Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 4, October 2026, pp: 1354-1370. e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

A Multimodal Graph Neural Network for Multiclass ADHD and ASD Classification with Leakage-Aware Evaluation

Chofifatul Hidayah[1], Wiharto[1], and Esti Suryani[2]

[1] Department of Informatics, Faculty of Information Technology and Data Science, University of Sebelas Maret, Surakarta, Indonesia
[2] Department of Data Science, Faculty of Information Technology and Data Science, University of Sebelas Maret, Surakarta, Indonesia

Corresponding author: Wiharto (e-mail: wiharto@staff.uns.ac.id), **Authors email:** Chofifatul Hidayah (e-mail: chofifatul23@student.uns.ac.id), Esti Suryani (e-mail: estisuryani@staff.uns.ac.id)

Abstract **Neurodevelopmental disorders such as attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) share overlapping clinical symptoms, complicating diagnosis and motivating objective, data-driven approaches using neuroimaging and machine learning. Graph neural networks (GNNs) have shown strong performance in this domain, yet many existing studies rely on single-modality data, transductive learning, and feature selection procedures that may introduce information leakage and inflate reported accuracy. This study proposes a multimodal graph learning framework that integrates resting-state fMRI (rs-fMRI), structural MRI (sMRI), and demographic data for classifying ADHD, ASD, and healthy controls (HC). The framework integrates temporal stability-based functional connectivity, hybrid feature selection, and adaptive multi-graph learning to exploit complementary information across these modalities. Using the ADHD-200 and ABIDE datasets, the framework is evaluated under three protocols that progressively tighten control over information leakage: transductive learning with global feature selection, inductive learning with global feature selection, and inductive learning with fold-wise feature selection. Results show that classification performance is highest under the transductive, globally-selected setting (85.5% accuracy for HC vs ADHD vs ASD, 92.5% for HC vs ASD, and 90.4% for HC vs ADHD), but decreases under the strictest leakage-aware protocol (70.9%, 81.7%, and 79.3%, respectively). This performance gap indicates that conventional evaluation protocols can substantially overestimate real-world generalization. Importantly, the proposed framework still achieves reasonable accuracy under the strictest setting, suggesting genuine discriminative capability beyond evaluation artifacts. These findings emphasize that leakage-aware evaluation, although yielding lower numbers, provides a more realistic and trustworthy estimate of model performance, highlighting its importance for developing reliable neuroimaging-based GNN models.**

Keywords **Multimodal learning; Neuroimaging; ChebGCN; Information leakage.**

I. Introduction

Neurodevelopmental Disorders (NDDs) are a group of neurological disorders that affect brain development and function, including attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) [1], [2]. These disorders generally emerge during childhood and significantly impact cognitive abilities, behavior, motor skills, and social interactions [3], [4]. The overlap of symptoms between ADHD and ASD makes the diagnostic process particularly challenging and often leads to misclassification, highlighting the need for more objective diagnostic approaches [5], [6]. Advances in neuroimaging technology have enabled the exploration of brain biomarkers associated with ADHD and ASD. One of the most widely used modalities is resting-state functional Magnetic Resonance Imaging (rs-fMRI), which represents brain activity through Blood Oxygen Level-Dependent (BOLD) signals and is commonly used to construct functional connectivity networks [7]. Meanwhile, Structural MRI (sMRI) provides anatomical information about the brain, such as regional volumes and cortical morphology [8]. In addition to neuroimaging data, demographic information such as age and sex has also been reported to improve classification performance when integrated with neuroimaging features [7]. Nevertheless, several limitations remain in neuroimaging-based classification studies. Many approaches still rely on single-modality data, limiting their ability to capture complementary functional,

Manuscript received 8 April 2026; Revised 10 June 2026; Accepted 25 July 2026; Available online 11 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1801

structural, and phenotypic information simultaneously [9], [10]. In addition, most studies focus on static functional connectivity despite the inherently dynamic nature of brain activity, which may provide more discriminative biomarkers [11]. Furthermore, suboptimal feature selection procedures may introduce redundant and noisy features that reduce classification performance [12], [13].

Several Graph Neural Network (GNN)-based methods have been proposed for the classification of neurodevelopmental disorders using neuroimaging data. Shao et al. [14] proposed the Heterogeneous Graph Convolutional Attention Network (HCAN), which constructs a heterogeneous population graph from fMRI and phenotypic data for ASD classification, arguing that the homogeneous graphs used in previous studies were unable to fully exploit phenotypic information. Another relevant approach is IFC-GNN, which explores interaction functional connectivity based on rs-fMRI and phenotypic data to distinguish ASD from Healthy Controls (HC), achieving an accuracy of 80.66% [7]. Meanwhile, DAGMNet proposes a dual-branch attention-pruned GNN that fuses multimodal sMRI and fMRI features, achieving 91.59% accuracy [15]. In addition, B2P-GL utilizes a population graph learning strategy for ADHD classification and reports competitive performance in a transductive graph setting [16]. Although these methods demonstrate promising performance, they share several limitations: HCAN and DAGMNet operate exclusively in transductive settings, B2P-GL constructs population graphs globally before the train-test split, and IFC-GNN applies global feature selection without discussing potential leakage implications all of which may contribute to optimistic performance estimates.

More broadly, in transductive settings, testing nodes may still be exposed during graph training, potentially leading to optimistic performance estimation [17]. Many neuroimaging-based GNN studies also do not explicitly distinguish between transductive and inductive evaluation, nor systematically assess how graph construction and feature selection strategy affect generalization [18], [19], which complicates fair comparison across studies and may overstate real-world performance.

To the best of the authors' knowledge, no prior study on multiclass ADHD-ASD-HC classification has jointly integrated temporal stability-based dynamic functional connectivity, multimodal rs-fMRI and sMRI features, phenotypic data, and adaptive graph fusion within a single framework that also systematically evaluates transductive versus inductive settings and global versus fold-wise feature selection. While the proposed framework builds on established components temporal stability features, DFS, mRMR, and ChebGCN its novelty lies in integrating them into a leakage-aware multimodal framework and in the systematic investigation of how the evaluation protocol affects model generalization.

To address these challenges, this study proposes an Adaptive Multi-Modal Graph Learning Framework based on Chebyshev Graph Convolutional Networks (ChebGCN). The framework is evaluated on ADHD-200 [20] and ABIDE datasets [21], which provide rs-fMRI, sMRI, and phenotypic data. Dynamic functional connectivity (dFC) features are extracted from rs-fMRI using a sliding-window approach [22], [23], followed by temporal-stability feature extraction adapted from the Generated Interaction Functional Connectivity framework [7]. Subsequently, Deep Feature Selection (DFS) is applied to select rs-fMRI features, while regional volume features from sMRI are optimized using Minimum Redundancy Maximum Relevance (mRMR) [8], [24]. All features are then integrated with demographic data to construct a multimodal inter-subject similarity graph. The constructed graphs are processed using a multigraph ChebGCN architecture [7], [25] equipped with an adaptive fusion mechanism to learn the relative contribution of each modality, trained under both transductive and inductive settings across multiclass (ADHD-ASD-HC) and binary classification tasks. The main contributions of this study are summarized as follows:

1. Systematically evaluating transductive and inductive settings alongside global and per-fold feature selection strategies to investigate the impact of information leakage on model generalization.
2. Adapting an existing temporal stability-based dynamic functional connectivity representation by employing a simplified edge-wise exponential kernel to improve computational efficiency.
3. Employing a hybrid modality-specific feature selection strategy that combines DFS for rs-fMRI and mRMR for sMRI.
4. Integrating rs-fMRI, sMRI, and phenotypic information within a leakage-aware multimodal Multigraph ChebGCN framework equipped with an adaptive fusion mechanism for unified ADHD-ASD-HC classification.

II. Method

This study proposes an adaptive multi-graph learning framework that integrates multimodal neuroimaging and demographic data for ASD and ADHD classification, as illustrated in Fig. 1. rs-fMRI data are processed to extract temporal stability-based dynamic functional connectivity features, while sMRI data are represented as ROI-based volumetric features and are optimized using mRMR feature selection prior to graph construction. Demographic attributes, including age

