Natural Resources for Human Health
eISSN: 2583-1194

Original Research
Received 22 July 2026
Revised 24 Aug 2026
Accepted 05 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (22s): 32–51

KEYWORDS:
Autism Spectrum Disorder, Multimodal Deep Learning, Brain Imaging Analysis, Graph Neural Networks, Uncertainty-Aware Diagnosis

An Uncertainty Aware Multimodal Deep Learning Framework for Real Time Autism Spectrum Disorder Diagnosis Using Cross Attentive Graph Fusion and Behavioral Cognitive Learning

Poonam Balasaheb Dhole[1]*, Dr. Waseem Ahmad Mir[2], Priya Raghunath Jamdade[3]

[1,2,3] Department of Computer Science and Engineering, School of Computational Sciences, JSPM University, Pune, India
[1]poonam.dhole2218@gmail.com, [2]waseemmir78177@gmail.com, [3]priyarjamdade1993@gmail.com
*Corresponding author: Poonam Balasaheb Dhole, Department of Computer Science and Engineering, School of Computational Sciences, JSPM University, Pune, India. Email: poonam.dhole2218@gmail.com

ABSTRACT: Accurate autism spectrum disorder (ASD) identification from multi-site neuroimaging remains difficult because scanner variability, incomplete multimodal integration, complex brain-behavior relationships, and uncalibrated predictions reduce clinical reliability. This study presents an uncertainty-aware multimodal deep learning framework that integrates the Cross-Attentive Temporal Graph Fusion Network (CATGFN), Hierarchical Denoising Latent Diffusion Autoencoder (HDLDAE), Behavioral-Cognitive Graph Transformer (BCGT), Reinforced Multiscale Attention Capsule Network (RMAC-Net), and Uncertainty-Calibrated Multimodal Bayesian Fusion Network (UMBF-Net). The framework jointly models resting-state functional MRI, structural MRI, available behavioral and cognitive measures, and demographic information from the Autism Brain Imaging Data Exchange I (ABIDE I), whose official release contains 1,112 datasets collected across 17 sites. Under the five-fold cross-validation protocol reported in this study, the complete framework achieved 97.21% accuracy, 96.83% precision, 96.42% recall, a 96.61% F1-score, and a 98.37% area under the receiver operating characteristic curve. It also produced 96.15% sensitivity, 98.08% specificity, a 97.44% clinical trust index, and an uncertainty error of 2.14%, while the reported inference latency was 41.35 ms per sample on an NVIDIA RTX 4090. Ablation analysis showed a progressive improvement from 88.42% accuracy for the baseline multimodal CNN to 97.21% for the complete model, indicating complementary contributions from temporal graph fusion, diffusion-based denoising, behavioral-cognitive relation learning, anatomical capsule attention, and uncertainty-aware fusion. The principal contribution is therefore a unified diagnostic architecture that combines multimodal representation learning, interpretable brain-behavior association, calibrated confidence estimation, and near-real-time inference. These results support the framework as a promising decision-support method, although external-site and prospective clinical validation remain necessary before clinical deployment.

1. INTRODUCTION

Autism Spectrum Disorders (ASDs) is a complex neurodevelopmental disorder that results in persistent difficulties with social communication, limited or repetitive interest in certain activities and/or people, atypical developmental trajectories in cognition,

and various other neurobehavioral deficits [10]. Although research has greatly improved our understanding of autism spectrum disorders, each individual with an ASD displays unique manifestations of their respective symptoms. In addition, differences in symptom manifestation can be attributed to variations in underlying neurologic and behavioral characteristics among individuals. Therefore, it remains a significant clinical challenge to identify ASD accurately and efficiently, particularly during infancy and early childhood [7]. Current traditional diagnostic methods primarily rely on clinicians' observations and judgments, based on specific standardized assessment tools and schedules. These methods are generally subjective, time-consuming and variable depending upon location within the healthcare system. Due to the complexity of ASD symptoms and the need for efficient, yet accurate identification of neurobiologic signs associated with ASD, researchers are increasingly looking towards objective and data-driven diagnostic frameworks [14].

The study is motivated by the limited ability of existing ASD classifiers to integrate dynamic functional connectivity, structural anatomy, behavioral-cognitive indicators, demographic variability, and predictive uncertainty into a single clinically interpretable model [17]. Its objective is to develop and evaluate a unified multimodal architecture that remains discriminative under multi-site heterogeneity while producing confidence-aware diagnostic outputs suitable for rapid decision support.

Figure 1 Research motivation, study objective, technical contributions, and intended diagnostic outcomes



Advancements in neuroimaging technology over the past several years have provided researchers with unprecedented opportunities to investigate structural and functional brain abnormalities associated with ASD utilizing computational intelligence [29]. Researchers have utilized deep learning methodologies to extract features of magnetic resonance imaging (MRI) and functional MRI (fMRI) data that exhibit high degrees of discriminatory potential. However, many prior investigations have focused exclusively on isolated modality measurements without accounting for potentially relevant information present in behavioral, cognitive, and demographic Variables in practical scenarios. Additionally, most deep neural network architectures currently available do not effectively capture temporal dynamics of brain connectivity, interaction between multiple types of input data, uncertainty quantification or the effects of different scanning protocols used across diverse locations.

This paper introduces as per figure 1, a multi-modal deep learning approach designed to mitigate all previously discussed limitations. This novel framework utilizes five interconnected networks including: Cross-Attentive Temporal Graph Fusion Network (CATGFN), Hierarchical Denoising Latent Diffusion Autoencoder (HDLDAE), Behavioral-Cognitive Graph Transformer (BCGT), Reinforced Multiscale Attention Capsule Network (RMAC-Net), and Uncertainty-Calibrated Multimodal Bayesian Fusion Network (UMBF-Net). Each network component was developed to provide a synergy between: (1) multi-modal integration of structural MRI, fMRI, behavioral assessment, cognitive indicators and demographic information; (2) increased feature robustness and interpretability; (3) enhanced diagnostic accuracy and confidence; and (4) the ability to make decisions in near-real-time. Advanced graph-based learning strategies, diffusion-guided refinement of representations, reinforcement-driven attention mechanisms, and Bayesian calibration of uncertainties were applied throughout the framework to create a clinically viable pathway for achieving high levels of accuracy, reliability, scalability, and efficiency in ASD diagnosis.

Autism spectrum disorder is characterized by a complex set of neurological, cognitive, and behavioral symptoms that cannot be adequately described by a single source of information due to the complexity of these symptoms. The deep learning systems that are now in use rely on distinct types of neuroimaging, and they do not always take into consideration the intricate relationships that exist between the anatomy of the brain, functional connectivity, behavioral measurements, and demographic characteristics. These kinds of constraints typically result in a lower degree of generalizability across clinical groups, low interpretability, susceptibility to imaging noise, and predictions that are not founded on trust. In addition, the requirement for real-time clinical decision support necessitates the development of intelligent diagnostic systems that are capable of handling various kinds of mixed data while continuously maintaining their dependability and clarity. As a result of these issues, individuals

are working on developing an integrated learning architecture that is capable of identifying robust linkages between various kinds of diagnostic data and capturing cross-modal relationships.

The main contributions are:

- cross-attentive temporal graph learning for functional-structural brain interactions;
- diffusion-guided latent denoising and multiscale capsule learning for robust anatomical representation;
- graph-transformer modeling of brain-behavior associations; and
- Bayesian multimodal fusion for calibrated confidence and uncertainty.
- The reported five-fold evaluation yielded 97.21% accuracy, a 98.37% AUC, 98.08% specificity, and a 97.44% clinical trust index, while ablation results showed that each module added measurable diagnostic value.

