Multi-site fMRI-based mental disorder detection using adversarial learning: an ABIDE study

Xin Wen[1] **(xwen@tyut.edu.cn), Shijie Guo**[1] **(2023521863@link.tyut.edu.cn)**
Yanqing Dong[2] **(dongyanqing0078@link.tyut.edu.cn)**, **Mengni Zhou**[1] **(zhoumengni@tyut.edu.cn)**
Jie Xiang[2*] **(xiangjie@tyut.edu.cn)**
[1]School of Software, Taiyuan University of Technology, Taiyuan, China
[2]School of Computer Science and Technology(Data Science), Taiyuan University of Technology, Taiyuan, China
[*] Corresponding author

Abstract

Heterogeneity in open fMRI datasets, caused by variations in scanning protocols, confounders, and population diversity, hinders representation learning and classification performance. To address these limitations, we propose a novel multi-site adversarial learning network (MSalNET) for fMRI-based mental disorder detection. Firstly, a representation learning module is introduced with a node information assembly (NIA) mechanism to extract features from functional connectivity (FC). This mechanism aggregates edge information from both horizontal and vertical directions, effectively assembling node information. Secondly, to generalize the feature across sites, we propose a site-level feature extraction module that can learn from individual FC. Lastly, an adversarial learning network is proposed to balance the trade-off between individual classification and site regression tasks. The proposed method was evaluated on Autism Brain Imaging Data Exchange (ABIDE). The results indicate that the proposed method achieves an accuracy of 75.56% with reducing variability from a data-driven perspective. The most discriminative brain regions revealed by NIA are consistent with statistical findings, uncovering the black box of deep learning to a certain extent. MSalNET offers a novel perspective on the detection of multi-site fMRI mental disorders and it considers the interpretability of the model, which is a crucial aspect in deep learning.

Keywords: Multi-site, Functional Connectivity, Adversarial Learning, Interpretability

Introduction

Functional connectivity (FC) derived from functional magnetic resonance imaging (fMRI) has proven to be a powerful tool for distinguishing various mental disorders(Kliemann et al., 2024; Jain, Rakshe, Sengar, Murugappan, & Ronickom, 2024). By analyzing the statistical dependencies in blood oxygen level-dependent (BOLD) signals across brain regions, FC provides critical insights into the brain's functional architecture and its disruptions in diseases(Song et al., 2022; Tang et al., 2022; Yang et al., 2022; Li et al., 2021). Despite its potential, challenges including site-specific biases(Jahani et al., 2024) and data heterogeneity(Alcaide, Illan, Ramirez, & Gorriz, 2024) hinder its broader application. Integrating multi-site fMRI and refining analytical methods are essential to enhance the reliability and generalizability of FC as a diagnostic biomarker in mental health research.

Aggregating multi-site fMRI data introduces site-specific discrepancies, prompting researchers to develop preprocessing-phase approaches to mitigate these effects. Common methods include ComBat (Zheng et al., 2024; Honnorat et al., 2024), which harmonizes site effects using empirical Bayesian framework, and the Subsampling Maximum-Mean-Distance based algorithm (Mueller et al., 2005), which addresses distribution shifts in pooled data. The generalized linear model (GLM) (Nelder & Wedderburn, 1972; Taki, 2024) has also been employed for its interpretability in assessing variable influences. These methods aim to eliminate site-related confounds before model training. However, preprocessing-phase approaches often rely on assumptions that may not hold for all datasets (Dinsdale, Jenkinson, & Namburete, 2021; Gallo et al., 2023) and struggle with nonlinear relationships or functional connectivity analysis (Plis et al., 2014).

In contrast, deep learning (DL)(Warren, Khan, & Moustafa, 2024) excels in handling large, complex datasets, making DL-based training-phase approaches for addressing site heterogeneity during model training.. These methods, such as Deep Neural Networks (DNNs) (Lamani, Julian Benadit, & Guruprasad, 2024) and Convolutional Neural Networks (CNNs)(Gupta et al., 2024; Nazir et al., 2024), leverage time series of regions of interest (ROIs) or functional connectivity to extract non-linear representations for classification or regression tasks. Advanced architectures like Graph Convolutional Networks (GCNs) model structural connectivity to aggregate node-level information (Parisot et al., 2018; Huynh et al., 2024), while autoencoder (AE) minimizes reconstruction errors to learn compact embeddings (Eslami, Mirjalili, Fong, Laird, & Saeed, 2019; Zhang, Feng, Han, Gong, & Duan, 2023). Although these approaches achieve strong performance in distinguishing disorders, they often mitigate site effects implicitly through representation learning, without directly modeling multi-site variability.

