Few-Shot Cross-Site Domain Generalization for Multi-Site Autism Brain Network Classification

Fatima Ez-Zahraa Bazay[1,2], Martin Barry[2], Antoine Widmer[2], Henning Müller[2], Rachid Jennane[3], Ben Glocker[4], and Ahmed Drissi El Maliani[1]

[1] LRIT, Faculty of Sciences in Rabat, Mohammed V University in Rabat, Rabat, Morocco
{fatimaezzahraa.bazay,a.elmaliani}@um5r.ac.ma
[2] School of Management, University of Applied Sciences and Arts Western Switzerland (HES-SO), Sierre, Switzerland
{martin.barry,antoine.widmer,henning.mueller}@hevs.ch
[3] Institut Denis Poisson, UMR CNRS 7013, University of Orleans, France
Rachid.Jennane@univ-orleans.fr
[4] Department of Computing, Imperial College London, UK
b.glocker@imperial.ac.uk

Abstract. Multi-site functional Magnetic Resonance Imaging (fMRI) datasets offer unprecedented opportunities for developing diagnostic biomarkers for neuropsychiatric disorders such as Autism Spectrum Disorder (ASD). However, site-induced distributional shifts across acquisition centers severely compromise model generalizability, rendering classifiers trained at one institution ineffective when deployed at others. In this work, we propose a few-shot cross-site domain generalization framework based on Model-Agnostic Meta-Learning (MAML) for multi-site Brain Network classification. By treating each acquisition site as a distinct meta-learning task, our framework meta-trains a linear classifier on top of a pre-trained backbone, learning an initialization that rapidly adapts to unseen sites using only $k = 3$ labeled support examples, achieving a 1.5% improvement in AUC across different backbone architectures. In comparison, achieving similar accuracy with classical approaches would require up to 15 labeled examples, demonstrating that MAML provides a strong site-generalizable initialization that enables rapid few-shot classifier adaptation to each unseen site. Our method consistently outperforms CORrelation ALignment (CORAL) and Meta-Learning for Domain Generalization (MLDG) across multiple backbone architectures, and performance gains further scale with the number of training sites. This study shows the necessity of focusing research on inter-site variability for fMRI studies, as classical methods poorly generalize to new sites. The implementation is available at https://github.com/fezz-bazay/MAML-TTA-Brain-Network.

Keywords: Meta-learning · Few-shot domain generalization · Brain network classification · Multi-site neuroimaging · Autism spectrum disorder

1 Introduction

Resting-state Functional Magnetic Resonance Imaging (rs-fMRI) has emerged as a powerful neuroimaging modality for investigating brain function and identifying diagnostic biomarkers for neuropsychiatric disorders, particularly Autism Spectrum Disorder (ASD) [1–4]. This non-invasive technique can measure spontaneous neural activity during task-free states, making it especially valuable for populations with cognitive limitations who may struggle with task-based paradigms. Recent advances in deep learning, particularly transformer architectures [5–8], have demonstrated exceptional promise for extracting discriminative patterns from complex brain network data, in which Regions of Interest (ROIs) form fully connected graphs, with connection profiles serving as informative node features.

However, the clinical deployment of these complex models faces a critical challenge: the dramatic performance degradation when applied across different data acquisition sites. The pursuit of larger datasets for improved statistical power has led to multi-institutional data aggregation, exemplified by datasets such as the Autism Brain Imaging Data Exchange (ABIDE) that encompasses 19 international sites [9,10]. Unfortunately, this introduces substantial technical heterogeneity, stemming from variations in scanner specifications, acquisition protocols, preprocessing pipelines and population demographics [11–14]. These site-specific variations create distributional shifts that severely compromise model generalizability, rendering diagnostic systems trained at one institution potentially ineffective when deployed at others. We distinguish between *intra-site* generalization, where the model is evaluated on held-out subjects from sites seen during training, and *inter-site* generalization, where the model is evaluated on completely unseen acquisition from other sites. The latter is a clinically relevant challenge for real-world deployment, as new clinical centers will inevitably differ in scanner hardware, acquisition protocols and patient demographics.

