency maps. In: ICLR (2014)
27. Yue, L., et al.: Prediction of 7-year's conversion from subjective cognitive decline to mild cognitive impairment. Hum. Brain Mapp. **42**(1), 192–203 (2021)
28. Felpete, A., et al.: Predicting progression in subjective cognitive decline (SCD) using a machine learning (ML) approach: the role of the complaint's severity. Alzheimer's Dementia **16**, e043492 (2020)