Inspired by the adversarial training strategy of Generative Adversarial Networks (GANs), we propose MSalNET (multi-site Adversarial Learning Network), an end-to-end framework designed to eliminate site-specific biases in functional connectivity from multi-site fMRI datasets. This framework addresses two core challenges: (1) effectively extracting site-related features from FC and (2) removing site confounders during training while ensuring model convergence. MSalNET integrates three key modules—a representation learning module for extracting task-relevant features, an adaptive site feature learning module for isolating site-specific patterns, and an adversarial learning module that combines site feature regression and main classification tasks trained adversarially. During training, the classification task maximizes diagnostic accuracy, while the regression task minimizes the

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

model's ability to predict site labels, thereby suppressing site-related biases and preserving disease-discriminative features. Experiments demonstrate MSalNET's effectiveness, achieving 75.56 ± 1.89% accuracy in Autism Spectrum Disorder (ASD) classification on multi-site data, highlighting its potential to enhance the reliability of computer-aided diagnosis systems in heterogeneous multi-site environments.

Method and Materials

Dataset and Preprocessing

We conducted a series of experiments on the ABIDE-I(Di Martino et al., 2014) dataset to validate the effectiveness of MSalNET. The ABIDE-I dataset contains rs-fMRI data and scale information from 17 sites, with a total of 1112 subjects, including 539 ASD patients and 573 normal controls(NC). We screened 1035 subjects from the official website which consist of 505 ASD subjects and 530 NC subjects with complete scale information.

Functional connectivity

A preprocessed fMRI data is a 4D time series including a 3D spatial dimension and a 1D temporal dimension. The time series were evaluated using the mean time series signal or BOLD signal of the voxels within the ROI in the brain atlas. Craddock 200 (CC200)(Craddock, James, Holtzheimer III, Hu, & Mayberg, 2012) with 200 ROIs defined was used. The FC between the mean time series of each ROI pair was evaluated using Pearson's correlation coefficient as shown in Equation (1).

$$FC = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{1}$$

Where n is the length of time series, X and Y are two time series, $\bar{X}$ and $\bar{Y}$ are the mean of time series X and Y , respectively.

Method

The MSalNET is designed with two core tasks (Figure 1): (1) site feature regression, where a regression model extracts site-specific information from the features; (2) adversarial learning, which combines classification loss and site feature regression loss to remove site-related biases. During training, we first train an independent site feature learning module to generate site feature vectors, which serve as labels for the two tasks in adversarial learning. Then, MSalNET alternates between these two tasks, with each task updating its own parameters to avoid interference. This approach effectively removes site-related differences, enabling the model to classify across multiple sites and ultimately converge on the target task.

Representation learning The input to the representation learning module is the functional connectivity (FC) matrix, which differs from images in computer vision due to its non-Euclidean nature. Conventional CNNs may struggle with FC, so we propose Node Information Assembly (NIA) which leverages 1D convolutional kernels(Equation 2). NIA consists of two convolutional layers and one fully connected layer (Figure 2). The first layer performs convolution between each target region and all others, generating region-specific features. The second layer performs convolution to extract whole-brain features. Finally, a fully connected layer refines the features for classification. The convolution operation is shown in Equation (2).

$$x_j^m = f(\sum_{i\in M_j} x_i^{m-1} \times k_{ij}^m + b_j^m) \tag{2}$$

To verify the effectiveness of NIA, we set up two feature extractors 2D-CNN and MLP to compare the classification results with NIA.

Multi-site feature extraction By leveraging the unsupervised training and nonlinear dimensionality reduction of autoencoders, subject-level feature vectors are extracted. The upper triangular part of the functional connectivity matrix is flattened into a one-dimensional vector and input into the autoencoder to obtain the subject-level feature vector. Average pooling is then used to compute the mean feature vector for each site (see Equation 3).

$$h_{enc} = \phi_{enc}(x) = f(W_{enc}x + b_{enc}) \tag{3}$$

where f is the activation function, W_{enc} and b_{enc} are the weight and bias of the encoder, respectively. The decoder reconstructs the encoder output back to the original input data, as shown in formula (4).

$$\acute{x} = \phi_{dec}(h_{enc}) = W_{dec}h_{enc} + b_{dec} \tag{4}$$

Adversarial learning The adversarial task (site feature regression) and the target classification task are unified into a joint objective function to achieve balanced optimization. The function combines classification loss L_{cls} and site feature regression loss L_{reg} aiming to minimize L_{cls} while maximizing L_{reg} to enforce adversarial learning. Let $X = \{X_1,...,X_N\}$, $Y = \{Y_1,...,Y_N\}$, $C = \{C_1,...,C_N\}$ represent the FC, labels, site labels of N subjects respectively. The model parameters include θ_E (feature extractor), θ_C(classifier), and θ_R (site regressor). For the regression task, we adopt mean squared error (MSE) loss:

$$L_{reg}(X_i,C_i;\theta_R) = \sum_{i=1}^{N}\frac{1}{m}\sum_{j=1}^{m}(C_{j,i}-\acute{C}_{j,i})^2 \tag{5}$$

The loss function of the main task classification is the CrossEntropy Loss, as show in Equation (6) where P_i represents the probability of a category.

