# Graph Construction Dominates Model Choice in GCN-Based Autism Classification from Resting-State fMRI

Elena Pitsik
*Research institute of applied artificial intelligence and digital solutions*
*Plekhanov Russian University of Economics*
Moscow, Russia
0000-0003-1850-2394

Aleksandra Antipova
*Research institute of applied artificial intelligence and digital solutions*
*Plekhanov Russian University of Economics*
Moscow, Russia

2026 10th Scientific School Dynamics of Complex Networks and their Applications (DCNA) | 979-8-3195-2599-4/26/$31.00 ©2026 IEEE | DOI: 10.1109/DCNA72004.2026.11733421

***Abstract*—Graph neural networks are widely applied to autism spectrum disorder (ASD) classification from resting-state fMRI, yet the upstream choice of how brain graphs are constructed is often treated as a fixed preprocessing step. We compare two functional-connectivity graph construction strategies: thresholded Pearson correlation and Ledoit-Wolf shrinkage precision. We use the full 884-subject ABIDE I sample, under both a topology-only support vector machine (SVM) baseline and a graph convolutional network (GCN). Across both classifiers, graph construction changed test AUROC far more than model choice: Pearson-based graphs reached 0.675 AUROC with the GCN versus 0.549 for Ledoit-Wolf, while the GCN improved on the SVM baseline by only 0.03-0.04 AUROC at fixed graph construction. Post-hoc GNNExplainer analysis of the Pearson model highlighted occipital regions in ASD and medial-frontal, cingulate and cerebellar regions in typically developing controls, consistent with prior literature; the same procedure applied to the near-chance Ledoit-Wolf model produced a uniform, non-anatomical offset, serving as a cautionary negative control.**

***Keywords—autism spectrum disorder, functional connectivity, graph neural networks, ABIDE, explainability***

## I. Introduction

Resting-state functional MRI (rs-fMRI) offers a non-invasive window onto the intrinsic functional organisation of the brain, and has become a central modality for data-driven classification of autism spectrum disorder (ASD). Network-based analysis of brain activity has been used to characterise and detect a wide range of psychiatric and neurological conditions, including major depressive disorder [1]–[3], schizophrenia [4], disorders of consciousness [5] and ASD [6], [7]. The methodology for assessing fMRI-based functional networks has itself evolved from classical pairwise connectivity to higher-order interactions [8]. Because functional connectivity is naturally expressed as a network of interacting regions, graph neural networks (GNNs) are an increasingly popular choice for this task, in line with the broader adoption of AI-based methods in mental-health research and rehabilitation [9], [10].

However, the analyst must choose how to estimate connectivity between regions and how to sparsify the resulting dense matrix into an adjacency structure [8]. These choices are frequently made in a single line of a methods section and then held fixed, while the bulk of modelling effort is directed at the network architecture. In this work, using ABIDE I sample, we compare two contrasting graph construction strategies under two classifiers of very different capacity, and find that the construction choice dominates. We further apply post-hoc explainability, and use a deliberately weak model to illustrate how explanation outputs can mislead.

This work has been supported by Russian Science Foundation (grant No 26-71-10028).

## II. Methods

### A. Data and preprocessing

We used rs-fMRI data from the Autism Brain Imaging Data Exchange I (ABIDE I) [11], obtained through the Preprocessed Connectomes Project. Data were preprocessed with the DPARSF pipeline with band-pass filtering and global signal regression. Regional time series were extracted using the Automated Anatomical Labelling (AAL) atlas [12], giving 116 regions of interest (ROIs) per participant. After excluding subjects lacking required derivatives, N = 884 remained (408 ASD, 476 typically developing [TD]).

### B. Graph construction

Each participant was represented as a weighted undirected graph over the 116 ROIs. Two edge-definition strategies were compared. Under the Pearson strategy, pairwise correlation coefficients and two-sided p-values were computed between all ROI time-series pairs; edges were retained where $p < 0.05$, with negative correlations and self-loops set to zero. Under the Ledoit-Wolf strategy, the covariance matrix was estimated with Ledoit-Wolf shrinkage [13] and inverted to a precision matrix encoding partial correlations. As shrinkage yields no exact zeros, entries with $|\theta| < 0.1$ and the diagonal were zeroed.

For both strategies, node features were the full connectivity profile of each ROI (a row of the subject's adjacency matrix), and edge weights were the absolute values of retained entries.

