

Original Research

Exploring Implicit Biological Heterogeneity in ASD Diagnosis Using a Multi-Head Attention Graph Neural Network

Hyung-Jun Moon¹, Sung-Bae Cho^{2,*}

¹Department of Artificial Intelligence, Yonsei University, 03722 Seoul, Republic of Korea

²Department of Computer Science, Yonsei University, 03722 Seoul, Republic of Korea

*Correspondence: sbcho@yonsei.ac.kr (Sung-Bae Cho)

Academic Editor: Woo-Yang Kim

Submitted: 2 January 2024 Revised: 13 April 2024 Accepted: 24 April 2024 Published: 17 July 2024

Abstract

Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder exhibiting heterogeneous characteristics in patients, including variability in developmental progression and distinct neuroanatomical features influenced by sex and age. Recent advances in deep learning models based on functional connectivity (FC) graphs have produced promising results, but they have focused on generalized global activation patterns and failed to capture specialized regional characteristics and accurately assess disease indications. **Methods:** To overcome these limitations, we propose a novel deep learning method that models FC with multi-head attention, which enables simultaneous modeling of the intricate and variable patterns of brain connectivity associated with ASD, effectively extracting abnormal patterns of brain connectivity. The proposed method not only identifies region-specific correlations but also emphasizes connections at specific, transient time points from diverse perspectives. The extracted FC is transformed into a graph, assigning weighted labels to the edges to reflect the degree of correlation, which is then processed using a graph neural network capable of handling edge labels. **Results:** Experiments on the autism brain imaging data exchange (ABIDE) I and II datasets, which include a heterogeneous cohort, showed superior performance over the state-of-the-art methods, improving accuracy by up to 3.7%p. The incorporation of multi-head attention in FC analysis markedly improved the distinction between typical brains and those affected by ASD. Additionally, the ablation study validated diverse brain characteristics in ASD patients across different ages and sexes, offering insightful interpretations. **Conclusion:** These results emphasize the effectiveness of the method in enhancing diagnostic accuracy and its potential in advancing neurological research for ASD diagnosis.

Keywords: autism spectrum disorder; dynamic functional connectivity; graph neural network; multi-head attention

1. Introduction

Autism spectrum disorder (ASD) presents a unique challenge in neurodevelopmental diagnostics due to its heterogeneous manifestations [1–3]. It is characterized by a spectrum of symptoms impacting communication, social interaction, and behavior. Key neurodevelopmental changes in ASD include structural brain changes and developmental changes, which can differ by sex.

Structural brain changes, including variations in the hippocampus, amygdala, cerebellum, and cortex, are notable in ASD. For instance, an enlarged hippocampus in children and adolescents, and differing amygdala sizes impacting emotional processing are common [4,5]. Moreover, alterations in the cortex thickness and cerebellum tissue underscore the complexity of ASD's neural basis [6,7]. Developmental changes in ASD are characterized by early brain overgrowth, followed by potential premature brain shrinkage. Excessive cerebrospinal fluid and enlarged head circumference in early development are also observed [8–10]. There are structural brain differences between males and females with ASD and, alongside variations in the amygdala, further complicate the subject's neural profile [7,11,12].

To overcome these challenges, extensive research has been conducted based on neuroimaging techniques such as functional magnetic resonance imaging (fMRI) with deep learning to interpret and extract patterns from the brain's complex features [13,14]. On the other hand, an approach based on functional connectivity (FC) has emerged, which aims to provide further insight into the brain's functional relationships [15–17]. Typically, FC-based methods utilize the brain's regional differences by designating each region as a biomarker and measuring the correlations between regions based on their temporal attributes. The graphs created by connecting each region according to these measured correlations explicitly represent the brain's spatial relationships, proving useful in various disease diagnoses when analyzed with tools such as graph neural networks (GNNs) [18].

However, many previous methods have focused solely on structural features or fail to consider the heterogeneous external characteristics of the data, leading to reduced efficiency in diagnosing ASD and overlooking the transient and localized features of the disorder [19,20]. Some symptoms may manifest in concentrated forms within specific timeframes and localized regions [21–23]. For instance, seizure disorders might exhibit changes in brain activity only at



Table 1. Studies related to deep learning models for the diagnosis of ASD.

Author	Feature	Method	Description
El-Gazzar A. <i>et al.</i> [30] (2019)	Spatio-Temporal	3D-CNN-LSTM	To overcome the challenges of extracting patterns from high-dimensional data with significant cardinality, the method uses 3D CNN and LSTM networks.
Eslami T. <i>et al.</i> [31] (2019)	Spatio-Temporal	ASD-DiagNet	To mitigate diagnostic limitations of behavior analysis through autoencoders combined with classifiers or sparse autoencoders.
Anirudh R. and Thiagarajan JJ. [39] (2019)	Graph Neural Net-work	Bootstrapping ensemble of graph convolutional neural networks (GCNs)	To address the challenges by integrating socio-cultural traits and brain activity patterns, a bootstrapped GCN reduces dependency on specific graph constructions, enhancing model robustness and offering improvements over existing graph-based neural networks even in the presence of noisy graphs.
Ahmed M.R. <i>et al.</i> [29] (2020)	Spatio	CNN Ensemble	This strategy combats data complexity and heterogeneity by deploying ensemble classifiers that learn from generative models in a post-learning phase.
Almuqhim F. and Saeed F. [32] (2021)	Spatio-Temporal	ASD-SAENet	To overcome the limitations of current clinical ASD diagnoses reliant on behavioral observations, ASD-SAENet utilizes a sparse autoencoder for feature optimization and classification.
Park K. <i>et al.</i> [41] (2021)	Spatio-Temporal	ResNet-LSTM with self-attention	To overcome the complexity of diagnosing ASD, the method leverages attention and convolutional recurrent neural networks to analyze dynamic connectivity between specific brain regions.
Kang, L. <i>et al.</i> [44] (2023)	Spatio-Temporal	Multi-view ensemble learning	To address inter-site heterogeneity in multi-site fMRI data, the authors utilized a multi-view ensemble learning network, demonstrating strong generalization capabilities.
Li W. <i>et al.</i> [43] (2022)	Spatio-Temporal	Self-supervised ensemble	This method addresses the issue of considering non-adjacent regions in CNNs by utilizing Vision Transformers.
Sotero R.C. <i>et al.</i> [45] (2023)	Feature Selection Algorithm	R-walk	Using a random walks technique for feature extraction and selection from resting-state fMRI data, this method demonstrates effectiveness in simplifying data and enhancing model performance.

ASD, Autism spectrum disorder; CNN, convolutional neural network; fMRI, functional magnetic resonance imaging; LSTM, long short-term memory.

distinct, fleeting moments, triggered by external or internal stimuli [24,25]. In other words, these methods fail to capture the temporal and spatial specificity essential for accurately representing and diagnosing ASD manifestations.

To cope with the challenges of capturing the transient and localized characteristics of ASD that are heterogeneous among ASD patients, we propose a novel deep learning method designed to model FC with multi-head attention. This model aims to extract transient and localized FC while simultaneously learning the structural and temporal features in fMRI images through the attention mechanism. Unlike conventional methods, the model not only identifies region-specific correlations across the overall spectrum but also highlights the connections at specific, transient time points. Understanding the nuanced characteristics of ASD necessitates an approach that can capture the unique traits of each patient, as well as the intricate temporal and spatial dynamics of their condition. By employing multi-head attention in our analysis of ASD, we enhance our ability to capture the unique and complex traits of each patient, including their temporal and spatial dynamics.

This method allows for the simultaneous processing of diverse ASD characteristics, improving the detection of

subtle patterns and offering deeper insights into the disorder. Multi-head attention thus plays a crucial role in facilitating more accurate and personalized diagnoses, marking a significant step forward in ASD research. After this extraction process, the FC is transformed into a graph. While traditional FC-based methods for ASD prediction typically represent this FC as a binary adjacency matrix, the method assigns weighted labels to the graph edges, reflecting the degree of correlation. To effectively process the graph with added edge information, the proposed method employs a GNN capable of handling edge labels to embed the FC and perform disease diagnosis. Due to its design to capture heterogeneous patterns, the proposed multi-head attention-based FC model is particularly adept at addressing the distinct neurodevelopmental changes characteristic of ASD's heterogeneity.

