



## OPEN ACCESS

## EDITED BY

Yuchou Chang,  
University of Massachusetts Dartmouth,  
United States

## REVIEWED BY

Abdulhamit Subasi,  
University of Turku, Finland  
Delaram Sadeghi,  
Islamic Azad University of Mashhad, Iran  
Genta Kulari,  
Autonomous University of Lisbon, Portugal

## \*CORRESPONDENCE

D. Darling Jemima  
✉ [darlingjemimai@gmail.com](mailto:darlingjemimai@gmail.com)

RECEIVED 27 August 2024

ACCEPTED 03 February 2025

PUBLISHED 20 February 2025

## CITATION

Jemima DD, Selvarani AG and Lovenia JDL  
(2025) Multi-view united transformer block of  
graph attention network based autism  
spectrum disorder recognition.  
*Front. Psychiatry* 16:1485286.  
doi: 10.3389/fpsy.2025.1485286

## COPYRIGHT

© 2025 Jemima, Selvarani and Lovenia. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Multi-view united transformer block of graph attention network based autism spectrum disorder recognition

D. Darling Jemima<sup>1\*</sup>, A. Grace Selvarani<sup>2</sup>  
and J. Daphy Louis Lovenia<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India, <sup>2</sup>Department of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India, <sup>3</sup>Department of Mathematics, Karunya Institute of Technology and Sciences, Coimbatore, India

**Introduction:** Autism Spectrum Disorder (ASD) identification poses significant challenges due to its multifaceted and diverse nature, necessitating early discovery for operative involvement. In a recent study, there has been a lot of talk about how deep learning algorithms might improve the diagnosis of ASD by analyzing neuroimaging data.

**Method:** To overrule the negatives of current techniques, this research proposed a revolutionary strategic model called the Unified Transformer Block for Multi-View Graph Attention Networks (MVUT\_GAT). For the purpose of extracting delicate outlines from physical and efficient functional MRI data, MVUT\_GAT combines the advantages of multi-view learning with attention processes.

**Result:** With the use of the ABIDE dataset, a thorough analysis shows that MVUT\_GAT performs better than Multi-view Site Graph Convolution Network (MVS\_GCN), outperforming it in accuracy by +3.40%. This enhancement reinforces our suggested model's effectiveness in identifying ASD. The result has implications over higher accuracy metrics. Through improving the accuracy and consistency of ASD diagnosis, MVUT\_GAT will help with early interference and assistance for ASD patients.

**Discussion:** Moreover, the proposed MVUT\_GAT's which patches the distance between the models of deep learning and medical visions by helping to identify biomarkers linked to ASD. In the end, this effort advances the knowledge of recognizing autism spectrum disorder along with the powerful ability to enhance results and the value of people who are undergone.

## KEYWORDS

MVUT\_GAT, transformer block, autism spectrum disorder, deep learning, neuroimaging

## 1 Introduction

Autism Spectrum Disorder (ASD) identification has emerged as a crucial issue within the field of neurodevelopmental research, as the prevalence of ASD grows worldwide. ASD is a complex and diverse neurodevelopmental syndrome marked by difficulties in social communication and repetitive activities. Identifying ASD early in a child's life is critical for appropriate intervention and support, yet it remains a significant difficulty owing to the complex and frequently subtle nature of the symptoms. Traditional diagnostic procedures depend mainly on clinical observations, behavioral evaluations, and interviews, which are subjective, time-consuming, and may lack the sensitivity required for early diagnosis (1). In recent years, the use of deep learning approaches into ASD detection has showed great potential and drew major interest from researchers and clinicians. Deep learning, a form of machine learning that uses artificial neural networks modelled after the human brain, provides a unique technique to identify the detailed patterns and traits linked with ASD in neuroimaging data. Magnetic resonance imaging (MRI) and functional MRI (fMRI) have emerged as critical methods for understanding the brain's anatomical and functional connections in a non-invasive manner (2). Deep learning methods, particularly neural networks with numerous layers, have shown the ability to automatically learn and extract nuanced patterns from neuroimaging data, allowing for the detection of minor anomalies linked with ASD. Minor anomalies are minor differences in the structure and function of the brain, undetectable using conventional methods and often too minor to be easily visible. Such minor differences include slight variations in the amygdala and cerebellum, abnormal connectivity patterns between areas of the brain, alterations in cortical thickness or volume, or microstructural changes in white matter. Advanced deep learning algorithms identify such minor differences when processing neuroimaging data. Thus, these minor differences are associated with some cognitive and behavioral features of ASD, which allows for earlier and more precise diagnosis. Deep learning has the ability to overcome various issues that standard approaches confront while identifying ASD. Deep learning algorithms can effectively handle large volumes of data, identifying nuanced patterns that would be difficult to detect manually. Furthermore, these models may incorporate multimodal information by incorporating data from a variety of sources, including structural and functional MRI, genetic markers, and behavioral evaluations. The end result is a more thorough and nuanced knowledge of ASD's neurological roots, which might lead to more accurate and early detection. Furthermore, the incorporation of attention processes into deep learning models enables the prioritizing of key variables, improving interpretability and enabling the discovery of biomarkers linked with ASD. Understanding how deep learning models make decisions is crucial for bridging the gap between their complex, 'black box' nature and the clinical insights needed to make informed decisions in practice. As deep learning advances, new architectures, optimization approaches, and hybrid models are being developed to improve the accuracy, resilience, and generalizability of ASD diagnosis models.

## 2 Related work

### 2.1 Recognizing ASD via various deep learning models

Functional Connectivity in ASD (3) is complex and shows hypo- and hyper-connectivity within many brain networks. Dimensional clustering can determine unique FC subtypes that relate to specific behavioral characteristics, hence unearthing a wide variety of connections between brain activity and behavior. Recognition of common patterns in FC would explain the heterogeneity of ASD better, facilitate better diagnosis, and develop tailored interventions that bridge neural changes to observable symptoms. The use of neuroimaging methods to better understand and diagnose neurodevelopmental disorders suggests a possible option for enhancing clinical practice. A technique for estimating effective connectivity in brain networks (BNs) (4) using EEG data, with a focus on children with attention-deficit hyperactivity disorder (ADHD). Their findings revealed substantial differences in directed information transmission across EEG electrodes in ADHD patients compared to healthy controls, with discriminative power notably high in the theta-band, which is associated with focus and motor activity. Similarly (5), used functional MRI (fMRI) data from the Autism Brain Imaging Data Exchange (ABIDE) dataset to solve the problem of multi-site data aggregation in autism diagnosis. They used data harmonization approaches to improve classification accuracy and discovered insights into ASD pathogenesis through network analysis. With machine learning techniques applied to fMRI functional connectivity, promising detection of ASD can be automated. The critical advances include integration of temporal dynamics, multiscale data, and focused analysis of brain networks relevant to ASD. Even though performance keeps improving, more work is needed to create robust and interpretable models for clinical application. Combining data modalities with larger datasets may further propel this field toward reliable computer-aided diagnosis of ASD (6). Their research produced encouraging results and revealed changes in brain network structure related to ASD. Furthermore (7), the significance of using sequential information from task-based fMRI for synthetic data augmentation in ASD diagnosis is demonstrated, demonstrating the potential for improving diagnostic accuracy and comprehending the underlying illness. Together, these studies highlight the value of neuroimaging techniques and advanced machine learning approaches in understanding the intricacies of neurodevelopmental disorders, opening the path for more effective diagnostic tools and deeper insights into the brain.