### C. Classifiers

As a topology-only reference, weighted node strength (sum of absolute incident edge weights) gave a 116-dimensional descriptor per subject, classified by an RBF-kernel SVM with balanced class weights under 10-fold stratified cross-validation, standardising features on training folds only. The regularisation constant C was swept over five values to confirm stability.

The GCN [14] applies GCNConv propagation with edge weights, batch normalisation, ReLU and feature dropout, then concatenates global mean and max pooling before a two-layer MLP head producing one logit. Edge dropout regularises training. Model selection used 10-fold stratified CV with an inner validation split for early stopping on validation AUROC (patience 25, maximum 200 epochs), giving approximately 675/120/89 subjects for train/validation/test per fold. Only test-fold metrics are reported, as validation metrics are optimistically biased by their role in model selection [15], [16]. Hyperparameters (hidden dimension, dropout, edge


Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:27:12 UTC from IEEE Xplore. Restrictions apply.

dropout, learning rate, weight decay, and number of graph-convolution layers) were tuned by exhaustive grid search over 48 configurations, run independently for each construction strategy.

### D. Explainability and statistics

GNNExplainer [17] was applied per subject on held-out test folds, so no explained subject contributed to the model explaining it. Node importance was the mean of the learned node mask across feature dimensions, giving a 116-dimensional vector per subject. Group differences in node strength and in importance were tested per ROI with two-sided Mann-Whitney U tests. Given the exploratory scope, regional p-values are reported uncorrected at $\alpha = 0.05$, and compared against the ≈5.8 of 116 regions expected under the null; the false discovery rate implications are addressed in Section IV.

## III. Results

### A. Graph construction versus model choice

Graph construction had a substantially larger effect on performance than model choice (Table I). Under both classifiers, Pearson-based graphs markedly outperformed Ledoit-Wolf precision graphs.

TABLE I. MEAN CROSS-VALIDATED TEST AUROC (10-FOLD CV, N = 884)

| Graph construction | SVM (node strength) | GCN |
|---|---|---|
| Pearson ($p < 0.05$) | 0.637 ± 0.053 | **0.675** |
| Ledoit-Wolf ($\|\theta\| > 0.1$) | 0.520 ± 0.059 | 0.549 |

Ledoit-Wolf representations were near chance under both classifiers, stable across the SVM regularisation sweep (0.518-0.528). Mass-univariate testing of Ledoit-Wolf node strength flagged only 2 of 116 ROIs at $p < 0.05$ uncorrected, indicating negligible node-level signal in that representation.

### B. GCN classification

Grid search identified a compact single-layer architecture as optimal for Pearson graphs (hidden dimension 16, dropout 0.3, edge dropout 0.2, learning rate 0.0001, weight decay 0.0001), reaching a mean test AUROC of 0.675. Single-layer models outperformed two-layer models throughout; at otherwise identical settings AUROC fell from 0.637 to 0.585 when a second convolution was added. The GCN improved on the SVM baseline by 0.038 AUROC on Pearson graphs and 0.029 on Ledoit-Wolf graphs, which is modest relative to fold-level standard deviations of 0.04-0.06, whereas the 0.126 gap between construction strategies is not. Training loss fell below 0.10 within 50-70 epochs while validation loss rose after roughly epoch 10, with early stopping typically between epochs 29 and 109.

### C. Regional importance (Pearson model)

GNNExplainer importance profiles were highly similar between groups (Fig. 1). Importance varied across ROIs (≈0.05-0.10) but this variation was almost entirely shared between ASD and TD. Group differences were small (maximum $|\Delta| = 0.0063$, ≈8% of the mean), yet the 20 most differentiating ROIs (Fig. 2) showed a coherent spatial pattern: regions weighted more heavily in ASD were exclusively occipital (bilateral cuneus, bilateral superior occipital gyrus, left calcarine, right inferior temporal), whereas regions weighted more heavily in TD comprised cerebellar (vermis IV-V, right Crus II, left lobule IX), medial-frontal (bilateral gyrus rectus, medial orbitofrontal, superior medial frontal), posterior cingulate, right thalamus and left postcentral regions.