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

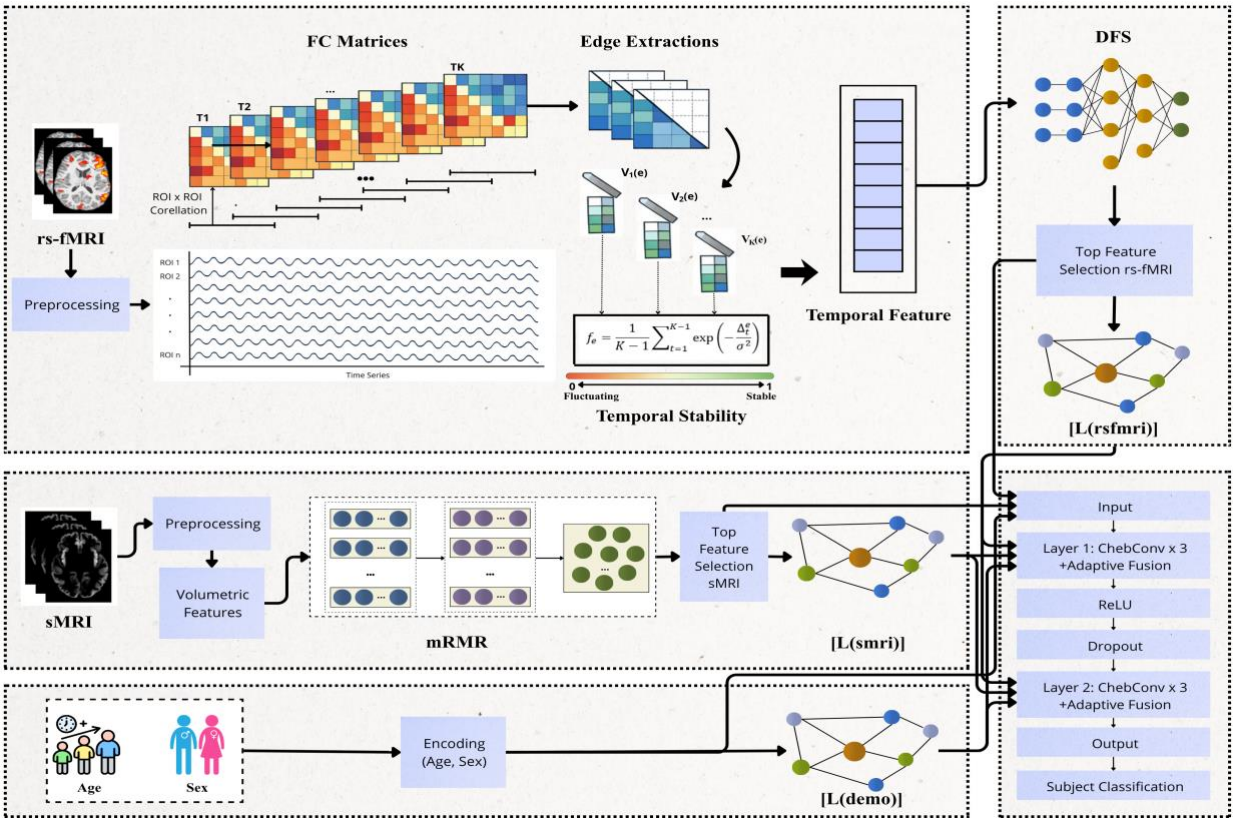

Fig. 1 Overview of the Proposed Multimodal Brain Graph Learning Framework with Adaptive Fusion.

and sex, are also encoded to construct complementary subject-similarity graphs. These heterogeneous graphs are jointly learned using ChebGCN, with an adaptive fusion mechanism that integrates shared and complementary multimodal information. The fused representations are then used for multi-class classification of HC, ADHD, and ASD subjects.

A. Dataset

This study utilizes publicly available neuroimaging datasets, namely ADHD-200 and ABIDE [21], which are part of the International Neuroimaging Data-sharing Initiative (INDI) [26]. These datasets provide multimodal data including rs-fMRI, sMRI, and phenotypic information. All neuroimaging data used in this study were obtained in preprocessed form from the Preprocessed Connectomes Project (PCP), where standardized preprocessing pipelines had been applied prior to public release. Specifically, rs-fMRI data were obtained from the fMRIPrep derivatives, which include slice-timing correction, head-motion correction, skull stripping, BOLD-to-T1w co-registration, spatial normalization to MNI space, brain masking, and susceptibility distortion correction when applicable [27]. Structural MRI (sMRI) data were processed using FreeSurfer [28] to extract cortical and subcortical morphometric features. The preprocessed rs-fMRI images were subsequently parcellated into 116 regions of interest (ROIs) using the Automated Anatomical Labeling (AAL) atlas [29]. ROI-wise BOLD time series were extracted using Nilearn's NiftiLabelsMasker, with z-score standardization (standardize=True) applied prior to functional connectivity estimation.

Subject selection was performed based on the availability of complete multimodal data and to obtain a balanced class distribution. In the multi-class classification scenario, the dataset consists of 188 ADHD subjects from ADHD-200, 188 ASD subjects from ABIDE, and 188 healthy control (HC) subjects, with 94 HC subjects from each dataset. In total, this study includes 564 subjects. In the binary classification scenario, HC subjects were selected from the same dataset as the corresponding disorder group to minimize potential inter-dataset bias. Specifically, the ADHD vs HC classification utilized 188 ADHD subjects and 188 HC subjects entirely from ADHD-200. Meanwhile, the ASD vs HC classification utilized 188 ASD subjects and 188 HC subjects entirely from ABIDE. This strategy was applied to maintain consistency in data distribution, acquisition characteristics, and preprocessing protocols across classification scenarios. Subjects were included only if complete rs-fMRI, sMRI, and phenotypic data were available. Subjects with missing imaging modalities or incomplete demographic information were excluded, and no data imputation was performed. Since

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

preprocessed ADHD-200 and ABIDE datasets were used, no additional motion-based exclusion, nuisance regression, or quality-control procedures were applied. The selected subjects were acquired from multiple sites within the ADHD-200 and ABIDE datasets. The demographic characteristics of the ADHD-200 dataset are summarized in Table 1, while those of the ABIDE dataset are presented in Table 2.

B. Functional Data Processing

Data from rs-fMRI are used to characterize functional connectivity (FC) between brain regions based on temporal correlations of BOLD signals. While conventional FC assumes stationarity, brain activity is inherently dynamic, motivating the use of dynamic functional connectivity (dFC) to capture time-varying patterns [11], [22]. To model these dynamics, a sliding-window approach is applied [30], in which the BOLD time series is segmented into overlapping windows, as illustrated in Fig. 1, which provides an overview of the proposed functional processing pipeline, including temporal stability feature extraction. For each window, an FC matrix is computed using the Pearson correlation coefficient, as defined in Eq. (1) [30].

Table 1. Demographic information of ADHD-200

Data Source	Sample type		Age range	Sex	
	ADHD	HC		M	F
KKI	22	34	8.0 – 13.0	33	23
Neuro-IMAGE	15	15	11.1 – 20.9	19	11
NYU	107	43	7.2 – 18.0	103	47
OHSU	13	5	7.3 – 11.3	10	8
Peking	31	41	8.4 – 15.8	48	24
Pittsburg	0	39	10.1 – 18.9	22	17
Washu	0	11	7.2 – 9.9	4	7

Table 2. Demographic information of ABIDE

Data Source	Sample type		Age range	Sex	
	ASD	HC		M	F
SDSU	8	20	8.7 – 17.2	22	6
Trinity	19	14	12.0 – 25.7	33	0
UM	15	74	8.2 – 28.8	70	19
Yale	22	28	7.0 – 17.8	34	16
Leuven	26	1	12.1 – 32.0	24	3
Pitt	23	25	9.3 – 35.2	42	6
Olin	14	14	10.0 – 24.0	23	5
OHSU	12	12	8.0 – 15.2	24	0
KKI	12	0	8.2 – 12.5	9	3
NYU	37	0	7.1 – 38.8	27	10

$$r_{ij} = \frac{\sum_{t=1}^{T}(x_i(t)-\bar{x}_i)(x_j(t)-\bar{x}_j)}{\sqrt{\sum_{t=1}^{T}(x_i(t)-\bar{x}_i)^2}\sqrt{\sum_{t=1}^{T}\left(x_j(t)-\bar{x}_j\right)^2}} \quad (1)$$

where r_{ij} denotes the correlation between the i-th and j-th ROIs, $x_i(t)$ and $x_j(t)$ represents the BOLD signals of the i-th and j-th ROIs at time t, respectively, $\bar{x}_i$ and $\bar{x}_j$ denote their mean signals, and T is the number of time points within a window. This process is repeated across all windows, resulting in a sequence of FC matrices that capture the temporal evolution of brain connectivity, forming a tensor of size $P \times R \times R$, where P is the total number of sliding windows generated for a subject and R is the number of ROIs.

A window length of 30 time points with a step size of 15 is used, resulting in 50% overlap [31]. Since FC matrices are symmetric, only the upper triangular elements are retained. For $R = 116$ ROIs, this produces $E = R(R-1)/2 = 6670$ unique FC per window. Each FC matrix is then vectorized by retaining these upper triangular elements, yielding an edge-wise feature vector $v_t \in R^E$ for the t-th window. Consequently, the dynamic FC of each subject is represented as a temporal sequence $\{v_t\}_{t=1}^{P}$, where $v_t(e)$ denotes the connectivity value of the e-th edge at the t-th sliding window and $e \in \{1, \ldots, E\}$.

C. Temporal Stability Feature Extraction

The temporal feature extraction stage in this study is inspired by the interaction functional connectivity framework proposed by [7], which models relationships between functional connectomes across different time points using an exponential kernel function. To improve computational efficiency and scalability, this study adopts a simplified formulation by applying the exponential kernel in a diagonal manner, comparing the same edge across consecutive time windows. Unlike the original GIFC framework, which models pairwise interactions between all connections and results in high computational complexity, proposed approach focuses on edge-wise temporal consistency, significantly reducing the feature dimensionality and computational cost. The temporal stability of each edge is computed based on the temporal differences between consecutive windows. The temporal difference is first defined as $\Delta_t^e = | v_t(e) - s \cdot v_{t+1}(e) |$ and the temporal stability is then calculated using an exponential kernel as shown in Eq. (2) [7].

$$f_e = \frac{1}{P-1}\sum_{t=1}^{P-1} \exp\left(-\frac{\Delta_t^e}{\sigma^2}\right) \quad (2)$$

where where f_e denotes the temporal stability of edge e, s is a scaling factor set to 1, and σ is determined in

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

a data-driven manner as the standard deviation of edge-wise temporal differences $| v_t(e) - v_{t+1}(e) |$ across all edges and consecutive windows for each subject individually. The normalization factor 1/(P − 1) ensures that the resulting feature is independent of the number of windows. This formulation provides a measure of temporal stability for each functional connection across the sequence of windows. Values closer to 1 indicate more stable connectivity patterns, while lower values reflect higher temporal variability. The final output is a one-dimensional feature vector of size $\mathbf{x}^{rsfmri} \in \mathbb{R}^{6670}$ for each subject, representing the temporal stability profile of all functional connections. These features are used as input for subsequent multimodal graph learning.