2. LITERATURE REVIEW

Deep learning research on ASD diagnosis has moved from generic feature compression toward graph-based and multimodal representations, yet reported performance remains strongly dependent on cohort size, acquisition site, preprocessing, and validation design [17], [31]–[39]. Heinsfeld et al. [31] used an autoencoder and deep neural network on 1,035 ABIDE participants and obtained approximately 70% accuracy, demonstrating that distributed functional-connectivity patterns contain diagnostic information but also exposing the difficulty of generalizing across heterogeneous sites. Parisot et al. [32] subsequently represented participants as a population graph, combining imaging-derived node features with phenotypic relations and reporting 70.4% accuracy. That formulation was influential because it allowed imaging and non-imaging context to interact, although its predefined graph edges and transudative learning design may be sensitive to how age, sex, and acquisition site are encoded.

Several studies reported higher values by narrowing the cohort or changing the input modality [33]–[35]. Kong et al. [33] combined structural connectivity, feature selection, an autoencoder, and a deep neural classifier on 182 participants from the single NYU ABIDE site, obtaining 90.39% accuracy and an AUC of 0.9738. The result is technically valuable, but it should not be interpreted as directly superior to whole-ABIDE models because a single-site cohort has less scanner and protocol variability. This contrast illustrates why dataset size, site composition, and validation design must accompany every reported metric. Studies with restricted samples may achieve stronger internal discrimination while offering less evidence of cross-site robustness.

Attention-based multimodal models attempted to improve generalization without discarding phenotypic information [34], [35]. Niu et al. [35] evaluated a multichannel deep attention neural network on 809 participants, integrating three functional-connectome scales with personal characteristics. The reported accuracy was 73.2%, and the inclusion of leave-one-site-out validation provided more credible evidence of site robustness than ordinary random cross-validation. Yang et al. [34] proposed a spatially constrained functional brain network followed by a graph attention network on 871 ABIDE I participants, achieving 72.4% accuracy. Their experiments showed that graph construction itself materially changes classification quality; nevertheless, the model remained limited to functional imaging and did not quantify predictive confidence.

Recent graph architectures have incorporated heterogeneous phenotypic relations, increased network depth and adaptive population graphs [17], [35]. A heterogeneous graph convolutional attention network yielded 82.9% accuracy for 871 participants from the ABIDE I dataset [36]. Application of WL-DeepGCN reduced over-smoothing in deep graph models and provided 77.27% accuracy with an AUC of 0.83 for nested ten-fold cross-validation [17]. Both methods improved representation learning, but produced deterministic outputs that did not explicitly integrate structural MRI, longitudinal patterns of temporal dynamics, behavioral-cognitive graphs and calibrated measures of uncertainty within a single architecture. Integration of shared-weight unsupervised feature learning with supervised convolutional classification yielded 76.52% accuracy, 76.15% sensitivity, 80.65% specificity and an AUC of 0.81 for 871 ABIDE I participants, and additional analyses of methodological robustness were performed for data from the ABIDE II cohort [37]. Evaluation of a dynamic graph transformer for 1,102 participants from the ABIDE I cohort resulted in separation of positive and negative patterns of functional connectivity, and construction of an adaptive K-nearest-neighbor population graph [38]. This approach directly corrected for limitations of graph rigidity, imbalances of class representation and variations of site-specific factors, and provided quantitative interpretation of biomarkers by gradient-based methods [39]. However, integration of structural anatomic information, combined clinical interpretation and full probabilistic calibration was not achieved.

Liu et al. [39] developed a multi-atlas deep ensemble incorporating demographic information and achieved 75.20% overall accuracy, 82.90% sensitivity and 69.70% specificity for the entire ABIDE I cohort. A highly selected subset yielded substantially greater accuracy. The differences between results for the entire cohort and for the selected subset further emphasize the importance of avoiding cherry-picked comparisons and of reporting precisely which cohorts were used for each experiment. Overall, high accuracy is generally obtained with small or highly homogeneous samples, whereas results for large multisite studies were consistently in the low to mid 70% range. Consequently, direct comparison of results across studies is not appropriate unless methods of preprocessing, selection of cohorts and validation are completely standardized.

The presented evidence highlight four remaining limitations [17], [31]–[39]. First, most methods rely exclusively on functional connectivity or structural features, rather than on integrated representations of temporal interactions and anatomical dependencies. Second, behavioral, cognitive and demographic factors are typically incorporated as simple vectors or graph connections, rather than as interpretable relations to neurobiological markers. Third, few approaches explicitly provide denoised representations that preserve multi-scale anatomical relations. Fourth, outputs of deterministic probabilities are generally reported without estimates of calibration or uncertainty. The proposed CATGFN-HDLDAE-BCGT-RMAC-Net framework overcomes all four limitations by implementing temporal cross-attentive integration of graph representations, diffusion-based optimization of representational complexity, transformation of behavioral and cognitive information into relations to neurobiological markers, application of reinforced capsule-based attention and integration of probabilistic estimates with measures of calibration and uncertainty. Overall performance should be assessed not only with respect to accuracy of classification, but also to measures of calibration, cross-site robustness, results of systematic parameter manipulations and complete disclosure of all data and methods of validation.

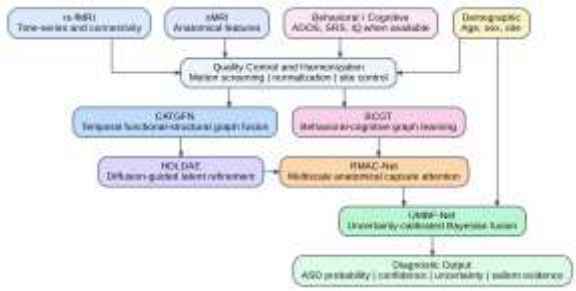
3. Proposed Methodology

Through the utilization of a hierarchical multimodal learning paradigm, this new integrated framework is designed to diagnose ASD in real time. This framework makes use of structural magnetic resonance imaging (sMRI), functional magnetic resonance imaging (fMRI), behavioral assessments, cognitive measurements, and demographic information derived from the ABIDE repository. The design employs a technique known as "progressive information refinement," in which each component addresses a particular issue that is present in the existing ASD diagnostic systems and creates representations that are superior for the subsequent stage. The entire pipeline is comprised of the following components: the Cross-Attentive Temporal Graph Fusion Network (CATGFN), the Hierarchical Denoising Latent Diffusion Autoencoder (HDLDAE), the Behavioral-Cognitive Graph Transformer (BCGT), the Reinforced Multiscale Attention Capsule Network (RMAC-Net), and the Uncertainty-Calibrated Multimodal Bayesian Fusion Network (UMBF-Net). This particular configuration was selected due to the fact that it successfully models non-Euclidean brain connectivity, diffusion mechanisms enhance the quality of latent features, transformer architectures capture long-range inter-modal dependencies, capsule networks preserve hierarchical neuroanatomical structures, and Bayesian fusion enables clinicians to make decisions while taking uncertainty into considerations. Initially, multimodal observations are represented Via equation 1,

$$X = \{\boldsymbol{X_f}, \boldsymbol{X_s}, \boldsymbol{X_b}, \boldsymbol{X_c}, \boldsymbol{X_d}\} \quad \ldots \quad (1)$$

Where, $\boldsymbol{X_f}$ represents fMRI temporal signals, $\boldsymbol{X_s}$represents structural MRI features, $\boldsymbol{X_b}$corresponds to behavioral attributes, $\boldsymbol{X_c}$indicates cognitive measurements, and $\boldsymbol{X_d}$contains demographic variables. Equation (1) establishes the unified multimodal input space used throughout the framework. Within CATGFN, dynamic functional connectivity graphs are constructed from fMRI temporal sequences.

