URLCF: Universal Representation Learning for Cross-Domain Few-Shot Autism Spectrum Disorder Detection

Zhenlin Mao[1], Qiyu Sun[1], Zhuo Wan[1], Jiashuang Huang[2], and Mingliang Wang[1,⋆]

[1] School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2] School of Artificial Intelligence and Computer Science, Nantong University, Nantong 226019, China
wml489@nuist.edu.cn

Abstract. Developing generalizable predictive models for brain disease detection critically depends on the availability of a large amount of fMRI data. To enable large-scale analysis, it is often necessary to aggregate data from multiple acquisition sites/domains. However, differences in scanners and acquisition protocols across sites lead to unfavorable heterogeneity in data distribution, which hinders reliable biomarker identification and undermines the generalizability of predictive models. In this paper, we propose URLCF, a novel few-shot learning framework that learns a set of universal representations by distilling knowledge from multiple separately trained models for cross-domain autism detection. Specifically, we first train multiple domain-specific networks to serve as teacher models for guiding the universal model. Subsequently, domain-specific data are simultaneously fed into both the universal and teacher models. By randomly masking the output features of the universal model and employing domain-specific generators to reconstruct these masked features, we align the learned representations of the universal model with those of the teachers. Concurrently, a domain discriminator is introduced to perform domain alignment, jointly facilitating the learning of universal representations. Finally, few-shot fine-tuning is conducted on the unseen domain to enable rapid adaptation to the unseen distribution. Experimental results on the ABIDE dataset demonstrate that URLCF achieves superior performance in autism detection under both zero-shot and five-shot settings, confirming its effectiveness in generalizing to unseen domains with limited data. The code is available at https://github.com/dwd111/URLCF.

Keywords: Universal Representation Learning · Few-Shot Learning · Cross-Domain Disease Detection · Autism Spectrum Disorder.

⋆ Corresponding author

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and atypical behavior patterns [1]. It is a multifaceted condition that profoundly impacts individual quality of life and broader societal development. Therefore, precise diagnosis of ASD is crucial for effective early intervention and societal care. Resting-state functional magnetic resonance imaging (rs-fMRI) captures spontaneous blood-oxygen-level-dependent (BOLD) fluctuations in a non-invasive and task-free manner, emerging as a crucial adjunct for the diagnosis of ASD. Functional connectivity (FC), derived from temporal correlations of BOLD signals [2,15], has been extensively investigated for the identification of various brain disorders.

Deep learning (DL) has recently shown great potential for FC analysis, owing to its capacity to learn informative feature representations. Despite these advances, DL models heavily rely on large amounts of labeled training data to achieve better performance. However, the scarcity of disease-related fMRI data at a single site limits the performance and generalizability of these models. Aggregating data from multiple sites is therefore commonly considered a solution to alleviate data insufficiency. Unfortunately, the data collected across multiple sites introduces substantial inter-site heterogeneity arising from variations in scanners and acquisition protocols, which hinders model performance and cross-domain generalization [19,21].

To mitigate the multi-site heterogeneity, a common approach is the ComBat harmonization method [20], which aims to eliminate site-specific differences from the data. However, since data needs to be reprocessed every time a new site is added, extra effort is required when dealing with unseen sites. Another approach is the domain adaptation (DA) method [3,12], which addresses the data heterogeneity problem by transferring knowledge from the source domain to the target domain during training. Nevertheless, this reliance on target domain data during training limits its generalization performance on unseen domains.

In this paper, we propose URLCF, a novel few-shot learning framework that learns a set of universal representations by distilling knowledge from multiple separately trained models for cross-domain ASD detection. Specifically, we first train multiple domain-specific networks as teacher models to guide the learning of the universal student model. Subsequently, to align the representations of the student model with those of the teacher models, domain-specific data are simultaneously fed into both models. This alignment is achieved by randomly masking the student model's output features and employing domain-specific generators to reconstruct the masked features. Meanwhile, we introduce a domain discriminator to perform domain alignment, thereby collectively promoting the learning of universal representations. Finally, we freeze the trained student model and integrate a lightweight adaptive module for few-shot fine-tuning, enabling rapid adaptation to unseen distributions. Our contributions are summarized as follows:

1) We propose URLCF, a novel few-shot learning framework that distills knowledge from multiple domain-specific teachers into a universal student model,

