Hierarchical Clustering Diffusion Model for fMRI Functional Connectivity to Enhance Autism Spectrum Disorder Diagnosis

Anonymous Author(s)
Affiliation
Address
email

Abstract

Functional magnetic resonance imaging (fMRI) data, particularly functional connectivity matrices, are crucial for studying brain disorders like Autism Spectrum Disorder (ASD). However, data scarcity often limits the performance of diagnostic models. We address this challenge by leveraging generative diffusion models for data augmentation. We introduce a novel transformer-based latent diffusion model, the Hierarchical Clustering Connective Diffusion Unit (HC-CDU), designed to synthesize realistic fMRI functional connectivity matrices. Our models effectively generate high-fidelity connectivity patterns, demonstrating an improvement of up to 3.61% in MAE reduction. In classification tasks on the ABIDE-I dataset, HC-CDU with ×1 augmentation demonstrated significant improvement, with AUC enhancing by up to 4.29% over baseline, showcasing enhanced discriminative power.

1 Background

Autism spectrum disorder (ASD) is a complex neurodevelopmental condition affecting communication, social interactions and behaviors (1). ASD diagnosis traditionally relies on clinical and behavioral assessments (2). Functional magnetic resonance imaging (fMRI) has emerged as an indispensable tool, offering insights into neural mechanisms underlying ASD via blood oxygen level-dependent (BOLD) signal analysis (3).

Machine learning has enabled numerous automated diagnostic frameworks based on functional connectivity analysis (6; 7). However, optimal performance requires large datasets. Collecting high-quality fMRI data is expensive, time-consuming and requires specialized equipment, leading to data scarcity that hampers ML model performance and generalizability.

fMRI data augmentation has explored traditional approaches such as adding Gaussian noise (8) or sliding windows on ROI signals (9). Sophisticated methods using Variational AutoEncoders (VAE) (10) and Generative Adversarial Networks (GAN) (11; 12) synthesize data but suffer from "posterior collapse" or "mode collapse" problems. Diffusion models (13; 14; 15) have emerged as powerful alternatives, demonstrating remarkable ability to generate high-quality medical imaging data.

2 Methods

The proposed HC-CDU is a transformer-based latent diffusion model enhanced by hierarchical clustering, as illustrated in Figure 1. For comparison, we also evaluate a non-hierarchical variant (CDU), which omits the hierarchical conditioning module. HC-CDU generates diverse and realistic fMRI functional connectivity matrices through three main components.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

2.1 Latent Connectivity Autoencoder

We adopt a Variational Quantized Autoencoder (VQ-VAE) to encode and discretize functional connectivity matrices into a structured latent space (16). The architecture consists of an encoder that maps a functional connectivity matrix $F \in \mathbb{R}^{V \times V}$ (where V is the number of ROIs) to a latent representation ζ_{enc}. Implemented as a CNN with convolutional layers and residual blocks, it projects F into a feature space of dimension $D_{lat} \times H_{lat} \times W_{lat}$. Each row of F represents one ROI's correlation with all others.

A vector quantization module discretizes the latent features, mitigating posterior collapse issues. A codebook $\mathcal{K} \in \mathbb{R}^{C \times D_{lat}}$ with C learnable vectors maps each latent feature $z_{enc,i} \in \zeta_{enc}$ to its closest codebook entry:

$$z_{quant,i} = \kappa_j, \quad where j = \arg\min_k \|z_{enc,i} - \kappa_k\|^2 \tag{1}$$

A decoder reconstructs the connectivity matrix $F' \in \mathbb{R}^{V \times V}$ from ζ_{quant} using deconvolutional layers. The latent autoencoder is trained by minimizing:

$$L_{VAE} = L_{recons} + L_{VQ} + L_{comm} \tag{2}$$

2.2 Conditional Diffusion Transformer with Hierarchical Conditioning

After training the latent connectivity autoencoder, learned latent representations ζ_{enc} serve as inputs to a conditional diffusion transformer, trained as a noise prediction network η_ψ. In HC-CDU, this model is conditioned on the subject's diagnostic state c, diffusion timestep k, and multi-scale hierarchical cluster embeddings derived via an integrated Hierarchical Clustering module, which performs two-level clustering to extract representations capturing functional brain organization at multiple scales.

The forward diffusion process progressively adds Gaussian noise to an initial latent representation ζ_0 over K time steps:

$$q(\zeta_k \mid \zeta_0) = \mathcal{N}(\zeta_k; \sqrt{\alpha_k}\zeta_0, (1 - \alpha_k)I) \tag{3}$$

where α_k is a scaling factor dependent on time step k.

