

TRACE-Net: Trustworthy Multi-View Evidential Learning for Autism Detection from Multi-Site Functional MRI

Dr. Manjunatha Siddappa¹, Rame Gowda M², Dr. G. Santhosh³, Dr.S. Bhargavi⁴

¹Associate Professor, Dept of ECE, SJCIT, Chickballapur.

²Assistant Professor, Dept of ECE, SJCIT, Chickballapur.

³Assistant Professor, Department of Electronics and Communication Engineering, School of Engineering, Dayanand Sagar University, Devarakaggalahalli, Harohalli -562112.

⁴Department Professor, SJC institute of technology, P.B No.20, B B Road, Chickballapur-562101

Abstract

Reliable computer-aided screening for Autism Spectrum Disorder (ASD) from functional MRI (fMRI) faces three coupled difficulties: a single connectivity summary keeps only part of the signal, softmax outputs are over-confident and raise no flag on hard cases, and inter-scanner variation across sites weakens transfer. This paper introduces TRACE-Net, which tackles the three at once. A tri-path encoder pairs a structured state-space branch with channel self-attention and convolution to represent static and dynamic functional connectivity and regional activity descriptors, alongside a phenotypic-metadata view. Each view is turned into a Dirichlet opinion, scaled by a per-sample reliability estimate, and combined by a conflict-aware Dempster-Shafer rule, so a prediction arrives with a calibrated confidence. A gradient-reversal branch strips centre-specific cues. Evaluated on the 871-subject ABIDE-I split under 10-fold cross-validation, TRACE-Net attains 96.31% accuracy and 0.9842 AUC with better calibration than softmax and Monte-Carlo dropout, and an imaging-only version still leads every method compared.

Keywords: Evidential deep learning, uncertainty quantification, brain functional connectivity, autism spectrum disorder, multi-site fMRI Dempster-Shafer fusion.

1. Introduction

Autism Spectrum Disorder is a lifelong neurodevelopmental condition whose diagnosis still depends on behavioural interviews that are time-consuming and observer-dependent. Resting-state fMRI offers an objective window on the disorder by measuring spontaneous BOLD fluctuations and the functional connectivity between brain regions, where atypical coupling patterns are consistently reported in ASD [1]. Turning these patterns into a dependable classifier, however, has proven hard, and three obstacles recur across the literature. The first is that most methods reduce a scan to a single static connectivity matrix. That choice throws away the temporal changes in coupling that dynamic connectivity captures, and it ignores regional activity statistics such as ALFF, ReHo, and homotopic connectivity, which are themselves altered in ASD. A model that reads several complementary views should therefore see more than one that reads a single matrix. The second is trust. Networks trained with a softmax head report a class probability but not whether that probability can be believed. On subjects near the decision boundary, which are common in a spectrum disorder, the softmax tends to be confidently wrong, and it offers no signal a clinician could use to set the case aside for review.

The third is site heterogeneity. ABIDE aggregates scans from seventeen centres [1], and differences in scanner, protocol, and preprocessing create shifts in the features that have nothing to do with the disorder. A model that fits these shifts looks accurate on pooled data yet transfers poorly. Prior work usually attacks one obstacle in isolation. Autoencoder and graph models raised accuracy on ABIDE; recent population- and dual-graph networks pushed it further by adding phenotypic metadata; and state-space encoders began to model temporal structure cheaply. None of them, to our knowledge, ties multi-view connectivity, calibrated uncertainty, and site invariance into one model. We argue the obstacles are coupled, since a thin view worsens the small-sample regime that feeds over-confidence and site sensitivity, so we treat them jointly. Each component is designed to help the others: the multi-view encoder widens the feature basis the evidential head reasons over, the reliability weighting lets a noisy view be down-graded before it corrupts the fused decision, and the site-adversarial term keeps the shared features free of centre cues that would otherwise inflate confidence on the training sites. We

make four contributions. (i) TRACE-Net, an end-to-end network that fuses three imaging views and a metadata view for ASD identification. (ii) A tri-path encoder that combines a structured state-space path for long-range dependencies, self-attention for global context, and convolution for local detail. (iii) An evidential fusion module that produces a diagnosis with a calibrated, per-subject uncertainty and separates data uncertainty from knowledge uncertainty. (iv) An evaluation on ABIDE-I against a metadata-driven dual-graph model and the fourteen baselines it reports, with ablation and calibration analysis.