To validate the effectiveness of the proposed method, we conduct experiments using the autism brainimaging data exchange (ABIDE) I and II datasets. We will show that this novel method significantly enhances the ability to differentiate between healthy brains and those affected by diseases in FC analysis experiments, aiming at ASD classification. Additionally, by categorizing patients according

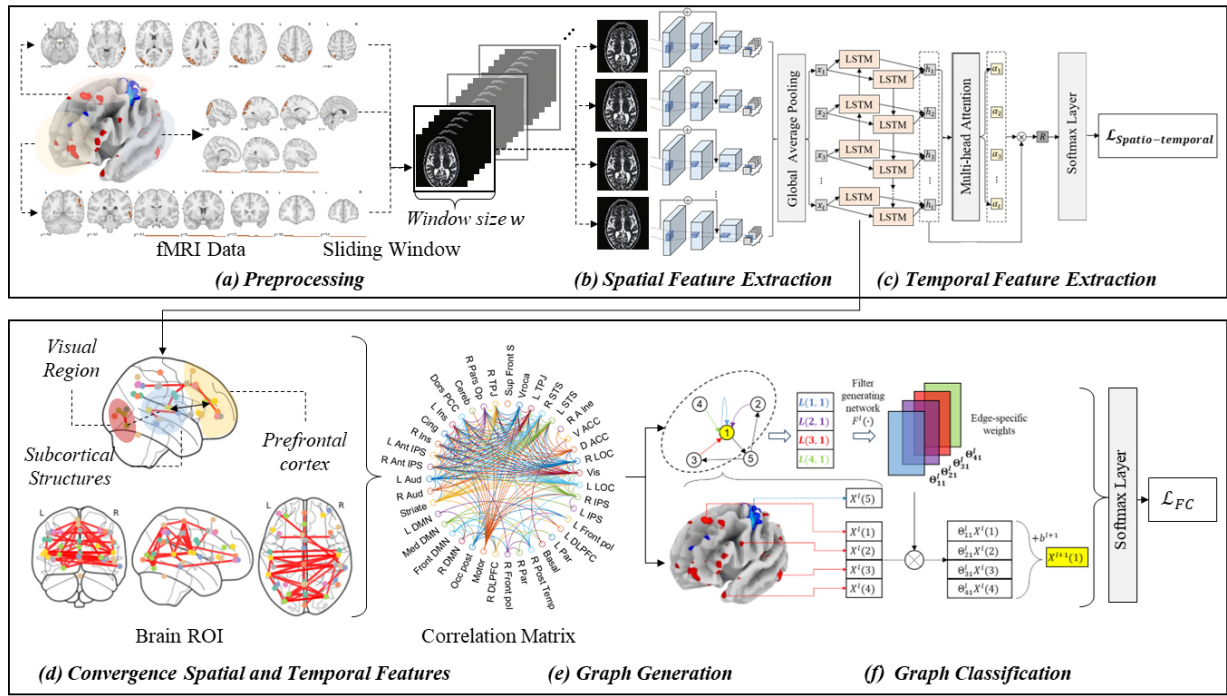


Fig. 1. Overview of the proposed method. (a–c) Steps in the pre-training of attention based spatio-temporal features. (d,e) Steps in the proposed brain network graph generation. (f) The final step to diagnosing the brain disorder with graph classifications. ROI, region of interest; FC, functional connectivity; DMN, Default Mode Network; DLPFC, Dorsolateral Prefrontal Cortex; LOC, Lateral Occipital Complex; IPS, Intraparietal Sulcus; ACC, Anterior Cingulate Cortex; STS, Superior Temporal Sulcus; TPJ, Temporoparietal Junction; PCC, Posterior Cingulate Cortex; L, Left; R, Right.

to age and sex as specified in the data, we will verify the observed heterogeneity in the brain imaging results of the patients. Furthermore, we will reaffirm that the proposed method achieves superior performance in datasets exhibiting this heterogeneity.

2. Related Works

An ensemble of convolutional neural networks (CNNs), encompassing architectures such as ResNet and Inception V3, has demonstrated proficiency in extracting spatial features from brain images, a critical factor in deciphering spatial conditions [26–28]. Such models have not only achieved high accuracy but also facilitated the analysis of both structural and functional brain aspects [29]. Studies that combine spatial features using convolutional neural networks and temporal features using recurrent neural networks have been validated for effectively capturing the brain's intricate spatio-temporal characteristics, which demonstrate superior performance in various disease diagnoses [30–32]. CNN and long short-term memory (LSTM) networks have been extensively used in brain disease diagnosis, particularly leveraging their capability to process spatio-temporal data effectively [33]. CNNs excel in extracting spatial features from complex datasets, such as fMRI images, by identifying patterns and structures within the brain. Meanwhile, LSTMs contribute by analyzing tem-

poral dynamics, which is crucial for understanding how brain activities evolve over time. This synergy allows for a comprehensive analysis of brain function and structure, leading to more accurate diagnosis and prognosis of various neurological conditions [34].

FC analysis has emerged as a fundamental tool in brain research, playing a critical role in identifying biomarkers for a wide range of neurological and mental diseases [35–37]. Traditional FC-based methods are particularly instrumental in mapping the brain's neural network [35]. They do so by creating graphs where nodes represent regions of interest (ROIs), specific areas within the brain that are of relevance to the study or disease in question [37]. The edges in these graphs symbolize the functional connections between these regions, essentially capturing the communication pathways or networks within the brain.

Moreover, the application of graph-based methods in fMRI analysis has significantly advanced understanding of the brain's structure and functionality [38–41]. These graph-based approaches provide a more comprehensive view of the interconnectedness of brain regions, going beyond traditional imaging techniques [38,42]. By examining the intricate network of connections, researchers can uncover patterns and associations that may not be immediately apparent through standard imaging analysis. This level of analysis is particularly valuable in the study of complex

brain disorders, where multiple regions of the brain may be affected simultaneously. Table 1 (Ref. [29–32,39,41,43–45]) summarizes the trends in the latest ASD-related foundational methods, including a classification and description of each method.

3. Proposed Method

Fig. 1 illustrates the proposed model that uses a residual convolutional bidirectional LSTM with a multi-head attention mechanism to extract spatial and temporal features. In the preprocessing phase (Fig. 1a), 4D fMRI images are segmented into 3D blocks at each time interval t , with each block containing k consecutive slices. These k -slice sliding windows move to the spatial feature extraction phase (Fig. 1b), creating a series of overlapping blocks that cover the entire fMRI image sequence by shifting v volumes at a time. The extracted spatial features are then passed to an LSTM in phase (Fig. 1c) for modeling the overall temporal attributes. During this phase, the proposed method employs multi-head attention to assign weights to the temporal attributes. The assigned weights enable the model to learn which temporal attributes are significant, in conjunction with $L_{spatio-temporal}$. At this phase, Fig. 1a–c outlines the classification training to extract spatial and temporal features from brain images through pre-learning. This allows the CNN-based spatial extractor to identify significant spatial features in complex images of the brain, and in part c, the relationships between each spatial feature extractor are learned through LSTM and attention, enabling the extraction of characteristic parts from brain images.

Subsequently, in phase (Fig. 1d), the model explicitly maps the dynamic (temporal) FC correlations of the input image sequence to regions of the brain using features extracted from the attention mechanism in phase (Fig. 1c). This mapping is utilized in phase (Fig. 1e) to create a graph, where the FCs are treated as the adjacency matrix of the graph, the brain regions as nodes, and the edges represent the degree of correlation. The constructed graph is utilized for ASD diagnosis using a GNN. We employ a GNN capable of processing edge labels to utilize the correlations of FC more effectively.

3.1 Residual CNN for Spatial Features

CNNs are widely utilized in analyzing fMRI imaging data for autism diagnosis. Typically, CNNs employ convolutional filters to detect structural features within magnetic resonance imaging (MRI) scans to classify whether the patient has the disease or not. However, this approach often falls short in capturing the unique spatial differences inherent to individual patients. To overcome this limitation, we use an enhanced residual convolutional neural network. This network is designed with a more extensive and deeper architecture compared with traditional CNNs, enabling it to discern spatial variances more effectively between regions, which vary significantly among patients. We have

defined spatial features not in 2-d space but in 3-d voxel space, hence we utilize a 3D ResNet structure.

We have devised a specific residual operation, denoted as $H(\cdot)$, alongside the convolutional operation $F(\cdot)$. These operations process the input at row i , column j , and layer l , represented as x_{ij}^l .

$$H(x) = F(x) + x. \quad (1)$$

The key innovation in the proposed method lies in the incorporation of a residual connection that maintains the gradient $F'(x) + 1$ across all layers.

$$F(x) = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab}^l x_{(i+a)(j+b)}^{l-1}. \quad (2)$$

This feature is particularly critical in addressing the issue of gradient vanishing, a common challenge in deep neural networks. It not only mitigates the gradient vanishing issue but also ensures minimal loss of dynamic FC inherent in convolutional networks. Simultaneously, it maximizes the extraction and analysis of spatial variations.

3.2 Multi-Head Attention for Temporal Features

The spatial features of the brain extracted through the residual network are fed into a BiLSTM-based multi-head attention mechanism for mapping time series attributes. One of the primary objectives of the proposed method is to extract and enhance transient brain features appearing in fMRI images, which is why we employ an attention-based mechanism to map these time series attributes. We use multi-head attention instead of single head attention because the latter is often insufficient to fully extract the dependencies between temporal features in the complex spatial context of the brain. Multi-head attention allows us to model these dependencies more effectively.