D. Structural Data Processing

Structural MRI (sMRI) data are used to capture anatomical characteristics of the brain that complement functional connectivity information. In this study, structural features are derived from preprocessed FreeSurfer [28] outputs, specifically the automated subcortical segmentation, which provides volumetric measurements of brain structures. For each subject, 45 regional volumetric features are extracted, forming a feature vector $\mathbf{x}^{sMRI} \in \mathbb{R}^{45}$. These features represent subcortical regions associated with neurodevelopmental disorders. Unlike the functional modality, which is based on the AAL atlas, the structural features are derived from FreeSurfer segmentation, resulting in heterogeneous feature representations across modalities. However, since both modalities are represented as subject-level feature vectors and integrated through a graph-based fusion mechanism, the difference in atlas space does not introduce misalignment during learning process. Finally, the features are normalized and refined using the mRMR method prior to multimodal integration.

E. Feature Selection

1. DFS

DFS is employed in this study to identify the most informative features from high-dimensional functional connectivity representations derived from rs-fMRI data. This approach is adopted following the framework proposed by [7], which integrates feature selection directly into a deep neural network architecture through an embedded learning mechanism. Unlike conventional feature selection methods, DFS incorporates a weighted input layer between the input layer and the first hidden layer of a Multilayer Perceptron (MLP). This layer establishes a one-to-one correspondence between each input feature and a learnable weight parameter. Given an input feature vector $\mathrm{x} = [x_1, x_2, \ldots, x_n]$and corresponding weight vector $\mathrm{w} = [w_1, w_2, \ldots, w_n]$ associated with the input (feature selection) layer, the transformed input is defined as xi′=xi·wi. where n denotes the number of input features, and w_i represents the importance weight associated with the i-th feature. This formulation enables the model to selectively emphasize relevant features while suppressing irrelevant ones during training. To promote sparsity and improve feature selection capability, an Elastic Net regularization is applied to the weight vector w, combining both L1 and L2 norms. The regularization term for the feature selection layer is defined as Eq. (3) [7].

$$l_{ri}(\mathrm{w}) = \mu_1 \left(\frac{1-\mu_2}{2} \parallel \mathrm{w} \parallel_2^2 + \mu_2 \parallel \mathrm{w} \parallel_1 \right) \quad (3)$$

where μ_1 controls the overall regularization strength and $\mu_2 \in [0,1]$ balances the contribution between L1 and L2 penalties. The L1 norm encourages sparsity by driving irrelevant feature weights toward zero, while the L2 norm stabilizes the learning process. In addition, a similar regularization is imposed on the weights of the hidden layers to reduce overfitting, as defined in Eq. (4) [7].

$$l_{rh}(\mathrm{W}) = \delta_1 \left(\frac{1-\delta_2}{2} \sum_{k=1}^{L-1} \left\| W^{(k)} \right\|_2^2 + \delta_2 \sum_{k=1}^{L-1} \left\| W^{(k)} \right\|_1 \right) \quad (4)$$

where $\mathrm{W}^{(k)}$denotes the weight matrix of the k-th layer, and δ_1, δ_2 are hyperparameters controlling the strength and balance of regularization. The overall loss function optimized during training is formulated as Eq. (5) [7].

$$l(\mathrm{W}) = l_{ce}(\mathrm{W}) + l_{ri}(\mathrm{w}) + l_{rh}(\mathrm{W}) \quad (5)$$

where l_{ce} represents the cross-entropy loss used for classification. The model parameters, including the feature weights w, are optimized using the Adam optimizer. After training, the learned weights $| w_i |$ are interpreted as feature importance scores. An elbow-point analysis of the sorted weight distribution identified 762 candidate features, which defined the search region for model selection. Candidate feature numbers (400, 500, 600, 700, and 800) were subsequently evaluated using the validation set, and the feature number yielding the highest validation accuracy was selected for each classification task. This end-to-end mechanism enables DFS to simultaneously perform feature selection and nonlinear representation learning while reducing redundant and less informative functional connectivity features.

2. mRMR

mRMR is employed to select informative structural features from sMRI data by balancing feature relevance to class labels with feature redundancy [24]. This is particularly important for sMRI, where regional brain volumes may exhibit anatomical correlations, introducing redundancy. The mRMR objective aims to select a feature subset $S \subset F$ that maximizes the trade-off between relevance and redundancy, as defined in Eq. (6) [24].

$$\max_{S \subset F} [D(S, y) - \mathcal{R}(S)] \quad (6)$$

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

where the relevance and redundancy terms are defined in Eq. (7) and Eq. (8) [24].

$$D(S,y) = \frac{1}{|S|}\sum_{x_i \in S} I(x_i;y) \quad (7)$$

$$\mathcal{R}(S) = \frac{1}{|S|^2}\sum_{x_i,x_j \in S} I(x_i;x_j) \quad (8)$$

where S denotes the selected feature subset from the original feature set F, x_i represents the i-th feature, and y denotes the class label. The function $I(a;b)$ indicates the mutual information between variables a and b, while $|S|$ represents the number of selected features. The term $\mathcal{R}(S)$ denotes the redundancy among features within the selected subset. In practice, mRMR employs a greedy incremental search strategy to iteratively select features by maximizing the trade-off between relevance and redundancy. The ranked structural features were then evaluated with candidate feature counts of 10, 20, 30, and 40, and the feature count that achieved the highest validation accuracy was selected for each classification task.

F. Multimodal Graph Construction

A population-based graph representation is constructed to model relationships between subjects using multimodal data, including rs-fMRI, sMRI, and demographic information. In this formulation, each node represents a subject, while edges encode similarity between subjects based on their feature representations. For each modality, a separate graph is constructed. Let $\mathrm{x}_i^{(m)}$ denote the feature vector of the i-th subject for modality m, where $m \in$ {rs-fMRI, sMRI}. The similarity between two subjects $i$and j is computed using a Gaussian similarity function, as defined in Eq. (9) [32], [33]

$$A_{ij}^{(m)} = \begin{cases} \exp\left(-\dfrac{\| \boldsymbol{x}_i^{(m)} - \boldsymbol{x}_j^{(m)} \|^2}{2\sigma_m^2}\right), & j \in \mathrm{kNN}(i) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $A_{ij}^{(m)}$denotes the edge weight between subjects i and j, $\mathrm{kNN}(i)$ denotes the set of k-nearest neighbors of node i, and σ_m controls the similarity scale for modality m. The parameter σ_m is determined in a data-driven manner as the median of pairwise distances between neighboring nodes, as defined in Eq. (10)

$$\sigma_m = \mathrm{median}\left(\{\| \mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)} \| \mid j \in \mathrm{kNN}(i)\}\right) \quad (10)$$

This strategy adapts the similarity scale to the intrinsic distribution of each modality and improves robustness to feature variability. The kNN strategy ensures that the resulting graph is sparse by retaining only the most similar neighbors, thereby reducing noise and preserving meaningful local structures for graph-based learning [32]. For demographic data, similarity is constructed based on attribute matching rather than Gaussian similarity. For each pair of subjects, similarity is computed by counting the number of demographic attributes they share in common, including age and gender. For age, values are rounded to the nearest integer prior to comparison. The aggregated matching scores are then normalized to the range [0,1] to ensure comparable scaling with other modalities. Unlike the rs-fMRI and sMRI graphs, the demographic graph does not apply kNN sparsification, as subjects with few shared attributes naturally receive low similarity scores, resulting in a sparse connectivity structure.

Although modality-specific graphs are constructed independently, the node features are integrated into a unified representation. Specifically, for each subject i, the feature vectors from rs-fMRI, sMRI, and demographic data are concatenated into a single feature vector $\mathrm{X}_i = [x_i^{rsfmri}, x_i^{smri}, x_i^{demo}]$. This unified representation is used as input to the graph neural network, while modality-specific relationships are preserved through separate graph Laplacians. Each modality produces an adjacency matrix $\mathrm{A}^{(m)} \in \mathbb{R}^{N\times N}$, where N denotes the number of subjects. The corresponding degree matrix $\mathrm{D}^{(m)}$ is defined as Eq. (11) [7].

$$D_{ii}^{(m)} = \sum_{j=1}^{N} A_{ij}^{(m)} \quad (11)$$

The normalized graph Laplacian for each modality is then computed as Eq. (12) [7].

$$\mathrm{L}^{(m)} = \mathrm{I} - (\mathrm{D}^{(m)})^{-1/2}\mathrm{A}^{(m)}(\mathrm{D}^{(m)})^{-1/2} \quad (12)$$

where I is the identity matrix. This formulation enables the construction of modality-specific graphs that capture complementary relationships across different data sources. By maintaining separate graph structures for each modality, the framework preserves distinct characteristics of functional, structural, and phenotypic information prior to integration in the subsequent graph learning stage [34], [35].

G. Adaptive Multi-Graph ChebGCN with Layer-wise Fusion Model

ChebGCN is selected due to its ability to capture multi-hop neighborhood information efficiently through polynomial approximation, making it suitable for sparse population graphs. To learn discriminative representations from multimodal graphs, a Multi-Graph ChebGCN is employed. This model operates on multiple modality-specific graphs constructed in the previous stage, enabling efficient learning of both local and global relationships between subjects [25]. Given a graph Laplacian $\mathrm{L}^{(m)}$for modality m, the Chebyshev graph convolution is defined using a truncated expansion of Chebyshev polynomials, as defined in Eq. (13) [25].

$$g_\theta^{(m)} * x = \sum_{k=0}^{K-1} \theta_k^{(m)} T_k(\tilde{\mathrm{L}}^{(m)})\mathrm{x} \quad (13)$$

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

where $g_\theta^{(m)} * x$ denotes the graph convolution output, x is the input node feature vector, $\theta_k^{(m)}$ are learnable parameters, K is the order of the polynomial, and $T_k(\cdot)$ denotes the Chebyshev polynomial of order k. The scaled Laplacian $\tilde{\mathrm{L}}^{(m)}$ is defined as Eq. (14) [25].

$$\tilde{\mathrm{L}}^{(m)} = \frac{2}{\lambda_{\max}} \mathrm{L}^{(m)} - \mathrm{I} \quad (14)$$

with $\lambda_{\max}$ being the largest eigenvalue of $\mathrm{L}^{(m)}$. The Chebyshev polynomials are computed recursively as $T_0(\mathrm{x}) = \mathrm{I}, T_1(\mathrm{x}) = \mathrm{x}, T_k(\mathrm{x}) = 2\mathrm{x}T_{k-1}(\mathrm{x}) - T_{k-2}(\mathrm{x})$. Using this formulation, the propagation rule of the ChebGCN layer for modality m is given by Eq. (15) [25].

$$\mathrm{H}^{(l+1,m)} = \sigma\left(\sum_{k=0}^{K-1} T_k\left(\tilde{\mathrm{L}}^{(m)}\right)\mathrm{H}^{(l,m)}\Theta_k^{(l,m)}\right) \quad (15)$$

where $\mathrm{H}^{(l,m)}$ denotes the node feature representation at layer l for modality m, $\Theta_k^{(l,m)}$ are learnable weight matrices, and $\sigma(\cdot)$ is a nonlinear activation function such as ReLU. The initial input is defined as $\mathrm{H}^{(0,m)} = \mathrm{X}$, where X denotes the input feature matrix.