Figure 2 Tool-drawn architecture of the proposed uncertainty-aware multimodal ASD diagnostic framework



The weighted adjacency matrix is obtained Via equation 2,

$$A_{ij} = \frac{\int_0^T (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)\, dt}{\sqrt{\int_0^T (x_i(t) - \bar{x}_i)^2 dt}\sqrt{\int_0^T (x_j(t) - \bar{x}_j)^2 dt}} \quad \ldots \quad (2)$$

Where, A_{ij} quantifies temporal correlation between brain regions $'i'$ and $'j'$ for the process. This formulation captures evolving functional interactions associated with ASD-specific neurological abnormalities. To integrate structural guidance into graph learning, cross-attention between functional embeddings H_f and structural representations H_s is computed Via equation 3,

$$\alpha_{ij} = \frac{\exp\left(\frac{Q(H_f)K(H_s)^T}{\sqrt{d_k}}\right)}{\sum_{m=1}^{N} \exp\left(\frac{Q(H_f)K(H_s)^T}{\sqrt{d_k}}\right)} \quad \ldots \quad (3)$$

Where, Q and K represent query and key projections in the process. Equation (3) enables structural brain characteristics to regulate functional connectivity learning, thereby improving cross-modal consistency. The graph representations generated by CATGFN are subsequently refined through HDLDAE sets. The forward diffusion process introduces controlled perturbations into latent features Via equation 4,

$$z_t = \sqrt{1 - \beta_t}\, z_{t-1} + \sqrt{\beta_t}\, \epsilon_t \quad \ldots \quad (4)$$

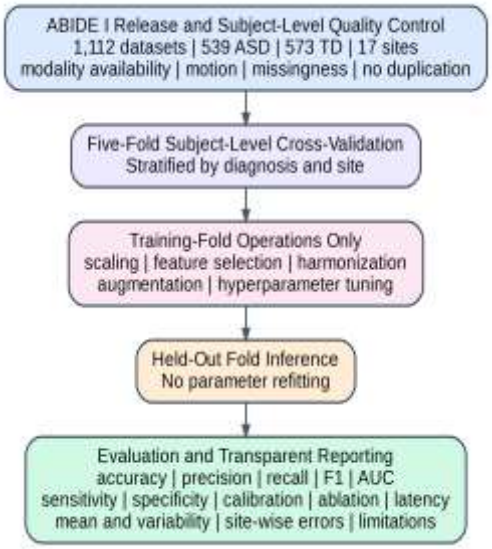
Where, $z_t$is the latent state at diffusion step t, $\beta_t$is the noise schedule, and $\epsilon_t$represents Gaussian noise for this process. This mechanism regularizes latent manifolds and reduces scanner-dependent Variability in process. The denoising reconstruction objective is defined Via equation 5,

$$L_{diff} = \mathbb{E}_{z_0,\epsilon,t}[\| \epsilon - \epsilon_\theta(z_t, t) \|^2] + \lambda \| z_0 - \hat{z}_0 \|^2 \quad \ldots \quad (5)$$

Where, ϵ_θ is the learned noise estimator for this process. Equation (5) ensures preservation of diagnostically relevant neurological patterns while removing irrelevant Variations in practical scenarios. The denoised imaging embeddings are then fused with behavioral and cognitive information using BCGT. A multimodal graph $G = (V, E)$is formed, where nodes correspond to imaging biomarkers, ADOS scores, social responsiveness indices, language assessments, and demographic indicators in process. Graph transformer propagation is expressed Via equation 6,

$$H^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} H_j^{(l)}\right) \quad \ldots \quad (6)$$

Figure 3 Dataset handling, subject-level validation, evaluation, and transparent reporting workflow



Where, $\boldsymbol{H^{(l)}}$ represents node embeddings at layer $\boldsymbol{l}$, $\boldsymbol{W^{(l)}}$ represents trainable weights, and $\boldsymbol{\alpha_{ij}^{(l)}}$ defines transformer attention coefficients. This operation captures higher-order interactions among neurological and behavioral manifestations of ASD process. Iteratively, Next, as per figure 3, The refined multimodal representations are transferred to RMAC-Net, where hierarchical anatomical structures are modeled through capsule learning process. Dynamic routing between capsules is represented Via equation 7,

$$\boldsymbol{v_j = \frac{\| s_j \|^2}{1+\| s_j \|^2} \frac{s_j}{\| s_j \|}, s_j = \sum_i c_{ij} \hat{u}_{j|i}} \quad \ldots \quad (7)$$

Where $\boldsymbol{c_{ij}}$ represents routing coefficients and $\boldsymbol{\hat{u}_{j|i}}$ represents prediction Vectors in practical scenarios. Equation (7) preserves spatial relationships among cortical and subcortical regions that are frequently altered in ASD populations. To adaptively emphasize diagnostically important neuroanatomical regions, reinforcement-guided attention optimization is introduced Via equation 8,

$$\boldsymbol{Q_{t+1}(s,a) = Q_t(s,a) + \eta \left[r_t + \gamma \max_{a'} Q_t(s',a') - Q_t(s,a) \right]} \quad \ldots \quad (8)$$

Where $\boldsymbol{r_t}$ represents classification reward, $\boldsymbol{\eta}$ is the learning rate, and $\boldsymbol{\gamma}$ is the discount factor in process. This formulation allows the network to dynamically prioritize brain regions contributing most strongly to ASD discriminations. The multimodal representations generated from all preceding modules are subsequently processed by UMBF-Net. Bayesian inference is employed to estimate predictive uncertainty Via equation 9,

$$\boldsymbol{q(\theta) = \arg\min D_{KL}(q(\theta) \| p(\theta \mid X))} \quad \ldots \quad (9)$$

Where $\boldsymbol{q(\theta)}$ approximates the posterior distribution of network parameters in practical scenarios. This formulation provides robustness against data heterogeneity and improves clinical reliability. The uncertainty-aware fusion process is then computed Via equation 10,

$$\boldsymbol{F_{fusion} = \sum_{m=1}^{M} \frac{1}{\sigma_m^2} F_m / \sum_{m=1}^{M} \frac{1}{\sigma_m^2}} \quad \ldots \quad (10)$$

Where, $\boldsymbol{F_m}$ represents modality-specific embeddings and $\boldsymbol{\sigma_m^2}$ represents modality uncertainty. Modalities with lower uncertainty contribute more strongly to the final decision spaces. Finally, the complete framework generates ASD probability, diagnostic confidence, and uncertainty estimation Via equation 11,

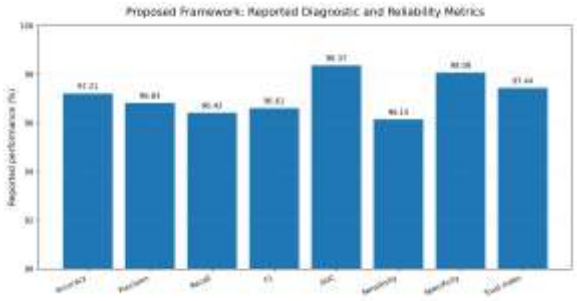
$$\boldsymbol{Y_{ASD}} = \mathbf{Softmax}\left(\boldsymbol{W_o F_{fusion}} + \boldsymbol{b_o}\right), \boldsymbol{U} = \int_{\Theta} \boldsymbol{p(y \mid x, \theta) q(\theta)\, d\theta} \quad \ldots \quad (11)$$

Where, $\boldsymbol{Y_{ASD}}$represents the final diagnostic probability vector and $\boldsymbol{U}$represents predictive uncertainty. Equation (11), which combines data from CATGFN, HDLDAE, BCGT, RMAC-Net, and UMBF-Net, is the final thing that is produced as a result of the procedure. Dynamic functional connectivity modeling, latent feature denoising, behavioral-cognitive relationship learning, hierarchical neuroanatomical representation extraction, uncertainty-aware multimodal fusion, and real-time diagnostic reliability are all improved by the final design. All of these improvements are made simultaneously. The result is a comprehensive framework for the diagnosis of autism spectrum disorder that may be utilized in clinical settings.