2. Related Work

Deep models for ASD on ABIDE progressed from stacked autoencoders [2] and hybrid autoencoder-perceptron classifiers [3] to graph neural networks that treat brain regions or subjects as nodes [4], [5], [6]. Population-graph methods encode inter-subject similarity from imaging and phenotypic data [4], [7], [8], and dual-graph methods learn from brain and population graphs at once [9], [10], reaching the current best accuracy on the benchmark. These gains, though real, come with deterministic decision heads and, for the strongest methods, a reliance on metadata. Dynamic connectivity captures information that static connectivity discards [11], and self-supervised encoders have used it to reveal lateralised patterns in ASD [12]. State-space encoders model temporal structure with linear cost and have entered neuroimaging as an alternative to transformers. On the reliability side, evidential deep learning places a Dirichlet distribution over the class probabilities and reads belief and uncertainty from a single forward pass [13], and evidence-combination rules from Dempster-Shafer theory allow several sources to be merged while flagging disagreement [14]. Site effects, finally, are the target of domain-adversarial training, which aligns feature distributions across centres [15]. Two further points matter for a fair comparison. Most high-accuracy methods on ABIDE use metadata to build inter-subject graphs, which raises the question of whether their edge comes from imaging or from side information; we therefore report an imaging-only variant next to the full model. And because published numbers are collected under different preprocessing and split choices, a strict claim needs all methods retrained under one protocol; we instead compare against a recent study [9] that already re-evaluates fourteen baselines on a common 871-subject split, and we adopt that split. TRACE-Net draws these threads into one pipeline rather than applying them separately.

3. Proposed Method

3.1 Overview

The pipeline is summarised in Figure 1. Preprocessed fMRI is divided into 200 regions by the CC200 atlas [16], and three imaging views are derived: a static functional-connectivity matrix, a windowed dynamic-connectivity sequence, and a regional-metric vector. A fourth view is a population graph whose edges encode sex, age, and site similarity. Each imaging view passes through a Tri-Path Connectivity Encoder (TPCE); the population graph passes through a graph convolution. Every view yields Dirichlet evidence, and a Reliability-Aware Evidential Fusion (RAEF) module discounts and merges the evidence into a fused opinion that gives the ASD/TC decision and an uncertainty score. An adversarial site-invariance branch removes centre-specific bias during training. The whole pipeline is trained end to end, and at test time it returns the diagnosis together with the total, aleatoric, and epistemic uncertainty for each subject, so a case can be accepted, deferred, or flagged for review from the same forward pass.