To integrate information from multiple modalities, the learned representations from each modality-specific graph are combined using an adaptive fusion mechanism, as defined in Eq. (16)

$$\mathrm{H}^{(l)} = \sum_{m=1}^{M} \alpha_m \mathrm{H}^{(l,m)} \quad (16)$$

where α_m denotes the learnable weight that controls the contribution of modality m, and M is the total number of modalities. This mechanism allows the model to automatically emphasize more informative modalities while reducing the influence of less relevant ones. The model is trained under a transductive learning setting, where the full graph structure is available during training, while labels are provided only for a subset of nodes.

H. Evaluation Strategy

Model performance is evaluated using a stratified K-fold cross-validation framework to ensure robust and unbiased estimation. In each fold, one subset is held out for testing while the remaining data are used for training and validation. A stratified split is maintained to preserve class distribution across all folds, with the test set kept completely unseen during training and model selection. All features are normalized using z-score standardization computed exclusively on the training data of each fold, and the resulting parameters are subsequently applied to the test data to prevent data leakage [36], [37]. The proposed model is optimized using the Adam optimizer with the categorical cross-entropy loss. The output logits are converted into posterior class probabilities using the softmax function in Eq. (17) [38]

$$\hat{p}_{i,r} = \frac{\exp(o_{i,r})}{\sum_{q=1}^{M} \exp(o_{i,q})} \quad (17)$$

where $o_{i,r}$ denotes the output logit of subject i for class r, M is the total number of classes, and $\hat{p}_{i,r}$ is the predicted probability that subject $i$belongs to class r. The network parameters are optimized by minimizing the categorical cross-entropy loss in Eq. (18) [38]

$$J = -\frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \sum_{r=1}^{M} t_{i,r} \log \hat{p}_{i,r} \quad (18)$$

where N_{tr} is the number of training subjects and $t_{i,r}$ is the one-hot encoded ground-truth label. Model performance is assessed using Accuracy, Macro F1-score, and AUC. The Macro F1-score is computed as Eq. (19) [39]

$$F_{macro} = -\frac{1}{M} \sum_{r=1}^{M} F_r \quad (19)$$

where F_r denotes the F1-score of class r, and M represents the number of classes. Finally, the average performance over all cross-validation folds is reported together with the 95% confidence interval in Eq. (20) [40]

$$CI_{95\%} = \bar{m} \pm t_{0.975,K_f-1} \frac{s_m}{\sqrt{K_f}} \quad (20)$$

where $\bar{m}$ is the mean performance metric across the K_f cross-validation folds, s_m is the corresponding standard deviation, and $t_{0.975,K_f-1}$ is the critical value of the Student's *t*-distribution with $K_f - 1$ degrees of freedom. It is important to note that the graph construction strategy and feature selection procedure vary across experimental settings, as described in the Experiment Setup section. These variations are intentionally designed to systematically analyze the impact of evaluation protocols on model generalization capability and potential information leakage.

III. Result

A. Experiment Setup and Implementation Details

This study evaluates ADHD, ASD, and HC classification under three experimental settings to analyze the effects of graph learning protocols and feature selection strategies on model generalization. The experiments are conducted for both multiclass (HC vs ADHD vs ASD) and binary classification tasks. Setting 1 (Transductive Learning with Global Feature Selection). Feature selection is performed globally using all subjects before cross-validation, and graphs are constructed using the complete dataset prior to the train-test split. Consequently, both training and test subjects are present in the graph during model training, allowing test subjects to indirectly influence the graph topology and feature space. This setting may therefore lead to optimistic performance estimates due to information leakage.

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

Algorithm 1. Leakage-aware training and evaluation procedure (Setting 3) of the proposed Multi-Graph ChebGCN

Input: Preprocessed subject dataset $\mathcal{D}=\{(X_i^{rsfmri}, X_i^{smri}, X_i^{demo})\}_{\{i=1\}^N}$
Number of folds K, candidate feature sizes $\mathcal{C}_{\text{rsfmri}}, \mathcal{C}_{\text{smri}}$, Validation split ratio p
Output: Aggregated performance metrics (Accuracy, Macro-F1, AUC) with 95% CI

(1) Generate stratified K-fold partition $\{(\mathcal{T}_k^{train}, \mathcal{T}_k^{test})\}_{\{k=1\}^K}$ such that $\mathcal{T}_k^{train} \cap \mathcal{T}_k^{test} = \emptyset$ and class proportions are preserved
(2) for k = 1 to K do
(3) Split $\mathcal{T}_k^{train}$ into $\mathcal{T}_k^{fit(training)}$ and $\mathcal{T}_k^{val(validation)}$, ratio (1-p):p
(4) Compute μ, σ from $\{X_i\}_{\{i\in\mathcal{T}_k^{fit}\}}$ for each modality m ∈ {rsfmri, smri}
(5) Apply z-score standardization to $\mathcal{T}_k^{fit}, \mathcal{T}_k^{val}$, and $\mathcal{T}_k^{test}$ using μ, σ
(6) for each candidate size c ∈ $\mathcal{C}_{\text{rsfmri}}$ do
(7) Fit DFS on $\{X_i^{rsfmri}\}_{\{i\in\mathcal{T}_k^{fit}\}}$ only → obtain weight vector w, select top-c features
(8) Evaluate resulting feature subset on $\mathcal{T}_k^{val}$, record validation accuracy
(9) End for
(10) Select c^*_{rsfmri} = argmax validation accuracy over $\mathcal{C}_{\text{rsfmri}}$
(11) for each candidate size c ∈ $\mathcal{C}_{\text{smri}}$ do
(12) Fit mRMR (MIQ) on $\{X_i^{smri}\}_{\{i\in\mathcal{T}_k^{fit}\}}$ only → select top-c features
(13) Evaluate resulting feature subset on $\mathcal{T}_k^{val}$, record validation accuracy
(14) end for
(15) Select c^*_{smri} = argmax validation accuracy over $\mathcal{C}_{\text{smri}}$
(16) Retain only the selected c^*_{rsfmri} and c^*_{smri} features for all subjects in $\mathcal{T}_k^{fit}, \mathcal{T}_k^{val}$, and $\mathcal{T}_k^{test}$ (no refitting on val/test data)
(17) Construct modality-specific graphs $\text{A}_{\text{train}}(m)$, $\text{A}_{val}(m)$, $\text{A}_{test}(m)$ independently for each modality.
(18) Compute normalized Laplacians $L_{\text{train}}(m)$, $L_{val}(m)$, $L_{test}(m)$
(19) Initialize the Multi-Graph ChebGCN using the predefined hyperparameter configuration summarized in Table 3
(20) Initialize parameters Θ and adaptive fusion

Setting 2 (Inductive Learning with Global Feature Selection). Feature selection is still performed globally before cross-validation; however, graph construction is performed independently within each fold using only training subjects. Test subjects are completely excluded from graph construction, and no edges are created between test and training subjects. During inference, the trained model is applied independently to each unseen test subject using the parameters learned from the training graph, without updating the graph topology or model parameters. This setting prevents information leakage arising from graph construction, although potential leakage from global feature selection remains.

Setting 3 (Inductive Learning with Fold-wise Stable Feature Selection). Feature selection is performed exclusively on the training subjects within each fold, and graphs are constructed using only the selected training features. Test subjects are excluded from both feature selection and graph construction, and no edges are established between test and training subjects. During inference, predictions are generated independently for each unseen test subject using the trained model, ensuring that no information from test subjects is exposed during feature selection, graph construction, or model optimization. Consequently, this setting represents a strictly leakage-aware evaluation protocol and provides a more realistic estimate of model generalization to unseen subjects. The complete leakage-aware training and evaluation procedure adopted in Setting 3 is summarized in Algorithm 1. The algorithm describes subject partitioning, train-only normalization, fold-wise feature selection, graph construction, model training, validation with early stopping, and final testing. All preprocessing operations that may introduce information leakage are performed exclusively on the training subset, while the validation subset is used for model selection and the test subset is reserved solely for final performance evaluation.

All experiments were conducted using Python in the Google Colab environment with GPU acceleration support. The proposed framework was evaluated using stratified 10-fold cross-validation to obtain a more robust and unbiased performance estimation. For functional data processing, a sliding window approach was applied with a window size of 30 and a step size of 15. Temporal stability features were computed using an exponential kernel with a scaling factor of σ = 1. Feature selection was performed using a hybrid strategy, where DFS was applied to rs-fMRI features, while mRMR with the Mutual Information Quotient (MIQ) criterion was used for sMRI features. Graphs were constructed using a k-nearest neighbor (kNN) approach followed by Gaussian similarity weighting. The main experimental settings and hyperparameters are summarized in Table 3.