Current approaches to address cross-site variability include naive data aggregation [15], domain adaptation methods such as CORrelation ALignment (CORAL) [16], and meta-learning for domain generalization such as Meta-Learning for Domain Generalization (MLDG) [17]. However, these strategies suffer from important limitations: aggregation methods fail to explicitly account for distributional differences [11, 12], domain adaptation approaches require target domain information and site-specific model customization [18], while MLDG lacks few-shot site adaptation capabilities for rapid adjustment to new acquisition sites. From a broader meta-learning and domain generalization perspective [19–22], existing approaches have shown limited effectiveness in the complex, high-dimensional space of neuroimaging data. While previous meta-learning approaches [23] employ modulation networks for feature calibration, these methods lack the systematic coupling of meta-initialization with few-shot site adaptation needed for effective cross-site generalization.

We address these limitations by proposing a novel few-shot cross-site domain generalization framework for multi-site neuroimaging classification. Our framework combines MAML with few-shot site adaptation, treating each acqui-

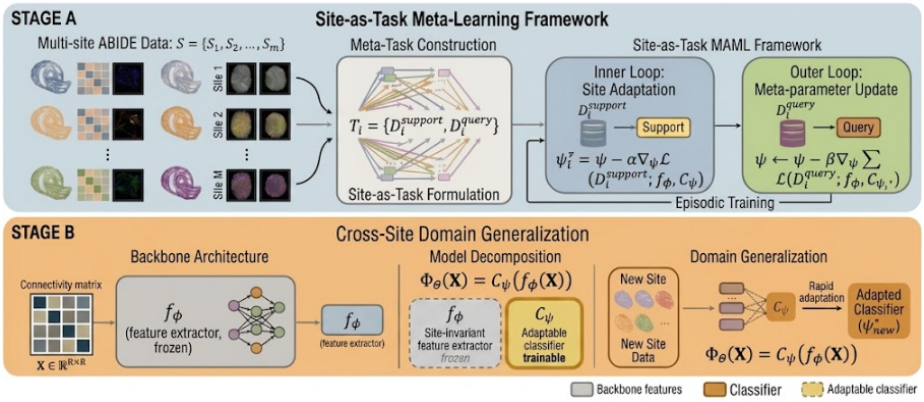

Fig. 1. Overview of our site-adaptive meta-learning framework for cross-site brain network classification. Stage A constructs cross-site meta-tasks using a site-as-task MAML formulation to meta-train an adaptable linear classifier C_ψ on top of a frozen pre-trained backbone f_ϕ, through episodic training across source sites. Stage B applies few-shot site adaptation to unseen sites, adapting only the classifier using $k = 3$ labeled support examples with $K = 3$ inner gradient steps.

sition site as a distinct meta-learning task. Only the linear classifier is adapted at inference time using k labeled support examples from the unseen site, while the backbone remains frozen, enabling efficient and rapid adaptation. This work addresses three research questions:

RQ1. Can treating acquisition sites as meta-learning tasks improve generalization to unseen sites?

RQ2. How many labeled examples per unseen site (support examples) does our method require compared to standard fine-tuning?

RQ3. Do performance gains scale with the number of training sites?

2 Method

We propose a two-stage framework for cross-site brain network classification, illustrated in Fig. 1. In Stage A, a linear classifier is meta-trained on top of a frozen pre-trained backbone across source sites using episodic training. In Stage B, the classifier is rapidly adapted to each unseen site at inference time using only $k = 3$ labeled support examples via few-shot site adaptation.

2.1 Problem Formulation

We formulate the multi-site neuroimaging problem as follows. Given a collection of neuroimaging sites $\mathcal{S} = \{S_1, S_2, ..., S_M\}$, each site S_m contains brain connectivity samples $\mathcal{D}_m = \{(\mathbf{X}_i^{(m)}, y_i^{(m)})\}_{i=1}^{n_m}$, where $\mathbf{X}_i^{(m)} \in \mathbb{R}^{R \times R}$ represents

functional connectivity between R brain regions and $y_i^{(m)} \in \{0, 1\}$ denotes the diagnostic label (ASD/Typical Development (TD)).