$$L_{cls}(X_i,Y_i;\theta_E,\theta_C) = -\sum_{i=1}^{N}[Y_i\log(P_i) + (1-Y_i)\log(1-P_i)] \tag{6}$$

The model is trained in alternating iterations, with each epoch optimizing the regression task loss L_{ref} firstly to train the regression model and then optimizing the objective function loss L_t to train the representation learning module and

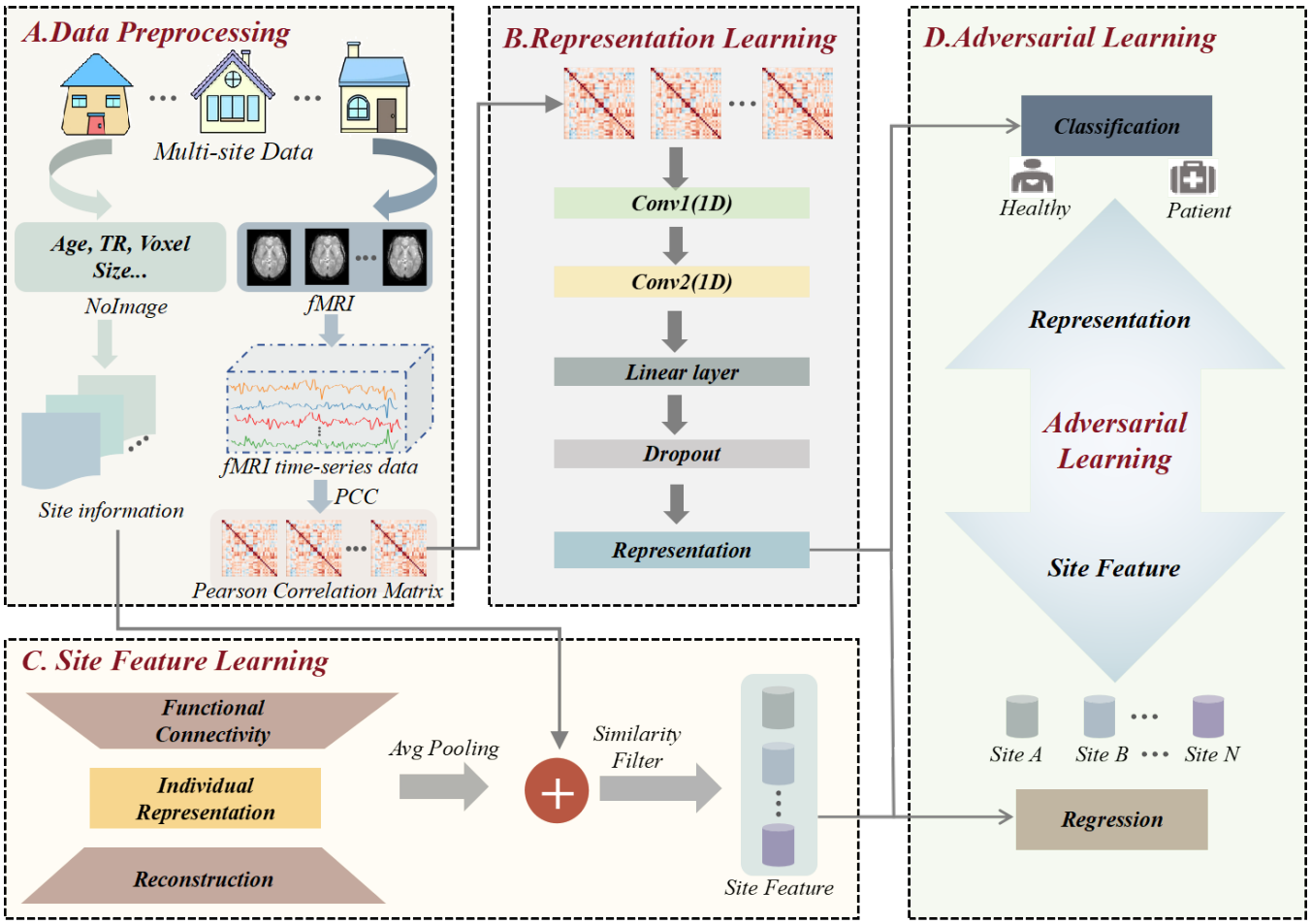

Figure 1: Overview of MSalNET. A) Data processing. B) Representation learning module. For each FC, NIA aggregates edges information to node through two 1D-CNN kernels sequentially from horizontal and vertical directions. C) Site feature learning module. AE is used to learn the representations of individuals and site labels are described by average pooling. Moreover, similarity filter is optional because it may not be executed due to the lack of non-image information. D) Adversarial Learning. A GAN-like adversarial learning network is designed to balance the tradeoff between site regression and disease classification.

classifier. To achieve adversity between the two training steps, the objective function needs to minimize the classification loss L_{cls} and maximize the regression task loss L_{ref} at the same time. Therefore, we design the objective loss function L_t as shown in Equation (7), where α is a hyperparameter, to balance the loss optimization of the two tasks and prevent L_{ref} increases excessively leading to too strong a restriction on the representation learning module to achieve the optimal classification performance.