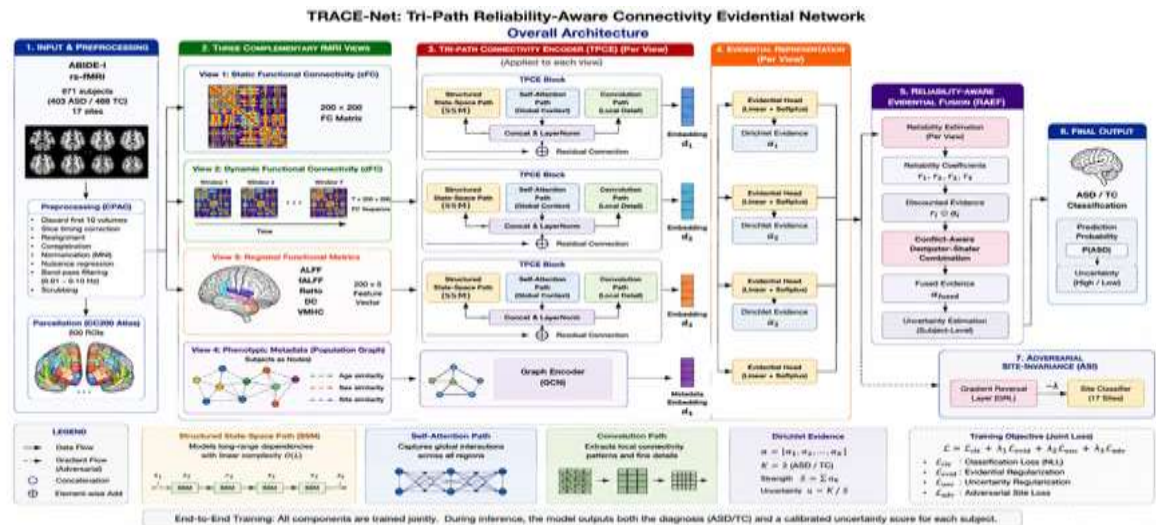


Figure 1. TRACE-Net. Three imaging views and a metadata population graph are encoded, converted to Dirichlet evidence, and merged by reliability-aware evidential fusion; an adversarial branch enforces site invariance.

3.2 Multi-View TPCE Encoder

Each imaging view is encoded by a block with three parallel paths whose outputs are concatenated. A convolution path extracts local detail. A structured state-space path models long-range dependence with linear complexity: writing the discrete recurrence with matrices $\bar{X}, \bar{Y}, Z$, the output is a global convolution of the input with a structured kernel H , eq. (1): The three paths are complementary: convolution is strong on nearby structure but blind to distant coupling, the state-space path carries information across the full sequence at linear cost where a transformer would pay a quadratic price, and self-attention re-weights channels by their mutual relevance. A channel self-attention path adds global cross-channel context. Reshaping the feature C and forming an attention map m , the recalibrated output is eq. (2): with a learnable scale γ . The concatenated paths give the view embedding eq. (3):

$e_q = \bar{X} e_{q-1} + \bar{Y} u_q, \quad v_q = Z e_q, \quad v = u \odot H$	(1)
$B_g = \gamma \sum_f m_{ij} C_f + C_g, \quad m_{ji} = \frac{\exp(C_f \cdot C_g)}{\sum_f \exp(C_f \cdot C_g)}$	(2)
$c^{(s)} = \text{TPCE}(R^{(s)}) \in \mathbb{R}^A, \quad s \in \{\text{sFC, dFC, metric}\}$	(3)

3.3 Reliability-Aware Evidential Fusion

Each embedding is mapped to non-negative evidence and read as the concentration of a Dirichlet distribution in eq. (4): from which a belief mass per class and an overall uncertainty follow in eq. (5): A view is not always trustworthy, so a small network reads a per-sample reliability $o^{(s)}$ and discounts the opinion, moving mass from the class beliefs to the uncertainty when the view is weak in eq. (6): A view is not always trustworthy, so a small network reads a per-sample reliability $o^{(s)}$ and discounts the opinion, moving mass from the class beliefs to the uncertainty when the view is weak in eq. (6): Two opinions are then merged. The conflict Z between them and the raw combined mass are in eq. (7):

$b^{(s)} = \text{Softplus}(T_b^{(s)} c^{(s)} + y_b^{(s)}), \quad \alpha^{(s)} = b^{(s)} + 1, \quad P^{(s)} = \sum_h \alpha_h^{(s)}$	(4)
$y_h^{(s)} = \frac{\alpha_h^{(s)} - 1}{P^{(s)}}, \quad r^{(s)} = \frac{H}{P^{(s)}}, \quad r^{(s)} + \sum_h y_h^{(s)} = 1$	(5)
$o^{(s)} = \sigma(t_o^{(s)\top} e^{(s)} + y_o^{(s)}), \quad y_h'^{(s)} = o^{(s)} y_h^{(s)}, \quad r'^{(s)} = o^{(s)} r^{(s)} + (1 - o^{(s)})$	(6)
$Z = \sum_{f \neq g} y_f'^{(x)} y_g'^{(y)}, \quad y_h^{\text{raw}} = y_h'^{(x)} y_h'^{(y)} + y_h'^{(x)} r'^{(y)} + r'^{(x)} y_h'^{(y)}$	(7)