The fundamental challenge arises from site-induced distribution shifts, where $P(\mathbf{X}|S_a) \neq P(\mathbf{X}|S_b)$ for different sites $S_a, S_b \in \mathcal{S}$, while the underlying diagnostic mapping $P(y|\mathbf{X})$ remains consistent across sites. Our objective is to learn a site-adaptive predictor $\Phi_\theta : \mathbb{R}^{R\times R} \to [0, 1]$ that minimizes the expected risk across all sites:

$$\min_\theta \mathbb{E}_{S\sim\mathcal{S}} \left[\mathbb{E}_{(\mathbf{X},y)\sim\mathcal{D}_S} \left[\mathcal{L}(\Phi_\theta(\mathbf{X}), y)\right]\right] \tag{1}$$

2.2 Site-as-Task Model-Agnostic Meta-Learning Framework

The core innovation of our approach lies in formulating cross-site adaptation as a meta-learning problem where each acquisition site becomes a distinct meta-learning task. We define a meta-task $\mathcal{T}_i$ for each site S_i as $\mathcal{T}_i = \{\mathcal{D}_i^{support}, \mathcal{D}_i^{query}\}$ where $\mathcal{D}_i^{support}$ and $\mathcal{D}_i^{query}$ represent support and query sets from the same site. Support and query sets are sampled randomly from each site using a fixed random seed, with k support examples and the remaining subjects serving as the query set.

To enable efficient adaptation, we decompose our model into complementary components:

$$\Phi_\theta(\mathbf{X}) = \mathcal{C}_\psi(f_\phi(\mathbf{X})) \tag{2}$$

where $\theta = \{\phi, \psi\}$ denotes all model parameters, $f_\phi : \mathbb{R}^{R\times R} \to \mathbb{R}^d$ serves as a site-invariant backbone feature extractor with parameters ϕ, and $\mathcal{C}_\psi : \mathbb{R}^d \to [0, 1]$ functions as an adaptable diagnostic classifier with parameters ψ that undergoes selective adaptation during meta-learning and few-shot site adaptation.

The backbone f_ϕ is pre-trained on source sites using standard supervised learning and kept frozen throughout both the inner and outer loops of meta-training. Only the linear classifier $\mathcal{C}_\psi : \mathbb{R}^d \to [0, 1]$ receives gradient updates during meta-learning and few-shot site adaptation. This two-stage design follows the feature reuse paradigm that was shown to be effective in few-shot classification [24], reducing the adaptation parameter space to the classifier only, which limits the risk of overfitting to the small support set ($k = 3$ examples) available at test time while preserving the feature representations learned during pre-training.

2.3 Site-as-Task MAML Algorithm

Our meta-learning algorithm adapts MAML [19,25] to the multi-site neuroimaging setting through episodic training where each episode simulates adaptation to a new acquisition site. The algorithm operates through inner and outer optimization loops:

Inner Loop (Site Adaptation): For each meta-task corresponding to site S_i, we adapt the classifier parameters using k support examples:

$$\psi_i^* = \psi - \alpha\nabla_\psi \mathcal{L}_{CE}(\mathcal{D}_i^{support}; f_\phi, \mathcal{C}_\psi) \tag{3}$$

Outer Loop (Meta-Update): We evaluate the adapted parameters on the query set and update only the classifier meta-parameters, while the backbone f_ϕ remains frozen:

$$\psi \leftarrow \psi - \beta \nabla_\psi \sum_{i=1}^{M} \mathcal{L}_{CE}(\mathcal{D}_i^{query}; f_\phi, \mathcal{C}_{\psi_i^*}) \tag{4}$$

f_ϕ remains frozen throughout; gradients flow only through ψ, making meta-training computationally efficient.

This constitutes our few-shot site adaptation procedure, where given k labeled support examples from an unseen site S_{new} at inference time, the same inner loop is applied to obtain site-adapted parameters ψ_{new}^*, which are then used to classify the remaining subjects of that site.

3 Experiments

This section evaluates the effectiveness of our proposed backbone-MAML framework with extensive experiments.