When considering the features $x_1, x_2, \dots, x_k$ extracted from the residual CNN blocks, their embedding through a BiLSTM and multi-head attention mechanism can be described by referencing the formula provided. At first, the features $x_1, x_2, \dots, x_k$ of the CNN are input into the BiLSTM network, which processes these features to capture the temporal dependencies, producing a series of hidden states $h_1, h_2, \dots, h_N$. In the context of the multi-head attention mechanism, each pair of hidden states h_i and h_j is evaluated to compute the attention weights a_{ij} . This is achieved through the following softmax function:

$$a_{ij} = \text{softmax}_j(h_{ij}) = \frac{\exp(\text{score}(h_i, h_j \cdot W_v h_i))}{\sum_{n=1}^N \exp(h_i, h_n \cdot W_v h_i)}. \quad (3)$$

Here, $\text{score}(h_i, h_j)$ is a function that measures the compatibility of the hidden states h_i and h_j , which is calculated as:

$$\text{score}(h_i, h_j) = f(W_q h_i, W_k h_j) \quad (4)$$

where W_q and W_k are weight matrices for the query and key in the attention mechanism, respectively.

Finally, the output representation r_i for each hidden state h_i is computed as a weighted sum of all hidden states, with the weights being the attention scores:

$$r_i = \sum_{n=1}^N a_{in} \cdot h_n. \quad (5)$$

This process is repeated independently K times, and the results are concatenated to form a single vector. The proposed multi-head attention's integration of spatial and temporal aspects enables it to focus on specific time frames and effectively model the transient characteristics of diseases. This mechanism allows for the analysis of disease characteristics from multiple perspectives, providing a detailed and multi-faceted understanding of the unique features of the disease.

3.3 Learning Dynamic Functional Connectivity

Park *et al.* [41] demonstrated that the extraction of spatio-temporal features while preserving the integrity of dynamic connectivity significantly enhances accuracy. Nevertheless, a more direct method to learn the dynamic connectivity across 39 brain regions could offer further improvements. As depicted in Fig. 1e, the method leverages the spatio-temporal model as a foundational, pre-trained model. We transform the output of the intermediate layer, which consists of time-series spatial feature maps, into a connectivity matrix that delineates the relationships between 39 distinct brain regions. Subsequently, this spatio-temporal feature map is merged with a connectivity adjacency matrix, incorporating attention weights between regions. The GNN is then employed to explicitly learn and articulate the connection information among these 39 brain regions.

Upon processing the fMRI data through the LSTM network, enhanced with an attention mechanism, the output consists of a series of feature representations, symbolized as $R = r_1, r_2, \dots, r_T$. Each element r_t in this series represents the features for different brain regions at time t , which are then fed into the 'ConnectivityMeasure' module to compute the connectivity matrices.

The connectivity matrix, denoted as C , is derived from these feature representations. Specifically, for each pair of brain regions, denoted as i and j , over the entire time series T , the connectivity metric is computed, resulting in a matrix C of $N \times N$ dimensions, where N is the number of brain regions (we use 39 regions). This can be formally represented as:

$$C_{ij} = \text{ConnectivityMeasure}(r_i, r_j). \quad (6)$$

Here, C_{ij} represents the connectivity value between regions i and j . The 'ConnectivityMeasure' method allows us to apply various metrics such as correlation, partial corre-

lation, tangent, or covariance, depending on analysis needs. This results in different forms of the matrix C , each encapsulating unique aspects of inter-regional brain dynamics:

$$C_{metric} = \text{metric}(F), \quad (7)$$

where $\text{metric} \in [\text{Corr}, \text{Pcorr}, \text{Tan}, \text{Cov}]$.

This method leads to the creation of a comprehensive connectivity matrix C , which captures the dynamic and complex inter-regional relationships within the brain. Such a matrix not only illustrates the direct interactions between different brain areas but also provides deeper insights into the overall functional network architecture.

3.4 Assigning Discrete Edge Label

Next, we interpret the connectivity matrix C obtained from fMRI data as the adjacency matrix A of a graph, with each element C_{ij} representing the strength of connection between brain regions i and j . This matrix is crucial for understanding the complex interconnections within the brain. To facilitate graph-based analysis, we categorize the continuous correlation values in matrix C into discrete edge labels based on quartiles, as follows:

$$\begin{cases} 1 & \text{if } C_{ij} \in Q_1 \\ 2 & \text{if } C_{ij} \in Q_2 \\ 3 & \text{if } C_{ij} \in Q_3 \\ 4 & \text{if } C_{ij} \in Q_4 \end{cases}, \quad \text{where } Q_n \text{ is } n\text{-th quartile.} \quad (8)$$

This labeling method allows us to transform the continuous brain connectivity data into a format suitable for GNN analysis.

3.5 Edge Conditioned Layer and Classification

In the proposed method, we translate the fMRI-derived connectivity matrix C into an adjacency matrix A for graph construction, with discrete edge labels derived from C 's quartile-based categorization. The edge-conditioned convolution (ECC) layer plays a crucial role in embedding this graph, capturing intricate connectivity patterns for ASD diagnosis. The ECC layer is defined by the following process:

Given the graph $G = (V, E)$ with nodes V (representing brain regions) and edges E (representing connections between regions), and an edge label function $L: E \rightarrow \mathbf{R}^s$ assigning labels to each edge, the ECC layer embeds the graph by computing a signal X^l at each node. This signal is a function of the node's features and its neighborhood structure, mathematically defined as:

$$X^l(v) = \sigma \left(\sum_{u \in N(v)} \Theta_{uv}^l \cdot X^{l-1}(u) \right). \quad (9)$$

Here, $X^l(v)$ is the feature vector of node v at layer l , $N(v)$ denotes the set of neighbors of v , Θ_{uv}^l is the edge-conditioned weight matrix for the edge from u to v at layer l ,

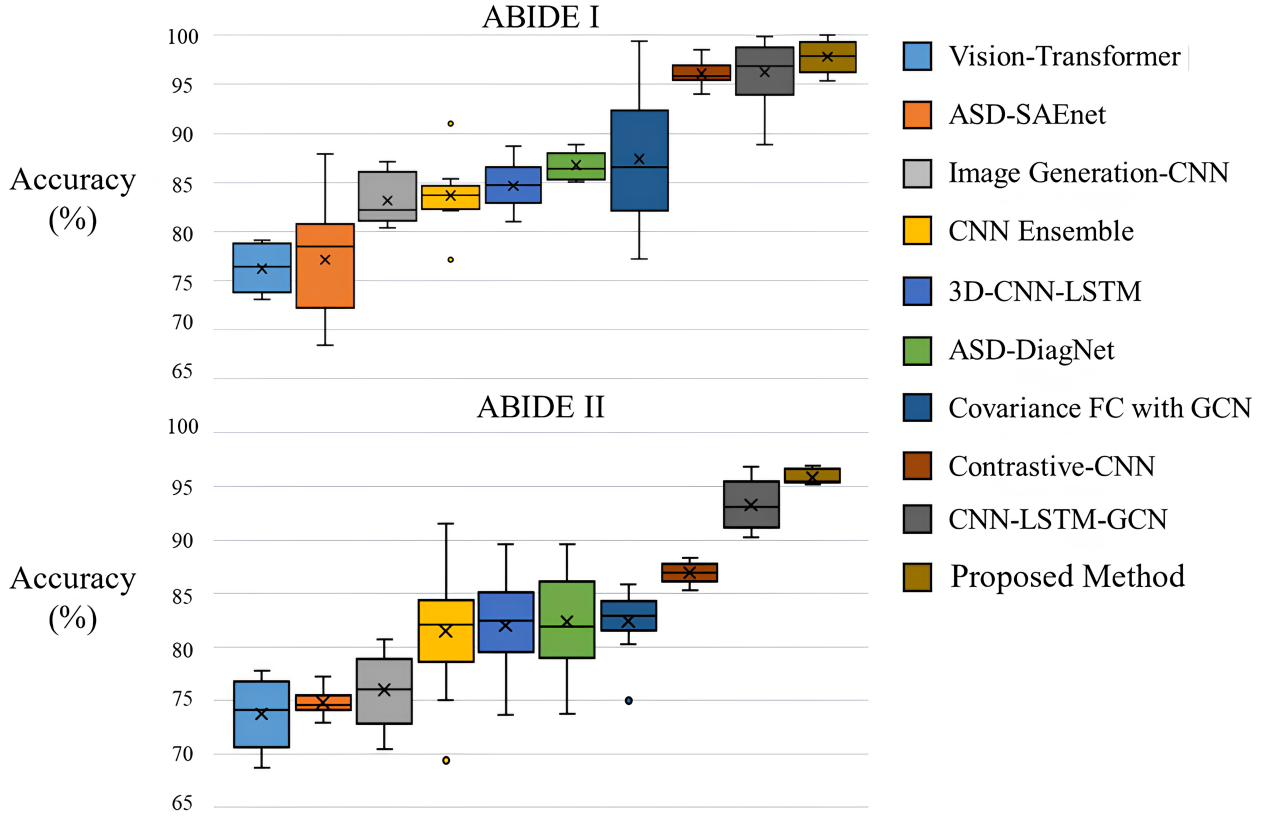


Fig. 2. Box plot of accuracy on the ABIDE I and II datasets. ABIDE, autism brainimaging data exchange.

and σ is a non-linear activation function. The weight matrix Θ_{uv}^l is computed as a function of the edge label $L(u, v)$, which reflects the strength and characteristics of the connectivity between regions u and v :

$$\Theta_{uv}^l = f^l(L(u, v)). \quad (10)$$

where f^l is a learnable function specific to layer l , typically implemented as a neural network.