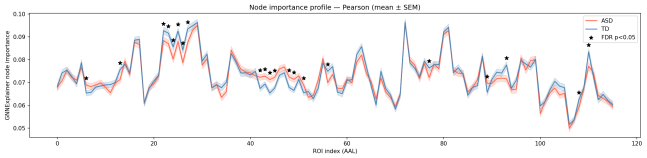



Fig. 1. Mean ± SEM GNNExplainer node importance across the 116 AAL ROIs for the Pearson-based GCN. ASD and TD profiles track one another closely; markers denote $p < 0.05$ (uncorrected).

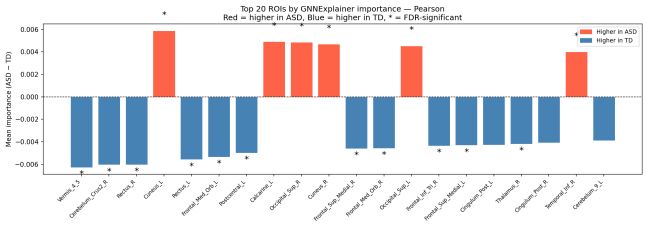



Fig. 2. The 20 ROIs with the largest ASD−TD importance difference (Pearson model). Positive (red) values indicate greater importance in ASD; markers denote $p < 0.05$ (uncorrected).

### D. Ledoit-Wolf explanations as a negative control

Applied to the near-chance Ledoit-Wolf model, the same procedure gave a qualitatively different, non-anatomical result: mean importance was higher in TD than ASD at all 116 regions, with a near-constant offset (≈0.267 vs ≈0.256) and no regional structure. Many regions reached uncorrected significance, as expected from a uniform group-level shift rather than regional specificity. The absolute importance scale also differed from the Pearson model (≈0.26 vs ≈0.07), tracking the greater edge density of the thresholded precision graphs.

## IV. Discussion

In this paper, we show that edge definition changed performance far more than classifier choice. Moving from Ledoit-Wolf to Pearson graphs improved AUROC by 0.126 (GCN) and 0.117 (SVM), while moving from SVM to GCN improved it by only 0.029-0.038 at fixed construction.

The 0.675 AUROC is consistent with connectome-based ASD classification on comparable ABIDE subsets: 67% accuracy in the benchmark of Abraham et al. [18] and 0.75 AUROC for the population-graph GCN of Parisot et al. [19], which additionally uses phenotypic metadata. Therefore, single-modality ABIDE classification remains an open problem.

The recovered pattern for Pearsons model agrees with reports of local visual overconnectivity with medial-frontal and cingulate underconnectivity [20], default-mode involvement of medial prefrontal and posterior cingulate hubs [21], and altered cerebellar connectivity of the posterior vermis and Crus II [22], [23]. The Ledoit-Wolf control shows the converse: a near-chance model yields explanations that are statistically flagged yet carry no regional information. The count of significant regions is therefore not a measure of explanation quality, and classifier competence is a precondition for interpretation.


Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:27:12 UTC from IEEE Xplore. Restrictions apply.

**Limitations.** Regional p-values are uncorrected; under Benjamini-Hochberg control no region survives, so the regional findings are exploratory.

## V. Conclusion

On the full ABIDE I sample, the choice of functional-connectivity graph construction affected ASD-classification performance far more than the choice between a linear baseline and a GCN. Thresholded Pearson correlation clearly outperformed Ledoit-Wolf precision under both classifiers. Explainability recovered anatomically plausible regions for the competent Pearson model but degenerated into a uniform artefact for the near-chance model, underscoring that explanations should be read only from models that first demonstrate predictive validity.

## References


[1] D. Stoyanov, V. Khorev, R. Paunova, S. Kandilarova, D. Simeonova, A. Badarin, A. Hramov, and S. Kurkin, "Resting-state functional connectivity impairment in patients with major depressive episode," Int. J. Environ. Res. Public Health, vol. 19, no. 21, art. 14045, 2022, doi: 10.3390/ijerph192114045.

[2] V. S. Khorev, S. A. Kurkin, G. Zlateva, R. Paunova, S. Kandilarova, M. Maes, D. Stoyanov, and A. E. Hramov, "Disruptions in segregation mechanisms in fMRI-based brain functional network predict the major depressive disorder condition," Chaos, Solitons & Fractals, vol. 188, art. 115566, 2024.

[3] S. A. Kurkin, N. M. Smirnov, R. Paunova, S. Kandilarova, D. Stoyanov, L. Mayorova, and A. E. Hramov, "Beyond pairwise interactions: higher-order Q-analysis of fMRI-based brain functional networks in patients with major depressive disorder," IEEE Access, vol. 12, pp. 197168-197186, 2024.