$$L_t(X_i, Y_i, C_i; \theta_E, \theta_c) = L_{cls}(X_i, Y_i; \theta_E, \theta_c) + \frac{\alpha}{L_{ref}(X_i, C_i)} \quad (7)$$

Interpretation

To identify potential biomarkers for brain disorder classification, we employ back-propagation to quantify the relative importance of individual regions of interest (ROIs). The Node Information Assembly is designed with a one-dimensional convolutional kernel sized to match the number of ROIs, ensuring a one-to-one mapping between kernel parameters and brain regions. This design allows each kernel parameter to represent the importance of a specific ROI, enhancing interpretability. To preserve this mapping during back-propagation, we avoid pooling layers and use non-overlapping receptive fields, minimizing redundant features while maintaining the integrity of ROI contributions. This approach facilitates the discovery of clinically relevant biomarkers.

Experiments and Results

To evaluate the effectiveness of the proposed method, we setup a comparison experiment and an ablation experiment. In all experiments, ten-fold cross-validation was performed. For each site, 90% of the data was used for training and

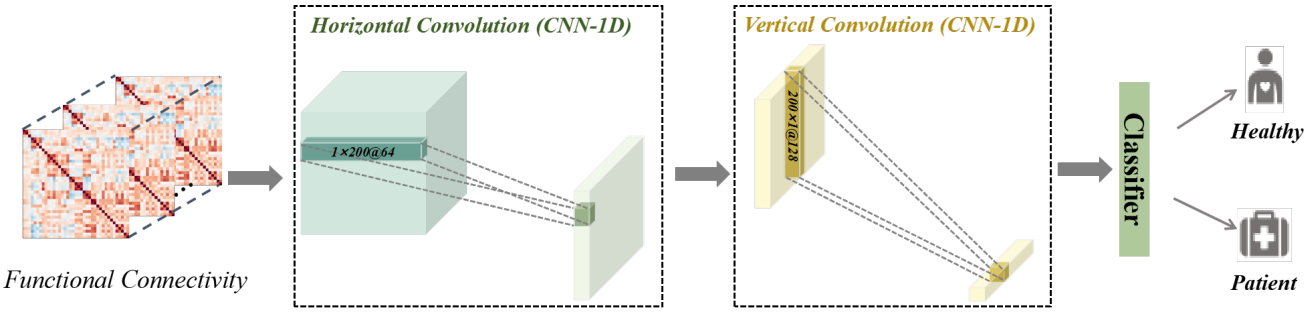

Figure 2: Structure diagram of NIA. The first convolution layer adopts a horizontal kernel and the result of each convolution can be seen as feature of a ROI. The second convolution layer adopts a vertical kernel to extract whole brain features from each ROI.

10% for testing. The classification performance of accuracy (ACC), area under the curve (AUC), precision, recall and F1 score were used as evaluation criteria. In the comparison experiment, we evaluated related methods that use functional connectivity as input and report results across as many sites as possible. In the ablation experiment, we compared different representation modules, such as a 2D-CNN and an MLP, under both adversarial and non-adversarial training settings. The proposed MSalNET could reach 75.56% ACC in ASD detection. Compared with the non-adversarial learning scenario, MSalNET showed nearly a 3% improvement.

Comparison Experiments The classification performance was compared with previous DL methods. Related works mainly focus on CNN, GCN and set their efforts on representation learning. The input data for these methods consisted of samples from all 17 sites. The results indicated that MSalNET achieved the best classification performance (Table 1).

Ablation Experiment In the adversarial case, NIA, MLP, and 2D-CNN, the accuracy rates were 75.56%, 74.20%, and 74.12%, respectively, with NIA showing a clear advantage (Table 2). In addition, when NIA was used and the hidden layer dimension of AE was set to 512, the accuracy reached 75.56%, which was a significant improvement compared to the accuracy of 73.81% without AE ($p < 0.05$). In the comparison experiment with and without adversarial, when NIA was used, the accuracy increased from 72.77% of the original one to 75.56% of the adversarial one, and no matter what representational learning module was selected, the adversarial ones could achieve significant improvement in classification performance ($p < 0.05$).

Interpretation The contribution of important ROIs was visualized (Figure 3). Additionally, a two-sample t-test between the patients and NCs was performed ($p < 0.05$, FWE corrected) where the ROIs with significant FC were labeled (Figure 4). In addition, we computed the clustering coefficient of ROIs to validate the important regions identified by the proposed method. The consistent areas across