Through successive ECC layers, the graph is embedded into a high-dimensional space that captures not only the features of individual brain regions but also the complex connectivity patterns among them. This embedded representation is then used for classification tasks, such as identifying ASD. The ECC's ability to condition on edge labels enables the model to adaptively learn different patterns of brain connectivity, crucial for accurate disease diagnosis.

The final classification is performed by feeding the embedded graph representation through a series of fully connected layers, culminating in a softmax layer for disease categorization:

$$Y = \text{softmax}(\text{FC}(X)^L). \quad (11)$$

where Y is the output classification, X^L is the feature representation from the final ECC layer, and FC denotes the fully connected layers. In the proposed method, we do not

use the typical graph classification Readout. Since we have predefined that each graph consists of the same number of nodes (39 in total), we do not employ an integrated graph embedding method that accounts for this.

4. Experimental Results

The experimental setup was conducted using a Nvidia DGX station (Nvidia, Santa Clara, CA, United States) equipped with 2560 Nvidia tensor cores, running on an Ubuntu desktop Linux OS. The hardware included four Tesla V100 GPUs (64 GB each) and 256 GB of LRDIMM DDR4 memory. For the software environment, we utilized Ubuntu with Python 3.x (<https://www.python.org/>) and TensorFlow version 2.3 (<https://www.tensorflow.org/>). Additionally, various Python libraries (<https://pypi.org/>) such as Scikit-learn, Nilearn, Nibabel, Monai, and Networkx were utilized for the experiments.

4.1 Dataset and Implementation

We employed the preprocessed ABIDE I and II datasets [44], accessible through a web-based platform. We preprocessed the dataset using the Configurable Pipeline for the Analysis of Connectomes (C-PAC) as our preprocessing pipeline. This version of the preprocessing steps includes slice timing and motion correction, intensity normalization (4D global mean = 1000), nuisance signal regression, band-pass filtering (0.01–0.1 Hz), and registration to the 3 mm

Table 2. ABIDE I and II data phenotypical information summary.

ABIDE I							ABIDE II						
Site	Count				Min age (y)	Max age (y)	Site	Count				Min age (y)	Max age (y)
	ASD	Control	Male	Female				ASD	Control	Male	Female		
CALTECH	19	19	30	8	17.0	56.2	BNI	29	29	58	0	18.0	64.0
CMU	14	13	21	6	19.0	40.0	EMC	27	27	44	10	6.2	10.7
KKI	22	33	42	13	8.07	12.77	ETH	13	24	37	0	13.8	30.7
LEUVEN_1	14	15	29	0	18.0	32.0	GU	51	55	71	35	8.1	13.9
LEUVEN_2	15	20	27	8	12.1	16.9	IP	22	34	26	30	6.1	46.6
MAX_MUN	24	33	50	7	7.0	58.0	IU	20	20	31	9	17.0	54.0
NYU	79	105	147	37	6.47	39.1	KKI	56	155	140	71	8.0	13.0
OHSU	13	15	28	0	8.0	15.23	KUL	28	0	28	0	18.0	35.0
OLIN	20	16	31	5	10.0	24.0	NYU_1	48	30	71	7	5.2	34.8
PITT	30	27	49	8	9.33	35.2	NYU_2	27	0	24	3	5.1	8.8
SBL	15	15	30	0	20.0	64.0	OHSU	37	56	57	36	7.0	15.0
SDSU	14	22	29	7	8.6	17.1	OILH	24	35	40	19	18.0	31.0
STANFORD	20	20	32	8	7.5	12.9	SDSU_1	33	25	49	9	7.4	18.0
TRINITY	24	25	49	0	12.0	25.9	SDSU_2	21	21	38	4	8.4	13.2
UCLA_1	49	33	71	11	8.4	17.9	TCD	21	21	42	0	10.0	20.0
UCLA_2	13	14	25	2	9.79	16.47	UCD	18	14	24	8	12.0	17.8
UM_1	55	55	84	26	8.2	19.2	UCLA	16	16	26	6	7.8	15.0
UM_2	13	22	33	2	12.8	28.8	USM	17	16	28	5	9.1	38.9
USM	58	43	101	0	8.8	90	U_MIA	13	15	22	6	7.1	14.3
YALE	28	28	40	16	7.0	17.8							

ABIDE, Autism Brain Imaging Data Exchange.

Montreal Neurological Institute (MNI) standard template. The fMRI volumes were down sampled to 4 mm in the MNI space. During the training phase, the data was augmented through intensity normalization, random Gaussian noise injection, and affine transformations, along with sequence sampling. During inference, a sliding window inference technique was employed, where the model made predictions for a continuous sequence of 10 frames (window size of 10) from user data using a sliding window approach. Soft voting was then conducted using the model's predictions across the sliding window to aggregate the final prediction results. Table 2 shows the data collection institutions for ABIDE I and II, along with the distributions of ASD/normal and male/female, and the minimum and maximum ages.

Given the variability in the size and length of each fMRI image, we standardized all images to a uniform size of $24 \times 24 \times 24$ and applied the sliding window technique to acquire serial images of fixed length. To ensure no overlap in patient information between training and testing, the dataset was partitioned based on individual patients rather than the images themselves. Consequently, the dataset was split into a training set comprising 90% of the cohort and a validation set consisting of the remaining 10%. The validation set was utilized for visualization and various analyses. Detailed information about the dataset used and the details of the proposed model are shown in Tables 3,4.

Table 3. Configuration data for training, validation, and testing by dataset (ABIDE I and ABIDE II).

Dataset	Type	Configure		
		Train	Valid	Test
ABIDE I	ASD	421	47	52
	Control	449	50	55
ABIDE II	ASD	422	47	52
	Control	481	53	59

4.2 Experimental Results

The results of the 10-fold cross-validation experiments in ABIDE I and II demonstrated the superiority of the proposed method. As shown in Fig. 2 and Table 5 (Ref. [29–32,41,44–47]) when compared with the graph-based approach Covariance FC with GCN, the method exhibited a performance improvement of approximately 3.7%p in the ABIDE I dataset and about 2.1%p in the ABIDE II dataset. This performance difference is attributed to two main reasons. Firstly, the use of covariance in FC for graph construction in traditional methods may not sufficiently capture the neurological characteristics of the brain. The proposed method, applying multi-head attention, diversifies perspectives in extracting correlations, thereby enhancing the ability to identify and emphasize heterogeneous data features from fMRI data. Secondly, unlike GCN which uses only node features as input, the method benefits

Table 4. Model structure of proposed pre-training method.

Operation	Time Distributed	In Channels	Out Channels	Kernel (Layer)	Stride	Padding	BatchNorm	Dropout	Activation
Input: $10 \times 1 \times 24 \times 24 \times 24$ (1 channel voxel with time)									
Input stem	True	1	32	$3 \times 3 \times 3$	$2 \times 2 \times 2$	$1 \times 1 \times 1$	True	0.3	ReLU
Layer 1	True	32	64	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	True	0.3	ReLU
Layer 2	True	64	128	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	True	0.3	ReLU
Max pooling	True	128	128	$2 \times 2 \times 2$	$2 \times 2 \times 2$	X	X	X	X
Layer 3	True	128	256	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	True	0.3	ReLU
Layer 4	True	256	512	$3 \times 3 \times 3$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	True	0.3	ReLU
Global Average Pooling	True	-	-	$3 \times 3 \times 3$	-	-	-	-	-
Bi-Directional LSTM	True	512	1024	2	-	-	-	X	ReLU
Multi-head Attention	-	1024	1024	8	-	-	-	X	ReLU
Fully Connected 1	-	1024	256	-	-	-	True	0.5	ReLU
Fully Connected 2	-	256	2	-	-	-	-	0.3	Softmax
Optimizer	Adam (lr = 0.0001, weight decay = 0.0005)								
Batch size	128								
Epochs	1000 epochs (early stopping 20)								
Lr scheduler	StepLR (step_size = 50, gamma = 0.5)								
Total Params: 20,236,146									

ReLU, Rectified Linear Unit.

Table 5. Accuracy on the ABIDE I and II datasets.