B. Results Under Different Evaluation Settings

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

weights α.
(21) for epoch = 1 to E do
(22) Forward propagation on $\mathcal{T}_k^{fit}$ using $L_{\text{train}}(m)$
(23) Compute cross-entropy loss
(24) Update Θ and α using Adam
(25) Evaluate on $\mathcal{T}_k^{val}$ using $L_{val}(m)$ for early stopping
(26) End for
(27) Apply the best saved model to $\mathcal{T}_k^{test}$ using $L_{test}(m)$
(28) Predict class labels for all test subjects without updating model parameters
(29) Compute fold-level Accuracy, Macro-F1, and AUC
(30) End for
(31) Compute mean ± SD and 95% CI.
(32) Return aggregated performance metrics.

Table 3. Experimental Settings and Reproducibility Details

Parameter	Value
Number of Feature rs-fMRI	6670
Number of Feature sMRI	45
Candidate Feature Selection rs-fMRI	[400, 500, 600, 700, 800]
Candidate Feature Selection sMRI	[10, 20, 30, 40]
Number of hidden layers	2
Hidden dimension	64
Chebyshev polynomial order	2
Dropout rate	0.3
Learning rate	0.001
Optimizer	Adam
Weight decay	0.0001
Batch handling	Full-batch
Validation split	0.15
Epochs	100
Random seed	42
Hardware	Google Colab + NVIDIA Tesla T4

Tables 4–6 summarize classification performance under three evaluation settings for multiclass (HC vs ADHD vs ASD) and binary (HC vs ASD and HC vs ADHD) tasks using different modality combinations. Overall, stricter evaluation protocols consistently reduced classification performance across all tasks and modalities. Setting 1 consistently achieved the highest performance across all tasks and modality combinations. As shown in Table 4, the multimodal configuration achieved the best accuracies of 85.5% for HC vs ADHD vs ASD, 92.5% for HC vs ASD, and 90.4% for HC vs ADHD. However, the transductive evaluation strategy combined with global feature selection may have contributed to optimistic performance estimates due to potential information leakage. As presented in Table 5, classification performance decreased under Setting 2 across most tasks and modality combinations, reflecting the effect of inductive evaluation on model generalization. Although global feature selection was still applied, the multimodal configuration remained the best-performing approach, achieving accuracies of 80.7% for HC vs ADHD vs ASD, 85.9% for HC vs ASD, and 86.7% for HC vs ADHD. Table 6 further demonstrates the impact of the strict leakage-aware evaluation adopted in Setting 3. When both inductive evaluation and fold-wise feature selection were applied, the multimodal configuration still achieved the highest performance among all modality combinations, with accuracies of 70.9% for HC vs ADHD vs ASD, 81.7% for HC vs ASD, and 79.3% for HC vs ADHD, although these values were lower than those obtained under Settings 1 and 2. Quantifying this effect for the multimodal configuration, accuracy declined from Setting 1 to Setting 3 by 14.6 percentage points (17.1% relative) for HC vs ADHD vs ASD (85.5%→70.9%), by 10.8 percentage points (11.7% relative) for HC vs ASD (92.5%→81.7%), and by 11.1 percentage points (12.3% relative) for HC vs ADHD (90.4%→79.3%).

The optimal numbers of selected features varied across evaluation settings and classification tasks. For Setting 1, the best-performing rs-fMRI/sMRI feature combinations were 500/40 for HC vs ADHD vs ASD, 500/40 for HC vs ASD, and 400/40 for HC vs ADHD. In Setting 2, the corresponding combinations were 500/30, 400/30, and 700/10, respectively. Under Setting 3, the optimal rs-fMRI feature numbers decreased substantially to 103, 220, and 252, while the optimal sMRI feature numbers were consistently 40. These results indicate that the optimal feature subset depends on both the classification task and the evaluation protocol, rather than a fixed number of selected features. Beyond the variation in optimal feature subsets, the relative contribution of each modality remained consistent across all evaluation settings. rs-fMRI consistently outperformed sMRI in nearly all tasks, indicating that functional connectivity provides more discriminative information than structural features for distinguishing HC, ADHD, and ASD subjects. Multimodal combinations, particularly those integrating rs-fMRI, sMRI, and demographic information, generally achieved better performance than any single modality, confirming that structural, functional, and demographic

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

Table 4. Performance under Setting 1 (Transductive Learning with Global Feature Selection).

Modality	HC vs ADHD vs ASD (%)			HC vs ASD (%)			HC vs ADHD (%)		
	Acc (95% CI)	F1-Score	AUC	Acc (95% CI)	F1-Score	AUC	Acc (95% CI)	F1-Score	AUC
sMRI	54.6 ± 4.4 (51.4–57.9)	53.4 ± 4.4	73.8 ± 3.4	64.3 ± 4.8 (60.7–68.0)	64.2 ± 4.9	73.3 ± 6.0	58.0 ± 5.0 (54.2–61.7)	57.5 ± 5.2	61.5 ± 7.3
rs-fMRI	75.9 ± 6.3 (71.2–80.7)	75.8 ± 6.2	88.6 ± 3.4	86.2 ± 4.5 (82.8–89.5)	86.1 ± 4.6	94.1 ± 4.0	85.7 ± 4.1 (82.6–88.7)	85.6 ± 4.1	93.8 ± 3.0
sMRI+ demo	65.8 ± 3.5 (63.2–68.4)	64.6 ± 3.8	82.0 ± 2.8	74.0 ± 3.5 (71.3–76.6)	73.9 ± 3.7	82.1 ± 4.2	65.1 ± 7.8 (59.3–71.0)	64.6 ± 8.3	70.4 ± 8.6
rs-fMRI+ dem	83.9 ± 2.2 (82.2–85.5)	83.9 ± 2.2	**95.5 ± 1.5**	**92.6 ± 4.1 (89.4–95.7)**	92.5 ± 4.2	97.8 ± 2.0	88.9 ± 5.6 (84.7–93.1)	88.8 ± 5.7	95.6 ± 2.6
sMRI+ rs-fMRI	77.7 ± 3.7 (74.9–80.4)	77.4 ± 3.6	92.4 ± 1.9	87.5 ± 6.6 (82.6–92.4)	87.4 ± 6.6	93.5 ± 4.8	88.9 ± 5.1 (85.0–92.7)	88.8 ± 5.1	94.6 ± 3.0
Multimodal	**85.5 ± 4.3 (82.2–88.7)**	**85.2 ± 4.2**	95.2 ± 1.2	92.5 ± 4.0 (89.5–95.6)	**92.5 ± 4.0**	**97.9 ± 2.4**	**90.4 ± 6.3 (85.7–95.2)**	**90.4 ± 6.4**	**96.5 ± 2.9**

features provide complementary information for classification. Furthermore, binary tasks consistently achieved higher performance than the multiclass task, suggesting that separating two groups is relatively easier

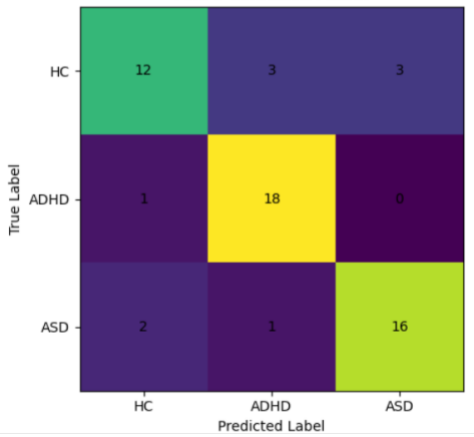

Fig. 2. Confusion Matrix for Multiclass HC–ADHD–ASD Classification (Setting 3, Best Fold)

than simultaneously distinguishing three clinically related classes. To statistically validate the superiority of the multimodal configuration, the Wilcoxon signed-rank test was performed on the fold-wise classification accuracies under Evaluation Setting 3, and Holm–Bonferroni correction was applied to account for multiple pairwise comparisons. The statistical comparison results are summarized in Table 7. The multimodal configuration significantly outperformed the sMRI, rs-fMRI, sMRI + demographic, and rs-fMRI + sMRI configurations after multiple-comparison correction (adjusted $p < 0.05$), with large effect sizes ($r = 0.886$). In contrast, the performance difference between the multimodal model and the rs-fMRI + demographic configuration was not statistically significant (adjusted $p = 0.25$), despite a moderate-to-large effect size ($r = 0.499$). These results indicate that multimodal integration provides a statistically significant performance improvement over most modality combinations under the leakage-aware evaluation protocol, whereas the additional contribution beyond the rs-fMRI + demographic configuration is comparatively limited. Beyond group-level statistical confirmation, examining class-wise classification patterns provides deeper diagnostic insight. Fig. 2 illustrates the confusion matrix for the best-performing fold under Setting 3 for the HC vs ADHD vs ASD task. ADHD exhibits the most stable classification, with 18 out of 19 samples correctly classified and no misclassifications as ASD, indicating strong feature separability. HC shows the highest misclassification rate, with 6 out of 18 samples incorrectly assigned to either ADHD or ASD, suggesting heterogeneous feature representations that partially overlap with both disorder classes. ASD samples are mostly correctly identified (16 out of 19) but show moderate confusion with HC. This asymmetric misclassification pattern where ADHD is the most discriminative class while HC and ASD exhibit partial feature overlap provides class-specific insights that aggregate metrics such as accuracy, F1-score, or AUC cannot capture. Overall, the performance gap between Setting 1 and Setting 3 highlights the impact of transductive learning and global feature selection on optimistic performance estimation. Nevertheless, the relatively stable performance under Setting 3, supported by statistical validation and class-wise analysis, indicates that the proposed framework has reasonable

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

Table 5. Performance under Setting 2 (Inductive Learning with Global Feature Selection).