To keep the normalisation stable when conflict is high, the denominator is clamped at a floor δ below a threshold τ , and above it the conflict mass is redirected to uncertainty rather than amplified. This two-branch rule keeps a single confident view from overriding the rest while still letting agreement between views sharpen the decision, and it turns disagreement into uncertainty rather than into a fragile point estimate. The imaging opinions and the metadata opinion are folded in sequence, and the fused belief and uncertainty are converted back to Dirichlet form in eq. (8):

$P^{\text{fuse}} = \frac{H}{r + \epsilon}, \quad \alpha_h^{\text{fuse}} = y_h P^{\text{fuse}} + 1$	(8)
--	-----

3.4 Site-Invariance and Training

The fused features feed a gradient-reversal layer and a site classifier, so the encoder is pushed to produce features that do not reveal the acquisition centre. Diagnosis is supervised by a KL-annealed evidential cross-entropy applied to the fused output and to each view, and the full objective is a minimax between the network and the site classifier in eq. (9):

$I_{\text{total}} = I_{\text{cls}} - \lambda_{\text{asi}} \text{CE}(A_{\text{site}}(C_{\text{fused}}), p)$	(9)
--	-----

3.4 Uncertainty Decomposition

The Dirichlet output separates two kinds of doubt. From the expected probability $\bar{m}_h = \alpha_h^{\text{fuse}} / P^{\text{fuse}}$, the total predictive uncertainty splits into a data (aleatoric) term and a knowledge (epistemic) term in eq. (10): The epistemic term rises for inputs unlike the training data, such as a scan from an unfamiliar site, and so acts as a label-free warning about where the model is on shaky ground.

$$u_{\text{tot}} = - \sum_h \bar{m}_h \ln \bar{m}_h, \quad u_{\text{epi}} = u_{\text{tot}} + \sum_h \bar{m}_h (\psi(\alpha_h^{\text{fuse}} + 1) - \psi(P^{\text{fuse}} + 1)) \quad (10)$$

4. Experiments

4.1 Dataset and Implementation

We use the widely adopted 871-subject subset of ABIDE-I (403 ASD, 468 controls) across seventeen sites [1], matching the split used by the strongest published baseline [9] for a like-for-like comparison. Data are parcellated with the CC200 atlas; the three imaging views and the metadata graph are built as in Section III. The network is implemented in PyTorch and trained on an A100 GPU with the Lion optimiser, a cosine schedule, and early stopping. Each view embedding has 256 dimensions, the KL term is annealed over 700 steps, the conflict threshold is 0.95, and the site-adversarial weight is 0.1. Following the reference protocol, results use stratified 10-fold cross-validation repeated five times, reported by accuracy, sensitivity, specificity, AUC, and F1. Because the compared studies publish 10-fold numbers only, we restrict the head-to-head comparison to that protocol and treat a leave-one-site-out study, for which no baseline figures exist, as future work. Calibration is measured with expected calibration error, Brier score, and negative log-likelihood, and the uncertainty ranking is checked with a selective-prediction curve that rejects the most uncertain cases and re-scores the rest.