[4] V. S. Khorev, S. A. Kurkin, D. Stoyanov, R. Paunova, S. Najar, S. Kandilarova, and A. E. Hramov, "Hierarchical network disruptions in schizophrenia: a multi-level fMRI study of functional connectivity," Psychiatry Res.: Neuroimaging, vol. 354, art. 112078, 2025, doi: 10.1016/j.pscychresns.2025.112078.

[5] S. A. Kurkin, L. A. Mayorova, V. S. Khorev, E. N. Pitsik, M. L. Radutnaya, E. L. Bondar, and A. E. Hramov, "Multiscale fMRI analysis reveals hierarchical network disruptions underlying disorders of consciousness," Chaos, Solitons & Fractals, vol. 200, art. 117008, 2025.

[6] E. Pitsik, S. Kurkin, O. Martynova, G. Portnova, and A. E. Hramov, "Hypergraph representation of multilayer brain network enhances autism spectrum disorder detection," Chaos, vol. 35, no. 7, art. 071104, 2025, doi: 10.1063/5.0279835.

[7] M. S. Kabir, S. Kurkin, G. Portnova, O. Martynova, Zh. Wang, and A. Hramov, "Contrastive machine learning reveals in EEG resting-state network salient features specific to autism spectrum disorder," Chaos, Solitons & Fractals, vol. 185, art. 115123, 2024, doi: 10.1016/j.chaos.2024.115123.

[8] S. A. Kurkin, A. N. Pisarchik, L. A. Mayorova, and A. E. Hramov, "Evolution of methods for assessing fMRI-based functional networks: from classical pairwise connectivity to higher-order interactions," Phys. Rep., vol. 1174, pp. 1-66, 2026.

[9] V. Khorev, A. Kiselev, A. Badarin, V. Antipov, O. Drapkina, S. Kurkin, and A. Hramov, "Review on the use of AI-based methods and tools for treating mental conditions and mental rehabilitation," Eur. Phys. J. Spec. Top., vol. 234, pp. 4139-4158, 2025, doi: 10.1140/epjs/s11734-024-01289-x.

[10] V. Khorev, S. Kurkin, A. Badarin, V. Antipov, E. Pitsik, A. Andreev, V. Grubov, O. Drapkina, A. Kiselev, and A. Hramov, "Review on the use of brain computer interface rehabilitation methods for treating mental and neurological conditions," J. Integr. Neurosci., vol. 23, no. 7, art. 125, 2024, doi: 10.31083/j.jin2307125.

[11] A. Di Martino et al., "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," Mol. Psychiatry, vol. 19, no. 6, pp. 659-667, 2014.

[12] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," NeuroImage, vol. 15, no. 1, pp. 273-289, 2002.

[13] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," J. Multivar. Anal., vol. 88, no. 2, pp. 365-411, 2004.

[14] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Represent. (ICLR), 2017.

[15] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," PLoS ONE, vol. 14, no. 11, e0224365, 2019.

[16] C. Flint et al., "Systematic misestimation of machine learning performance in neuroimaging studies of depression," Neuropsychopharmacology, vol. 46, pp. 1510-1517, 2021.

[17] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNNExplainer: generating explanations for graph neural networks," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 32, pp. 9240-9251, 2019.

[18] A. Abraham et al., "Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example," NeuroImage, vol. 147, pp. 736-745, 2017.

[19] S. Parisot et al., "Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease," Med. Image Anal., vol. 48, pp. 117-130, 2018.

[20] J. O. Maximo, C. L. Keown, A. Nair, and R. A. Müller, "Approaches to local connectivity in autism using resting state functional connectivity MRI," Front. Hum. Neurosci., vol. 7, art. 605, 2013.

[21] M. Assaf et al., "Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients," NeuroImage, vol. 53, no. 1, pp. 247-256, 2010.

[22] G. Olivito et al., "Lobular patterns of cerebellar resting-state connectivity in adults with autism spectrum disorder," Eur. J. Neurosci., vol. 47, no. 6, pp. 729-735, 2018.

[23] S. Arnold Anteraper et al., "Abnormal cortico-cerebellar functional connectivity in autism spectrum disorder," Front. Syst. Neurosci., vol. 12, art. 74, 2018.




Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:27:12 UTC from IEEE Xplore. Restrictions apply.