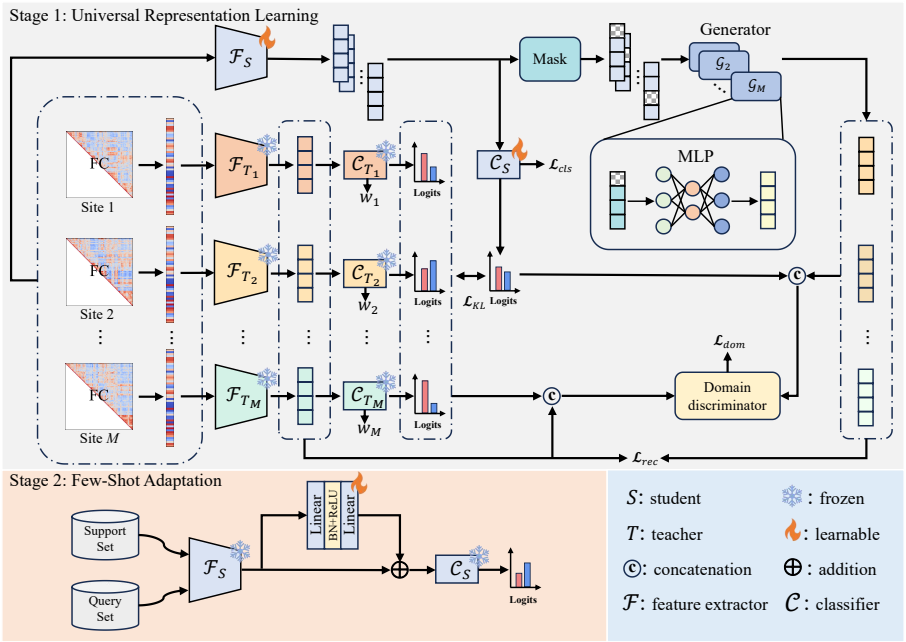


Fig. 1. The framework of our proposed URLCF.

thereby alleviating multi-site fMRI heterogeneity for cross-domain ASD detection.

2) We introduce an alignment strategy that integrates masked generation with domain discrimination to promote the learning of universal representations. Moreover, a lightweight adaptive module is employed to enable rapid few-shot fine-tuning (e.g., 5-shot) to unseen domains.
3) Extensive experiments on the ABIDE dataset demonstrate that our URLCF consistently outperforms state-of-the-art methods in cross-domain ASD detection under both zero-shot and five-shot settings.

2 Method

Given a multi-site dataset $\{D_i\}_{i=1}^{N}$ with $D_i = \{(\mathbf{x}_j^{(i)}, y_j^{(i)})\}_{j=1}^{P_i}$, where P_i denotes the sample size of the i-th site, and N denotes the number of sites. Our objective is to learn a mapping function $f : \mathcal{X} \to \mathcal{Y}$ that accurately predicts the diagnostic label $y \in \mathcal{Y}$ (e.g., ASD or NC) from the input FC feature $\mathbf{x} \in \mathcal{X}$. As shown in Fig. 1, our framework consists of two stages: **1)** We learn a universal student model by distilling knowledge from multiple separately trained models. **2)** We freeze the learned student model, add a trainable lightweight module, and fine-tune it with a few samples to rapidly adapt to the target domain distribution.

2.1 Domain-Specific Networks Pre-Training

To leverage domain-specific knowledge, we first train N domain-specific networks as teacher models, denoted as $\{T_i\}_{i=1}^{N}$, on their respective domains. Each teacher T_i consists of a feature extractor $\mathcal{F}_{T_i}$ and a classifier $\mathcal{C}_{T_i}$. The validation accuracy of each teacher classifier is adopted to determine its confidence weight w_i during distillation, formulated as follows:

$$w_i = \frac{\exp(s_i/\tau)}{\sum_{j=1}^{M} \exp(s_j/\tau)}, \tag{1}$$

where τ is a parameter to control the smoothness of the weight distribution (e.g., $\tau = 0.9$), and M is the number of teacher models in the distillation stage.

2.2 Learning Universal Representation

Given M training datasets $\{D_i\}_{i=1}^{M}$, known as source domains $\mathcal{D}_b$. In this phase, we freeze the teacher models $\{T_i\}_{i=1}^{M}$ and distill their knowledge into the student model S during training. Specifically, the student model S consists of a feature extractor $\mathcal{F}_S$ and a general classifier $\mathcal{C}_S$. As illustrated in Fig. 1, we feed domain-specific data into $\{T_i\}_{i=1}^{M}$ and S simultaneously. Then, we employ Kullback-Leibler (KL) divergence to obtain prediction logits based on the logits of the teacher models Z_{T_i}, supplemented by cross-entropy classification loss based on the true labels. In addition, we also integrate a mask generative strategy for feature alignment and a conditional domain discriminator for domain alignment to jointly promote the learning of universal representations.