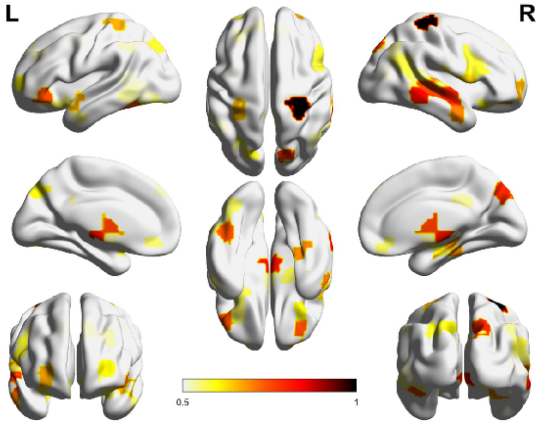

Figure 3: Classification visualization of brain region contribution with adversarial architecture. The darker the color, the higher the relative importance of the brain region.

both approaches included the middle/inferior/superior temporal gyrus, left triangular inferior frontal gyrus, right middle frontal gyrus, right dorsal/medial superior frontal gyrus, inferior orbitofrontal cortex, left calcarine cortex, right lingual gyrus, right fusiform gyrus, right hippocampus, right parahippocampal gyrus, thalamus, right caudate, putamen, left insula, left inferior parietal lobule, right precentral gyrus, precuneus, right cuneus, left angular gyrus, right superior occipital gyrus, and right postcentral gyrus. Notably, MSalNET uniquely identified additional critical regions, such as the left middle/superior temporal pole, right middle/superior orbitofrontal cortex, right medial superior frontal gyrus, left superior medial/inferior orbitofrontal cortex, left lingual gyrus, left parahippocampal gyrus, left caudate, right insula, and left cuneus.

Discussion

Harmonizing the confounding effects Previous studies have primarily relied on cross-validation and mathematical approaches, such as the generalized linear model (GLM)

Table 1: The comparison with the existing methods for ASD(%)

Ref.	Sites	Method	ACC	SEN	SPE
(Parisot et al., 2018)	17	GCN	70.4	-	-
(Eslami et al., 2019)	17	ASD-DiagNet	70.3	68.3	72.2
(Y. Liu, Xu, Li, Yu, & Yu, 2020)	17	Attention selection based on Extra-Tree	72.2	68.8	75.4
(Wang, Li, & Hu, 2021)	17	cGCN	71.6	-	-
(Zhang et al., 2023)	17	AE+F-score	70.9	70.7	75.5
(Xu et al., 2024)	17	ContrastPool	68.63	-	-
(R. Liu et al., 2024)	17	ST-HAG	71.9	-	-
Ours	17	MSalNET	**75.6**	-	-

Table 2: Ablation results for MSalNET in multi-sites for ASD(%)

Adversarial	Kernel	ACC	AUC	Precision	Recall	F1Score	Site ACC
×	MLP	72.46±3.23	75.70±3.43	73.90±2.05	71.88±3.48	71.37±4.04	37.04±6.50
	2D-CNN	71.37±3.45	76.17±3.60	73.47±3.56	71.01±3.47	70.41±3.91	37.54±7.47
	NIA	72.77±3.08	77.85±2.25	73.59±3.17	72.49±3.24	72.32±3.28	21.39±3.14
✓	MLP	74.20±3.08	76.51±3.68	75.55±2.72	73.79±3.38	73.52±3.70	18.73±5.23
	2D-CNN	74.12±2.61	77.10±3.15	74.79±1.96	73.97±2.97	73.78±2.95	18.58±1.89
	NIA	**75.56±1.89**	**78.99±2.11**	**76.60±2.25**	**75.36±1.89**	**75.20±2.00**	**15.76±2.17**

(Nelder & Wedderburn, 1972) and ComBat, to address the classification challenges of multi-site resting-state fMRI (rs-fMRI) datasets and mitigate site heterogeneity. While GLM is limited in handling nonlinear relationships (Reardon, Li, & Hu, 2021), ComBat harmonizes fMRI data across sites by adjusting statistical values, thereby reducing dataset heterogeneity and enhancing group difference detection. However, subsequent studies have shown that ComBat's effectiveness in improving multi-site dataset classification is limited beyond its original context (Gallo et al., 2023; Chen et al., 2022). Although site harmonization can significantly impact domain discrimination, its effectiveness tends to decrease when datasets are balanced and site information is independent of group membership (Gallo et al., 2023). Instead of relying on single-task discrimination, our approach introduces an adversarial relationship between the two tasks. The objective function balances their contributions, constraining automatic feature learning to reduce the impact of site factors on the main classification task. Experimental results demonstrate that this architecture effectively reduces the influence of site-related information in the learned representations, decreases the model's ability to predict site labels (indicating successful bias removal), and significantly improves diagnostic classification performance.