An Unsupervised Contrastive Graph Learning (UCGL) framework for resting-state fMRI research (8) underlines the difficulty of acquiring labelled training data in clinical practice and suggests a pretext model trained on unlabeled data for subsequent illness diagnosis tasks. The UCGL framework is tested on three rs-fMRI datasets, and it outperforms current techniques in

the automated diagnosis of major depressive illness, ASD, and Alzheimer's disease. ASD is diagnosed using structural (sMRI) and functional (fMRI) MRI modalities (9). To solve the low data availability, transfer learning is used in conjunction with four vision transformers and a 3D-CNN model. The investigations use several ways to generate data and extract slices from raw 3D sMRI and 4D fMRI images, yielding ground breaking findings. The brain disease categorization in resting-state functional magnetic resonance imaging (rs-fMRI) data study (10). Using multi-omics data, the team created eleven networks that depicted various facets of the brain. Kullback-Leibler divergence is used in their methodology, Graph convolution techniques are then applied to learn gene-disease connections. Functional magnetic resonance imaging (fMRI) data has been more important in neuroscience research recently for the identification of neurological illnesses and for the comprehension of cognitive processes. Many research has looked at various approaches to deal with issues such noise, limited sample numbers, and the requirement for interpretability that arise while processing fMRI data. Applying fMRI data to categorize neurological disorders or cognitive function has been the subject of several research. Using fMRI-derived brain graphs, GroupINN (11)—a grouping-based interpretable neural network—classifies cognitive performance well. Node grouping is included into the architecture of this model, which jointly learns these groups and extracts graph properties. Using resting-state fMRI time-series data, networks of long- and short-term memories (LSTMs) (12) can be used to directly diagnose autism spectrum disorders (ASD). They used the extensive, multi-site Autism Brain Imagery Data Exchange (ABIDE) I database for testing and training, and as a result, their classification accuracy was higher than that of earlier techniques. By applying algorithms that use deep learning to the ABIDE dataset (13), it was possible to identify ASD patients only by looking at patterns of brain activation. This approach achieved a remarkable degree of accuracy and revealed functional connectivity patterns linked to ASD. ASD-DiagNet (14), a system for ASD identification using fMRI data, was developed utilizing the 1,035 participants from 17 imaging facilities in the Autism Brain Imaging Information Exchange dataset. Techniques include a hybrid strategy for feature extraction that uses a single layer the perceptron and an autoencoder, as well as a linear interpolation-based data augmentation technique. In order to overcome the shortcomings of conventional behavioral observation-based diagnostic techniques, the research emphasizes the importance of developing machine learning infrastructures for qualitative diagnosis of varied mental disorders like autism spectrum disorder.

In these studies, the ABIDE dataset has become a notable resource offering a big, multi-site collection of fMRI data for research on ASD. By using this dataset, researchers may test and train their models on a variety of samples, which improves the generalizability of their conclusions. This research uses a range of methodologies, from deep learning architectures to conventional machine learning approaches. While deep learning models frequently give better performance but lack transparency, classical models could offer interpretability. Thus, by putting forth models which are both practical and understandable, recent initiatives have sought to close this gap.

## 2.2 GNN-based ASD recognition

Adversarial Graph Contrastive Learning (A-GCL) (15) is a method for identifying neurodevelopmental problems using fMRI data. The model employs a graph neural network (GNN) based on graph contrastive learning, with graphs generated from fMRI data. A-GCL outperforms three datasets, Autism Brain Imaging Data Exchange (ABIDE) I, ABIDE II, and attention deficit hyperactivity disorder (ADHD), over three atlases. While an A-GCL model may provide a good theoretical fit, this may be underpinned with several limitations including high-quality data in fMRI, inability to generalize across very diverse populations, low interpretability, and computing demands. Specifically, the atlases used need to be determined and only the resting-state fMRI is adopted. A Multi-Scale Dynamic Graph Learning (MDGL) (16) framework for detecting brain disorders utilizing resting-state functional magnetic resonance imaging (rs-fMRI) data. They use various brain atlases to build multi-scale dynamic functional connectivity networks (FCNs) and graph neural networks to extract spatiotemporal information from them. ASD categorization based on brain functional activity and gene expression using an attention-based graph neural network (GNN) (17). Their findings highlighted the significance of customized information in ASD diagnosis and biomarker identification. By combining individual brain topology and graph data. Furthermore, they identified brain regions important for ASD, such as the social-brain circuit and default-mode network, and discovered ASD-related genes using functional MRI data and gene expression analysis, demonstrating the potential of their approach for effective ASD diagnosis and biomarker identification. Present the Autism Spectrum Disorder-based Attention GNN and Crossover Boosted Meerkat Optimization (ASD-AttGCBMO) (18) algorithm. The suggested technique uses structural Magnetic Resonance Imaging (sMRI) data from the ABIDE 1 dataset for preprocessing to improve picture quality. Surface-based analysis and voxel-based morphometry (VBM) both extract significant information such surface area, cortical thickness, shape descriptors, and brain volumes. To address issues such as overfitting and class imbalance, the model utilizes attention GNNs with crossover-boosted meerkat optimization. PLSNet (19) is a position-aware graph-convolution-network-based model for ASD diagnosis that uses functional MRI (fMRI) data. PLSNet includes a time-series encoder for feature extraction and a connection generator to represent long-term relationships. PLSNet includes a time-series encoder for feature extraction and a connection generator to represent long-term relationships. Position embedding and a rarefying approach are used to solve challenges such as brain region variation and dimensionality complexity. The work delivers cutting-edge performance on the Autism Brain Imaging Data Exchange dataset.

The neuroimaging methods, namely magnetic resonance imaging (MRI) and functional MRI (fMRI), to diagnose Autism Spectrum Disorder (ASD). Each research suggests a unique methodology, such as machine learning techniques, generative adversarial networks (GANs), unsupervised contrasting graph learning, and adversarial self-supervised GNNs. A Conditional Generative Adversarial Network (cGAN) for predicting ASD (20). The researchers emphasize the limits of classic supervised machine learning

techniques when dealing with tiny datasets, and they provide a cGAN that surpasses normal GANs in terms of prediction accuracy. A graph attention network (GAT) (21), based on spatially restricted sparse functional brain networks (FBNs), was used to diagnose ASD. They developed a unique approach, Pearson's correlation-based Spatial Constraints Representation (PSCR), for estimating FBN structures and feeding them into a GAT for classification. Their trials using the ABIDE I dataset demonstrated the superiority of the PSCR technique as well as the influence of various FBNs on classification outcomes. Their suggested system produced encouraging classification results, surpassing rival approaches and offering insights for future illness detection using FBN and GNN frameworks.