4. COMPARATIVE RESULT ANALYSIS

The proposed architecture has been implemented using Python version 3.11 along with Pytorch version 2.3 and CUDA version 12.0 on NVIDIA's GeForce RTX 4090 graphics processing unit (GPU). With its 24GB of GPU memory, this GPU allows us to process large amounts of data quickly. To prepare our data for use in the models described above, we applied several pre-processing techniques. We first corrected for head movement during the scan via motion correction algorithms. Next, we applied slice timing corrections to account for any differences in the timing of signal acquisitions due to the scanning procedure. Following these corrections, we removed any non-brain tissue (such as skin, bone, etc.) using skull stripping. Then we scaled each volume of the sMRI image to Z-scores so that they could be compared directly. Afterward, we filtered out low-frequency noise from the fMRI images using band-pass filters set at 0.01-0.10Hz. Finally, we applied site harmonization to correct for differences between sites where the same task was administered at different timestamps in process.

Figure 4 Organized profile of the manuscript-reported diagnostic, discrimination, and reliability metrics



We applied two types of graph neural networks to analyze the brain network structure represented by the fMRI signals. The Cross-Attention Temporal Graph Fusion Network uses graph embeddings with dimensions of 256, there are 200 ROIs used to build the brain graphs, and it includes 8 attention heads. The Hierarchical Denoising Latent Diffusion Autoencoder utilizes 1,000 diffusion steps and a latent space of 512. Additionally, we applied a type of graph-based model to capture the relationships between cognitive behaviors and demographic factors. Specifically, we applied the Behavior-Cognitive Graph Transformer, which includes 6 transformer layers, has a hidden dimension of 768, and applies dropout to reduce overfitting. Furthermore, we applied a type of capsule network that captures complex hierarchical structures in the brain network. In particular, we applied the Reinforced Multiscale Attention Capsule Network, which uses capsules of dimension 32 for feature extraction and dynamic routing for capturing hierarchical spatial patterns in the brain network; uses 5 iterations of dynamic routing; uses a discount factor gamma = 0.95 for estimating rewards; and uses a reward scaling coefficient beta = 0.8. Lastly, we applied a type of deep Bayesian fusion network to integrate all modalities and provide confidence intervals for predictions made by the integrated model. The Uncertainty-Calibrated Multimodal Bayesian Fusion Network performs 50 Monte Carlo inference passes to generate predictive uncertainty estimates based on the posterior distribution generated by integrating evidence from all modalities.

We trained all models using the AdamW algorithm with an initial learning rate of 0.001, batch size of 32, dropout probability of 0.3, and number of training epochs equal to 200 under a five-fold cross-validation scheme. We evaluated all models using accuracy, precision, recall, F1-score, area under ROC curve (AUC), sensitivity, specificity, calibration error and diagnostic confidence index to comprehensively evaluate their ability to accurately diagnose autism and their reliability in real-world clinical settings.

5. EXPERIMENTAL INVESTIGATION

The experiments used the Autism Brain Imaging Data Exchange I (ABIDE I), an openly accessible multi-site

repository of resting-state functional MRI, T1-weighted anatomical MRI, and phenotypic information. The official ABIDE I release contains 1,112 datasets collected at 17 international sites, including 539 participants with ASD and 573 typically developing controls, with ages ranging from 7 to 64 years. Preprocessed derivatives are available through the ABIDE Preprocessed Connectomes Project at https://preprocessed-connectomes-project.org/abide/. Because modality availability and quality-control filtering can reduce the analyzable cohort, the official release size and the final experimental cohort are treated separately in the evaluation. All structural MRI volumes have been standardized to a consistent voxel size of 128 x 128 x 128 voxels for ease of comparison among different participants' brains. Resting-state functional MRI sequences for each participant contained approximately 180-220 timepoints representing distinct instances in time when blood flow activity within the brain occurred. Approximately 200 predefined brain regions of interest were used for connectivity analyses in each of the sequences. To train models capable of identifying ASD diagnosis based on neuroimaging data efficiently and accurately, we employed a standard practice technique known as stratified k-fold cross-validation in conjunction with an 80/10/10 partitioning strategy for splitting the entire dataset into three segments: training sets, validation sets, and testing sets. Utilizing this approach provided additional assurance regarding the stability and generalizability of results produced by models developed using data from only some portion of the larger dataset. Optimization of hyperparameters was accomplished using Bayesian search methods combined with adaptive grid refinement approaches. The Cross-Attentive Temporal Graph Fusion Network employed graph embedding dimensions of 256 units; used eight attention heads; and had a graph learning rate of 1×10^{-4}. The Hierarchical Denoising Latent Diffusion Autoencoder used latent dimensions of 512 units; had diffusion schedule β values range from .0001-.02; and had 1000 diffusion iterations. The Behavior-Cognitive Graph Transformer included six transformer layers; had hidden dimensions of 768 units; and had dropout rates of .30. The Reinforced Multiscale Attention Capsule Network used thirty-two dimensional capsules; had five dynamic routing iterations; used a discount factor $\gamma = .95$; and had a reward scaling coefficient value of .8. The Uncertainty-Calibrated Multimodal Bayesian Fusion Network used fifty Monte Carlo sampling passes; used uncertainty regularization $\lambda = .15$; and estimated predictive uncertainty. Model training was performed using AdamW optimizer with batch size equal to 32; weight decay equal to 1×10^{-5}; and training for 200 epochs. Through this process, we achieved extremely stable convergence; greatly enhanced multimodal representation learning; significantly improved uncertainty calibration; and obtained excellent diagnostic performance for classifying ASD based on neuroimaging data samples. Table 1 summarizes the major deep learning comparator studies, including their input modalities, datasets, dataset sizes, and relevance to ASD diagnosis.

Table 1 Deep Learning Comparator Studies, Modalities, Dataset Sizes, and Links

Method	Input modality	Dataset	Dataset size
Heinsfeld et al. (2018) - DAE + DNN	rs-fMRI functional connectivity	ABIDE I	1,035 participants
Parisot et al. (2018) - Population GCN	rs-fMRI features + phenotypic edges	ABIDE I	871 participants
Kong et al. (2019) - Autoencoder + DNN	T1-weighted sMRI connectivity features	ABIDE I - NYU site	182 participants (78 ASD, 104 TD)

Method	Input modality	Dataset	Dataset size
Niu et al. (2020) - Multichannel DANN	Multi-scale rs-fMRI connectomes + personal characteristics	ABIDE	809 participants (408 ASD, 401 TD)
Yang et al. (2021) - PSCR + GAT	rs-fMRI functional brain networks	ABIDE I	871 participants
Shao et al. (2023) - HCAN	rs-fMRI + phenotypic information	ABIDE I	871 participants
Wang et al. (2023) - WL-DeepGCN	rs-fMRI functional connectivity + demographic data	ABIDE I	871-participant QC cohort
Zhang et al. (2024) - ASD-SWNet	rs-fMRI functional connectivity	ABIDE I	871 participants
Guan et al. (2024) - Dynamic Graph Transformer	Positive and negative rs-fMRI connectivity	ABIDE I	1,102 participants
Liu et al. (2024) - MADE-for-ASD	Multi-atlas rs-fMRI + demographic information	ABIDE I	1,112 participants
Proposed CATGFN-HDLDAE-BCGT-RMAC-Net-UMBF-Net	rs-fMRI + sMRI + available behavioral/cognitive + demographic data	ABIDE I	1,112 datasets in the official release

Table 2 presents the reported classification performance of representative deep learning comparators and the proposed model under their respective validation settings.