Task (dataset)	Method	Accuracy
ABIDE I	Bootstrapping ensemble of GCNNs [47]	68.46 ± 7.94
	ASD-DiagNet [31]	76.22 ± 2.29
	Multi-view ensemble learning [44]	74.16 ± 1.31
	ASD-SAEEnet [32]	77.13 ± 5.44
	R-walk [45]	73.45 ± 2.81
	CNN Ensemble [29]	83.67 ± 3.24
	3D-CNN-LSTM [30]	84.65 ± 2.21
	ResNet-LSTM with self-attention [41]	86.74 ± 1.34
	Covariance FC with GCN	87.66 ± 6.84
	Self-supervised ensemble [46]	94.13 ± 3.12
	Proposed Method	97.88 ± 1.64
ABIDE II	Multi-view ensemble learning [44]	73.75 ± 2.97
	ASD-SAEEnet [32]	74.81 ± 1.13
	R-walk [45]	75.99 ± 3.28
	CNN Ensemble [29]	81.48 ± 5.66
	3D-CNN-LSTM [30]	82.38 ± 4.74
	ResNet-LSTM with self-attention [41]	82.39 ± 2.93
	Covariance FC with GCN	86.94 ± 0.91
	Self-supervised ensemble [46]	93.23 ± 2.18
	Proposed Method	95.35 ± 0.12

from the Edge conditioned layer that models the correlations between nodes, leading to more effective embedding of the constructed FC and subsequently improved performance (detailed in section 4.3). Due to these characteristics,

the proposed method outperformed existing 3D CNN neural network-based approaches that achieved state-of-the-art (SOTA) results through additional parameters or enhanced learning algorithms.

The receiver operating characteristic (ROC) curves of the five models are shown in Fig. 3. In contrast to alternative network models, our proposed model demonstrates a more balanced trade-off between accuracy and recall rates. Additionally, the area under the ROC curve (AUC) index of our model surpasses that of the others, affirming its capacity to capture the intrinsic features of the data. This underscores the superior generalization ability of our model.

4.3 Analysis of Correlations of ROI

Fig. 4 illustrates the brain regions activated as identified through the proposed method's model internal attention values, based on the presence or absence of various diseases, age, and sex derived from data and additional analyses. Notably, in the activation maps of Fig. 4a–c, where ASD is present, each map shows different regional activations, highlighting that the areas highlighted by the proposed method vary with age and sex. Furthermore, the activation maps of Fig. 4c,d for similar age groups and the same sex show generally similar patterns of highlighted regions, despite the presence or absence of disease. This underlines the necessity for models diagnosing ASD to capture not only specific regions and times of brain activity

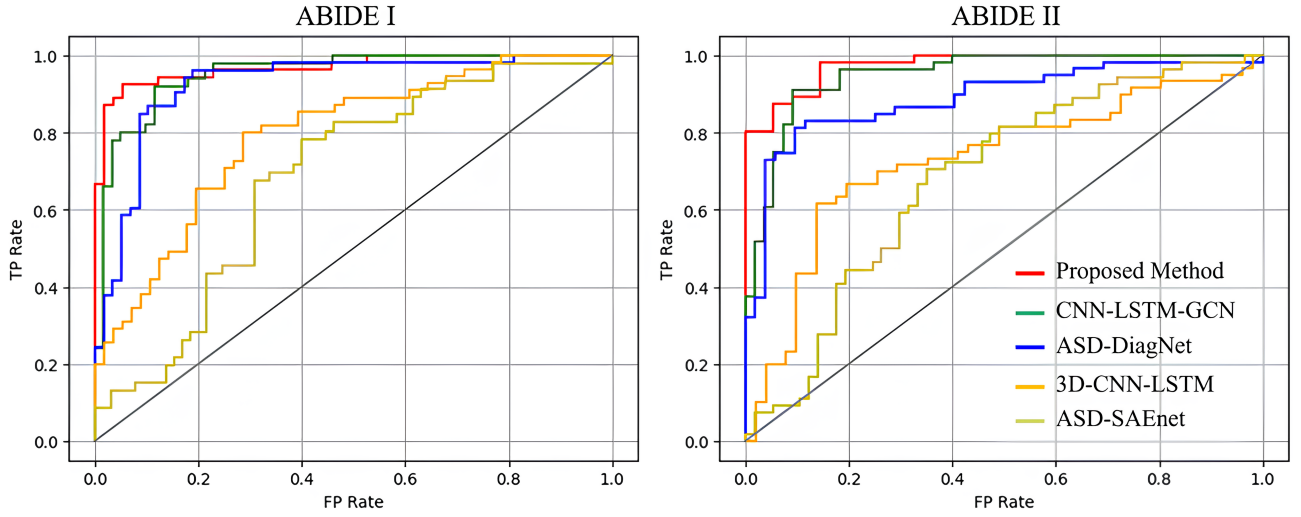


Fig. 3. ROC curves on the ABIDE I and II datasets. ROC, receiver operating characteristic; FP, False Positive; TP, True Positive.

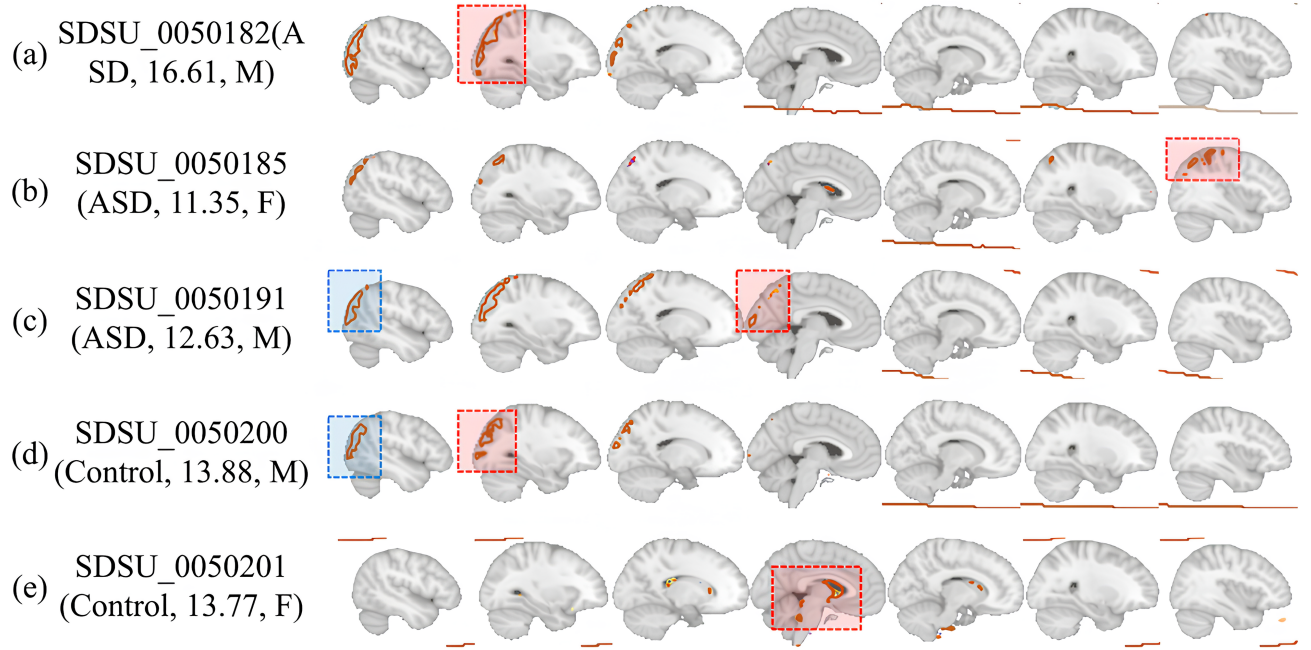


Fig. 4. Brain ROI activations depend on the presence or absence of disease, age, and sex. (a–c) represent brain scans of patients with ASD, while (d,e) depict brain scans of controls. The blue boxes denote common features visible in both individuals with the disease and those without it, while the red boxes indicate the level of activation varying according to age, gender, and disease status of each patient and control. ROI, represent regions of interest; M, male; F, female.

but to also comprehend the extent of each disease present in the dataset and the unique characteristics of each patient. In essence, this disorder exhibits heterogeneity based on disease progression, sex, and age, indicating that various complex external factors must be considered in diagnosing ASD. Based on our analysis of Fig. 4 and the detailed observations of brain region activations through the proposed method's attention values, it becomes evident that our approach excels at extracting the nuanced characteristics of ASD and its variability across different ages and sexes. This

proficiency is particularly crucial given the heterogeneity of ASD, which manifests distinctly in each individual based on a complex interplay of factors including disease progression, age, and sex.