Masked Generative Alignment. Let $\{\mathbf{f}_S^{(i)}\}_{i=1}^{M} = \{\mathbf{f}_S^{(1)}, \mathbf{f}_S^{(2)}, \ldots, \mathbf{f}_S^{(M)}\}$ denote the set of feature batches produced by the student model, where $\mathbf{f}_S^{(i)}$ represents the batch of feature vectors corresponding to the site D_i. We generate masked representations $\tilde{\mathbf{f}}_S^{(i)}$ by applying a binarized mask $\mathbf{M}$ to the student features $\mathbf{f}_S^{(i)}$. The process of mask definition and feature masking is formulated as:

$$\mathbf{M} = \begin{cases} 0, & \text{if } \mathbf{R} \leq \rho \\ 1, & \text{otherwise} \end{cases}, \qquad \tilde{\mathbf{f}}_S^{(i)} = \mathbf{f}_S^{(i)} \odot \mathbf{M}, \tag{2}$$

where $\mathbf{R}$ is a random matrix with elements in the range $(0, 1)$ and the same dimensionality as $\mathbf{f}_S^{(i)}$, ρ is a hyperparameter controlling the masking ratio (e.g., 0.25), and $\odot$ denotes the element-wise multiplication.

Subsequently, given the domain-specific generators $\{\mathcal{G}_i\}_{i=1}^{M}$, the masked features are reconstructed using their respective generator $\mathcal{G}_i$. Specifically, we employ $\mathcal{G}_i$ to project the incomplete student features $\tilde{\mathbf{f}}_S^{(i)}$ into the teacher's feature space, thereby aligning the reconstructed representations with the teacher features. Notably, the parameters of these generators are optimized jointly with the student model. The reconstruction process is formulated as:

$$\hat{\mathbf{f}}_S^{(i)} = \mathcal{G}_i(\tilde{\mathbf{f}}_S^{(i)}) = \mathbf{W}_2 \cdot Leaky\left(BN(\mathbf{W}_1 \cdot \tilde{\mathbf{f}}_S^{(i)})\right), \tag{3}$$

where $\mathcal{G}_i$ denotes the domain-specific generator of the i-th site, which includes two linear layers parameterized by $\mathbf{W}_1$ and $\mathbf{W}_2$, a Batch Normalization layer, and a LeakyReLU activation.

Conditional Domain Discriminator. To align the reconstructed representations with domain-specific distributions, we introduce a conditional domain discriminator D, implemented as a Multi-layer Perceptron (MLP). First, D is optimized on the frozen teachers' features and predictions to establish precise decision boundaries for each source domain. Subsequently, the features produced by the domain-specific generators are combined with the student predictions and predicted by D as site D_i, which can be formulated as:

$$y_d = D(\hat{\mathbf{f}}_S^{(i)} \| \ \mathbf{p}_S^{(i)}) = D\left(\hat{\mathbf{f}}_S^{(i)} \| \ \sigma(Z_S)\right), \tag{4}$$

where $\|$ denotes the concatenation operation, $\mathbf{p}_S^{(i)}$ is the class probability vector for the i-th site, $\sigma(\cdot)$ represents the softmax function, Z_S corresponds to the student's logits, and y_d indicates the domain scores generated by the discriminator.

2.3 Few-Shot Adaptation

A strong general-purpose feature extractor is expected to learn representations that generalize well across many previously unseen domains. Nevertheless, this task becomes significantly more demanding when a substantial domain gap exists (e.g., variations in scanner protocols) between the training set $\mathcal{D}_b$ and the test set $\mathcal{D}_t$, necessitating further adaptation to the target context. Each few-shot task in test set $\mathcal{D}_t$ consists of a support set $S_i = \{(\mathbf{x}_i, y_i)\}_{i=1}^{|\mathcal{S}|}$ with $|\mathcal{S}|$ sample and label pairs respectively and a query set $Q = \{(\mathbf{x}_j)\}_{j=1}^{|\mathcal{Q}|}$ with $|\mathcal{Q}|$ samples to be classified. We introduce a lightweight trainable module $\mathcal{R}_\psi$ with parameters ψ. We append $\mathcal{R}_\psi$ to the frozen $\mathcal{F}_S$ and optimize ψ using only the support set $\mathcal{S}_i$. The process is formulated as follows:

$$\mathbf{z_i} = \mathcal{F}_S(\mathbf{x}_i), \quad \hat{\mathbf{z}}_\mathbf{i} = \mathbf{z_i} \oplus \mathcal{R}_\psi(\mathbf{z_i}), \quad \mathbf{p}(y_i|\mathbf{x}_i) = \sigma\left(\mathcal{C}_S(\hat{\mathbf{z}}_\mathbf{i})\right), \tag{5}$$

where $\oplus$ denotes the element-wise addition, $\mathcal{R}_\psi$ consists of two linear layers with batch normalization and ReLU activation, initialized to perform an identity mapping at the start of adaptation and $\mathbf{p}(y_i|\mathbf{x}_i)$ represents the probability assigned to the ground-truth class y_i by the frozen general classifier.

2.4 Loss Function

Our framework is jointly optimized with four loss functions: (1) a cross-entropy loss $\mathcal{L}_{cls}$ to supervise ASD classification; (2) a mask reconstruction loss $\mathcal{L}_{rec}$ that enables teachers to improve the student's representational capacity through guided feature recovery; (3) a distillation loss $\mathcal{L}_{KL}$ to transfer prediction logits

from the teacher models; and (4) a domain alignment loss $\mathcal{L}_{dom}$ to mitigate domain-specific biases. These loss functions are defined as:

$$\mathcal{L}_{cls} = \sum_{i=1}^{M} w_i CE(\mathcal{C}_S(\mathbf{f}_S^{(i)}), y^{(i)}), \quad \mathcal{L}_{rec} = \sum_{i=1}^{M} w_i \left\| \hat{\mathbf{f}}_S^{(i)} - \mathcal{F}_{T_i}(\mathbf{x}^{(i)}) \right\|_2^2, \quad (6)$$

$$\mathcal{L}_{KL} = \sum_{i=1}^{M} w_i KL\left(\sigma\left(\frac{Z_{T_i}}{\tau}\right), \sigma\left(\frac{Z_S^{(i)}}{\tau}\right)\right), \quad \mathcal{L}_{dom} = \sum_{i=1}^{M} w_i CE\left(y_d, i\right), \quad (7)$$

where w_i denotes the confidence weight assigned to the i-th teacher, and $Z_S^{(i)}$ is the predictive logits produced by the student classifier corresponding to the input samples from the i-th site. Finally, the total loss is computed as:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{dom} + \beta \mathcal{L}_{KL} + \gamma \mathcal{L}_{res}, \quad (8)$$

where λ, β and γ are trade-off hyperparameters for balancing different losses.

For adaptation to an unseen domain, we optimize the lightweight module $\mathcal{R}_\psi$ by minimizing the cross-entropy loss over the support set S_i:

$$\mathcal{L}_{\text{adapt}}(\psi) = -\frac{1}{|S|} \sum_{i=1}^{|S|} \log \mathbf{p}(y_i|\mathbf{x}_i), \quad (9)$$

where $|S|$ is the number of samples in the support set.

3 Experiments

3.1 Experimental Settings

Datasets and Preprocessing. We evaluate URLCF on the ABIDE [5] dataset, which contains functional MRI data from 17 acquisition sites. For reproducibility, we downloaded the version provided by the Preprocessed Connectome Project. After preprocessing, we collected 1102 subjects, comprising 531 ASD patients and 571 normal controls (NCs). For brain region parcellation, we employed the Anatomical Automatic Labeling (AAL) atlas with 116 ROIs [18].

Baselines. The URLCF is compared with three types of baselines to verify its effectiveness, including: (1) Basic methods: GCN [9], BrainNetTF [11], and MLP [14]. (2) Few-shot learning methods: Prototypical Networks [16], Meta-Baseline [4], and MAML [7]. (3) Domain-invariant methods: DSR [3], DVAE [13], and ComBat [10].

Implementation Details. The models are trained on an NVIDIA 2080Ti GPU. We build our method on a 3-layer MLP backbone for both the teacher models and the student model. Each classifier is implemented as a single fully connected layer. During the universal representation learning phase, we use the Adam optimizer with a learning rate of 5e-4 and a weight decay of 5e-4. The batch size is 16 and the epoch is 500. For the subsequent few-shot adaptation stage, we use

Table 1. Experimental results of the comparison methods on the ABIDE dataset.