Modality	HC vs ADHD vs ASD (%)			HC vs ASD (%)			HC vs ADHD (%)		
	Acc (95% CI)	F1-Score	AUC	Acc (95% CI)	F1-Score	AUC	Acc (95% CI)	F1-Score	AUC
sMRI	49.7 ± 5.7 (45.4–53.9)	48.3 ± 6.6	70.5 ± 5.2	62.2 ± 8.1 (56.1–68.3)	61.5 ± 8.9	66.1± 12.3	55.1 ± 5.2 (51.1–59.0)	52.5 ± 6.3	58.1 ± 6.9
rs-fMRI	72.3 ± 8.6 (65.8–78.8)	72.1 ± 8.7	88.6 ± 5.2	82.2 ± 6.4 (77.4–87.0)	82.1 ± 6.5	89.2 ± 6.3	82.7 ± 4.6 (79.2–86.2)	82.6 ± 4.7	89.1 ± 4.7
sMRI+ demo	56.4 ± 5.7 (52.1–60.7)	55.6 ± 6.0	75.3 ± 5.6	64.1 ± 8.6 (57.6–70.6)	63.9 ± 8.7	66.9 ± 9.7	60.7 ± 7.2 (55.3–66.1)	59.9 ± 7.4	63.4 ± 5.3
rs-fMRI+ demo	77.5 ± 8.9 (70.8–84.2)	77.3 ± 9.2	91.6 ± 4.2	84.1 ± 7.5 (78.5–89.7)	84.0 ± 7.5	91.9 ± 7.2	85.1 ± 6.2 (80.5–89.8)	85.1 ± 6.2	92.2 ± 4.2
sMRI+ rs-fMRI	75.2 ± 5.7 (70.9–79.5)	75.1 ± 5.6	90.3 ± 3.2	84.3 ± 8.2 (78.1–90.5)	84.2 ± 8.3	91.2 ± 6.9	83.0 ± 5.9 (78.6–87.4)	82.8 ± 6.0	91.3 ± 3.2
Multimodal	**80.7 ± 4.5 (77.3–84.1)**	**80.7 ± 4.5**	**93.6 ± 2.0**	**85.9 ± 5.7 (81.6–90.2)**	**85.9 ± 5.7**	**92.9 ± 5.9**	**86.7 ± 3.9 (83.8–89.6)**	**86.7 ± 4.0**	**92.7 ± 2.3**

Table 6. Performance under Setting 3 (Inductive Learning with Fold-wise Stable Feature Selection).

Modality	HC vs ADHD vs ASD (%)			HC vs ASD (%)			HC vs ADHD (%)		
	Acc (95% CI)	F1-Score	AUC	Acc (95% CI)	F1-Score	AUC	Acc (95% CI)	F1-Score	AUC
sMRI	53.7 ± 5.8 (49.3–58.1)	53.0 ± 6.0	73.2 ± 3.6	60.6 ± 6.3 (55.9–65.3)	60.1 ± 6.3	65.3 ± 6.1	58.5 ± 7.5 (52.8–64.1)	57.9 ± 7.6	59.3 ± 9.3
rs-fMRI	54.6 ± 6.0 (50.1–59.1)	54.2 ± 6.0	73.6 ± 5.0	75.8 ± 7.3 (70.3–81.3)	75.7 ± 7.3	81.9 ± 8.4	74.0 ± 7.7 (68.2–79.8)	73.2 ± 8.6	85.6 ± 5.6
sMRI+ demo	57.4 ± 5.0 (53.7–61.2)	56.5 ± 5.2	76.6 ± 3.6	60.9 ± 6.4 (56.1–65.7)	60.4 ± 6.7	69.6 ± 6.9	58.0 ± 6.1 (53.4–62.6)	57.1 ± 6.8	62.9 ± 7.6
rs-fMRI+ demo	63.7 ± 7.1 (58.3–69.1)	63.5 ± 7.2	80.3 ± 5.7	79.2 ± 8.3 73.0–85.5)	79.1 ± 8.3	**88.7 ± 7.2**	**79.8 ± 5.3 (75.8–83.8)**	**79.7 ± 5.3**	87.0 ± 5.7
sMRI+ rs-fMRI	64.0 ± 7.2 (58.6–69.5)	63.7 ± 7.2	83.2 ± 5.1	81.6 ± 8.4 (75.3–88.0)	81.5 ± 8.5	88.0 ± 7.0	79.8 ± 6.8 (74.7–85.0)	79.6 ± 7.1	**88.8 ± 4.9**
Multimodal	**70.9 ± 6.7 (65.9–76.0)**	**70.9 ± 6.5**	**86.6 ± 3.7**	**81.7 ± 5.3 (77.7–85.6)**	**81.6 ± 5.3**	87.3 ± 6.4	79.3 ± 6.4 (74.4–84.1)	79.2 ± 6.5	88.4 ± 5.0

generalization capability on unseen subjects.

C. Analysis of Adaptive Modality Fusion

Table 8 presents the average adaptive fusion weights learned by the proposed model under Setting 3. Overall, the learned weights are relatively balanced across modalities, indicating that the model effectively integrates complementary information from fMRI, sMRI, and demographic data. For the HC vs ADHD vs ASD and HC vs ASD tasks, demographic features receive slightly higher weights, whereas for HC vs ADHD, sMRI contributes marginally more in the first fusion layer. These results suggest that the proposed adaptive fusion mechanism dynamically adjusts modality importance according to the classification task rather than relying on a single dominant modality.

IV. Discussion

A. Discussion of Main Findings

The results demonstrate that evaluation protocols and feature selection strategies substantially influence the reported performance of graph-based neuroimaging models. Across all classification tasks, Setting 1 consistently achieved the highest performance, whereas Setting 3 produced lower but potentially more

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

Table 7. Statistical Comparison of Modality Configurations for HC–ADHD–ASD Classification Under Evaluation Setting 3 Using the Wilcoxon Signed-Rank Test

Comparison	Effect Size(r)	Adjusted p-value	Significant
Multimodal vs sMRI	0.886	0.0098	Yes
Multimodal vs rsfMRI	0.886	0.0156	Yes
Multimodal vs sMRI+demo	0.886	0.0098	Yes
Multimodal vs rsfMRI+demo	0.499	0.2500	No
Multimodal vs rsfMRI+sMRI	0.886	0.0098	Yes

realistic estimates of model generalization capability. The superior performance observed in Setting 1 may be attributed to the transductive graph construction and global feature selection strategy, where all subjects are involved during graph construction prior to the train-test split, potentially allowing indirect information sharing between training and testing data. In contrast, Setting 3 employs inductive learning with fold-wise stable feature selection, minimizing potential information leakage and providing a stricter evaluation framework. The considerable performance gap between Setting 1 and Setting 3 highlights that evaluation strategies can substantially affect reported classification accuracy and emphasizes the importance of leakage-aware protocols in neuroimaging-based graph learning studies. Across all evaluation settings, rs-fMRI consistently outperformed sMRI, indicating that functional connectivity provides more discriminative information than structural brain features. Nevertheless, multimodal integration consistently improved performance, suggesting that rs-fMRI, sMRI, and demographic variables provide complementary information. Although demographic variables alone have limited predictive value, the adaptive fusion mechanism assigned relatively high weights because age and sex complement neuroimaging features rather than acting as primary discriminative variables. Consequently, modality weights reflect each modality's contribution to the combined prediction rather than its standalone performance. However, because the ADHD-200 and ABIDE datasets differ in acquisition sites, scanner characteristics, and demographic distributions, age and sex may partially encode site-specific rather than disease-specific information. Therefore, the relatively high adaptive fusion weights assigned to demographic variables should be interpreted cautiously, and future studies should

Table 8. Learned Adaptive Fusion Weights under Setting 3

Task	Layer	$\overline{\alpha_{fMRI}}$	$\overline{\alpha_{sMRI}}$	$\overline{\alpha_{demo}}$
HC vs ADHD vs ASD	1	0.3244	0.3299	0.3457
	2	0.3222	0.3290	0.3488
HC vs ASD	1	0.3311	0.3318	0.3371
	2	0.3305	0.3297	0.3398
HC vs ADHD	1	0.3286	0.3367	0.3347
	2	0.3299	0.3332	0.3369

employ harmonization techniques (e.g., ComBat) together with external validation to better separate disease-related from site-related effects.

B. Comparison with Previous Studies

Table 9 presents the comparison between the proposed framework and previous studies for multiclass and binary classification tasks. Overall, the proposed framework achieved competitive performance under both conventional and stricter evaluation settings. Although the best results were obtained under Setting 1, the model maintained competitive accuracies under the leakage-aware Setting 3, achieving 70.9% for HC vs ADHD vs ASD, 81.7% for HC vs ASD, and 79.3% for HC vs ADHD. These findings suggest that the proposed multimodal graph learning framework retains useful discriminative information even when strict evaluation procedures are applied. Table 9 further demonstrates that differences among previous studies extend beyond classification accuracy. Earlier graph-based approaches such as IFC-GNN [7], HCAN [14], DAGMN [15], B2P-GL [16], and MMGCN [41] differ substantially in sample size, preprocessing pipeline, feature selection strategy, graph construction, and validation protocol. Several previous studies adopted transductive graph construction or did not explicitly report whether graph construction and feature selection were performed independently within each training fold, which may increase the risk of information leakage. In contrast, the proposed framework explicitly separates feature selection and graph construction within each training fold under Setting 3, thereby providing a more rigorous leakage-aware evaluation. Regarding preprocessing specifically, HCAN [14] and IFC-GNN both relied on data preprocessed through the C-PAC pipeline (via the Preprocessed Connectomes Project) [7], consistent with the PCP-preprocessed derivatives used in our framework, whereas DMDC [42] and ESTGCN [43] each applied independent, non-PCP preprocessing pipelines (FSL-based processing for DMDC; SPM-12 combined with the Athena toolbox and the Craddock 200 atlas for ESTGCN). This heterogeneity in

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

preprocessing pipelines across compared studies further limits the extent to which absolute accuracy values can be directly attributed to modeling choices alone.