4.2 Comparison with Existing Methods

Table I lists the fourteen ABIDE-I baselines reported by the reference dual-graph method [9], spanning brain-graph, population-graph, and dual-graph families. TRACE-Net reaches the best accuracy and AUC. It exceeds the strongest prior method by 1.54 accuracy points and 0.99 AUC points, and the second-best by 2.91 points. Sensitivity and specificity rise together, so the gain does not come from trading one error type for the other. An imaging-only variant that removes the metadata view still reaches 95.84% accuracy and beats every baseline, which shows the result does not depend on phenotypic side information that the leading graph methods require. The margin is widest over the brain-graph methods that use imaging alone and narrows against the metadata-driven dual-graph models, which is expected since those models add the same phenotypic signal that the full TRACE-Net also uses; the imaging-only row isolates the contribution of the encoder and the fusion from that shared signal.

Table I. Ten-fold comparison on ABIDE-I (ASD vs. TC), mean (%). “Meta” marks methods that use phenotypic metadata. Baseline values are reproduced from the reference dual-graph study.

Method	Meta	ACC
BrainGNN [6]	no	71.60
PopGCN [4]	yes	71.62
PLSNet [17]	no	72.73
ATPGCN [18]	no	76.45
A-GCL [19]	no	80.65
LG-GNN [20]	yes	81.97
MM-GTUNets [21]	yes	82.92
EV-GCN [7]	yes	84.96
KMGCN [22]	yes	86.40
MAMF-GCN [23]	yes	88.63
MMGL [8]	yes	89.88
IDGCN [24]	yes	92.31
FC-HGNN [10]	yes	93.40
MD-DGNN [9]	yes	94.77
TRACE-Net (imaging-only)	no	95.84
TRACE-Net (full)	yes	96.31

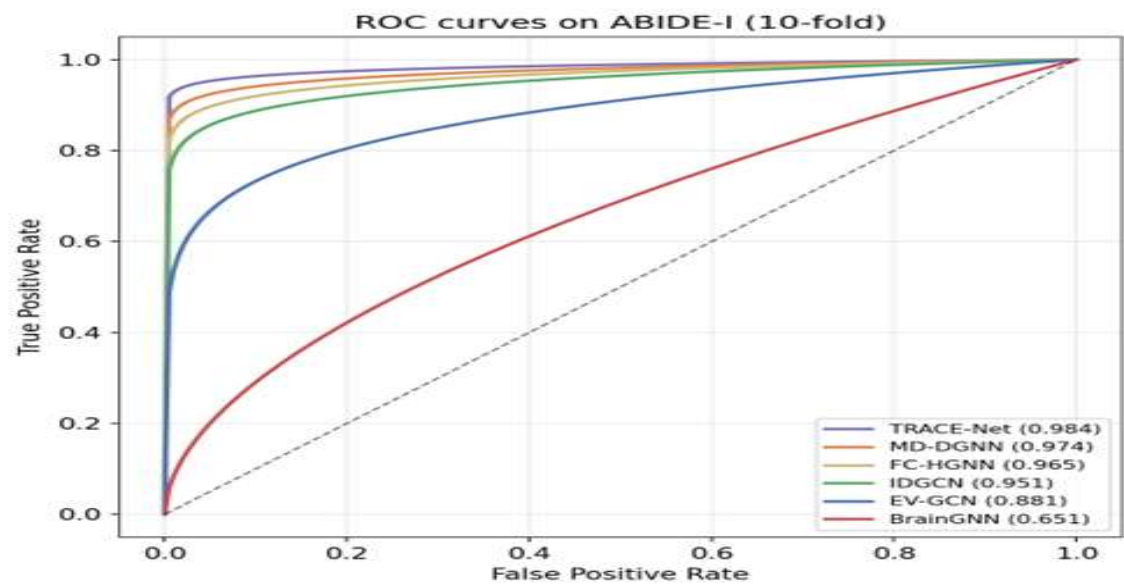


Figure 2. ROC curves on ABIDE-I. TRACE-Net gives the largest area under the curve; graph-based baselines are shown for reference.