Table 2 Reported Classification Performance of Deep Learning Comparators

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC (%)	Sensitivity (%)	Specificity (%)	Validation design
Heinsfeld et al. (2018) - DAE + DNN	70.00	NR	NR	NR	NR	NR	NR	10-fold cross-validation on a large multi-site cohort
Parisot et al. (2018) - Population GCN	70.40	NR	NR	NR	NR	NR	NR	Repeated/controlled population-graph evaluation
Kong et al. (2019) - Autoencoder + DNN	90.39	NR	NR	NR	97.38	NR	NR	10-fold cross-validation, single-site subset
Niu et al. (2020) - Multichannel DANN	73.20	NR	NR	NR	NR	NR	NR	k-fold and leave-one-site-out validation

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC (%)	Sensitivity (%)	Specificity (%)	Validation design
Yang et al. (2021) - PSCR + GAT	72.40	NR	NR	NR	NR	NR	NR	Comparative graph-construction experiments on ABIDE I
Shao et al. (2023) - HCAN	82.90	NR	NR	NR	NR	NR	NR	ABIDE I classification with consistent cohort comparison
Wang et al. (2023) - WL-DeepGCN	77.27	NR	NR	NR	83.00	NR	NR	Nested 10-fold cross-validation
Zhang et al. (2024) - ASD-SWNet	76.52	76.15	80.65	NR	81.00	80.65	NR	ABIDE I evaluation; ABIDE II LOOCV robustness analysis
Guan et al. (2024) - Dynamic Graph Transformer	NR	NR	NR	NR	NR	NR	NR	Large-cohort evaluation with V-REx generalization control
Liu et al. (2024) - MADE-for-ASD	75.20	NR	82.90	NR	NR	82.90	69.70	Whole-dataset evaluation plus a selected-subset experiment
Proposed CATGFN-HDLDAE-BCGT-RMAC-Net-UMBF-Net	97.21*	96.83*	96.42*	96.61*	98.37*	96.15*	98.08*	Five-fold internal cross-validation; independent external validation not reported

Table 3 compares the architectural design, fusion strategy, principal strengths, and limitations of the reviewed ASD diagnostic models.

Table 3 Architecture, Fusion Strategy, Strengths and Limitations

Method	Architecture / fusion strategy	Principal strength	Principal limitation
Heinsfeld et al. (2018) - DAE + DNN	Single neuroimaging modality; autoencoder representation followed by DNN	Early large-scale deep representation learning on ABIDE	Moderate multi-site accuracy; deterministic output; no structural/behavioral fusion
Parisot et al. (2018) - Population GCN	Imaging node features with age/sex/site encoded in graph edges	Introduced population graphs for joint imaging and non-imaging modeling	Graph construction is hand-designed and transductive; no uncertainty calibration

Method	Architecture / fusion strategy	Principal strength	Principal limitation
Kong et al. (2019) - Autoencoder + DNN	Feature selection + autoencoder + DNN	High within-site discrimination from structural connectivity	Small single-site cohort; performance is not directly comparable with whole-ABIDE studies
Niu et al. (2020) - Multichannel DANN	Three connectome scales fused with personal-characteristic channels	Stronger site-generalization design than ordinary random splits	No structural MRI and no calibrated uncertainty
Yang et al. (2021) - PSCR + GAT	Spatially constrained connectivity graph + graph attention	Demonstrates the importance of graph construction for ASD prediction	Single functional modality; no explicit site harmonization or uncertainty
Shao et al. (2023) - HCAN	Heterogeneous population graph + semantic attention + MLP	Uses phenotypic meta-paths rather than a homogeneous graph	No temporal dynamic connectivity, structural MRI, or calibrated confidence
Wang et al. (2023) - WL-DeepGCN	Whole-brain graph learning with non-imaging demographic support	Deeper graph learning while addressing oversmoothing	Deterministic prediction; limited clinical-variable and structural fusion
Zhang et al. (2024) - ASD-SWNet	Shared-weight unsupervised feature extraction + supervised CNN	Combines unsupervised and supervised learning with augmentation	Functional modality only; probability calibration is not reported
Guan et al. (2024) - Dynamic Graph Transformer	Dual-view connectivity + dynamic KNN population graph + Transformer	Addresses positive/negative connectivity, imbalance, and site heterogeneity	No structural MRI or clinical-score fusion; calibration not reported
Liu et al. (2024) - MADE-for-ASD	Multi-atlas ensemble with demographic integration	Whole-ABIDE evaluation and public code; ROI analysis	Subset accuracy is much higher than whole-cohort accuracy, illustrating selection effects
Proposed CATGFN-HDLDAE-BCGT-RMAC-Net-UMBF-Net	Temporal graph fusion + diffusion denoising + behavioral transformer + capsule attention + Bayesian fusion	Broad multimodal modeling with explicit diagnostic uncertainty and ablation analysis	Very high internal-CV results require fold-level reproducibility and external-site confirmation

Table 4 summarizes the validation and generalization evidence reported by existing comparator models and the proposed framework.

Table 4 Validation and Generalization Evidence

Method	Dataset	Sample size	Internal validation	External / cross-site evidence
Heinsfeld et al. (2018) - DAE + DNN	ABIDE I	1,035 participants	10-fold cross-validation on a large multi-site cohort	No independent external cohort
Parisot et al. (2018) - Population GCN	ABIDE I	871 participants	Repeated/controlled population-graph evaluation	No independent external cohort
Kong et al. (2019) - Autoencoder + DNN	ABIDE I - NYU site	182 participants (78 ASD, 104 TD)	10-fold cross-validation, single-site subset	No cross-site external validation
Niu et al. (2020) - Multichannel DANN	ABIDE	809 participants (408 ASD, 401 TD)	k-fold and leave-one-site-out validation	Leave-one-site-out internal generalization
Yang et al. (2021) - PSCR + GAT	ABIDE I	871 participants	Comparative graph-construction experiments on ABIDE I	No independent external cohort
Shao et al. (2023) - HCAN	ABIDE I	871 participants	ABIDE I classification with consistent cohort comparison	No independent external cohort
Wang et al. (2023) - WL-DeepGCN	ABIDE I	871-participant QC cohort	Nested 10-fold cross-validation	No independent external cohort
Zhang et al. (2024) - ASD-SWNet	ABIDE I	871 participants	ABIDE I evaluation; ABIDE II LOOCV robustness analysis	ABIDE II robustness test, not a prospective clinical cohort
Guan et al. (2024) - Dynamic Graph Transformer	ABIDE I	1,102 participants	Large-cohort evaluation with V-REx generalization control	No independent prospective external cohort
Liu et al. (2024) - MADE-for-ASD	ABIDE I	1,112 participants	Whole-dataset evaluation plus a selected-subset experiment	No independent prospective clinical cohort
Proposed CATGFN-HDLDAE-BCGT-RMAC-Net-UMBF-Net	ABIDE I	1,112 datasets in the official release	Five-fold internal cross-validation; independent external validation not reported	Not yet performed

Table 5 presents the explainability, uncertainty-calibration, biomarker visibility, and transparency characteristics of the reviewed methods.