C. Limitations and Clinical Implications

Despite the promising results, several limitations should be acknowledged. First, combining subjects from the ADHD-200 and ABIDE datasets may introduce site- and dataset-specific bias because the images were acquired using different scanners and acquisition protocols. Although demographic variables (age and sex) improved multimodal classification, they may only partially capture acquisition-site characteristics, as ADHD-200 and ABIDE differ in scanner protocols and demographic composition. Since no explicit harmonization technique (e.g., ComBat) was applied, residual site effects cannot be

Table 9. Comparison of Classification Performance Under Different Evaluation Protocols for HC vs ADHD vs ASD, HC vs ASD, and HC vs ADHD Tasks

Study	Dataset	Sample Size	Modality	Validation Protocol	Evaluation Setting	ACC (%)	AUC (%)
HC vs ADHD vs ASD							
DMDC [42]	ADHD-200 + ABIDE	890	rs-fMRI	Hold-out split	Subject-level train/test split	62.0	-
Set 1 (Ours)	ADHD-200 + ABIDE	564	rs-fMRI + sMRI + demography	10-fold CV	Transductive and global feature selection	85.5	95.2
Set 3 (Ours)	ADHD-200 + ABIDE	564	rs-fMRI + sMRI + demography	10-fold CV	Inductive and fold-wise stable feature selection	70.9	86.6
HC vs ASD							
HCAN [14]	ABIDE	871	rs-fMRI + demography	10-fold CV	Transductive population graph and feature selection procedure not explicitly reported	82.9	84.6
IFC-GNN [7]	ABIDE	871	rs-fMRI + demography	10-fold CV	Global feature selection setting and unclear graph construction protocol	80.7	-
MMGCN [41]	ABIDE	949	rs-fMRI + Demography	10-fold CV	Per-fold feature selection and graph construction protocol not explicitly reported	78.3	84.0
DAGMNet [15]	ABIDE	871	rs-fMRI + sMRI + demography	10-fold CV	Transductive population graph setting	91.6	96.8
B2P-GL [16]	ABIDE	850	rs-fMRI + Demography	10-fold CV	Transductive population graph with global graph construction	83.4	83.0
Set 1 (Ours)	ABIDE	564	rs-fMRI + sMRI + demography	10-fold CV	Transductive and global feature selection	92.5	97.9
Set 3 (Ours)	ABIDE	564	rs-fMRI + sMRI + demography	10-fold CV	Inductive and fold-wise stable feature selection	81.7	87.3
HC vs ADHD							
B2P-GL [16]	ADHD-200	938	rs-fMRI + Demography	10-fold CV	Transductive population graph with global graph construction	78.6	78.8
ESTGCN [43]	ADHD-200	216	rs-fMRI	10-fold CV	Graph construction and feature selection not explicitly reported	75.4	-
dGCN [44]	ADHD-200	603	rs-fMRI + Demography	10-fold CV	Inductive subject-level graph classification	72.0	-
Set 1 (Ours)	ADHD-200	564	rs-fMRI + sMRI + demography	10-fold CV	Transductive and global feature selection	90.4	96.5
Set 3 (Ours)	ADHD-200	564	rs-fMRI + sMRI + demography	10-fold CV	Inductive and fold-wise stable feature selection	79.3	88.4

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

excluded and may have influenced both the learned representations and the adaptive fusion weights. Second, this study relied on PCP-preprocessed rs-fMRI data and predefined structural MRI features without investigating alternative preprocessing strategies. Furthermore, although DFS and mRMR identified discriminative features, their biological relevance was not systematically analyzed, limiting the interpretability of the proposed framework. The confusion matrix was also presented only for the best-performing cross-validation fold, which may not fully represent the average classification behavior across all folds. Finally, the proposed framework was evaluated only through internal cross-validation on 564 subjects and has not yet been validated using an independent external cohort. Therefore, the reported performance should be interpreted as an estimate of internal generalization rather than definitive clinical applicability. Future work should incorporate larger multi-site datasets, harmonization methods, external validation, and explainable graph learning to improve robustness and biological interpretability.

Despite these limitations, the proposed leakage-aware multimodal graph learning framework has important practical implications. The results demonstrate that leakage-aware evaluation provides more reliable performance estimates, while integrating rs-fMRI, sMRI, and demographic information consistently improves classification compared with single-modality models. Although further external validation is required, the proposed framework provides a promising basis for computer-aided decision support in distinguishing ADHD, ASD, and healthy controls, while encouraging more rigorous and reproducible evaluation practices in neuroimaging-based machine learning.

V. Conclusion

This study proposed a multimodal Multi-Graph ChebGCN framework for the classification of HC, ADHD, and ASD and investigated the effects of evaluation protocols and feature selection strategies on model performance and information leakage. Results showed that Setting 1 achieved the highest accuracies, while performance decreased under the leakage-aware Setting 3, which yielded multimodal accuracies of 70.9% for HC vs ADHD vs ASD, 81.7% for HC vs ASD, and 79.3% for HC vs ADHD, highlighting the impact of evaluation protocols on reported model performance Across all settings, rs-fMRI outperformed sMRI, and multimodal integration consistently achieved the best performance. Overall, the proposed framework maintained competitive performance under leakage-aware evaluation. Future work will explore larger multi-site datasets, fully inductive graph learning, and explainable graph learning to improve robustness and interpretability.

Acknowledgment

The authors would like to express sincere gratitude to Universitas Sebelas Maret for providing research funding through the Hibah Riset Group scheme as stipulated in contract No. 462/UN27.22/PT.01.03/2026.

Funding

This research was funded by Universitas Sebelas Maret through the Hibah Riset Group scheme under contract No. 462/UN27.22/PT.01.03/2026. The funder had no role in the study design, data collection, analysis, interpretation of data, manuscript preparation, or the decision to publish.

Author Contribution

Chofifatul Hidayah contributed to model implementation, experimentation, evaluation, and manuscript drafting. Wiharto supervised the study, contributed to the research design, and provided critical revisions to the manuscript. Esti Suryani contributed to methodological development, analysis support, and manuscript review. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Data Availability

The ADHD-200 and ABIDE I datasets analyzed in this study are publicly available through their respective repositories. The ADHD-200 dataset is available from the ADHD-200 Consortium, and the ABIDE I dataset is available through the Autism Brain Imaging Data Exchange (ABIDE). Access to these datasets is subject to the respective data-use agreements and repository policies. The processed derivatives generated during this study are available from the corresponding author upon reasonable request.

Code Availability

The code and experimental scripts used in this study are available from the corresponding author upon reasonable request

Declarations

Ethics approval and consent to participate

This study is a secondary analysis of publicly available, de-identified neuroimaging datasets from the ADHD-200 Consortium [20] and the Autism Brain Imaging Data Exchange (ABIDE I) [21]. Ethical approval and written informed consent were obtained by the original data contributors, as described in the respective dataset publications. All analyses in this study were conducted in accordance with the data-use agreements and policies

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

governing these datasets. No attempt was made to re-identify individual participants.

Consent for Publication
Not applicable.

Competing Interests
The authors declare no competing interests.

Declaration of Generative AI Use
During the preparation of this manuscript, the authors used an AI-based language tool solely for grammar refinement

References

[1] E. David, B. Eva, and G. Christopher, "Neurodevelopmental disorders and comorbidity in young adults attending a psychiatric outpatient clinic," *Psychiatry Res.*, vol. 313, p. 114638, Jul. 2022, doi: 10.1016/j.psychres.2022.114638.

[2] F. Panvino *et al.*, "Neurotrophins in Neurodevelopmental Disorders: A Narrative Review of the Literature," *Int. J. Mol. Sci.*, vol. 26, no. 17, p. 8335, Aug. 2025, doi: 10.3390/ijms26178335.

[3] E. Mark Mahone, S. Warschausky, and T. Andrew Zabel, "Introduction to the JINS Special Issue: Neurodevelopmental Disorders," *Journal of the International Neuropsychological Society*, vol. 24, no. 9, pp. 893–895, Oct. 2018, doi: 10.1017/S1355617718000905.

[4] I. Parenti, L. G. Rabaneda, H. Schoen, and G. Novarino, "Neurodevelopmental Disorders: From Genetics to Functional Pathways," *Trends Neurosci.*, vol. 43, no. 8, pp. 608–621, Aug. 2020, doi: 10.1016/j.tins.2020.05.004.

[5] K. Kaur, V. Chawla, and B. Rana, "Phase Synchronization-based Topological Features for Autism and ADHD Classification," *in Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2024), Bengaluru, India, Dec. 13–15, 2024, New York, NY, USA: ACM, 8 pp.,* doi: 10.1145/3702250.3702295.

[6] S. Martinez, K. Stoyanov, and L. Carcache, "Unraveling the spectrum: overlap, distinctions, and nuances of ADHD and ASD in children," *Front. Psychiatry*, vol. 15, p. 1387179, Sep. 2024, doi: 10.3389/ fpsyt.2024.1387179

[7] X. Wang, X. Zhang, Y. Chen, and X. Yang, "IFC-GNN: Combining interactions of functional connectivity with multimodal graph neural networks for ASD brain disorder analysis," *Alexandria Engineering Journal*, vol. 98, pp. 44–55, Jul. 2024, doi: 10.1016/j.aej.2024.04.023.

[8] D. C. Lohani and B. Rana, "ADHD diagnosis using structural brain MRI and personal characteristic data with machine learning framework," *Psychiatry Res. Neuroimaging*, vol. 334, Art. no. 111689, Sep. 2023, doi: 10.1016/j.pscychresns.2023.111689.

[9] Y. Zhao, L. Li, and B. S. Caffo, "Multimodal Neuroimaging Data Integration and Pathway Analysis," *Biometrics*, vol. 77, no. 3, pp. 879–889, Sep. 2021, doi: 10.1111/biom.13351.