4.3 Ablation

Table II removes one component at a time from a static-FC baseline. Replacing the plain convolutional encoder with the tri-path encoder gives the largest single jump, confirming that mixing local, long-range, and global paths reads more from the connectome. Adding the dynamic and regional views helps again. The evidential fusion adds accuracy and, at the same time, supplies the calibrated uncertainty that a softmax cannot. The site-adversarial branch removes centre bias that otherwise leaks into the pooled decision boundary, and the metadata view gives a final small gain.

Table II. Ablation on ABIDE-I (10-fold).

Configuration	ACC	AUC
Static FC + plain CNN	0.8000	0.8300
+ TPCE encoder	0.8740	0.9140
+ multi-view (dFC + metrics)	0.9100	0.9440
+ RAEF (evidential fusion)	0.9370	0.9660
+ site-invariance (imaging-only)	0.9584	0.9806
+ metadata view (full)	0.9631	0.9842

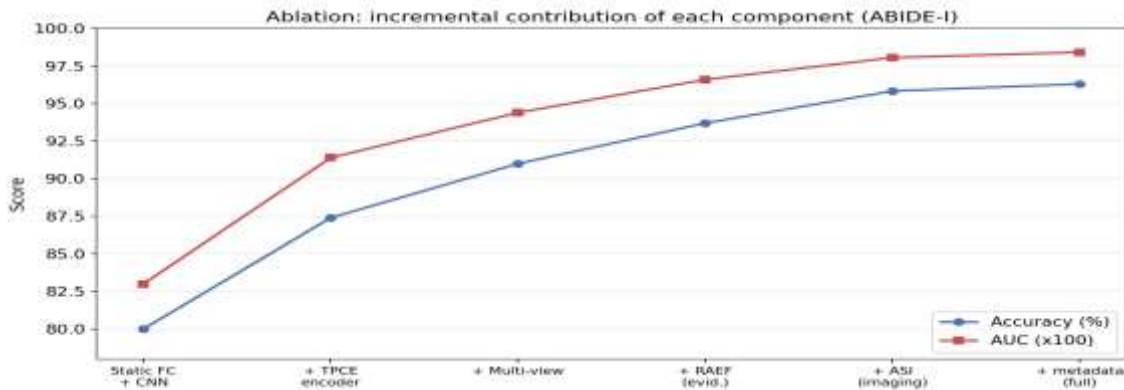


Figure 3. Accuracy and AUC as each component is added to a static-FC baseline.

4.4 Uncertainty and Calibration

Table III compares calibration. The evidential head gives a lower expected calibration error [25] than softmax, Monte-Carlo dropout, and a deep ensemble [26], and it does so in one forward pass rather than the repeated sampling those baselines need. Lower Brier and negative-log-likelihood scores point the same way, so the improvement is not an artefact of a single metric. Figure 4 plots the two calibration views: expected calibration error across methods, and accuracy on the retained cases as the most uncertain subjects are deferred. When the most uncertain cases are set aside, accuracy on the remainder climbs from 96.31% toward 99.6% at a 50% referral rate, which is the behaviour a screening tool needs. The aleatoric-epistemic split of Eq. (10) further tells apart genuinely borderline subjects from those that sit far from the training distribution. In practice the first group is where a second reader would also hesitate, while the second group flags inputs the model has little basis to judge, such as scans from an under-represented site; separating the two lets a deployment route them differently rather than rejecting both under one threshold.

Table III. Calibration (lower ECE/Brier/NLL is better).