Interpretation of "black-box" and biomarkers Unlike traditional regression models and decision trees, DL models are inherently less interpretable, necessitating post hoc interpretability methods such as hotspot maps and class activation maps (He, Zhang, Ren, & Sun, 2016). Functional connectivity, as a form of semi-structured data, poses unique challenges for interpretation. To address this, MSalNET employs the Node Information Assembly module for representation learning, balancing model performance with interpretability. The NIA module is designed with convolutional kernels sized to match the number of ROIs, establishing a one-to-one mapping between kernel weights and brain regions. By omitting pooling layers, this design preserves the integrity of weight-

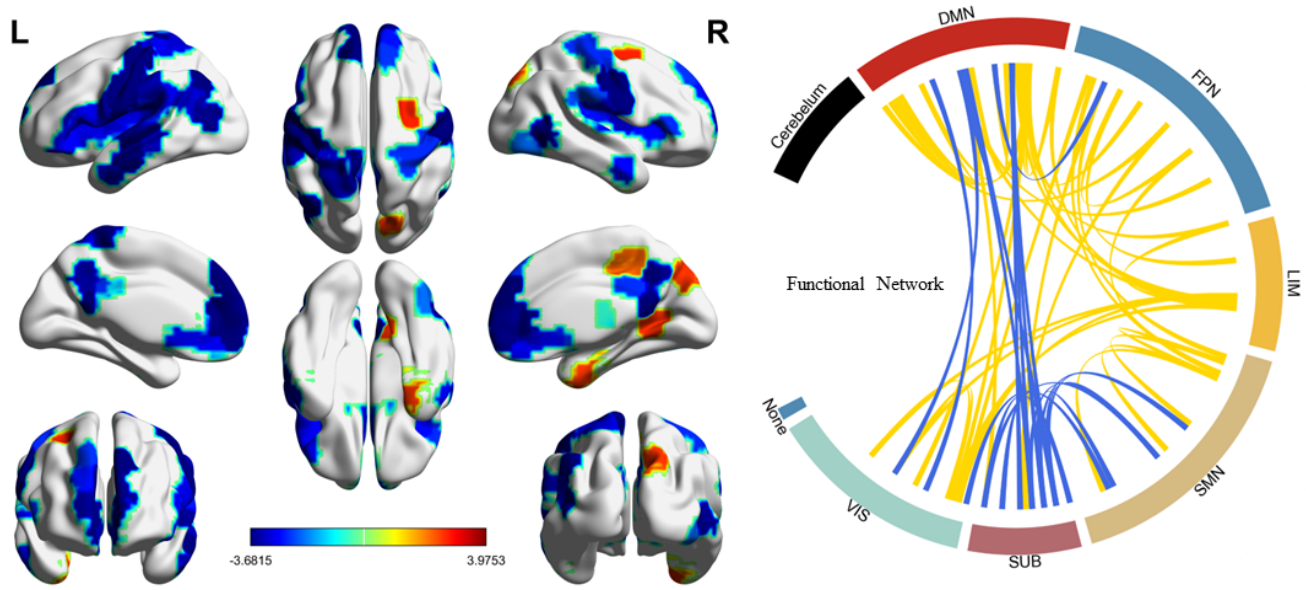

Figure 4: Left: clustering coefficient of ROIs. Red means the clustering coefficient of ASD group is higher than it in NC group. And the blue means the opposite. Right: ASD statistical result of brain region contribution. The blue lines indicate that the mean of t values in the ASD group is higher than it in NC group. The yellow lines indicate the mean of t values in the ASD group is lower than it in the NC group.

to-region mappings during back-propagation, ensuring that the relative importance of each brain region can be directly quantified through model weights. This approach enables the identification of potential disease-related biomarkers by analyzing region-specific contributions to classification outcomes, thereby demystifying the "black box" nature of DL in FC analysis.

The interpretability analysis demonstrates that the top-ranked ROIs contributing to the classification of ASD versus NC individuals not only align with established findings but also reveal novel biomarkers. The overlapping regions, including insula, temporal lobe, and prefrontal-default mode network (DMN) circuits, corroborate prior resting-state fMRI studies. Specifically, the insula exhibits atypical FC in ASD, particularly in internal-external cognitive switching (Nomi, Molnar-Szakacs, & Uddin, 2019; Nomi & Uddin, 2015), while reduced temporal-frontoparietal FC correlates with impaired social communication in adults with ASD (Holiga et al., 2019; Watanabe & Rees, 2017). Additionally, our method uniquely identifies critical regions, including the left middle/superior temporal pole, dorsolateral superior frontal gyrus, orbitofrontal cortex, and deep nuclei such as the right insula, left lingual gyrus, parahippocampal gyrus, cuneus, and caudate. These regions form three functional clusters: (1) temporal pole-related areas, which modulate auditory/visual processing and social behavior (Joshi et al., 2017); (2) frontal lobe circuits, which influence emotion regulation and executive function; and (3) deep nuclei, including the insula and caudate, which regulate interhemispheric communication, reward processing, and motor planning (Ameis & Szatmari, 2012; Kana & Wadsworth, 2012). Importantly, statistical methods failed to detect FC anomalies in deep nuclei, which task-based fMRI and structural MRI studies associate with ASD-related hypoactivation (Blakemore, 2008; Kleinhans et al., 2008). Our approach bridges rs-fMRI, task-fMRI, and structural findings, identifying biomarkers that reflect both functional and white matter alterations (Ameis & Szatmari, 2012). This multi-modal consistency underscores the validity of our interpretability framework in uncovering ASD pathophysiology beyond conventional statistical analyses.