Type	Method	ACC(%)	SEN(%)	SPE(%)	AUC(%)
Basic	GCN	65.87	53.77	77.07	67.61
	BrainNetTF	66.72	57.97	74.45	69.62
	MLP	66.70	66.33	67.26	69.19
Few-shot learning	Prototypical Networks	58.71	57.59	60.23	61.07
	Meta-Baseline	59.16	55.19	62.43	62.35
	MAML	62.17	64.53	61.37	67.37
Domain-invariant	ComBat	68.76	69.09	67.33	69.57
	DSR	70.75	66.57	74.00	72.61
	DVAE	71.09	67.24	74.10	73.42
Ours	URLCF zero-shot	72.72	68.81	75.83	73.59
	URLCF five-shot	**75.48**	**71.93**	**78.20**	**75.99**

Table 2. Experimental results of the ablation study.

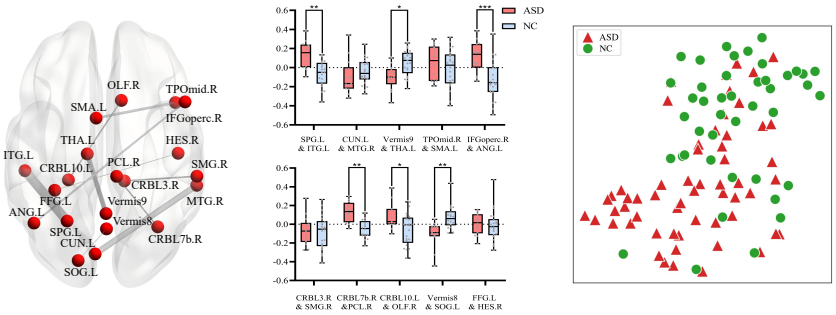
Setting	Method	ACC(%)	SEN(%)	SPE(%)	AUC(%)
zero-shot	URLCF w/o $M_{\mathcal{R}}$	70.76	67.64	72.99	72.96
	URLCF w/o D	71.13	**68.93**	72.73	71.99
	URLCF w/o w_i	71.42	68.86	73.86	72.75
	URLCF	**72.72**	68.81	**75.83**	**73.59**
five-shot	URLCF w/o $M_{\mathcal{R}}$	72.76	69.11	75.27	74.98
	URLCF w/o D	72.10	69.66	73.80	74.36
	URLCF w/o w_i	73.91	**72.03**	75.67	74.90
	URLCF	**75.48**	71.93	**78.20**	**75.99**

the Adadelta optimizer with a learning rate of 1e-3 and train the model for 10 iterations.

Evaluation Metrics. In this study, we employ the leave-one-site-out cross-validation strategy to assess the performance of different methods. Specifically, one site is used as the test domain for fine-tuning, while the remaining sites are used as the training domains for distillation. To evaluate the performance, four evaluation metrics are used: accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC).

3.2 Results Analysis

From Table 1, we observe that URLCF outperforms all baseline methods. Specifically, under the five-shot setting, the ACC of URLCF reaches 75.48%, which is 4.39% higher compared to the suboptimal method and 13.31% higher compared to the classic meta-learning method (MAML). Notably, even under the zero-shot setting, URLCF achieves an ACC of 72.72%, which still exceeds the domain adaptation method DVAE (71.09%), demonstrating exceptional gener-



(a) Top 10 brain FC patterns (b) Statistical analysis of these patterns (c) t-SNE visualization of features

Fig. 2. Visualization of the learned features by our URLCF.

alization ability to unseen domains. These significant improvements validate our core contributions: 1) The framework effectively captures universal representations from multiple source sites; 2) The few-shot adaptation effectively mitigates the domain bias, enabling precise diagnosis even with limited target data.

3.3 Ablation Study

Three variants of the URLCF model are constructed to assess the role of the key modules: URLCF w/o $M_{\mathcal{R}}$, removing the masked reconstruction for feature alignment; URLCF w/o D, removing the conditional domain discriminator; and URLCF w/o w_i, removing the confidence weights for the distillation stage.

As shown in Table 2, removing $M_{\mathcal{R}}$ significantly reduces performance, indicating that teacher-guided feature recovery is crucial for learning universal representations. Without D, the model's performance also decreases, suggesting that domain alignment effectively mitigates site heterogeneity. Moreover, removing w_i negatively impacts the model. This demonstrates the vital role of confidence weights in the distillation stage. The ablation experiments show that each module actively contributes to its overall performance.