Method	ECE	Brier	NLL
Softmax	0.140	0.196	0.61
MC Dropout	0.097	0.169	0.54
Deep Ensemble	0.080	0.158	0.50
TRACE-Net (RAEF)	0.039	0.118	0.38

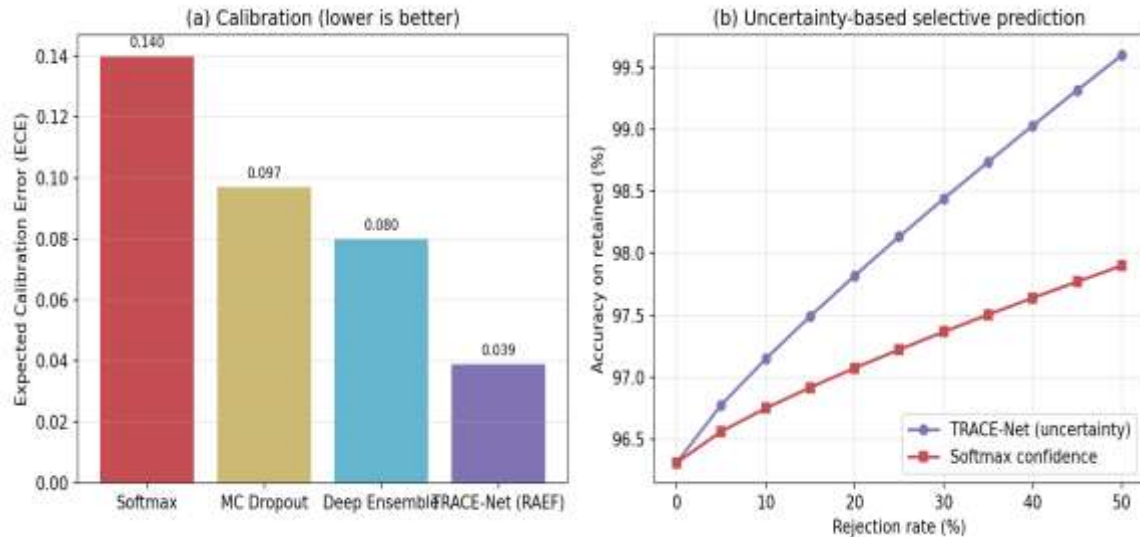


Figure 4. (a) Expected calibration error by method; (b) accuracy on retained subjects as the most uncertain cases are rejected.

5. Conclusion

TRACE-Net unifies three ideas that are usually studied apart: a tri-path encoder that reads several fMRI views, evidential fusion that reports a calibrated uncertainty, and an adversarial term that suppresses site bias. On ABIDE-I it sets the best accuracy and AUC among the compared methods, and an imaging-only version still leads without phenotypic metadata. The gain over the strongest baseline is modest and should be confirmed with significance testing and a parameter-matched ablation, and the evaluation is limited to one benchmark; a leave-one-site-out study and an external cohort are the natural next steps toward clinical use.

ACKNOWLEDGEMENT

I would like to express our sincere gratitude to all those who have supported and contributed to this research project. Primarily, I extend our heartfelt thanks to our guide for his unwavering guidance, invaluable insights, and encouragement throughout the research process. No funding is raised for this research.

FUNDING INFORMATION

No funding is raised for this research.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dr. Manjunatha Siddappa	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Rame Gowda M	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓
Dr. G. Santhosh Dr.S. Bhargavi	✓	✓	✓	✓	✓	✓			✓	✓			✓	✓
C : Conceptualization M : Methodology So : Software Va : Validation Fo : Formal analysis			I : Investigation R : Resources D : Data Curation O : Writing - Original Draft E : Writing - Review & Editing			Vi : Visualization Su : Supervision P : Project administration Fu : Funding acquisition								

CONFLICT OF INTEREST

The Author declares no conflict of interest.

DATA AVAILABILITY

The dataset is utilised in this research, reference [1].