The focus is on ASD, and task-based functional magnetic resonance imaging (fMRI) data is used. The authors focus on data-driven learning algorithms for biomarker identification and outcome prediction. Their deep learning methods use GNNs (22) for spatial variables and Long Short-Term Memory (LSTM) networks for temporal features. The chapter emphasizes the significance of dynamic connectivity changes and provides a more comprehensive, integrated model that includes spatiotemporal aspects as well as neural ordinary differential equations. ASD (23), utilizing a thorough examination of multi-modal imaging markers. Their dual-branch GNN performs a major diagnosis by extracting and combining data from structural and functional magnetic resonance imaging. The study also uses a perturbation model to find brain imaging signals and a neuro-transcriptomic joint analysis to reveal putative genetic biomarkers related with ASD brain development. Contribute to psychiatric diagnosis (24) using brain-networks by presenting a Granger causality-inspired GNN (CI-GNN). The model strives for interpretability without resorting to *post-hoc* interpretative paradigms.

A graph neural network architecture called BrainGNN (25) was used to analyze fMRI datasets from the Human Connectome Program (HCP) 900 Participant Release and Autism Spectrum Disorder (ASD). Their solution leverages both functional and topological knowledge gathered from fMRI data by using new ROI-aware graph convolution layers (Ra-GConv). Furthermore, BrainGNN uses ROI-selection layer pooling (R-pool) to emphasize important brain areas, making it easier to read. Regularization terms are suggested to promote flexible modeling of single or group-level patterns and fair ROI selection, such as units loss, topK pooled (TPK), which loss, and group-level constancy (GLC) loss.

## 2.3 GCN-based ASD recognition

In recent research, there has been an increasing emphasis on using functional brain networks (FBN) to classify neurological illnesses, particularly Autism Spectrum Disorder (ASD). MVS-GCN (2), a multi-view graph convolution network guided by previous brain structure learning, to overcome the issues presented by subject heterogeneity and noise correlations in brain networks. Their machine learning technique not only helps to classify neurological illnesses, but also gives an interpretable framework for deeper insights into the brain network. A joint learning architecture of multi-level dynamic brain networks for

the diagnosis of ASD (26). They overcome the constraints of previous graph convolutional network (GCN)-based techniques by allowing bidirectional information sharing across brain networks and adding edge weight information via an edge self-attention mechanism. It allows for information complementarity across different layers of brain networks. The Autism Spectrum Disorder-based Attention Graph Neural Network and Crossover Boosted Meerkat Optimization (ASD-AttGCBMO) method (27) detects ASD using structural MRI data. Their approach uses attention graph neural networks and a crossover boosted optimization strategy to improve feature categorization between ASD and control participants. Graph Neural Network (GNN) topologies and machine learning models for analyzing rs-fMRI data to better understand schizophrenia (28). They train deep graph convolutional neural networks (DGCNNs) and machine learning models with graph-theoretical measurements based on functional correlations between brain areas of interest CI-GNN finds influential subgraphs associated with choices (e.g., major depressive disorder), meeting the demand for interpretable graph neural networks and highlighting the significance of causal linkages in explainability. Moving on to neurodegenerative illnesses (29), investigates early dementia prediction using fMRI data and a Graph Convolutional Network (GCN) technique A multi-task learning strategy using a knowledge graph attention network to identify both mental and physical diseases (MPD) (30) simultaneously.

Investigating disparities in brain activity to differentiate between individuals with Autism Spectrum Disorder (ASD) and those without aids in understanding the root causes of ASD, leading to enhanced diagnosis and treatment strategies. As a result, functional connectivity (FC) analysis (31) derived from resting-state functional magnetic resonance imaging (rs-fMRI) data has emerged as a potent method for assessing and charting brain activity.

Using a supervised the siamese graph convolution neural network (s-GCN) (32) as a foundation, this approach learns a graph similarity measure with a specific emphasis on comparing brain connection networks. The model takes graph structure into account by using spectral graph convolutions, which function in the graph spectral domain. This improves results on the ABIDE database. Graph neural networks, with applications in neurology and other domains, have shown promise as a method for processing graph-structured data. During the graph representation learning process, local structural information is preserved by the use of EigenPooling (33), a pooling operation based on the graph Fourier transform. By fusing layers for pooling based on EigenPooling with conventional graph convolutional layers, they create EigenGCN, a graph neural network architecture for graph classification, and demonstrate its efficacy on six widely used benchmarks. To overcome the shortcomings of current approaches in the field of resting-state functional magnetic resonance imaging (rs-fMRI) (34) research, formulate functionally connected networks as spatiotemporal graphs. To simulate the non-stationary character of functional connectivity, they present a spatio-temporal graphing convolutional network (ST-GCN) built on brief sub sequences of BOLD time series. When it comes to predicting age and gender from BOLD data, ST-GCN performs better than standard methods.

It also finds key brain areas and functional linkages that are involved in the predictions.

Functional brain networks (FBN) have attracted a lot of interest as a means of diagnosing neurological disorders including autism spectrum disorders (ASD). Accurate categorization is difficult due to noisy correlations in brain networks and considerable subject heterogeneity. MVS-GCN (2), which combines graph neural networks to get efficient end-to-end representations for brain networks. To improve classification performance and find possible functional subnetworks, this approach combines multi-view graph convolutional neural networks with previous knowledge of brain anatomy. Using the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and Autism Brain Imaging Data Exchange (ABIDE) datasets, the authors assess the MVS-GCN model and show that it is more effective than current techniques. This concept motivates to generate a integrating the views in a combined form with the help of transformer and introduced a new Multi-View United Transformer Block (MVUTB). The proposed model also improves the performance of class discrimination using Graph Attention Network.

### 3 Background

#### 3.1 Graph attention layer on ASD

The Graph Attention Layer (GAT) is critical in identifying essential patterns and characteristics in neuroimaging data for recognizing Autism Spectrum Disorder (ASD). The application of GAT is discussed here, with equations to show how it works in the context of ASD recognition.

The GAT layer in the ASD identification task receives a collection of node characteristics reflecting neuroimaging data connected with brain areas (34). Let  $h = \{h_1, h_2, \dots, h_N\}$  N is the number of brain regions, and F is the number of related characteristics for each area. The goal is to create a new set of node functionalities  $h_0 = \{h_{01}, h_{02}, \dots, h_{0N}\}$ , with  $h_{01}$  in  $\mathbb{R}^{F_0}$  as Output.

The transformation starts with a common linear transformation, parameterized by a weight matrix.  $W$  in  $\mathbb{R}^{F_0 \times F}$ , applied to each node’s features:

$$\hat{h}'_i = W\hat{h}_i \tag{1}$$

A self-attention mechanism is used to produce attention coefficients ( $e_{ij}$ ) that indicate the significance of characteristics from node j to node i for ASD detection. The attention coefficients are generated using a shared attentional mechanism and masked attention, focusing solely on first-order neighbors.

$$e_{ij} = a(\hat{h}'_i, \hat{h}'_j) \tag{2}$$

To introduce the graph structure, masked attention is conducted, only for the node j which is a neighbor of node i, that is taking just the neighborhood  $N_i$  of node i. To make coefficients comparable across nodes, a softmax function is applied:

$$a_{ij} = \text{Softmax}\left(\frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})}\right) \tag{3}$$

Finally, the new node includes  $\hat{h}_{0i}$  is calculated as a weighted sum of the original characteristics, using attention coefficients as weights.

$$\hat{h}_{0i} = \sum_{j \in N_i} a_{ij} \hat{h}'_j \tag{4}$$

This method efficiently captures the significance of characteristics from surrounding brain areas in the context of ASD identification (35) by utilizing the self-attention mechanism and graph structure information. The GAT layer’s capacity to selectively aggregate information depending on attention coefficients improves the overall efficacy of ASD detection in neuroimaging data.