Fig. 5 illustrates the analysis of the correlation matrix extracted by the proposed method. As observed in this correlation matrix, the time-series features extracted from ASD patients and processed through the attention mechanism closely connect various regions of the brain, reflecting their interdependencies and enabling effective prediction of

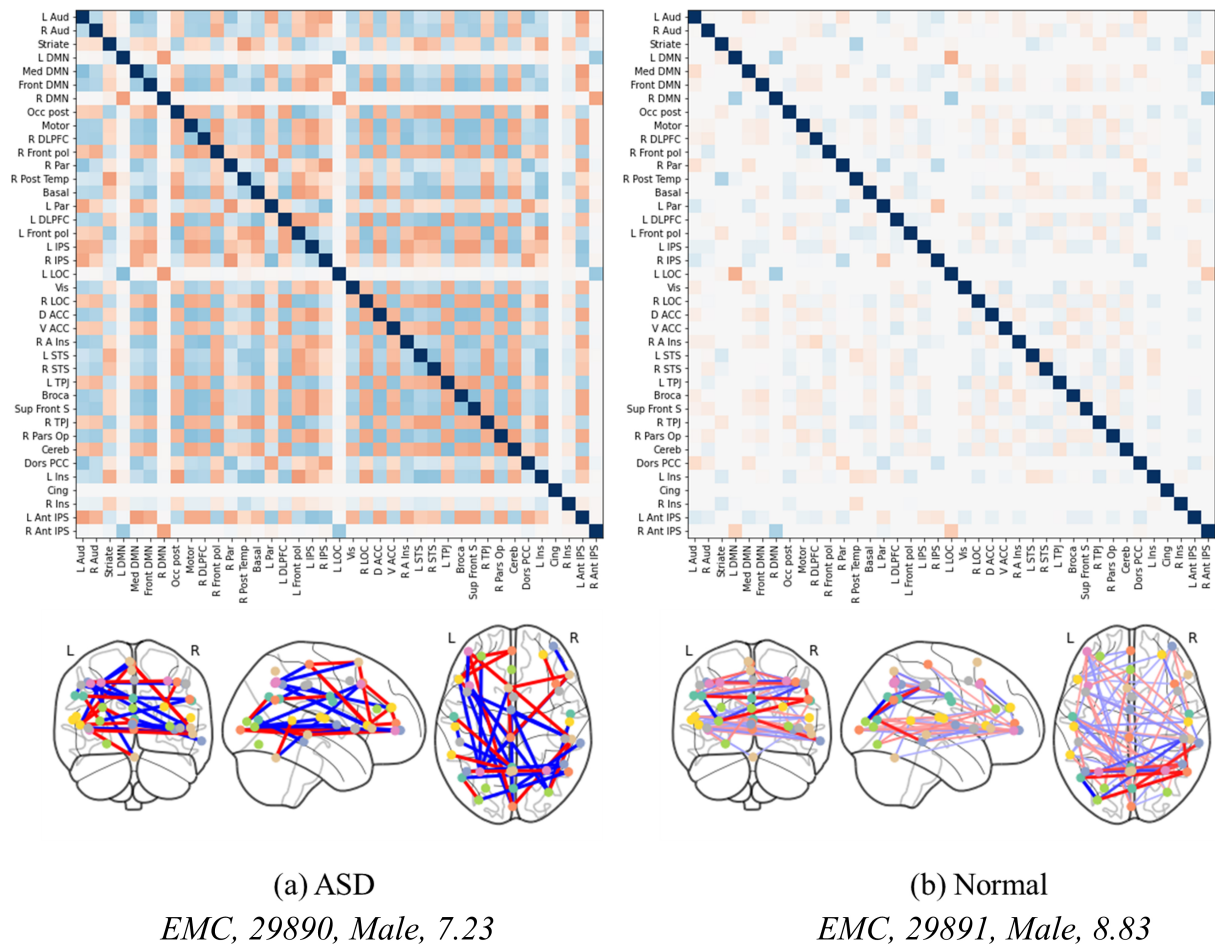


Fig. 5. Correlation matrices and brain regional connectivity graphs in ASD and normal brains. (a,b) show the level of brain activation between normal individuals and patients of similar age groups.

ASD. In contrast, in a normal brain, the attention mechanism does not establish as strong connections between regions, indicating that the model fails to identify significant characteristics that could highlight connections across different brain areas. This distinction highlights the strength of the proposed method, which leverages attention-based feature extraction, graph transformation, and a GNN. These steps enhance the differentiation of disease-specific features, thereby improving the accuracy of disease diagnosis. The method's effectiveness is rooted in its ability to accentuate the contrast between the neurobiological patterns of ASD and those of a typical brain, illustrating the nuanced complexities captured by the proposed method. The integration of the attention mechanism with graph-based analyses provides a more comprehensive understanding of brain connectivity patterns, crucial for discerning the subtle yet significant differences in neurological conditions. This advanced methodology demonstrates a significant stride in enhancing diagnostic precision through sophisticated analysis of fMRI data.

Fig. 6 depicts FC overlaid on brain images to analyze differences based on the presence of disease, as well

as sex and age. Below each image, details such as the collection site, subject ID, sex, age, and disease status are documented. This allows for a detailed examination of patterns in ASD across different age groups, as observed in the differences between images Fig. 6a,b. This not only suggests that FC characteristics in ASD can vary with age and between individuals but also highlights the importance of deep learning models in capturing these individual variations. Furthermore, a comparison between images Fig. 6a,c demonstrates sex-based variations in brain characteristics. For instance, the posterior region of the left hemisphere is almost inactive in image Fig. 6c compared with Fig. 6a, highlighting the diversity in brain activation patterns in ASD. The differences among images Fig. 6a–d, where no similar patterns are observed, underscore the necessity of the proposed model. Additionally, the similarities observed in images Fig. 6d–f, regardless of the presence or absence of disease, present a unique challenge. Considering that d represents ASD while Fig. 6e,f are from healthy brains, distinguishing these subtle differences becomes a primary function of the model. Therefore, the effectiveness of our proposed method, which employs enhanced FC with multi-

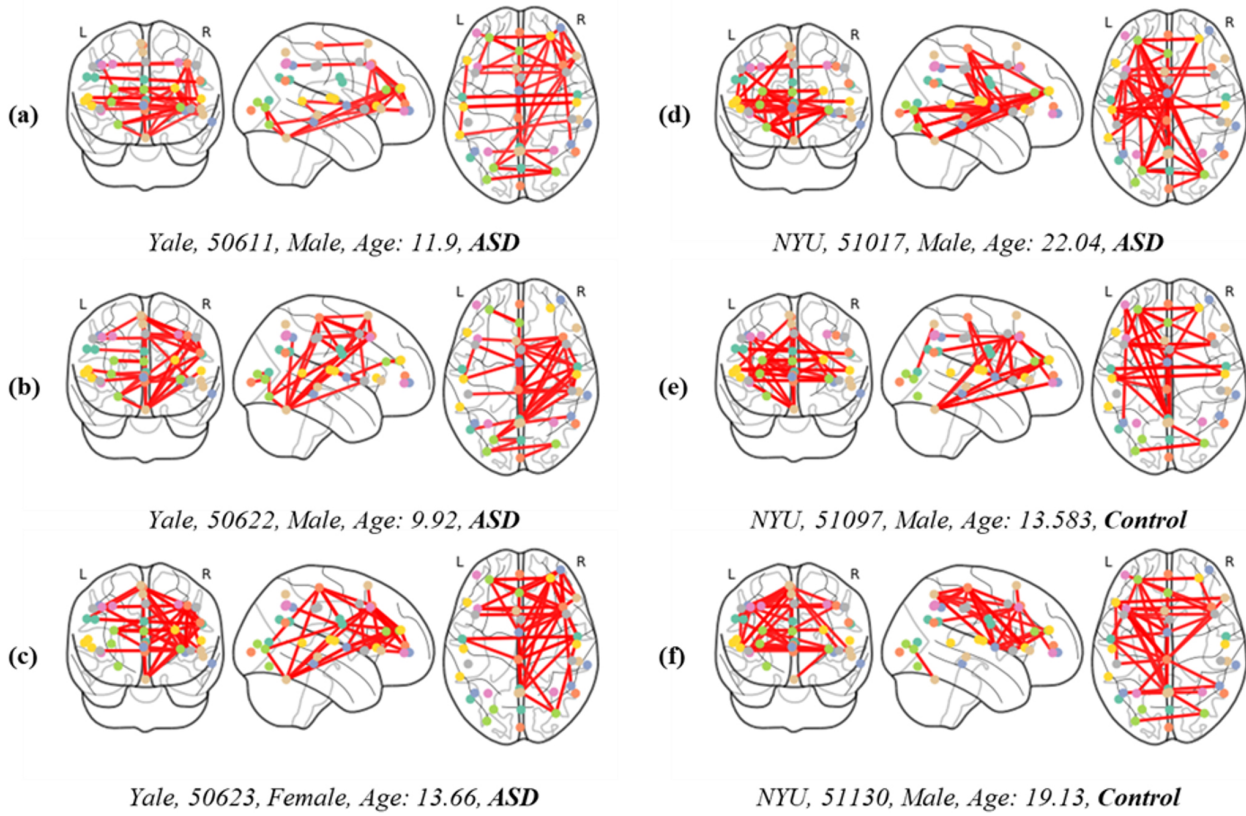


Fig. 6. FC displayed over brain images, categorized by the presence or absence of disease, and differentiated by sex and age (measured in years). Subgraphs (a–f) represent the interconnectivity between brain regions of both patients and control individuals.

head attention and an edge label-based GNN, is validated by its superior performance.