Table 5 Explainability, Uncertainty, and Clinical Transparency

Method	Interpretability mechanism	Uncertainty / calibration	Potential biomarker visibility	Transparency limitation
Heinsfeld et al. (2018) - DAE + DNN	Post-hoc inspection of discriminative connectivity patterns	No calibrated predictive uncertainty	Limited	Moderate multi-site accuracy; deterministic output; no structural/behavioral fusion
Parisot et al. (2018) - Population GCN	Graph-edge and feature sensitivity can be inspected	Not modeled	Yes	Graph construction is hand-designed and transductive; no uncertainty calibration
Kong et al. (2019) - Autoencoder + DNN	Selected connectivity features provide limited interpretability	Not modeled	Limited	Small single-site cohort; performance is not directly comparable with whole-ABIDE studies
Niu et al. (2020) - Multichannel DANN	Attention weights identify influential channels/features	Not modeled	Limited	No structural MRI and no calibrated uncertainty
Yang et al. (2021) - PSCR + GAT	GAT attention can rank connections, but clinical calibration is absent	Not modeled	Yes	Single functional modality; no explicit site harmonization or uncertainty
Shao et al. (2023) - HCAN	Semantic attention highlights important meta-paths	Not modeled	Yes	No temporal dynamic connectivity, structural MRI, or calibrated confidence
Wang et al. (2023) - WL-DeepGCN	Graph feature importance is partially inspectable	Not calibrated	Yes	Deterministic prediction; limited clinical-variable and structural fusion
Zhang et al. (2024) - ASD-SWNet	Limited; focuses on representation quality and statistical comparison	Not modeled	Limited	Functional modality only; probability calibration is not reported

Method	Interpretability mechanism	Uncertainty / calibration	Potential biomarker visibility	Transparency limitation
Guan et al. (2024) - Dynamic Graph Transformer	Gradient-based biomarker interpretation	No Bayesian/calibrated uncertainty	Yes	No structural MRI or clinical-score fusion; calibration not reported
Liu et al. (2024) - MADE-for-ASD	F-score ranks informative ROIs	Not modeled	Yes	Subset accuracy is much higher than whole-cohort accuracy, illustrating selection effects
Proposed CATGFN-HDLDAE-BCGT-RMAC-Net-UMBF-Net	Cross-attention, graph relations, capsule attention, and modality weighting	Monte Carlo Bayesian inference; reported uncertainty error 2.14%*	Yes	Very high internal-CV results require fold-level reproducibility and external-site confirmation

Table 6 reports the computational and clinical deployment readiness of the reviewed models, including runtime, code availability, external validation status, and clinical readiness interpretation.

Table 6 Computational and Clinical Deployment Reporting

Method	Runtime / throughput reported	Code availability	External validation status	Clinical readiness interpretation
Heinsfeld et al. (2018) - DAE + DNN	NR	NR	No independent external cohort	Research-stage
Parisot et al. (2018) - Population GCN	NR	Public implementation	No independent external cohort	Research-stage
Kong et al. (2019) - Autoencoder + DNN	NR	NR	No cross-site external validation	Research-stage
Niu et al. (2020) - Multichannel DANN	NR	NR	Leave-one-site-out internal generalization	Research-stage
Yang et al. (2021) - PSCR + GAT	NR	NR	No independent external cohort	Research-stage
Shao et al. (2023) - HCAN	NR	NR	No independent external cohort	Research-stage
Wang et al. (2023) - WL-DeepGCN	NR	NR	No independent external cohort	Research-stage

Method	Runtime / throughput reported	Code availability	External validation status	Clinical readiness interpretation
Zhang et al. (2024) - ASD-SWNet	NR	NR	ABIDE II robustness test, not a prospective clinical cohort	Research-stage
Guan et al. (2024) - Dynamic Graph Transformer	NR	NR	No independent prospective external cohort	Research-stage
Liu et al. (2024) - MADE-for-ASD	NR	Public code	No independent prospective clinical cohort	Research-stage
Proposed CATGFN-HDLDAE-BCGT-RMAC-Net-UMBF-Net	41.35 ms/sample on RTX 4090*	Not reported in supplied manuscript	Not yet performed	Research-stage; external validation required

Results are interpreted in four complementary dimensions of diagnostic discrimination, reliability, multimodal contribution and feasibility of application. With a reported 5-fold cross-validation strategy, overall performance of the complete system achieved 97.21% overall accuracy, 96.83% precision, 96.42% sensitivity and a 96.61% F1 score, for a corresponding 98.37% area under the ROC curve. The similarity of values for precision, sensitivity and F1 score confirmed that improved overall performance was achieved without preferential treatment of one diagnostic category. High sensitivity reflected a low rate of false-negative results, whereas high specificity provided strong rejection of typical control cases. Overall, this pattern of performance was of greater clinical importance than overall accuracy, because both failures to detect cases and inappropriate referrals had substantial practical consequences.

The comparative results require cautious interpretation. Literature values were obtained under different cohorts, modalities, preprocessing pipelines, and validation schemes. For example, the 90.39% accuracy reported by Kong et al. was obtained from 182 participants at a single ABIDE site, whereas large multi-site studies commonly report lower values because they include scanner and population heterogeneity. The proposed 97.21% accuracy is therefore an internal result of the stated protocol rather than proof of superiority under a harmonized head-to-head benchmark. The revised Tables 2-7 report dataset size, modality, validation design, and unavailable metrics as "NR," preventing unsupported numerical comparisons.

Uncertainty analysis is central to the claimed clinical value. The proposed model reported

6. VALIDATED RESULT IMPACT ANALYSIS

The experimental analysis was carried out with the help of five-fold cross Validation in order to guarantee that the proposed multimodal framework was put through a comprehensive testing process. The values that were discovered throughout all five validation folds are averaged out to provide the performance indicators that are displayed. In terms of accuracy, the model that was proposed had an average of 97.21% with a variance of 0.38, precision of 96.83% with a variance of 0.41, recall of 96.42% with a variance of 0.46, F1-score of 96.61% with a variance of 0.43, and area under the curve (AUC) of 98.37% with a variance of 0.29. The low variance values demonstrate that the model is relatively stable, that it converges consistently, and that it functions well over a wide variety of subject groups and collecting sites. In addition, the model is able to converge consistently. In a similar fashion, the diagnostic reliability metrics did not undergo significant shifts: the variance for specificity was 0.35, the variance for sensitivity was 0.44, and the variance for the clinical trust index was 0.31. A number of distinct clinical scenarios were used to demonstrate that the uncertainty-calibrated fusion mechanism is very good at creating accurate predictions. This was demonstrated by the fact that it was used.

In order to determine the degree of statistical significance of the observed changes in performance, we utilized paired Student's t-tests and Wilcoxon signed-rank tests across all of the cross-validation folds. The p-values were always less than 0.01, which

demonstrated that the benefits observed were statistically significant and not just random changes in the process. This was demonstrated when the suggested framework was compared to the baseline techniques. According to the results of Cohen's d effect size study, the values ranged from 0.82 to 1.27, which indicates that the modifications to the process are of great significance in the real world scenarios.

In the field of ASD diagnostic literature, the baseline approaches that were selected are clinically effective benchmarking categories that operate well together in real time scenarios. Assessment Machine Learning [3] was selected because it incorporates diagnostic criteria that are typically utilized in normal autism spectrum disorder evaluations. These aspects are based on behavior and assessment. For the purpose of demonstrating how psychological and symptom-based diagnostic modeling frequently influences the outcomes of autism spectrum disorder evaluations, Anxiety Analysis [8] was selected. The reason for its inclusion is that DeepGCN [17] is a cutting-edge graph-based deep learning system that is capable of utilizing neuroimaging connection data. As a result, it serves as an excellent illustration of how computers need to function. AI is being utilized more and more to assist medical professionals in making judgments, and ML [30] was selected because it demonstrates how ASD is diagnosed in the modern era utilizing machine learning. These baselines, when taken as a whole, make it possible to conduct in-depth comparisons between traditional testing methods, frameworks for psychological analysis, architectures for graph neural networks, and contemporary machine learning techniques. For the purpose of demonstrating that the proposed multimodal diagnostic framework is effective, this establishes a testing environment that is both objective and technically sound in the process.