#### 3.2 Transformer-based ASD recognition

Transformer-based models have gained popularity in recent neuroimaging studies for detecting Autism Spectrum Disorder (ASD) (36, 37). Transformers, which were first introduced for natural language processing, have shown amazing ability to capture subtle patterns and relationships within sequential data, making them ideal for analyzing complex brain imaging data linked with ASD.

##### 3.2.1 Overview of transformer model

The Transformer design includes an encoder-decoder structure, however in the context of ASD identification, we concentrate on the encoder. The Transformer encoder’s main components include self-attention mechanisms (38), multi-head self-attention (MHA), and feedforward neural networks (FFN).

##### 3.2.2 Self-attention mechanisms

The self-attention mechanism is a critical component of the Transformer model, allowing it to weigh various elements of the input sequence independently. Given an input sequence, the self-attention mechanism calculates attention ratings for each element in the sequence in comparison to all other items. The attention scores are then utilized to calculate a weighted total of the input sequence, which results in context-aware representations for each piece. The mathematical expression for the self-attention process is as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

The query, key, and value matrices are represented by  $Q, K, V$ , respectively, and  $d_k$  denotes the dimension of the key vectors.

##### 3.2.3 Multiple-head self-attention

To increase model expressiveness, the Transformer utilizes many attention heads in tandem. Each head gains a separate understanding of the relationships within the input sequence.

The outputs from the several heads are concatenated and linearly processed to generate the final attention output. The formula for multi-head self-attention is as follows:

$$MHA(Q, K, V) = \text{Concat}(\text{Head}_1, \dots, \text{Head}_n)W_o \quad (6)$$

Where  $\text{Head}_i = \text{Attention}(QW_{Qi}, KW_{Ki}, VW_{Vi})$  and  $W_o$  is the output projection matrix.

### 3.2.4 Feedforward neural network

After the attention process, the model uses feedforward neural networks to perform further nonlinear transformations. The FFN is made up of two linear layers separated by a non-linear activation function (often a ReLU). The mathematical expression for the FFN operation is as follows:

$$FFN(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \quad (7)$$

The weight matrices and bias terms are represented by  $W_1$ ,  $b_1$ ,  $W_2$  and  $b_2$ .

In the context of ASD detection, the Transformer encoder examines neuroimaging data, such as MRI scans, to identify ASD-related characteristics and patterns. The self-attention mechanism allows the model to focus on specific areas of interest, whilst the multi-head mechanism improves the model's capacity to collect varied and complicated correlations in brain data. The ensuing feed forward neural networks modify the representations for downstream tasks, such as distinguishing between ASD and control participants.

## 4 Proposed model of ASD multi-view united transformer block

In this section, we introduce the ASD Multi-View United Transformer Block (ASD-MVUTB), a new model for detecting Autism Spectrum Disorder (ASD). Our suggested design uses transformer blocks to combine multi-view information, taking use

of varied views from neuroimaging data. The model effectively captures subtle patterns from several perspectives, encouraging complete feature extraction for accurate ASD detection. Experimental validation shows that the proposed ASD-MVUTB outperforms existing techniques.

The workflow for detecting Autism Spectrum Disorder (ASD) in Figure 1, is structured around N distinct views ( $V_1, V_2$ , and  $V_N$ ), each comprising both positive (+V) and negative (-V) perspectives. In the  $V_1$ -Positive View (+V), the process initiates with a Graph Attention Network (GAT-1), wherein neuroimaging data undergoes attention-based feature extraction. The output of GAT-1 is then subject to a Residual Addition operation, combining the GAT-1 output and the original positive view data. Subsequently, another Graph Attention Network (GAT-2) is applied, repeating the attention-based feature extraction process. The results of GAT-2 for both +V and -V are concatenated, and the concatenated output undergoes two linear layers (linear (128,16) and linear (16,2)) to determine the binary classification output, indicating whether the patient exhibits ASD (39, 40).

Mathematically, the process can be expressed as follows:

1. GAT-1 Process:

$$GAT - 1(+V) = GAT - 1 \text{ Output} \quad (8)$$

2. Residual Addition:

$$GAT - 1 \text{ Output} + +V = \text{Residual Addition} \quad (9)$$

3. GAT-2 Process:

$$GAT - 2(\text{Residual Addition Output}) = GAT - 2 \text{ Output} \quad (10)$$

4. GAT-2 Layer Concatenation:

$$\text{Concatenate}(GAT - 2 \text{ Output} (+V), GAT - 2 \text{ Output} (-V)) \quad (11)$$

5. Linear Layer Application:

$$\text{Linear}(128, 16) \rightarrow \text{Lineary}(16, 2) = \text{View}_1 \text{ Output} \quad (12)$$

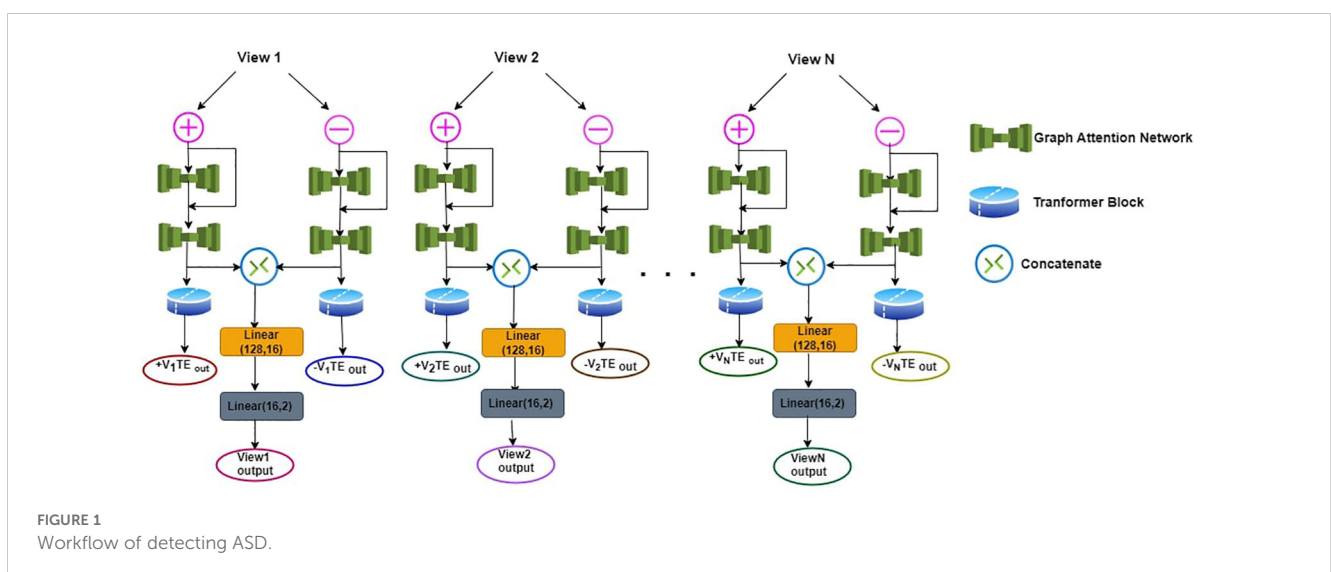


FIGURE 1 Workflow of detecting ASD.