Table 6 shows the statistical significance of various methods in comparison with the proposed method across a range of experimental settings. Each row represents a different method, while the columns delineate the p -values obtained through Wilcoxon tests and their corresponding false discovery rate (FDR)-corrected values. These metrics serve as indicators of the likelihood that the observed differences between the proposed method and its counterparts are due to chance. Upon examining the table, it is evident that the p -values associated with the proposed method consistently exhibit a substantial disparity compared with other techniques. For instance, the p -values for CNN Ensemble, Covariance FC with GCN, ASD-SAE_{net}, Vision-Transformer, 3D-CNN-LSTM, ASD-DiagNet, Contrastive CNN, and CNN-LSTM-GCN are 3.5×10^{-9} , 6.58×10^{-8} , 2.2×10^{-6} , 1.3×10^{-5} , 3.4×10^{-5} , 4.1×10^{-5} , 1.1×10^{-3} , and 8.4×10^{-3} , respectively, with FDR-corrected values in a similar range. This stark contrast suggests that the differences observed between the proposed method and each alternative are highly unlikely to have occurred by random chance alone. Furthermore, the FDR-corrected p -values reinforce the robustness of these findings, indicating that even after adjusting for multiple comparisons, the statistical significance of the disparities remains pronounced.

Thus, based on these analyses, it can be confidently concluded that the proposed method exhibits statistically significant differences compared with the other methodologies evaluated in the study. Such findings underscore the efficacy and potential superiority of the proposed approach in the context of the addressed problem domain.

Our experimental analysis confirms that the proposed method significantly outperforms existing models in detecting ASD, showcasing its potential to improve early diagnosis and contribute to personalized healthcare strategies. The attention-based feature extraction mechanism effectively identifies the most relevant features across diverse patient profiles, illustrating a deep understanding of the disorder's complexity. This capability is a testament to the model's design, which integrates advanced machine learning techniques to navigate the intricate patterns of brain activity unique to ASD.

In conclusion, the proposed method introduces a more complex yet significantly more effective approach to ASD diagnosis. Despite the increase in complexity, this method marks a substantial advancement by offering unparalleled precision in understanding and diagnosing the disorder. Its sophisticated use of advanced techniques, including but not limited to multi-head attention, allows for the precise extraction of features from brain activation maps, enhancing diagnostic accuracy and our understanding of ASD.

Table 6. Results of statistical analyses for the proposed method and other methods using the same dataset.

Method	p -value (Wilcoxon)	Corrected p -value (FDR)
**CNN Ensemble	3.5×10^{-9}	2.8×10^{-8}
**Covariance FC with GCN	6.58×10^{-8}	2.6×10^{-7}
**ASD-SAEnet	2.2×10^{-6}	5.9×10^{-6}
**Vision-Transformer	1.3×10^{-5}	52.6×10^{-5}
**3D-CNN-LSTM	3.4×10^{-5}	5.4×10^{-5}
**ASD-DiagNet	4.1×10^{-5}	5.4×10^{-5}
*Contrastive CNN	1.1×10^{-3}	1.3×10^{-3}
CNN-LSTM-GCN	8.4×10^{-3}	8.4×10^{-3}

* $p < 0.005$. ** $p < 0.0005$. FDR, false discovery rate.

This method's ability to navigate the disorder's complexities provides clear advantages, including the potential for more accurate diagnoses, the development of personalized treatment plans, and the facilitation of further research into ASD. By leveraging these advanced techniques, the proposed method sets a new standard for ASD diagnosis, promising improved outcomes for individuals with ASD and paving the way for more effective, personalized care.

4.4 Limitations

Although our method has shown promising results in effectively extracting brain networks and has undergone rigorous performance and qualitative evaluations for disease identification, we acknowledge the existence of multiple challenges that impede our progress towards a truly comprehensive diagnostic tool. In practical scenarios, it is highly unlikely to encounter situations where only a single disease needs detection. This reality encourages us towards the ambitious goal of developing a universal diagnostic model that can seamlessly transition between diagnosing various neurological conditions, such as Alzheimer's disease, attention deficit/hyperactivity disorder (ADHD), and beyond.

However, a significant hurdle that currently limits our research is the focused nature of existing studies on particular diseases. This specificity necessitates a unique pre-processing routine for each disease dataset, complicating the development of a universal model. The need for a simplified and standardized preprocessing procedure cannot be overstated, as it would significantly facilitate the adaptation of models to handle diverse datasets efficiently.

Moreover, the advent of transformer-based feature extractors, renowned for their efficacy in numerous fields, particularly due to their strengths in transfer learning, presents an opportunity for innovation in fMRI data analysis. These models' ability to adapt and learn from vast amounts of data could revolutionize how we approach fMRI data, making them an excellent candidate for future research. Nonetheless, developing a transformer-based model that caters specifically to the unique characteristics of fMRI data, such as its high dimensionality and the subtle nature of signal variations indicative of different neuro-

logical conditions, poses an additional layer of complexity. Such a model would need to not only extract relevant features from the fMRI data effectively but also be capable of doing so in a way that is computationally feasible and clinically relevant.

To evolve into a universal model that is adaptable to various downstream tasks derived from fMRI data, we must also consider the integration of multimodal data sources, including genetic information, clinical assessments, and behavioral data. This integration would undoubtedly enhance the model's diagnostic capabilities but also introduces the challenge of managing and analyzing heterogeneous data types.

5. Conclusion

In this paper, we present a significant advancement in understanding and diagnosing neurological conditions, using advanced neural network techniques applied to fMRI data. The proposed method, which incorporates attention-based feature extraction, graph transformation, and GNNs, has demonstrated superior accuracy in identifying ASD and differentiating it from normal (neurotypical) brain patterns. The SOTA models achieve 94.13% and 93.23% accuracy on the ABIDE I and II datasets, whereas our proposed method has achieved 97.88% and 95.35% accuracy on the same datasets. Thus, we achieved an accuracy increase of up to 3.7%p over other SOTAs, outperforming the latest model methods. Our proposed method has not only enhanced the accuracy of ASD diagnosis but also provides deeper insights into the underlying neurobiological mechanisms.

Future research will focus on developing brain feature extraction methods using contrastive learning. Although the proposed method is effective, it is essential that the feature extraction process is enhanced, especially for identifying patterns in continuous images. The goal is to synergistically integrate feature extraction with graph learning, aiming for more refined and accurate neural network models for use in neurological diagnostics.

Abbreviations

ASD, Autism spectrum disorder; FC, functional connectivity; SOTA, state-of-the-art; CNNs, convolutional neural networks; ROIs, represent regions of interests; GNN, graph neural network; LSTM, long short-term memory; GCN, graph convolutional neural networks; ROC, receiver operating characteristic; AUC, area under the ROC curve; ADHD, attention deficit/hyperactivity disorder.

Availability of Data and Materials

The datasets used in this study can be found in the ABIDE repository (<http://preprocessed-connectomes-project.org/abide/>).

Author Contributions

HJM and SBC designed the research study, and wrote the manuscripts. HJM conducted the implementation of the method and analyzed the results. SBC guided the overall research including the implementation and the analysis. Both authors contributed to editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Sung-Bae Cho appreciates the Yonsei Fellow Program funded by Lee Youn Jae and Air Force Defense Research Sciences Program funded by Air Force Office of Scientific Research.

Funding

This work was supported by IITP grant funded by the Korean government (MSIT) (No. RS-2020-II201361, Artificial Intelligence Graduate School Program (Yonsei University); No. RS-2022-0-II220113, Developing a Sustainable Collaborative Multi-modal Lifelong Learning Framework) and ETRI grant funded by the Korean government (24ZB1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System).