3.1 Experimental Settings

Dataset. We conduct experiments on the Autism Brain Imaging Data Exchange (ABIDE) dataset [26, 27], which collects resting-state fMRI from 19 international sites containing 1,009 subjects (516 ASD, 493 TD). Brain networks are defined using the Craddock 200 atlas [9] with 200 ROIs. We apply stratified sampling during data splitting to ensure balanced representation of both diagnostic groups. We adopt a strict leave-sites-out evaluation protocol. Four large sites serve as source domains for training NYU (175 subjects), UM1 (106), UCLA1 (72), and USM (71), totaling 424 subjects. The remaining 15 sites (585 subjects) are held out as completely unseen test sites. For the scaling experiments, training sites are selected in decreasing order of size from all 19 ABIDE sites, ranging from $N = 1$ to $N = 16$, with the remaining sites serving as unseen test sites at each operating point. Within source sites, subjects are split into 70% for MAML training, 10% for intra-site validation, and 20% for intra-site testing using stratified sampling to ensure balanced ASD/TD representation in each split.

Implementation Details. The backbone is pre-trained on the source sites and frozen during meta-training. Only the linear classifier ($800 \rightarrow 2$) undergoes meta-learning and few-shot site adaptation. We set the inner learning rate $\alpha = 0.001$, meta learning rate $\beta = 0.001$, inner loop steps $K = 3$, support size $k = 3$, meta-batch size 4, and train for 200 epochs with Adam optimizer. The model is trained on a single NVIDIA V100 GPU. All results are averaged over 5 seeds (43–47).

Baselines. We compare against three methods. (1) BNT source-only [6] is a standard, state-of-the-art, model trained on source sites without domain

adaptation. (2) BNT-CORAL [16] performs domain adaptation via second-order statistics alignment. (3) BNT-MLDG [17] applies meta-learning for domain generalization without few-shot site adaptation. All methods use the same BNT backbone for fair comparison. We report AUC, Accuracy (ACC), Sensitivity (SEN), and Specificity (SPEC) on the 15 unseen inter-sites.

3.2 Performance Analysis (RQ1)

Table 1 presents inter-site results on ABIDE. To study the framework's consistency, we use 3 different backbones: a multi-layer perceptron (MLP), a Graph-Transformer [5], and the state-of-the-art Brain Network Transformer (BNT). In this section, the MAML-FewShot uses $k = 3$ support examples per site and $K = 3$ inner gradient steps. A thorough study of these parameters is presented in Sec. 3.4. Our framework outperforms all baselines. Notably, it achieves an inter-site AUC of $72.39 \pm 1.30\%$ with BNT as the backbone. Comparatively, CORAL and MLDG yield negligible or negative improvements over the BNT source-only model (-0.54% and -0.55%, respectively). Table 1 further shows that MAML-FewShot consistently improves over source-only across all three backbones (MLP: $+1.45\%$, GraphTransformer: $+1.54\%$, BNT: $+1.73\%$), demonstrating the generality of the framework. Regarding clinical metrics, BNT-MAML-FewShot achieves 75.31% sensitivity and 57.39% specificity on inter-site. This asymmetry reflects the class imbalance in unseen sites and the heterogeneity of ASD presentations. Notably, GT-MAML-FewShot yields a specificity below chance (44.24%), indicating that backbone choice critically affects clinical reliability, and motivating our focus on BNT as the primary backbone. The inconsistency between sensitivity and specificity likely reflects class imbalance in unseen sites and the inherent heterogeneity of ASD presentations across acquisition centers, where the model learns to prioritize ASD detection at the cost of reduced specificity. We conduct the rest of our study on BNT, as the improvement is consistent across backbones and BNT achieves the highest accuracy.

3.3 Ablation Study and Backbone Generalizability (RQ2)

Table 1 *Ablation* shows the impact of both MAML and few-shot adaptation independently, using BNT as the backbone. Three observations emerge. First, Few-shot adaptation alone without MAML degrades inter-site AUC (-1.19% below BNT source-only), confirming that SGD fine-tuning on $k = 3$ supports examples without a meta-learned initialization fails to improve generalization to unseen sites, and even hurts performance compared to no adaptation at all. Second, MAML alone, without few-shot adaptation, also yields a marginal degradation (-0.11%), as MAML learns a strong site-generalizable initialization but does not adapt the classifier at test time to the specific distribution of the new site. Third, BNT-MAML-FewShot achieves the best inter-site AUC (72.39%, a 1.73% improvement), confirming that MAML and few-shot adaptation are complementary; MAML provides a well-initialized starting point from which few-

Table 1. Inter-site and intra-site results on ABIDE I. Mean ± std over 5 seeds (43–47). For MAML-FewShot, we used $k = 3$ support examples per site and $K = 3$ inner gradient steps.