Next, after the GAT-2 process for each view, a Transformer Encoder (TE) is applied. The TE operation is conducted separately for both the +V and -V outcomes of each view. This multi-view approach enhances the model’s ability to capture diverse features and patterns related to ASD within the neuroimaging data.

The Transformer Encoder for all View (V’s) is expressed as:

$$TE(GAT - 2 \text{ Output}( + V)) = +V \text{ TE Output} \quad (13)$$

$$TE(GAT - 2 \text{ Output}( - V)) = -V \text{ TE Output} \quad (14)$$

This entire process is replicated for View<sub>2</sub> (V<sub>2</sub>), View<sub>3</sub> (V<sub>3</sub>) and upto View<sub>N</sub>(V<sub>N</sub>). In this work three views are used to construct the ASD model. The utilization of Graph Attention Networks and Transformer Encoders facilitates capturing intricate patterns in neuroimaging data, and the linear layers aid in synthesizing and classifying the information for ASD detection. The concatenated results from the GAT-2 process ensure a comprehensive representation of both positive and negative perspectives in each view, contributing to the overall diagnostic outcome.

The proposed ASD-MVUTB model architecture, as illustrated in Figure 2, involves distinct processing of positive and negative TE outputs from Views (+Vs and -Vs) separately. Specifically, the TE outputs +V<sub>i</sub>TEout (X1) and +V<sub>i+1</sub>TEout (X2) are processed independently in the first MVUTB layer. The resulting outputs, denoted as X1 and X2, are then combined with +V<sub>i+2</sub>TEout in the second MVUTB layer. These intermediate results are flattened to create feature vectors. A parallel process is applied to the negative TE outputs, resulting in three sets of flattened feature vectors. Subsequently, the flattened results from both +Vs and -Vs are concatenated, and the concatenated feature vector undergoes three linear transformations: Linear(384, 32), Linear(32, 16), and Linear(16, 2).

The ASD-MVUTB working process is represented as follows:

1. MVUTB layers:

• Positive TE Outputs Processing.

$$X1 = MVUTB(+ V_i TE_{OUT}, +V_{i+1} TE_{OUT}) \quad (15)$$

$$X2 = MVUTB(X1, +V_{i+2} TE_{OUT}) \quad (16)$$

• Negative TE Outputs Processing.

$$Y1 = MVUTB(- V_i TE_{OUT}, -V_{i+1} TE_{OUT}) \quad (17)$$

$$Y2 = MVUTB(Y1, -V_{i+2} TE_{OUT}) \quad (18)$$

2. +Vs and -Vs Flatteninng.

$$F_X = Flatten(X1, X2) \quad (19)$$

$$F_Y = Flatten(Y1, Y2) \quad (20)$$

3. Concatenation:

$$F_{Concat} = Concat(F_X, F_Y) \quad (21)$$

4. Linear Transformations:

$$Z1 = Linear (F_{Concat}, 384, 32) \quad (22)$$

$$Z2 = Linear (Z1, 32, 16) \quad (23)$$

$$Z3 = Linear (Z2, 16, 2) \quad (24)$$

The motivation behind this architecture lies in the enhanced feature extraction from both positive and negate+ve TE outputs independently, and the subsequent concatenation allows the model to capture and leverage information from both perspectives.

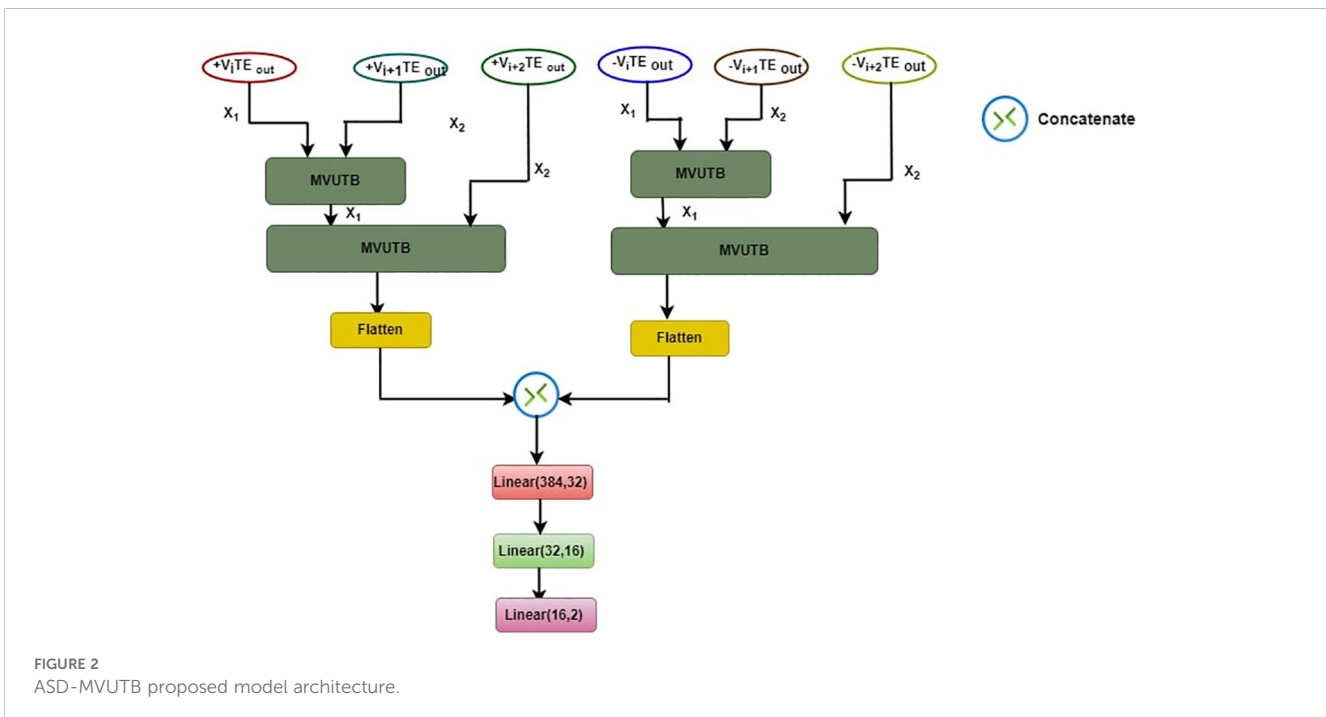


FIGURE 2 ASD-MVUTB proposed model architecture.

The linear layers at the end serve to further distill and map the concatenated features into a final prediction space for ASD classification. This layered approach enhances the model’s ability to discern intricate patterns and relationships within the TE outputs, contributing to its effectiveness in ASD recognition.

In the proposed framework depicted in Figure 3, the Multi-View Transformer Block operates through a meticulous process involving two views of features, denoted as X1 and X2. These views undergo Layer Normalization simultaneously. Subsequently, a linear mapping is applied to the query (q) and key (k) from X1, and the value (v) from X2, which are then fed into the Multi-Head Attention (MHA) mechanism. The output of MHA, along with the original features from X1, undergoes an element-wise addition (A1) to yield a preliminary result (R1). Following this, another Layer Normalization is applied, succeeded by a Multi-Layer Perceptron (MLP). The output of the MLP, combined with the result from the previous addition (R1), undergoes a second addition (A2), culminating in the generation of the final outcome. This iterative and additive process allows for the effective integration of information from multiple views.