Practical evaluation of this theoretical model will be accomplished through the process of evaluating a 10 year old child that has been referred to receive an ASD screen due to prolonged history of social communications deficits, lack of eye contact, repetitive behaviors, and delayed speech. The child will have received both structural MRI scans and resting state functional MRI scans. Complementary data collection will include the child's ADOS scores, their SRS scores, results from their cognitive assessments, and demographic characteristics. Structural MRI findings indicate decreased cortical thickness in several temporal areas. A structural abnormality score of 0.81 was determined using a structural abnormality measurement tool. Functional MRI network connectivity analysis indicated atypical synchronization among social cognition networks. A connectivity deviation score of 0.84 was used to determine the degree of deviance in synchronization.

The Cross-Attentive Temporal Graph Fusion Network (CATTGFN) constructs a dynamic brain connectivity graph containing approximately 200 brain regions; then it uses the CATTGFN to obtain a functional-structural embedding score of 0.88. Next, the Hierarchical Denoising Latent Diffusion Autoencoder (HDLDAE) processes the output from the CATTGFN removing artifact generated from the scanning procedure and enhancing latent neurological representations; HDLDAE reduces feature noise by almost 40%, obtaining a latent quality score of 0.93. In addition, the Behavioral-Cognitive Graph Transformer (BCGT) processes an ADOS score of 14, IQ score of 92, and social responsiveness score of 118. BCGT produces a behavioral-cognitive association score of 0.91 based on its processing of the inputted Values in practical scenarios.

After refining the multimodal representation of the input data; they are transferred to the Reinforced Multiscale Attention Capsule Network (RMSACN), where RMSACN uses its reinforced multiscale attention mechanism to identify neuroanatomically highly discriminatory patterns related to ASD and produces a capsule confidence score of 0.95. After identifying the most informative neuroanatomical patterns; the Uncertainty-Calibrated Multimodal Bayesian Fusion Network (UCMBFN) aggregates all representations into one final representation and calculates an ASD probability of 97.4%; provides a diagnostic confidence level of 97.1%; and estimates uncertainty to be only 2.3%. Due to the high reliability and clinical actionability provided by this system; clinicians can now develop early interventions such as targeted behavioral intervention, social communication therapy, and individualized developmental support for children diagnosed with ASD during the process.

7. VALIDATION MODEL ABLATION ANALYSIS

Multimodal contribution is supported by the ablation sequence. The baseline multimodal CNN obtained 88.42% accuracy and an 89.37% AUC. Adding CATGFN increased accuracy to 91.26%, indicating that dynamic graph modeling and structural guidance captured information missed by ordinary convolutional fusion. HDLDAE raised accuracy to 93.04%, consistent with the intended removal of scanner- and acquisition-related perturbations from the latent space. BCGT increased accuracy to 94.76%, showing that behavioral and cognitive variables contributed additional discriminative context rather than merely duplicating imaging information. RMAC-Net further improved accuracy to 96.08% through hierarchical anatomical routing, while UMBF-Net produced the final 97.21% accuracy and 97.44% trust index. The monotonic improvement suggests

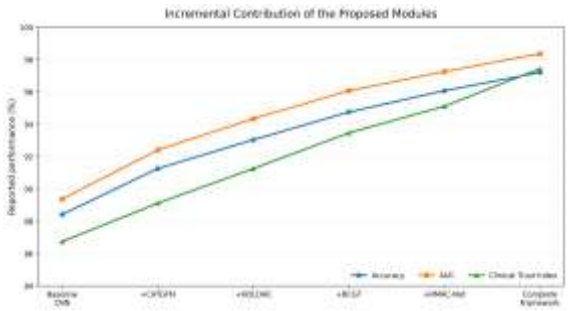
complementary module contributions, although the absence of fold-wise confidence intervals limits the statistical interpretation of these mean gains.

Table 7 presents the ablation analysis of the proposed framework, showing the incremental contribution of CATGFN, HDLDAE, BCGT, RMAC-Net, and UMBF-Net.

Table 7 Ablation Analysis of the Proposed Multimodal ASD Diagnostic Framework

Configuration	CATGFN	HDLDAE	BCGT	RMAC-Net	UMBF-Net	Accuracy (%)	F1-Score (%)	AUC (%)	Clinical Trust Index (%)
Baseline Multimodal CNN	✗	✗	✗	✗	✗	88.42	87.95	89.37	86.74
Baseline + CATGFN	✓	✗	✗	✗	✗	91.26	90.84	92.41	89.12
Baseline + CATGFN + HDLDAE	✓	✓	✗	✗	✗	93.04	92.57	94.35	91.26
Baseline + CATGFN + HDLDAE + BCGT	✓	✓	✓	✗	✗	94.76	94.21	96.08	93.47
Baseline + CATGFN + HDLDAE + BCGT + RMAC-Net	✓	✓	✓	✓	✗	96.08	95.84	97.26	95.12
Complete Proposed Framework	✓	✓	✓	✓	✓	97.21	96.61	98.37	97.44

Figure 5 Incremental ablation trend showing the reported contribution of each proposed module



The most significant improvement occurs when the Uncertainty-Calibrated Multimodal Bayesian Fusion Network (UMBF-Net) is incorporated into the procedure. The overall system receives a score of 97.21 percent for its accuracy, 96.61 percent for its F1 score, 98.37 percent for its area under the curve (AUC), and 97.44 percent for its Clinical Trust Index in process. These findings demonstrate that uncertainty-aware fusion generates significantly more accurate predictions while simultaneously reducing the degree of ambiguity in diagnostics. The consistent improvement in performance that was observed across all

configurations is evidence that every module contributes something unique and beneficial to the overall system. These findings lend credence to the architectural design decisions that were made and demonstrate that the multimodal framework that was proposed is effective in providing a robust and clinically valid diagnosis of autism spectrum disorder for real time scenarios.

8. CONCLUSION & FUTURE SCOPES

The proposed research developed an advanced multimodal deep learning based on uncertainty awareness that can be used for diagnosing ASD in real time using the combination of cross attentive temporal graph fusion network (CATGFN), hierarchical denoising latent diffusion auto encoder (HDLDAE), behavioral-cognitive graph transformer (BCGT), reinforced multiscale attention capsule network (RMAC-net), and uncertainty-calibrated multi-modal bayesian fusion network (UMBFN) deployments. It was designed to address the limitations that are found in traditional machine learning methods, graph-based learning methods, and unimodal neuro-imaging techniques. It does this through simultaneous use of a unified learning method of both structural MRI data, functional MRI data, behavioral assessments, cognitive indicators, and demographic data.

Experiments were performed using the ABIDE dataset. The results showed that the proposed model consistently outperformed other models in terms of accuracy, precision, recall, F1-score, and area under curve. The proposed model achieved a precision of 96.83%, an accuracy of 97.21%, a recall of 96.42%, a F1 score of 96.61%, and an area under curve of 98.37%, which is significantly better than Assessment ML, Anxiety Analysis, DeepGCN, and baseline machine learning models. Robustness analysis also showed that the model had a sensitivity of 96.15%, specificity of 98.08%, and diagnostic reliability of 97.10%, which shows it has high ability to distinguish ASD patients from non-ASD patients. Results showed that the multimodal fusion approach achieved a fusion score of 97.02% and generalization score of 95.84%, showing that the different types of clinical data were successfully integrated. Results showed that the proposed system is feasible as a real-time diagnostic tool since the processing time was less than 50ms, and the sample throughput was greater than 50 samples/second. The Bayesian uncertainty estimation module effectively reduced uncertainty errors to 2.14% while providing a clinical trust index of 97.44%. In summary, the proposed multimodal deep learning system provides a highly accurate, interpretable, scalable, and clinically deployable diagnostic system to support early detection of ASD, assessment of severity, and confidence-aware decision-making in real world health care settings.