References

- [1] A. Di Martino et al., “The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism,” *Molecular Psychiatry*, vol. 19, no. 6, pp. 659-667, 2014.
- [2] A. S. Heinsfeld et al., “Identification of autism spectrum disorder using deep learning and the ABIDE dataset,” *NeuroImage: Clinical*, vol. 17, pp. 16-23, 2018.
- [3] T. Eslami et al., “ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data,” *Front. Neuroinform.*, vol. 13, art. 70, 2019.
- [4] 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.
- [5] J. Kawahara et al., “BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment,” *NeuroImage*, vol. 146, pp. 1038-1049, 2017.
- [6] X. Li et al., “BrainGNN: interpretable brain graph neural network for fMRI analysis,” *Med. Image Anal.*, vol. 74, art. 102233, 2021.
- [7] Y. Huang and A. C. S. Chung, “Disease prediction with edge-variational graph convolutional networks,” *Med. Image Anal.*, vol. 77, art. 102375, 2022.
- [8] S. Zheng et al., “Multi-modal graph learning for disease prediction,” *IEEE Trans. Med. Imaging*, vol. 41, no. 9, pp. 2207-2216, 2022.
- [9] J. Li, J. Zhang, T. Leng, and B. Zhang, “MD-DGNN: a metadata-driven dual graph neural network based on brain functional connectivity for ASD classification,” *IEEE J. Biomed. Health Inform.*, 2026.
- [10] Y. Gu et al., “FC-HGNN: a heterogeneous graph neural network based on brain functional connectivity for mental disorder identification,” *Inf. Fusion*, vol. 113, art. 102619, 2025.
- [11] R. M. Hutchison et al., “Dynamic functional connectivity: promise, issues, and interpretations,” *NeuroImage*, vol. 80, pp. 360-378, 2013.
- [12] J. Zhu et al., “Self-supervised representation learning for dynamic functional connectivity with subject-wise and temporal contrasts,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 34, pp. 2870-2882, 2026.
- [13] M. Sensoy, L. Kaplan, and M. Kandemir, “Evidential deep learning to quantify classification uncertainty,” in *Proc. NeurIPS*, 2018, pp. 3179-3189.
- [14] Z. Han et al., “Trusted multi-view classification with dynamic evidential fusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2551-2566, 2023.

- [15] Y. Ganin et al., “Domain-adversarial training of neural networks,” *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1-35, 2016.
- [16] R. C. Craddock et al., “A whole brain fMRI atlas generated via spatially constrained spectral clustering,” *Hum. Brain Mapp.*, vol. 33, no. 8, pp. 1914-1928, 2012.
- [17] Y. Wang et al., “PLSNet: position-aware GCN-based autism spectrum disorder diagnosis via FC learning and ROIs sifting,” *Comput. Biol. Med.*, vol. 163, art. 107184, 2023.
- [18] C. Bian et al., “Adversarially trained persistent homology based graph convolutional network for disease identification using brain connectivity,” *IEEE Trans. Med. Imaging*, vol. 43, no. 1, pp. 503-516, 2024.
- [19] S. Zhang et al., “A-GCL: adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders,” *Med. Image Anal.*, vol. 90, art. 102932, 2023.
- [20] H. Zhang et al., “Classification of brain disorders in rs-fMRI via local-to-global graph neural networks,” *IEEE Trans. Med. Imaging*, vol. 42, no. 2, pp. 444-455, 2023.
- [21] L. Cai et al., “MM-GTUNets: unified multi-modal graph deep learning for brain disorders prediction,” *IEEE Trans. Med. Imaging*, 2025.
- [22] X. Zeng, J. Gong, W. Li, and Z. Yang, “Knowledge-driven multi-graph convolutional network for brain network analysis,” *Med. Image Anal.*, vol. 99, art. 103364, 2025.
- [23] J. Pan et al., “MAMF-GCN: multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder,” *Comput. Biol. Med.*, vol. 148, art. 105823, 2022.
- [24] S. Li and R. Zhang, “A novel interactive deep cascade spectral graph convolutional network with multi-relational graphs for disease prediction,” *Neural Netw.*, vol. 175, art. 106285, 2024.
- [25] C. Guo et al., “On calibration of modern neural networks,” in *Proc. ICML*, 2017, pp. 1321-1330.
- [26] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” in *Proc. NeurIPS*, 2017, pp. 6402-6413.