1. Layer Normalization:

$$LayerNorm(X) = \frac{(X - \mu)}{\sqrt{\sigma^2 + \epsilon}} \times \gamma + \beta \quad (25)$$

Where  $\mu$  and  $\sigma$  represent the mean and standard deviation,  $\epsilon$  is a small constant, and  $\gamma$  and  $\beta$  are learnable parameters.

2. Multi-Head Attention (MHA):

$$MHA = (q_{x1}, k_{x1}, V_{x2}) = \text{Softmax} \left( \frac{q_{x1} \cdot k_{x1}^T}{\sqrt{d_k}} \right) V_{x2} \quad (26)$$

3. Addition (A1):

$$R1 = X1 + MHA(q_{x1}, k_{x1}, V_{x2}) \quad (27)$$

4. Multi-Layer Perceptron (MLP):

$$MLP(X) = \text{ReLU}(XW_1 + b_1)W_2 + b_2 \quad (28)$$

5. Addition (A2):

$$R2 = R1 + MLP(R1) \quad (29)$$

The adoption of Multi-Head Attention allows the model to capture intricate relationships within and between X1 and X2, enhancing representational capacity. The subsequent addition of the MHA output and the original X1 features promotes the fusion of learned contextual information with the original content. The Layer Normalization following each operation ensures stability and normalization of the intermediate results. The introduction of the MLP introduces non-linear transformations, further enhancing the model’s ability to capture complex patterns. The final addition consolidates the MLP’s output with the result from the initial MHA addition, producing a comprehensive and refined representation that incorporates information from both views. This iterative

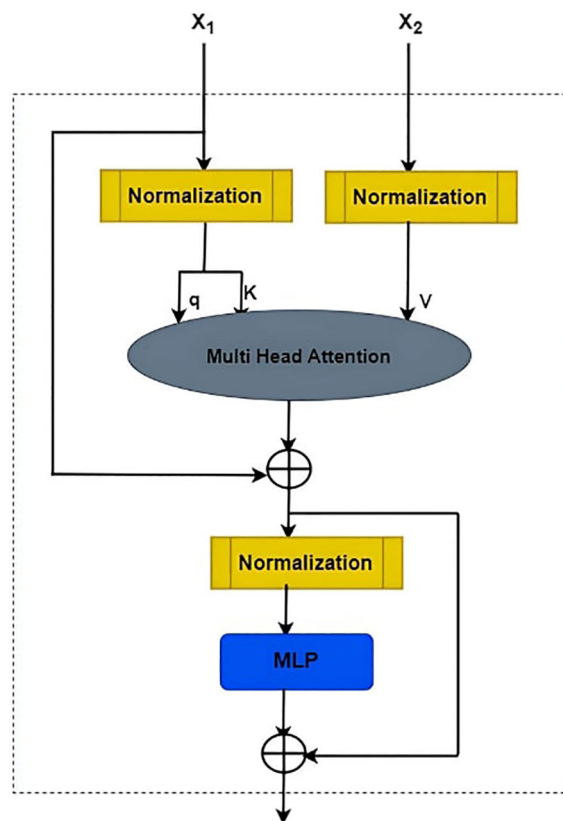


FIGURE 3 Framework of multi-views in transformer block.



process aligns with the principles of transformer-based architectures, facilitating multi-view feature integration for improved learning and representation.

## 5 Experimental results

The experimental findings in this section show how well the Multi-View United Transformer Block of the Graph Attention Network detects ASD. Using a variety of neuroimaging modalities, including as MRI and fMRI data, the model performs exceptionally well in identifying subtle patterns linked to ASD. In addition, the inclusion of attention processes improves interpretability and makes it possible to find significant biomarkers. These findings highlight the potential of deep learning techniques to transform the knowledge and early identification of ASD.

In **Table 1**, a comprehensive comparison between the proposed Multi-View United Transformer Block of Graph Attention Network (MVUT\_GAT) and several existing methods reveals notable advancements in accuracy. Compared to the MVS-GCN (1) approach the MVUT\_GAT exhibits a superior performance, surpassing MVS-GCN by +3.4%, ASD-DiagNet by +4.98%, and DAE by +5.68%. These substantial accuracy improvements underscore the efficacy of the proposed method in outperforming various state-of-the-art approaches in Autism Spectrum Disorder identification.

A comparative examination of the Area Under the Curve (AUC) values for several approaches on the ABIDE dataset is shown in **Figure 4**. An extensive summary of each technique's efficacy in diagnosing Autism Spectrum Disorder (ASD) is given by the AUC, a statistic used to evaluate the performance of classification models. The graph provides information about the relative advantages and disadvantages of the models by graphically representing the discriminative capacity of different methodologies. The most promising techniques for precise ASD diagnosis are chosen by academics and clinicians with the help of this

**TABLE 1** Comparison of existing methods for sensitivity, specificity, accuracy, and AUC.

	Accuracy	Sensitivity	Specificity
DAE (12)	67.61	78.7	53.2
ASD-DiagNet (13)	68.31	60.31	67.76
GroupINN (10)	63.6	61.52	57.36
ST-GCN (33)	57.29	54.78	48.91
Eigenpooling GCN (32)	57.5	58.81	59.94
LSTM-ASD (11)	68.5	-	-
BrainGNN (24)	61.84	61.65	60.79
sGCN (31)	67.54	64.73	60.12
MVS-GCN (1)	69.89	70.18	63.05
MVUT_GAT	73.29	73.86	69.67

- means sensitivity and specificity are not applicable.

quantitative assessment based on the ABIDE dataset. When evaluating how well various algorithms handle neuroimaging data for ASD identification, the figure is a useful point of comparison.

**Figure 5** displays the values of the super node number for the graph structure learning and the corresponding outcomes. Findings indicate that the number of nodes in the coarsened graph significantly affects classification performance, indicating the importance of the brain network's coarsening level for graph structure learning. More specifically, the coarsened graph's topology information is condensed if there are fewer super nodes than there are, and richer if there are more super nodes than there are.

$V \in \{1, 2, 3, 4, 5, 6\}$  was changed, and the performance variation with different values was examined in order to examine the impact of variability throughout the number of views. As seen in **Figure 6**, the performance gets better the more views there are, up to a maximum of three. This indicates that distinct views have built-in correlations that help capture topological details. Further evidence that more views will unavoidably contribute duplicate data to the model and lower its classification performance comes from the observation that the model's performance tends to stabilize as the number of views rises.

## 6 Conclusion and discussion

Identification of Autism Spectrum Disorder (ASD) may advance with the incorporation of deep learning models, such as the Multi-View United Transformer Block of Graph Attention Network. These algorithms' capacity to automatically identify subtle abnormalities in multimodal neuroimaging data improves diagnostic precision and makes it easier to identify ASD early on. By including attention processes, interpretability is further enhanced and the gap between clinically applicable discoveries and complicated models is closed. The proposed model Multi-View United Transformer Block of Graph Attention Network (MVUT\_GAT) outperforms the existing model by +3.4% which improves the robustness and generalizability of ASD diagnostic models. In order to improve the robustness and generalizability of ASD diagnostic models, continued research efforts should concentrate on honing architectures, optimizing models, and investigating hybrid techniques as deep learning continues to develop. To further enhance our knowledge of the neurological foundations of ASD, future research may investigate the integration of other data sources, including as genetic markers and behavioral assessments. These developments might transform early diagnosis and intervention approaches, ultimately leading to better outcomes for people with ASD worldwide.