Method	Protocol	AUC (%)	ACC (%)	SEN (%)	SPEC (%)
Source-only baselines					
MLP	Intra	77.18 ± 4.59	69.38 ± 5.56	75.73 ± 17.37	62.73 ± 19.39
	Inter	67.60 ± 0.60	62.50 ± 1.44	73.36 ± 14.52	50.50 ± 16.62
GraphTransformer	Intra	71.94 ± 3.56	66.25 ± 2.72	68.82 ± 8.52	63.79 ± 3.08
	Inter	66.14 ± 0.62	61.74 ± 0.86	64.63 ± 5.35	58.56 ± 4.73
BNT	Intra	77.34 ± 7.83	69.88 ± 6.34	73.33 ± 9.33	66.51 ± 11.54
	Inter	70.66 ± 0.86	65.47 ± 1.18	71.79 ± 3.92	58.49 ± 4.05
Domain adaptation/generalization baselines					
BNT-CORAL	Intra	76.03 ± 5.15	69.38 ± 5.98	72.42 ± 8.22	66.14 ± 6.73
	Inter	70.12 ± 0.23	65.03 ± 1.26	70.94 ± 3.69	58.49 ± 5.16
BNT-MLDG	Intra	77.36 ± 5.42	70.62 ± 5.17	85.09 ± 10.91	54.92 ± 15.47
	Inter	70.11 ± 1.19	64.65 ± 2.66	$\mathbf{77.92 \pm 8.40}$	50.00 ± 13.97
Ablation (BNT backbone)					
BNT-FewShot (w/o MAML)	Intra	89.49 ± 3.88	79.06 ± 4.26	80.13 ± 7.39	78.34 ± 7.82
	Inter	69.47 ± 1.76	64.63 ± 1.88	69.01 ± 4.39	$\mathbf{59.85 \pm 3.46}$
BNT-MAML (w/o adaptation)	Intra	89.49 ± 5.60	77.81 ± 3.88	76.67 ± 8.59	80.30 ± 9.34
	Inter	70.55 ± 0.75	63.35 ± 3.15	71.79 ± 3.92	58.49 ± 4.05
Our method					
MLP-MAML-FewShot	Intra	77.18 ± 4.59	69.38 ± 5.56	75.73 ± 17.37	62.73 ± 19.39
	Inter	69.05 ± 1.20	63.02 ± 1.80	69.20 ± 12.37	56.30 ± 10.79
GT-MAML-FewShot	Intra	71.94 ± 3.56	66.25 ± 2.72	68.82 ± 8.52	63.79 ± 3.08
	Inter	67.68 ± 0.83	61.29 ± 2.89	$\mathbf{77.00 \pm 12.40}$	44.24 ± 18.53
BNT-MAML-FewShot	Intra	89.49 ± 5.32	78.44 ± 6.43	76.67 ± 8.59	80.30 ± 9.34
	Inter	$\mathbf{72.39 \pm 1.30}$	$\mathbf{66.75 \pm 2.22}$	$\mathbf{75.31 \pm 5.14}$	57.39 ± 7.96

shot adaptation successfully bridges the remaining domain gap with just $K = 3$ gradient steps.

3.4 Data Efficiency and Scaling Analysis (RQ3)

Figure 2(a) reveals a persistent generalization gap between intra-site AUC (source sites, $\sim$ 87–90%) and inter-site AUC (unseen sites, $\sim$ 70–79%) across all values of N. This gap of $\sim$ 15–20% confirms that models trained on source sites generalize poorly to unseen acquisition centers, regardless of the number of training sites. This motivates our focus on inter-site evaluation as the primary metric throughout this study. Importantly, the gap narrows as N increases, suggesting that exposure to more diverse training sites improves cross-site generalization. Figure 2(b) shows that as training sites increase from $N = 1$ to 16, MAML-FewShot consistently outperforms BNT source-only in terms of AUC at every operating point (+1.22% at $N = 4$, +3.50% at $N = 16$). Importantly, the performance gap widens as the number of training sites increases, suggesting that meta-learning benefits compound with dataset diversity. Notably, even at $N = 1$, MAML-