3.4 Visualization Analysis

We analyze the top 10 selected FC patterns and conduct statistical comparisons between the ASD and NC groups using a standard t-test algorithm. From Fig. 2 (a), we observe that these key FCs are mainly distributed in the frontoparietal, temporal, and cerebellar regions, which is highly consistent with previous studies [6,17]. As shown in Fig. 2 (b), these FCs exhibit significant group differences between ASD and NC, confirming them as reliable biomarkers. In addition, the t-SNE visualization in Fig. 2 (c) demonstrates that the learned features effectively separate ASD and NC subjects into clear clusters. These results validate that our model captures highly interpretable biomarkers for ASD diagnosis [8].

4 Conclusion

This study presents URLCF, a novel few-shot learning framework designed to extract universal representations for cross-domain ASD detection. Our URLCF learns universal representations by distilling knowledge from multiple teacher models under masked generative alignment and performs lightweight few-shot adaptation on a frozen student network. Extensive experimental results on the ABIDE dataset demonstrate that URLCF outperforms existing methods in both zero-shot and five-shot settings, indicating that the learned universal representations effectively enhance cross-site generalization.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (Nos. 62472228 and 62471259), and the Natural Science Foundation of Jiangsu Province of China (No. BK20230422).

Disclosure of Interests. The authors have no competing interests to declare relevant to this article's content.

References

1. Anagnostou, E., Taylor, M.J.: Review of neuroimaging in autism spectrum disorders: what have we learned and where we go from here. Molecular Autism **2**(1), 4 (2011)
2. Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., et al.: Toward discovery science of human brain function. Proceedings of the National Academy of Sciences **107**(10), 4734–4739 (2010)
3. Cai, R., Li, Z., Wei, P., Qiao, J., Zhang, K., Hao, Z.: Learning disentangled semantic representation for domain adaptation. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. pp. 2060–2066 (2019)
4. Chen, Y., Liu, Z., Xu, H., Darrell, T., Wang, X.: Meta-baseline: Exploring simple meta-learning for few-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9062–9071 (2021)
5. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry **19**(6), 659–667 (2014)
6. D'Mello, A.M., Stoodley, C.J.: Cerebro-cerebellar circuits in autism spectrum disorder. Frontiers in Neuroscience **9**, 408 (2015)
7. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning. pp. 1126–1135. PMLR (2017)
8. Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., Meneguzzi, F.: Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical **17**, 16–23 (2018)
9. Jiang, B., Zhang, Z., Lin, D., Tang, J., Luo, B.: Semi-supervised learning with graph learning-convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11313–11320 (2019)

10. Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics **8**(1), 118–127 (2007)
11. Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y., Yang, C.: Brain network transformer. Advances in Neural Information Processing Systems **35**, 25586–25599 (2022)
12. Li, X., Gu, Y., Dvornek, N., Staib, L.H., Ventola, P., Duncan, J.S.: Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. Medical Image Analysis **65**, 101765 (2020)
13. Lian, J., Zhang, C., Yu, D.: Robust disentangled variational speech representation learning for zero-shot voice conversion. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 6572–6576. IEEE (2022)
14. Rossi, F., Conan-Guez, B.: Functional multi-layer perceptron: a non-linear tool for functional data analysis. Neural Networks **18**(1), 45–60 (2005)
15. Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., Nichols, T.E., Robinson, E.C., Salimi-Khorshidi, G., Woolrich, M.W., et al.: Functional connectomics from resting-state fMRI. Trends in Cognitive Sciences **17**(12), 666–682 (2013)
16. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in Neural Information Processing Systems **30** (2017)
17. Song, R., Cao, P., Wen, G., Zhao, P., Huang, Z., Zhang, X., Yang, J., Zaiane, O.R.: BrainDAS: Structure-aware domain adaptation network for multi-site brain network analysis. Medical Image Analysis **96**, 103211 (2024)
18. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage **15**(1), 273–289 (2002)
19. 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 Transactions on Medical Imaging **39**(3), 644–655 (2020)
20. Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunimatsu, A., Okada, N., et al.: Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. PLoS Biology **17**(4), e3000042 (2019)
21. Yousefnezhad, T.: Orthogonal contrastive learning for multi-representation fMRI analysis. Advances in Neural Information Processing Systems **38**, 115689–115708 (2025)