## 7 Future work

Future studies on MVUT\_GAT for ASD diagnosis will focus on some key issues to improve the robustness and generalizability of the model. Among these, the main focuses should be on optimizing the model, hyperparameters refinement, an increase in efficiency

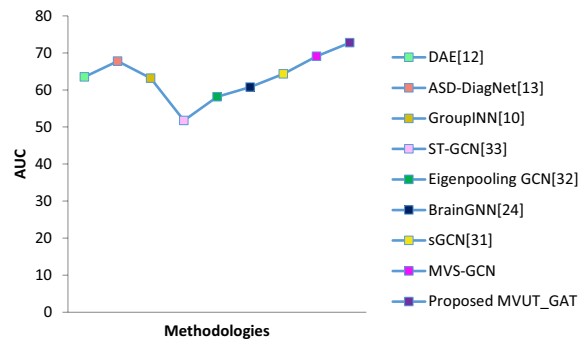


FIGURE 4 The comparison of various methods AUC on the ABIDE dataset.

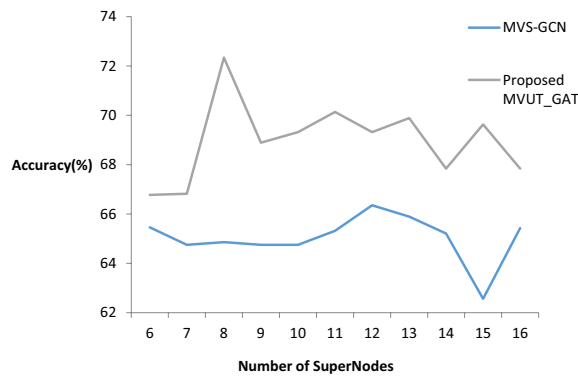


FIGURE 5 Performance for different number of super nodes of coarsened graphs.

levels along with transfer learning, and other advanced techniques that can help it handle such datasets. Additionally, MVUT\_GAT could be combined with other deep learning models, like CNNs or RNNs, to enhance performance. Multimodal data integration, such as genetic markers, behavioral assessments, and neuroimaging, would give a better understanding of ASD and could also enhance diagnostic accuracy. Improving interpretability is yet

another important area. Mechanisms of attention and explainability techniques, such as the saliency maps used by the model, will clarify the decision-making process for clinicians, so they can better trust and implement deep learning models in practice while filling a gap between advanced algorithms and clinical use. Longitudinal studies should be pursued to track the changes in ASD over time with neuroimaging data. This will ensure

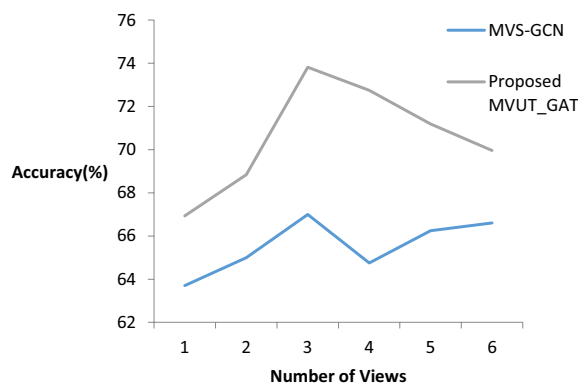


FIGURE 6 Performance variation as the number of views increases.

earlier detection of ASD and real-time tracking of interventions. A greater population from more diverse demographics would increase the number of datasets, which would be more generalizable for the model and its applications in different settings of healthcare. Finally, collaboration among deep learning researchers, clinicians, and neuroscientists will be very critical to align the models with clinical needs and ensure that these advances are translated into better ASD diagnosis and intervention, hence improving outcomes for individuals around the world.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: [https://fcon\\_1000.projects.nitrc.org/indi/abide/](https://fcon_1000.projects.nitrc.org/indi/abide/).

## Author contributions

DJ: Methodology, Writing – original draft. AS: Supervision, Writing – review & editing. JL: Validation, Writing – review & editing.

## References

1. First LL, Gagnon M. Challenges in early diagnosis of neurodevelopmental disorders: Implications for assessment and intervention. *J Clin Child Psychol.* (2017) 46:451–65. doi: 10.1080/15374416.2016.1234376
2. Wen G, Cao P, Bao H, Yang W, Zheng T, Zaiane O. MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis. *Comput Biol Med.* (2022) 142:105239. doi: 10.1016/j.compbiomed.2022.105239
3. Easson AK, Fatima Z, McIntosh AR. Functional connectivity-based subtypes of individuals with and without autism spectrum disorder. *Network Neurosci.* (2019) 3:344–62. doi: 10.1162/netn\_a\_00067
4. Abbas AK, Azemi G, Amiri S, Ravanshadi S, Omidvarnia A. Effective connectivity in brain networks estimated using EEG signals is altered in children with ADHD. *Comput Biol Med.* (2021) 134:104515. doi: 10.1016/j.compbiomed.2021.104515
5. Ingallhalikar M, Shinde S, Karmarkar A, Rajan A, Rangaprakash D, Deshpande G. Functional connectivity-based prediction of Autism on site harmonized ABIDE dataset. *IEEE Trans Biomed Eng.* (2021) 68:3628–37. doi: 10.1109/TBME.2021.3080259
6. Zhang J, Feng F, Han T, Gong X, Duan F. Detection of autism spectrum disorder using fMRI functional connectivity with feature selection and deep learning. *Cogn Comput.* (2023) 15:1106–17. doi: 10.1007/s12559-021-09981-z
7. Wang J, Dvornek NC, Staib LH, Duncan JS. Learning sequential information in task-based fMRI for synthetic data augmentation. In: *International Workshop on Machine Learning in Clinical Neuroimaging*. Springer Nature Switzerland, Cham (2023). p. 79–88.
8. Wang X, Chu Y, Wang Q, Cao L, Qiao L, Zhang L, et al. Unsupervised contrastive graph learning for resting-state functional MRI analysis and brain disorder detection. *Hum Brain Mapp.* (2023) 44:5672–92. doi: 10.1002/hbm.v44.17
9. Alharthi AG, Alzahrani SM. Multi-Slice Generation sMRI and fMRI for Autism Spectrum Disorder Diagnosis Using 3D-CNN and Vision Transformers. *Brain Sci.* (2023) 13:1578. doi: 10.3390/brainsci13111578
10. Zhang H, Song R, Wang L, Zhang L, Wang D, Wang C, et al. Classification of brain disorders in rs-fMRI via local-to-global graph neural networks. *IEEE Trans Med Imaging.* (2022) 42:444–55. doi: 10.1109/TMI.2022.3219260
11. Yan Y, Zhu J, Duda M, Solarz E, Sripada C, Koutra D. (2019). Groupinn: grouping-based interpretable neural network for classification of limited, noisy brain data, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (New York, NY, USA: ACM). pp. 772–82.
12. Dvornek NC, Ventola P, Pelphrey KA, Duncan JS. Identifying autism from resting-state fmri using long short-term memory networks. In: *International Workshop on Machine Learning in Medical Imaging*. Cham, Switzerland Springer (2017). p. 362–70.
13. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the abide dataset. *Neuroimage: Clin.* (2018) 17:16–23. doi: 10.1016/j.nicl.2017.08.017