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Jönemo J, Abramian D, Eklund A. Evaluation of Augmentation Methods in Classifying Autism Spectrum Disorders from fMRI Data with 3D Convolutional Neural Networks. *Diagnostics*. 2023; 13: 2773.
- [2] Thomas RM, Gallo S, Cerliani L, Zhutovsky P, El-Gazzar A, van Wingen G. Classifying Autism Spectrum Disorder Using the Temporal Statistics of Resting-State Functional MRI Data With 3D Convolutional Neural Networks. *Frontiers in Psychiatry*. 2020; 11: 440.
- [3] Park K, Cho S. A residual graph convolutional network with spatio-temporal features for autism classification from fMRI brain images. *Applied Soft Computing*. 2023; 142: 110363.
- [4] Barnea-Goraly N, Frazier TW, Piacenza L, Minshew NJ, Keshevan MS, Reiss AL, *et al.* A preliminary longitudinal volumetric MRI study of amygdala and hippocampal volumes in autism. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2014; 48: 124–128.
- [5] Schumann CM, Hamstra J, Goodlin-Jones BL, Lotspeich LJ, Kwon H, Buonocore MH, *et al.* The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. *The Journal of Neuroscience*. 2004; 24: 6392–6401.
- [6] Nordahl CW, Scholz R, Yang X, Buonocore MH, Simon T, Rogers S, *et al.* Increased rate of amygdala growth in children aged 2 to 4 years with autism spectrum disorders: a longitudinal study. *Archives of General Psychiatry*. 2012; 69: 53–61.
- [7] Nordahl CW, Iosif AM, Young GS, Hechtman A, Heath B, Lee JK, *et al.* High Psychopathology Subgroup in Young Children With Autism: Associations With Biological Sex and Amygdala Volume. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2020; 59: 1353–1363.e2.
- [8] Stoodley CJ. Distinct regions of the cerebellum show gray matter decreases in autism, ADHD, and developmental dyslexia. *Frontiers in Systems Neuroscience*. 2014; 8: 92.
- [9] Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, *et al.* Early brain development in infants at high risk for autism spectrum disorder. *Nature*. 2017; 542: 348–351.
- [10] Ohta H, Nordahl CW, Iosif AM, Lee A, Rogers S, Amaral DG. Increased Surface Area, but not Cortical Thickness, in a Subset of Young Boys With Autism Spectrum Disorder. *Autism Research*. 2016; 9: 232–248.
- [11] Thion MS, Low D, Silvén A, Chen J, Grisel P, Schulte-Schrepping J, *et al.* Microbiome Influences Prenatal and Adult Microglia in a Sex-Specific Manner. *Cell*. 2018; 172: 500–516.e16.
- [12] Erny D, Hrabě de Angelis AL, Jaitin D, Wieghofer P, Staszewski O, David E, *et al.* Host microbiota constantly control maturation and function of microglia in the CNS. *Nature Neuroscience*. 2015; 18: 965–977.
- [13] Hu J, Cao L, Li T, Dong S, Li P. GAT-LI: a graph attention network based learning and interpreting method for functional brain network classification. *BMC Bioinformatics*. 2021; 22: 379.
- [14] Chen Z, Qing J, Xiang T, Yue WL, Zhou JH. Seeing beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023; 22710–22720.
- [15] Abrol A, Fu Z, Du Y, Calhoun VD. Multimodal Data Fusion of Deep Learning and Dynamic Functional Connectivity Features to Predict Alzheimer's Disease Progression. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*. 2019; 2019: 4409–4413.
- [16] Shao L, Fu C, You Y, Fu D. Classification of ASD based on fMRI data with deep learning. *Cognitive Neurodynamics*. 2021; 15: 961–974.
- [17] Wadhwa T, Mahmud M. Computational Model of Functional Connectivity Distance Predicts Neural Alterations. *IEEE Transactions on Cognitive and Developmental Systems*. 2023.
- [18] Wadhwa T, Kakkar D. Social cognition and functional brain network in autism spectrum disorder: Insights from EEG graph-

theoretic measures. *Biomedical Signal Processing and Control*. 2021; 67: 102556.

- [19] Uddin LQ, Supekar K, Lynch CJ, Khouzam A, Phillips J, Feinstein C, *et al.* Salience network-based classification and prediction of symptom severity in children with autism. *JAMA Psychiatry*. 2013; 70: 869–879.
- [20] Lynch CJ, Uddin LQ, Supekar K, Khouzam A, Phillips J, Menon V. Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. *Biological Psychiatry*. 2013; 74: 212–219.
- [21] Uddin LQ, Supekar K, Menon V. Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in Human Neuroscience*. 2013; 7: 458.
- [22] Supekar K, Musen M, Menon V. Development of large-scale functional brain networks in children. *PLoS Biology*. 2009; 7: e1000157.
- [23] Moon H, Bu S, Cho S. A graph convolution network with sub-graph embedding for mutagenic prediction in aromatic hydrocarbons. *Neurocomputing*. 2023; 530: 60–68.
- [24] Pan C, Yu H, Fei X, Zheng X, Yu R. Temporal-spatial dynamic functional connectivity analysis in schizophrenia classification. *Frontiers in Neuroscience*. 2022; 16: 965937.
- [25] Takagi Y, Nishimoto S. High-resolution image reconstruction with latent diffusion models from human brain activity. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023; 14453–14463.
- [26] Guo X, Dominick KC, Minai AA, Li H, Erickson CA, Lu LJ. Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method. *Frontiers in Neuroscience*. 2017; 11: 460.
- [27] Zhao Y, Ge F, Zhang S, Liu T. 3D Deep Convolutional Neural Network Revealed the Value of Brain Network Overlap in Differentiating Autism Spectrum Disorder from Healthy Controls. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. 2018; 2: 172–180.
- [28] Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP. Clinical applications of the functional connectome. *NeuroImage*. 2013; 80: 527–540.
- [29] Ahmed MR, Zhang Y, Liu Y, Liao H. Single Volume Image Generator and Deep Learning-Based ASD Classification. *IEEE Journal of Biomedical and Health Informatics*. 2020; 24: 3044–3054.
- [30] El-Gazzar A, Quaak M, Cerliani L, Bloem P, van Wingen G, Mani Thomas R. A Hybrid 3DCNN and 3DC-LSTM Based Model for 4D Spatio-Temporal fMRI Data: an ABIDE Autism Classification Study. or 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging. 2019; 147: 95–102.
- [31] 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. *Frontiers in Neuroinformatics*. 2019; 13: 70.
- [32] Almuqhim F, Saeed F. ASD-SAENet: A Sparse Autoencoder, and Deep-Neural Network Model for Detecting Autism Spectrum Disorder (ASD) Using fMRI Data. *Frontiers in Computational Neuroscience*. 2021; 15: 654315.
- [33] Dvornek NC, Ventola P, Duncan JS. COMBINING PHENOTYPIC AND RESTING-STATE FMRI DATA FOR AUTISM CLASSIFICATION WITH RECURRENT NEURAL NETWORKS. *Proceedings. IEEE International Symposium on Biomedical Imaging*. 2018; 2018: 725–728.
- [34] Deng X, Zhang J, Liu R, Liu K. Classifying ASD based on time-series fMRI using spatial-temporal transformer. *Computers in Biology and Medicine*. 2022; 151: 106320.
- [35] Wang C, Xiao Z, Wu J. Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data. *Physica Medica*. 2019; 65: 99–105.
- [36] Aghdam MA, Sharifi A, Pedram MM. Diagnosis of Autism Spectrum Disorders in Young Children Based on Resting-State Functional Magnetic Resonance Imaging Data Using Convolutional Neural Networks. *Journal of Digital Imaging*. 2019; 32: 899–918.
- [37] Ingahlalikar M, Shinde S, Karmarkar A, Rajan A, Rangaprakash D, Deshpande G. Functional Connectivity-Based Prediction of Autism on Site Harmonized ABIDE Dataset. *IEEE Transactions on Bio-medical Engineering*. 2021; 68: 3628–3637.
- [38] Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, *et al.* Distance Metric Learning Using Graph Convolutional Networks: Application to Functional Brain Networks. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. 2017; 147: 469–477.
- [39] Anirudh R, Thiagarajan JJ. Bootstrapping Graph Convolutional Neural Networks for Autism Spectrum Disorder Classification. *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019; 3197–3201.
- [40] Chen L, Huang Y, Liao B, Nie K, Dong S, Hu J. Graph Learning Approaches for Graph with Noise: Application to Disease Prediction in Population Graph. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2020; 2724–2729.
- [41] Park K, Bu S, Cho S. Learning Dynamic Connectivity with Residual-Attention Network for Autism Classification in 4D fMRI Brain Images. *Intelligent Data Engineering and Automated Learning – IDEAL 2021*. 2021; 8: 387–396.
- [42] Bu S, Cho S. Triplet-trained graph transformer with control flow graph for few-shot malware classification. *Information Sciences*. 2023; 649: 119598.
- [43] Li W, Wang S, Liu G. Transformer-based Model for fMRI Data: ABIDE Results. 2022 7Th International Conference on Computer and Communication Systems (ICCCS). 2022; 162–167.
- [44] Kang L, Chen J, Huang J, Jiang J. Autism spectrum disorder recognition based on multi-view ensemble learning with multi-site fMRI. *Cognitive Neurodynamics*. 2023; 17: 345–355.
- [45] Sotero RC, Sanchez-Bornot JM, Iturria-Medina Y. Improving fMRI-based Autism Spectrum Disorder Classification with Random Walks-informed Feature Extraction and Selection. *Proceedings of the 2023 10th International Conference on Bioinformatics Research and Applications*. 2023.
- [46] Gaur M, Chaturvedi K, Vishwakarma DK, Ramasamy S, Prasad M. Self-supervised ensembled learning for autism spectrum classification. *Research in Autism Spectrum Disorders*. 2023; 107: 102223.
- [47] Chen Z, Ji J, Liang Y. Convolutional neural network with an element-wise filter to classify dynamic functional connectivity. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2019; 643–646.