Conclusion

Aiming at the multi-site problem existing in fMRI research, we propose a new method to alternately train the disease classification task and the site classification task in two steps during the model training process with the adversarial idea to eliminate site confounding. In MSalNET, we introduced 1) node information assembly module, 2) site level feature extraction module, 3) GAN-like adversarial learning module. Moreover, we also try to reveal the "black-box" of deep learning in a neuroimage perspective, the result indicate that patterns via our interpretation method could match biological explanations from another research. The comparison result indicate that the proposed method has superior performance than the existing ones. The ablation results indicate that the proposed three mechanisms have positive effect on the whole method. Our work mainly focuses on fMRI, multi-modal data fusion is the next attempt which has been proved effective in classification tasks. In the feature level, besides FC more measures should be considered.

Acknowledgments

This work was supported by the following grants: the National Natural Science Foundation of China 62206196 (Xin Wen), 62376184 (Jie Xiang) and 62303445 (Mengni Zhou).

References

Alcaide, F., Illan, I., Ramirez, J., & Gorriz, J. (2024). Unraveling the autism spectrum heterogeneity: Insights from abide i database using data/model-driven permutation testing approaches. *arXiv preprint arXiv:2405.12225*.

Ameis, S. H., & Szatmari, P. (2012). Imaging-genetics in autism spectrum disorder: advances, translational impact, and future directions. *Frontiers in psychiatry*, *3*, 46.

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267–277.

Chen, A. A., Beer, J. C., Tustison, N. J., Cook, P. A., Shinohara, R. T., Shou, H., & Initiative, A. D. N. (2022). Mitigating site effects in covariance for machine learning in neuroimaging data. *Human brain mapping*, *43*(4), 1179–1195.

Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, *33*(8), 1914–1928.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... others (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, *19*(6), 659–667.

Dinsdale, N. K., Jenkinson, M., & Namburete, A. I. (2021). Deep learning-based unlearning of dataset bias for mri harmonisation and confound removal. *NeuroImage*, *228*, 117689.

Eslami, T., Mirjalili, V., Fong, A., Laird, A. R., & Saeed, F. (2019). Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data. *Frontiers in neuroinformatics*, *13*, 70.

Gallo, S., El-Gazzar, A., Zhutovsky, P., Thomas, R. M., Javaheripour, N., Li, M., ... others (2023). Functional connectivity signatures of major depressive disorder: machine learning analysis of two multicenter neuroimaging studies. *Molecular Psychiatry*, *28*(7), 3013–3022.

Gupta, S., Bhuiyan, M. R. I., Chowa, S. S., Montaha, S., Rahman, R., Mehedi, S. T., & Rahman, Z. (2024). Enhancing autism spectrum disorder classification with lightweight quantized cnns and federated learning on abide-1 dataset. *Mathematics*, *12*(18), 2886.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Holiga, Š., Hipp, J. F., Chatham, C. H., Garces, P., Spooren, W., D'Ardhuy, X. L., ... others (2019). Patients with autism spectrum disorders display reproducible functional connectivity alterations. *Science Translational Medicine*, *11*(481), eaat9223.

Honnorat, N., Seshadri, S., Killiany, R., Blangero, J., Glahn, D. C., Fox, P., & Habes, M. (2024). Riemannian frameworks for the harmonization of resting-state functional mri scans. *Medical image analysis*, *91*, 103043.

Huynh, N., Yan, D., Ma, Y., Wu, S., Long, C., Sami, M. T., ... others (2024). The use of generative adversarial network and graph convolution network for neuroimaging-based diagnostic classification. *Brain Sciences*, *14*(5), 456.

Jahani, A., Jahani, I., Khadem, A., Braden, B. B., Delrobaei, M., & MacIntosh, B. J. (2024). Twinned neuroimaging analysis contributes to improving the classification of young people with autism spectrum disorder. *Scientific Reports*, *14*(1), 20120.

Jain, V., Rakshe, C. T., Sengar, S. S., Murugappan, M., & Ronickom, J. F. A. (2024). Age-and severity-specific deep learning models for autism spectrum disorder classification using functional connectivity measures. *Arabian Journal for Science and Engineering*, *49*(5), 6847–6865.

Joshi, G., Arnold Anteraper, S., Patil, K. R., Semwal, M., Goldin, R. L., Furtak, S. L., ... others (2017). Integration and segregation of default mode network resting-state functional connectivity in transition-age males with high-functioning autism spectrum disorder: a proof-of-concept study. *Brain Connectivity*, *7*(9), 558–573.

Kana, R. K., & Wadsworth, H. M. (2012). "the archeologist's career ended in ruins": hemispheric differences in pun comprehension in autism. *NeuroImage*, *62*(1), 77–86.

Kleinhans, N. M., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., ... Aylward, E. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain*, *131*(4), 1000–1012.