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

14. Eslami T, Mirjalili V, Fong A, Laird AR, Saeed F. ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data. *Front Neuroinformatics.* (2019) 13:70. doi: 10.3389/fninf.2019.00070
15. Zhang S, Chen X, Shen X, Ren B, Yu Z, Yang H, et al. A-GCL: Adversarial graph contrastive learning for fMRI analysis to diagnose neurodevelopmental disorders. *Med Image Anal.* (2023) 90:102932. doi: 10.1016/j.media.2023.102932
16. Lang J, Wylie G, Haig C, Gillberg C, Minnis H. Towards system redesign: An exploratory analysis of neurodivergent traits in a childhood population referred for autism assessment. *PLoS One.* (2024) 19:e0296077. doi: 10.1371/journal.pone.0296077
17. Wang Z, Xu Y, Peng D, Gao J, Lu F. Brain functional activity-based classification of autism spectrum disorder using an attention-based graph neural network combined with gene expression. *Cereb Cortex.* (2023) 33:6407–19. doi: 10.1093/cercor/bhac513
18. Goel L, Gupta S, Gupta A, Rajan SN, Gupta VK, Singh A, et al. Advancing ASD detection: novel approach integrating attention graph neural networks and crossover boosted meerkat optimization. *Int J Mach Learn Cybernetics.* (2024) 1:1–19. doi: 10.1007/s13042-023-02092-1
19. Wang Y, Long H, Zhou Q, Bo T, Zheng J. PLSNet: Position-aware GCN-based autism spectrum disorder diagnosis via FC learning and ROIs sifting. *Comput Biol Med.* (2023) 107184. doi: 10.1016/j.compbiomed.2023.107184
20. Raja KC, Kannimathu S. Conditional generative adversarial network approach for autism prediction. *Comput Syst Sci Eng.* (2023) 44. doi: 10.32604/csse.2023.025331
21. Yang C, Wang P, Tan J, Liu Q, Li X. Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks. *Comput Biol Med.* (2021) 139:104963. doi: 10.1016/j.compbiomed.2021.104963
22. Duncan J, Staib LH, Dvornek N, Li X, Zhuang J, Wang J, et al. Data-driven learning strategies for biomarker detection and outcome prediction in Autism from task-based fMRI. In: *Deep Learning for Medical Image Analysis*. Cambridge, MA, USA: Academic Press (2024). p. 357–93.
23. Gao J, Xu Y, Li Y, Lu F, Wang Z. Comprehensive exploration of multi-modal and multi-branch imaging markers for autism diagnosis and interpretation: insights from an advanced deep learning model. *Cereb Cortex.* (2024) 34:bbad521. doi: 10.1093/cercor/bhad521
24. Zheng K, Yu S, Chen B. Ci-gnn: A granger causality-inspired graph neural network for interpretable brain network-based psychiatric diagnosis. *Neural Networks.* (2024) 106147. doi: 10.1016/j.neunet.2024.106147
25. Li X, Zhou Y, Dvornek N, Zhang M, Gao S, Zhuang J, et al. Brainngn: interpretable brain graph neural network for fmri analysis. *Med Image Anal.* (2021) 67:102233. doi: 10.1016/j.media.2021.102233
26. Zhang P, Zhang W, Sun W, Xu J, Hu H, Wang L, et al. Identification of gene biomarkers for brain diseases via multi-network topological semantics extraction and graph convolutional network. *BMC Genomics.* (2024) 25:1–16. doi: 10.1186/s12864-024-09967-9

27. Xie H, Yang Y, Cui H, Yang C. Federated learning for cross-institution brain network analysis. *NeuroImage*. (2024) 252.
28. Xu J, Yang Y, Huang D, Gururajapathy SS, Ke Y, Qiao M, et al. Data-driven network neuroscience: On data collection and benchmark. *Adv Neural Inf Process Syst*. (2024) 36. doi: 10.17608/k6.auckland.21397377
29. Han S, Sun Z, Zhao K, Duan F, Caiafa CF, Zhang Y, et al. Early prediction of dementia using fMRI data with a graph convolutional network approach. *J Neural Eng*. (2024) 21:016013. doi: 10.1088/1741-2552/ad1e22
30. Zhang W, Kong L, Lee S, Chen Y, Zhang G, Wang H, et al. Detecting mental and physical disorders using multi-task learning equipped with knowledge graph attention network. *Artif Intell Med*. (2024), 102812. doi: 10.1016/j.artmed.2024.102812
31. Jiang H, Cao P, Xu M, Yang J, Zaiane O. Hi-gcn: a hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction, *Comput. Biol Med*. (2020) 127:104096. doi: 10.1016/j.compbiomed.2020.104096
32. Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, et al. Metric learning with spectral graph convolutions on brain connectivity networks. *Neuroimage*. (2018) 169:431–42. doi: 10.1016/j.neuroimage.2017.12.052
33. Ma Y, Wang S, Aggarwal CC, Tang J. (2019). Graph convolutional networks with eigenpooling, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (New York, NY, USA: ACM). pp. 723–31.
34. Gadgil S, Zhao Q, Pfefferbaum A, Sullivan EV, Adeli E, Pohl KM. (2020). Spatio-temporal graph convolution for resting-state fmri analysis, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020)*. Cham, Switzerland: Springer. pp. 528–38. Springer.
35. Xue Y, Dong HY, Feng JY, Bai MS, Li D, Yang H, et al. Parent-child interaction related to brain functional alterations and development outcomes in autism spectrum disorder: A study based on resting state-fMRI. *Res Dev Disabil*. (2024) 147:104701. doi: 10.1016/j.ridd.2024.104701
36. Xue Y, Bai MS, Dong HY, Wang TT, Mohamed ZA, Jia FY. Altered intra- and inter-network brain functional connectivity associated with prolonged screen time in pre-school children with autism spectrum disorder. *Eur J Pediatr*. (2024) 183:1–9. doi: 10.1007/s00431-024-05500-y
37. Qian S, Yang Q, Cai C, Dong J, Cai S. Spatial-temporal characteristics of brain activity in autism spectrum disorder based on hidden Markov model and dynamic graph theory: a resting-state fMRI study. *Brain Science*. (2024) 14(5):507–12. doi: 10.3390/brainsci14050507
38. Wang Y, Long H, Bo T, Zheng J. Residual graph transformer for autism spectrum disorder prediction. *Comput Methods Programs Biomed*. (2024) 224:108065. doi: 10.1016/j.cmpb.2024.108065
39. Zhao F, Feng F, Ye S, Mao Y, Chen X, Li Y, et al. Multi-head self-attention mechanism-based global feature learning model for ASD diagnosis. *Biomed Signal Process Control*. (2024) 91:106090. doi: 10.1016/j.bspc.2024.106090
40. Le Cunff AL, Dommert E, Giampietro V. Neurophysiological measures and correlates of cognitive load in attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD) and dyslexia: A scoping review and research recommendations. *Eur J Neurosci*. (2024) 59:256–82. doi: 10.1111/ejn.16201