[10] A. Porter, S. Fei, K. S. F. Damme, R. Nusslock, C. Gratton, and V. A. Mittal, "A meta-analysis and systematic review of single vs. multimodal neuroimaging techniques in the classification of psychosis," *Molecular Psychiatry 2023 28:8*, vol. 28, no. 8, pp. 3278–3292, Aug. 2023, doi: 10.1038/s41380-023-02195-9.

[11] E. C. Peterson, H. R. Smolker, A. D. Moser, and R. H. Kaiser, "Review of Dynamic Resting-State Methods in Neuroimaging: Applications to Depression and Rumination," *Hum. Brain Mapp.*, vol. 46, no. 15, p. e70377, Oct. 2025, doi: 10.1002/hbm.70377.

[12] S. Pau, A. Perniciano, B. Pes, and D. Rubattu, "An Evaluation of Feature Selection Robustness on Class Noisy Data," *Information*, vol. 14, no. 8, p. 438, Aug. 2023, doi: 10.3390/info14080438.

[13] J. Ding, J. Du, H. Wang, and S. Xiao, "A novel two-stage feature selection method based on random forest and improved genetic algorithm for enhancing classification in machine learning," *Sci. Rep.*, vol. 15, no. 1, Art. No. 16828, May 2025, doi: 10.1038/s41598-025-01761-1.

[14] L. Shao, C. Fu, and X. Chen, "A heterogeneous graph convolutional attention network method for classification of autism spectrum disorder," *BMC Bioinformatics*, vol. 24, no. 1, p. 363, Dec. 2023, doi: 10.1186/S12859-023-05495-7.

[15] L. Wang, X. Li, J. Yuan, and Y. Chen, "DAGMNet: Dual-Branch Attention-Pruned Graph Neural Network for Multimodal sMRI and fMRI Fusion in Autism Prediction," *Biomedicines*, vol. 13, no. 9, p. 2168, Sep. 2025, doi: 10.3390/biomedicines13092168.

[16] Q. Liao *et al.*, "A brain-to-population graph learning framework for diagnosing brain disorders," *Neural Networks*, vol. 197, p. 108526, May 2026, doi: 10.1016/j.neunet.2025.108526.

[17] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, "Data leakage inflates prediction performance in connectome-based machine learning models," *Nature Communications 2024 15:1*, vol. 15, no. 1, Art. No. 1829, Feb. 2024, doi: 10.1038/s41467-024-46150-w.

[18] H. Mohammadi and W. Karwowski, "Graph Neural Networks in Brain Connectivity Studies: Methods, Challenges, and Future Directions,"

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

Brain Sci., vol. 15, no. 1, Art. no. 17, Jan. 2025, doi: 10.3390/brainsci15010017.

[19] Y. Yang, C. Ye, and T. Ma, "A deep connectome learning network using graph convolution for connectome-disease association study," *Neural Networks*, vol. 164, pp. 91–104, Jul. 2023, doi: 10.1016/j.neunet.2023.04.025.

[20] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, "The Neuro Bureau ADHD-200 Preprocessed repository," *Neuroimage*, vol. 144, no. Pt B, pp. 275–286, Jan. 2017, doi: 10.1016/j.neuroimage.2016.06.034.

[21] A. Di Martino *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol. Psychiatry*, vol. 19, no. 6, pp. 659–667, Jun. 2014, doi: 10.1038/mp.2013.78.

[22] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking Whole-Brain Connectivity Dynamics in the Resting State," *Cerebral Cortex (New York, NY)*, vol. 24, no. 3, pp. 663–676, Mar. 2014, doi: 10.1093/cercor/bhs352.

[23] R. M. Hutchison *et al.*, "Dynamic functional connectivity: Promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, Oct. 2013, doi: 10.1016/j.neuroimage.2013.05.079.

[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/tpami.2005.159.

[25] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering," *in Advances in Neural Information Processing Systems*, vol. 29, D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, Eds., pp. 3844–3852, 2016. Available: https://proceedings.neurips.cc/paper_files/paper/2016/file/04df4d434d481c5bb723be1b6df1ee65-Paper.pdf

[26] M. Mennes, B. B. Biswal, F. X. Castellanos, and M. P. Milham, "Making data sharing work: The FCP/INDI experience," *Neuroimage*, vol. 82, pp. 683–691, Nov. 2013, doi: 10.1016/j.neuroimage.2012.10.064.

[27] O. Esteban *et al.*, "fMRIPrep: a robust preprocessing pipeline for functional MRI," *Nat. Methods*, vol. 16, no. 1, pp. 111–116, Jan. 2019, doi: 10.1038/S41592-018-0235-4.

[28] B. Fischl, "FreeSurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, Aug. 2012, doi: 10.1016/j.neuroimage.2012.01.021.

[29] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002, doi: 10.1006/nimg.2001.0978.

[30] B. Qiu, Q. Wang, X. Li, W. Li, W. Shao, and M. Wang, "Adaptive spatial-temporal neural network for ADHD identification using functional fMRI," *Front. Neurosci.*, vol. 18, Art. no. 1394234, May 2024, doi: 10.3389/fnins.2024.1394234

[31] A. D. Savva, G. D. Mitsis, and G. K. Matsopoulos, "Assessment of dynamic functional connectivity in resting-state fMRI using the sliding window technique," *Brain Behav.*, vol. 9, no. 4, p. e01255, Apr. 2019, doi: 10.1002/BRB3.1255.

[32] S. Parisot *et al.*, "Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease," *Med. Image Anal.*, vol. 48, pp. 117–130, Aug. 2018, doi: 10.1016/j.media.2018.06.001.

[33] S. I. Ktena *et al.*, "Metric learning with spectral graph convolutions on brain connectivity networks," *Neuroimage*, vol. 169, pp. 431–442, Apr. 2018, doi: 10.1016/j.neuroimage.2017.12.052.

[34] Y. D. Zhang *et al.*, "Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation," *Information Fusion*, vol. 64, pp. 149–187, Dec. 2020, doi: 10.1016/j.inffus.2020.07.006.

[35] M. Vaida and Z. Huang, "Multimodal graph neural networks in healthcare: a review of fusion strategies across biomedical domains," *Front. Artif. Intell.*, vol. 8, p. 1716706, Jan. 2026, doi: 10.3389/frai.2025.1716706.

[36] J. Wen *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, Art. no. 101694, Jul. 2020, doi: 10.1016/j.media.2020.101694.

[37] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *Gigascience*, vol. 6, no. 5, pp. 1–9, May 2017, doi: 10.1093/gigascience/gix019.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. *Cambridge*, MA, USA: MIT Press, 2016.

[39] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948

classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
[40] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 7th ed. Hoboken, NJ, USA: John Wiley & Sons, 2018.
[41] T. Song, Z. Ren, J. Zhang, and M. Wang, "Multi-View and Multimodal Graph Convolutional Neural Network for Autism Spectrum Disorder Diagnosis," *Mathematics*, vol. 12, no. 11, p. 1648, May 2024, doi: 10.3390/math12111648.
[42] Z. Zolghadr, S. A. H. Batouli, H. A. Majd, L. Shafaghi, and Y. Mehrabi, "fMRI-Based Multi-class DMDC Model Efficiently Decodes the Overlaps between ASD and ADHD," *Basic Clin. Neurosci.*, vol. 15, no. 3, pp. 367–382, May 2024, doi: 10.32598/bcn.2023.4302.1.
[43] L. Chen *et al.*, "An explainable spatio-temporal graph convolutional network for the biomarkers identification of ADHD," *Biomed. Signal Process. Control*, vol. 99, p. 106913, Jan. 2025, doi: 10.1016/j.bspc.2024.106913.
[44] K. Zhao, B. Duka, H. Xie, D. J. Oathes, V. Calhoun, and Y. Zhang, "A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD," *Neuroimage*, vol. 246, p. 118774, Feb. 2022, doi: 10.1016/j.neuroimage.2021.118774.

Author Biography

Chofifatul Hidayah is an undergraduate Informatics student at Universitas Sebelas Maret with research interests in artificial intelligence, machine learning, biomedical informatics, and computational intelligence. Her current work focuses on applying graph neural networks and deep learning methods for neurodevelopmental disorder classification, particularly ADHD and ASD. She has served as a teaching assistant for Computer Networks, Software Engineering, and Numerical Methods courses. In addition, she has participated in research on autonomous ship prototype development, contributing to the intelligent systems and computational aspects of the project. She actively engages in scientific forums, workshops, and collaborative research activities to expand her academic and research experience.

Wiharto is a senior lecturer in the Department of Informatics, Faculty of Information Technology and Data Science at Universitas Sebelas Maret, Surakarta. He has a strong academic background and broad experience in computational science and engineering. He obtained his doctoral degree from Universitas Gadjah Mada with a specialization in biomedical informatics. His research interests include biomedical informatics, artificial intelligence, and computational intelligence, particularly in developing innovative computational methods to address real-world challenges. In addition to his research activities, he actively mentors undergraduate and postgraduate students, publishes scientific papers in reputable national and international journals, and collaborates on interdisciplinary research projects in health informatics and artificial intelligence.

Esti Suryani is an active researcher in the Computational Science and Engineering research group. She earned her Master's degree in Computer Science from Universitas Gadjah Mada, where she developed a solid background in advanced computing methods. Her research interests cover several areas, particularly data analysis and the development of computational models for addressing complex engineering problems in real-world applications. She is also passionate about interdisciplinary collaboration, working with researchers from various fields to design innovative and data-driven solutions, especially in algorithm optimization and computational modeling. In addition, she actively contributes to student mentoring and research supervision, supporting the development of future computational science researchers.

Manuscript received 8 April 2025; Revised 10 June 2025; Accepted 15 June 2025; Available online 27 June 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.948