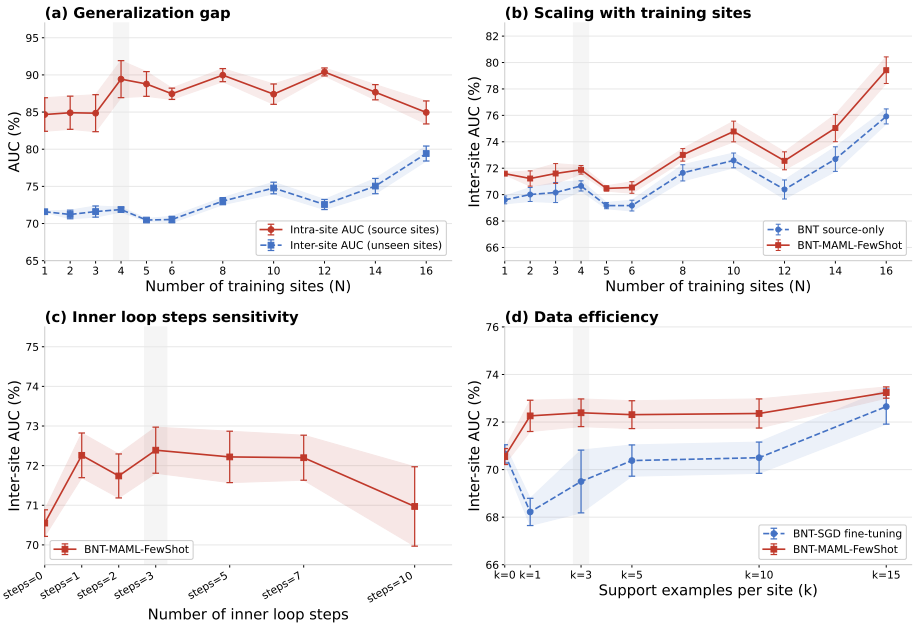

Fig. 2. (a) Generalization gap: Intra-site AUC (source sites) vs. inter-site AUC (unseen sites) as a function of number of training sites N. **(b) Scaling with training sites:** Inter-site AUC vs. number of training sites N for BNT-MAML-FewShot and BNT source-only. **(c) Inner loop steps sensitivity:** Inter-site AUC vs. number of inner loop steps K with $k = 3$ fixed. **(d) Data efficiency:** Inter-site AUC vs. number of support examples k for BNT-MAML-FewShot and BNT-SGD fine-tuning.

FewShot outperforms BNT source-only, suggesting that the episodic training objective encourages a more generalizable initialization even from a single site. The number of inner loop steps K is the main hyperparameter of our framework. As shown in Fig. 2(c), Steps= 3 (as reported in Table 1) achieves strong performance (72.39%), with performance slightly degrading for steps$\geq$ 5, suggesting that a small number of gradient steps is sufficient and that excessive adaptation steps are not beneficial. Figure 2(d) shows that BNT-MAML-FewShot consistently outperforms SGD fine-tuning across all values of k. Notably, even with $k = 1$, MAML-FewShot already achieves 72.26% AUC, a performance that SGD fine-tuning only matches at $k = 15$, demonstrating that the meta-learned initialization substantially reduces the need for labeled support examples at deployment time.