The inverse diffusion process systematically denoises ζ_k to recover ζ_0 through a transformer-based noise prediction network. The architecture utilizes transformer blocks that predict noise present in ζ_k. To incorporate conditioning information (timestep k, class label), we employ adaptive layer normalization (adaLN):

$$adaLN(\zeta_k, k, C_{cond}) = \gamma(k, C_{cond}) \cdot \frac{\zeta_k - mean(\zeta_k)}{std(\zeta_k)} + \beta(k, C_{cond}) \tag{4}$$

The noise prediction network is trained by minimizing the L1 loss between predicted and true noise:

$$L_{diffusion} = \mathbb{E}_{\zeta_0,\eta,k}[\|\eta_k - \eta_\psi(\zeta_k, k, C_{cond})\|^2] \tag{5}$$

New functional connectivity matrices are generated by integrating the trained VQ-VAE and conditional diffusion transformer, modulated by hierarchical cluster embeddings. Starting from initial noise and target class label, the model iteratively denoises via reverse diffusion process, with hierarchical conditioning applied at each step.

3 Experiments and Results

Dataset: We developed and evaluated our methodology using resting-state fMRI data from the ABIDE-I dataset (17), with 505 ASD and 530 controls. The brain was segmented into 200 ROIs using the CC200 atlas (18). We adopted a stratified 5-fold cross-validation strategy with approximately 60% training, 20% validation, and 20% testing data.

Implementation: VQ-VAE trained 800 epochs (batch size 64, learning rate 3e-4, commitment cost 0.25). Architecture: 128 internal channels, 2 residual blocks, embedding dim 16, codebook size 768. Diffusion models trained 600 epochs (batch size 4, learning rate 3e-4, 100 timesteps, L1 loss).

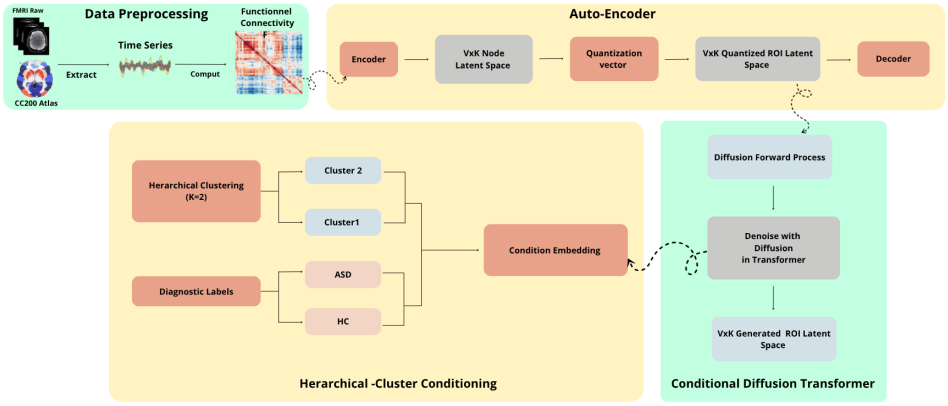

Figure 1: The comprehensive HC-CDU pipeline includes: (1) Data preprocessing to extract ROI signals and compute functional connectivity matrices; (2) a Latent Connectivity Autoencoder (VQ-VAE) that encodes inputs into structured latent space; and (3) a Conditional Diffusion Transformer that generates novel latent representations, guided by hierarchical clustering to capture multi-scale brain network structures.

Table 1: 5-Fold Cross-Validation Performance

Fold	No Augmentation				HC-CDU x1				HC-CDU x4				CDU x1				CDU x4			
	Acc	AUC	Sen	Spe	Acc	AUC	Sen	Spe	Acc	AUC	Sen	Spe	Acc	AUC	Sen	Spe	Acc	AUC	Sen	Spe
1	69.57	74.19	73.58	65.35	**71.50**	**78.48**	**76.42**	**66.34**	67.15	**75.39**	**79.25**	54.46	69.08	**76.66**	72.64	65.35	64.25	72.86	**77.36**	50.50
2	68.12	76.78	70.75	65.35	**70.05**	**79.28**	**71.70**	**68.32**	**69.08**	75.55	**73.58**	64.36	**69.08**	**76.81**	67.92	**70.30**	65.22	73.21	**77.36**	52.48
3	67.63	72.70	64.15	71.29	67.63	**73.68**	**66.98**	68.32	59.90	66.91	**79.25**	39.60	66.18	**73.70**	**66.98**	65.35	59.42	66.97	40.59	**77.36**
4	70.53	76.26	71.70	69.31	61.35	69.48	64.15	58.42	62.32	66.58	**72.64**	51.49	62.32	69.88	67.92	56.44	60.39	66.77	**79.25**	40.59
5	66.18	74.39	67.92	64.36	**67.63**	**75.68**	**70.75**	64.36	**66.67**	**75.74**	**69.81**	63.37	**66.67**	74.38	**70.75**	62.38	66.18	74.25	**81.13**	50.50