Kliemann, D., Galdi, P., Van De Water, A. L., Egger, B., Jarecka, D., Adolphs, R., & Ghosh, S. S. (2024). Resting-state functional connectivity of the amygdala in autism: a preregistered large-scale study. *American Journal of Psychiatry*, *181*(12), 1076–1085.

Lamani, M. R., Julian Benadit, P., & Guruprasad, C. (2024). Advandnn: Deep neural network analysis of neuroimaging for identifying vulnerable brain regions in autism spectrum disorder. In *International conference on power engineering and intelligent systems (peis)* (pp. 497–510).

Li, P., Zhou, M., Yan, W., Du, J., Lu, S., Xie, S., & Zhang, R. (2021). Altered resting-state functional connectivity of the right precuneus and cognition between depressed and non-depressed schizophrenia. *Psychiatry Research: Neuroimaging*, *317*, 111387.

Liu, R., Huang, Z.-A., Hu, Y., Huang, L., Wong, K.-C., & Tan, K. C. (2024). Spatio-temporal hybrid attentive graph network for diagnosis of mental disorders on fmri time-series data. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *8*(6), 4046-4058. doi: 10.1109/TETCI.2024.3386612

Liu, Y., Xu, L., Li, J., Yu, J., & Yu, X. (2020). Atten-

tional connectivity-based prediction of autism using heterogeneous rs-fmri data from cc200 atlas. *Experimental neurobiology*, *29*(1), 27.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., ... Beckett, L. (2005). The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, *15*(4), 869.

Nazir, M. I., Mazumder, T., Islam, M. M., Ehsan, M. A., Rahman, R., & Helaly, T. (2024). Enhancing autism spectrum disorder diagnosis through a novel 1d cnn-based deep learning classifier. In *2024 3rd international conference on advancement in electrical and electronic engineering (icaeee)* (pp. 1–6).

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *135*(3), 370–384.

Nomi, J. S., Molnar-Szakacs, I., & Uddin, L. Q. (2019). Insular function in autism: Update and future directions in neuroimaging and interventions. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *89*, 412–426.

Nomi, J. S., & Uddin, L. Q. (2015). Developmental changes in large-scale network connectivity in autism. *NeuroImage: Clinical*, *7*, 732–741.

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., & Rueckert, D. (2018). Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical image analysis*, *48*, 117–130.

Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., ... Calhoun, V. D. (2014). Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, *8*, 229.

Reardon, A. M., Li, K., & Hu, X. P. (2021). Improving between-group effect size for multi-site functional connectivity data via site-wise de-meaning. *Frontiers in computational neuroscience*, *15*, 762781.

Song, X., Zhou, F., Frangi, A. F., Cao, J., Xiao, X., Lei, Y., ... Lei, B. (2022). Multicenter and multichannel pooling gcn for early ad diagnosis based on dual-modality fused brain network. *IEEE Transactions on Medical Imaging*, *42*(2), 354–367.

Taki, B. A. (2024). *Low separation rank in tensor-variate generalized linear models*. Unpublished doctoral dissertation, Rutgers The State University of New Jersey, School of Graduate Studies.

Tang, S., Wu, Z., Cao, H., Chen, X., Wu, G., Tan, W., ... Liu, Z. (2022). Age-related decrease in default-mode network functional connectivity is accelerated in patients with major depressive disorder. *Frontiers in Aging Neuroscience*, *13*, 809853.

Wang, L., Li, K., & Hu, X. P. (2021). Graph convolutional network for fmri analysis based on connectivity neighborhood. *Network Neuroscience*, *5*(1), 83–95.

Warren, S. L., Khan, D. M., & Moustafa, A. A. (2024). Assistive tools for classifying neurological disorders using fmri and deep learning: A guide and example. *Brain and Behavior*, *14*(6), e3554.

Watanabe, T., & Rees, G. (2017). Brain network dynamics in high-functioning individuals with autism. *Nature communications*, *8*(1), 16048.

Xu, J., Bian, Q., Li, X., Zhang, A., Ke, Y., Qiao, M., ... Gulyás, B. (2024). Contrastive graph pooling for explainable classification of brain networks. *IEEE Transactions on Medical Imaging*, *43*(9), 3292-3305. doi: 10.1109/TMI.2024.3392988

Yang, L., Jin, C., Qi, S., Teng, Y., Li, C., Yao, Y., ... Wei, X. (2022). Alterations of functional connectivity of the lateral habenula in subclinical depression and major depressive disorder. *BMC psychiatry*, *22*(1), 588.

Zhang, J., Feng, F., Han, T., Gong, X., & Duan, F. (2023). Detection of autism spectrum disorder using fmri functional connectivity with feature selection and deep learning. *Cognitive Computation*, *15*(4), 1106–1117.

Zheng, X., Ravid, O., Barry, R. A., Kim, Y., Wang, Q., Kim, Y.-g., ... He, X. (2024). Denoising variational autoencoder as a feature reduction pipeline for the diagnosis of autism based on resting-state fmri. *arXiv preprint arXiv:2410.00068*.