4 Conclusion

We proposed a site-adaptive meta-learning framework combining MAML with few-shot site adaptation for multi-site autism brain network classification. By

treating each acquisition site as a meta-learning task, our method achieves 72.39% inter-site AUC on 15 completely unseen ABIDE sites, outperforming CORAL, MLDG, and BNT source-only across multiple backbone architectures. The ablation confirms that MAML and few-shot adaptation are complementary, with the framework requiring only $k = 3$ support examples compared to $k = 15$ for standard fine-tuning, and performance scaling consistently from $N = 1$ to $N = 16$ training sites. **Strengths.** Unlike domain adaptation methods [16] requiring target data during training, our framework requires no target-domain data during meta-training, adapting to unseen sites using only $k = 3$ labeled support examples at inference. Unlike MLDG [17], it enables rapid site-specific adjustment in $K = 3$ gradient steps, and unlike [23], it is backbone-agnostic requiring no architectural modifications. **Limitations.** A persistent $\sim$ 15–20% gap between intra- ($\sim$ 87–90%) and inter-site AUC (72.39%) highlights that cross-site generalization remains an open challenge. Experiments are limited to ABIDE I with $N = 4$ source sites, though performance improves consistently with more training sites. Furthermore, we did not explicitly control for demographic confounds such as age and sex distributions across sites, which may contribute to the observed inter-site variability. Evaluation is restricted to ABIDE I, and generalization to independent cohorts or other neuropsychiatric conditions such as ADHD remains to be established. **Future Work.** Future directions include extending to ABIDE II and other disorders such as schizophrenia and ADHD, investigating structural and diffusion MRI modalities, and extending to multi-class classification beyond binary ASD/TD diagnosis.

References

1. Michael Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology*, 21(4):424–430, 2008.
2. Daniel P Kennedy and Eric Courchesne. The intrinsic functional organization of the brain is altered in autism. *Neuroimage*, 39(4):1877–1885, 2008.
3. Jocelyn V Hull, Lisa B Dokovna, Zachary J Jacokes, Carinna M Torgerson, Andrei Irimia, and John Darrell Van Horn. Resting-state functional connectivity in autism spectrum disorders: a review. *Frontiers in psychiatry*, 7:205, 2017.
4. Jared A Nielsen, Brandon A Zielinski, P Thomas Fletcher, Andrew L Alexander, Nicholas Lange, Erin D Bigler, Janet E Lainhart, and Jeffrey S Anderson. Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience*, 7:599, 2013.
5. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
6. Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.
7. Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical image analysis*, 74:102233, 2021.

8. Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrerro Moreno, Ben Glocker, and Daniel Rueckert. Spectral graph convolutions for population-based disease prediction. In *International conference on medical image computing and computer-assisted intervention*, pages 177–185. Springer, 2017.
9. Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7(27):5, 2013.
10. Thomas E Nichols. The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior*, 8(2):153–182, 2014.
11. Alexandre Abraham, Michael P Milham, Adriana Di Martino, R Cameron Craddock, Dimitris Samaras, Bertrand Thirion, and Gael Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 147:736–745, 2017.
12. Jean-Philippe Fortin, Nicholas Cullen, Yvette I Sheline, Warren D Taylor, Irem Aselcioglu, Philip A Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J McGrath, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*, 167:104–120, 2018.
13. Ben Glocker, Robert Robinson, Daniel C Castro, Qi Dou, and Ender Konukoglu. Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects. *arXiv preprint arXiv:1910.04597*, 2019.
14. Christian Wachinger, Anna Rieckmann, Sebastian Pölsterl, Alzheimer's Disease Neuroimaging Initiative, et al. Detect and correct bias in multi-site neuroimaging datasets. *Medical Image Analysis*, 67:101879, 2021.
15. Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: clinical*, 17:16–23, 2018.
16. Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016.
17. Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
18. Mingliang Wang, Daoqiang Zhang, Jiashuang Huang, Pew-Thian Yap, Dinggang Shen, and Mingxia Liu. Identifying autism spectrum disorder with multi-site fmri via low-rank domain adaptation. *IEEE transactions on medical imaging*, 39(3):644–655, 2019.
19. Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.
20. Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5149–5169, 2021.
21. Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip S Yu. Generalizing to unseen domains: A survey on domain generalization. *IEEE transactions on knowledge and data engineering*, 35(8):8052–8072, 2022.
22. Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. *arXiv preprint arXiv:2007.01434*, 2020.

23. Jaein Lee, Eunsong Kang, Eunjin Jeon, and Heung-Il Suk. Meta-modulation network for domain generalization in multi-site fmri classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 500–509. Springer, 2021.
24. Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019.
25. Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.
26. Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, 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.
27. Autism brain imaging data exchange. https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html. Last accessed 2025/06/09.