Table 2: Mean Absolute Error Assessment for Generated Functional Connectivity Matrices

Scenario	**MAE (Mean $\pm$ Std)**
Real Data Baseline	0.194 ± 0.001
HC-CDU x1	$\mathbf{0.187 \pm 0.003}$
HC-CDU x4	$\mathbf{0.187 \pm 0.003}$
CDU x1	$\mathbf{0.177 \pm 0.001}$
CDU x4	$\mathbf{0.178 \pm 0.001}$

Transformer: hidden size 128, depth 14, 8 heads. HC-CDU used 20 and 8 clusters with temperature 0.5. SVM classifier with RBF kernel for evaluation.

HC-CDU ×1 (Table 1) achieved balanced improvements. Fold 1: 78.48% AUC (+4.29% over baseline). Average: 67.63% accuracy, 75.32% AUC. CDU ×1: 74.29% AUC, 66.67% accuracy. Higher augmentation (×4) reduced performance but increased sensitivity. All models (Table 2) achieved lower MAE than baseline, confirming high-fidelity synthetic data generation.

4 Conclusion

We introduced HC-CDU, a hierarchical clustering-enhanced diffusion model for fMRI connectivity augmentation in ASD diagnosis. Moderate synthetic data augmentation significantly improves diagnostic performance while maintaining data fidelity. Hierarchical clustering provides benefits over non-hierarchical approaches, establishing a framework for addressing data scarcity in psychiatric neuroimaging research.

5 Broader Impact

Potential negative impacts include privacy concerns from synthetic data generation, bias amplification, and over-dependence on automated tools. We emphasize careful validation and human oversight.

References

[1] Lord, C., Elsabbagh, M., Baird, G., Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *Lancet*, 392(10146), 508–520.

[2] Vahia, V.N. (2013). Diagnostic and statistical manual of mental disorders 5: a quick glance. *Indian Journal of Psychiatry*, 55(3), 220–223.

[3] Ecker, C., Bookheimer, S.Y., Murphy, D.G. (2015). Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *The Lancet Neurology*, 14(11), 1121–1134.

[4] Assaf, M., et al. (2010). Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *NeuroImage*, 53(1), 247–256.

[5] Han, S., et al. (2024). Early prediction of dementia using fMRI data with a graph convolutional network approach. *Journal of Neural Engineering*, 21(1), 016013.

[6] Wang, C., Xiao, Z., Wu, J. (2019). Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data. *Physica Medica*, 65, 99–105.

[7] Yan, Y., et al. (2019). GroupINN: grouping-based interpretable neural network for classification of limited, noisy brain data. In *Proceedings of the 25th ACM SIGKDD International Conference*, pages 772–782.

[8] Pei, S., et al. (2022). Data augmentation for fMRI-based functional connectivity and its application to cross-site ADHD classification. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–15.

[9] Wang, X., et al. (2022). Contrastive functional connectivity graph learning for population-based fMRI classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 221–230.

[10] Qiang, N., et al. (2023). Functional brain network identification and fMRI augmentation using a VAE-GAN framework. *Computers in Biology and Medicine*, 165, 107395.

[11] Li, C., et al. (2021). BrainNetGAN: data augmentation of brain connectivity using generative adversarial network for dementia classification. In *Deep Generative Models and Data Augmentation Workshop*, pages 103–111.

[12] Zhang, X., et al. (2023). Diffusion model-based data augmentation for lung ultrasound classification with limited data. In *ML4H@ NeurIPS*, pages 664–676.

[13] Ho, J., Jain, A., Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 33, 6840–6851.

[14] Rombach, R., et al. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695.

[15] Zhao, H., et al. (2025). Diffusion transformer-augmented fMRI functional connectivity for enhanced autism spectrum disorder diagnosis. *Journal of Neural Engineering*, 22(1), 016044.

[16] van den Oord, A., et al. (2017). Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, 30.

[17] Autism Brain Imaging Data Exchange, `https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html`, last accessed 2024/05/22

[18] Craddock, R.C., et al. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8), 1914–1928.