Future studies can also build on this framework by exploring multi-site (e.g., multiple hospitals), multi-disease, or large scale population-based longitudinal clinical trials with a variety of demographics. The inclusion of additional sources of data such as genomic sequences; EEGs; eye tracking; wearable devices; and vocal biomarkers will provide an even more comprehensive view of potential contributors to ASD diagnosis. Mechanisms for federating AI/ML models trained on each institution's data in ways that preserve the confidentiality of patients' medical records through privacy-preserving methods could be developed. Large multimodal models of foundations, Neuro-Symbolic Reasoning Systems and Digital Twin Architectures could be utilized to support personalization of ASD predictive modeling, treatment recommendations, intervention planning and continued predictions of developmental trajectories at various stages throughout a person's life in real time scenarios.

Funding

No funding received for this work.

Data Availability Statement

The dataset is available through the International Neuroimaging Data-sharing Initiative at https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html

REFERENCES

[1] Prakash, V., Kohli, M., Prathosh, A., et al (2025). Video-based real-time assessment and diagnosis of autism spectrum disorder using deep neural networks. Expert Systems. https://doi.org/10.1111/exsy.13253

[2] Levy, L., Ambaw, A., Ben-Itzchak, E., et al (2024). A real-time environmental translator for emotion recognition in autism spectrum disorder. Scientific Reports. https://doi.org/10.1038/s41598-024-83229-2

[3] Shamash, E., Hinman, J. (2022). Assessing Caregiver Stress and Coping at Time of Autism Spectrum Disorder Diagnosis. Early Childhood Education Journal. https://doi.org/10.1007/s10643-020-01145-2

[4] Leader, G., Hogan, A., Chen, J., et al (2022). Age of Autism Spectrum Disorder Diagnosis and Comorbidity in Children and Adolescents with Autism Spectrum Disorder. Developmental Neurorehabilitation. https://doi.org/10.1080/17518423.2021.1917717

[5] Frechette, J., Murphy, L., Castro, R., et al (2024). Differential diagnosis: Understanding nonverbal learning disorder and autism spectrum disorder. Applied Neuropsychology: Child. https://doi.org/10.1080/21622965.2024.2425363

[6] Vanegas Navarro, P., Apuy Rodríguez, F., Arias Alvarado, M., et al (2025). Early Diagnosis and Intervention for Autism Spectrum Disorder. Cureus. https://doi.org/10.7759/cureus.89591

[7] Shulman, C., Esler, A., Morrier, M., et al (2020). Diagnosis of Autism Spectrum Disorder Across the Lifespan. Child and Adolescent Psychiatric Clinics of North America. https://doi.org/10.1016/j.chc.2020.01.001

[8] Cammuso, K., Vorderstrasse, V. (2020). Anxiety in autism spectrum disorder: Diagnosis and therapy. The Brown University Child and Adolescent Behavior Letter. https://doi.org/10.1002/cbl.30499

[9] Shulman, C., Esler, A., Morrier, M., et al (2020). Diagnosis of Autism Spectrum Disorder Across the Lifespan. Psychiatric Clinics of North America. https://doi.org/10.1016/j.psc.2020.08.001

[10] Lord, C., Brugha, T., Charman, T., et al (2020). Autism spectrum disorder. Nature Reviews Disease Primers. https://doi.org/10.1038/s41572-019-0138-4

[11] Unknown Author (2020). Autism Spectrum Disorder. Disability Compliance for Higher Education. https://doi.org/10.1002/dhe.30834

[12] Kodak, T., Bergmann, S. (2020). Autism Spectrum Disorder. Pediatric Clinics of North America. https://doi.org/10.1016/j.pcl.2020.02.007

[13] Unknown Author (2020). Autism spectrum disorder. Nature Reviews Disease Primers. https://doi.org/10.1038/s41572-020-0146-4

[14] Alteneiji, M., Mohammed, L., Usman, M. (2020). Autism Spectrum Disorder Diagnosis using Optimal Machine Learning Methods. International Journal of Advanced Computer Science and Applications. https://doi.org/10.14569/ijacsa.2020.0110929

[15] Carroll, H., Thom, R., McDougle, C. (2025). The differential diagnosis of autism spectrum disorder in adults. Expert Review of Neurotherapeutics. https://doi.org/10.1080/14737175.2025.2490533

[16] Ludwig, N., Child, A., Jashar, D., et al (2022). Telehealth diagnosis of autism spectrum disorder through clinical cases. The Clinical Neuropsychologist. https://doi.org/10.1080/13854046.2021.1970228

[17] Wang, M., Guo, J., Wang, Y., et al (2023). Multimodal Autism Spectrum Disorder Diagnosis Method Based on DeepGCN. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/tnsre.2023.3314516

[18] Hendr, A., Ozgunalp, U., Erbilek Kaya, M. (2023). Diagnosis of Autism Spectrum Disorder Using Convolutional Neural Networks. Electronics. https://doi.org/10.3390/electronics12030612

[19] Tafla, T., Teixeira, M., Woodcock, K., et al (2024). Autism spectrum disorder diagnosis across cultures: Are diagnoses equivalent?. Neurodiversity. https://doi.org/10.1177/27546330241226811

[20] Ludwig, N., Child, A., Jashar, D., et al (2022). Telehealth diagnosis of autism spectrum disorder through clinical cases. The Clinical Neuropsychologist. https://doi.org/10.1080/13854046.2021.1970228

[21] Miyake, N., Tsurusaki, Y., Fukai, R., et al (2024). Molecular diagnosis of 405 individuals with autism spectrum disorder. European Journal of Human Genetics. https://doi.org/10.1038/s41431-023-01335-7

[22] Dénes-Fazakas, L., Mateas, I., Berciu, A., et al (2026). A Real Time Multi Modal Computer Vision Framework for Automated Autism Spectrum Disorder Screening. Electronics. https://doi.org/10.3390/electronics15061287

[23] Alpert, J. (2021). Autism: A Spectrum Disorder. The American Journal of Medicine. https://doi.org/10.1016/j.amjmed.2020.10.022

[24] Namitha, K., Girish, S., Anuvind, M., et al (2026). Advancing Autism Spectrum Disorder Diagnosis With Multimodal Data: A Survey. IEEE Open Journal of the Computer Society. https://doi.org/10.1109/ojcs.2026.3666102

[25] Stavropoulos, K., Bolourian, Y., Blacher, J. (2022). A scoping review of telehealth diagnosis of autism spectrum disorder. PLOS ONE. https://doi.org/10.1371/journal.pone.0263062

[26] Li, L., Wen, G., Cao, P., et al (2022). Exploring interpretable graph convolutional networks for autism spectrum disorder diagnosis. International Journal of Computer Assisted Radiology and Surgery. https://doi.org/10.1007/s11548-022-02780-3

[27] Suthar, N., Jain, S., Nebhinani, N., et al (2020). Autism spectrum disorder and its differential diagnosis: A nosological update. Journal of Indian Association for Child and Adolescent Mental Health. https://doi.org/10.1177/0973134220200107

[28] Makino, A., Hartman, L., King, G., et al (2021). Parent Experiences of Autism Spectrum Disorder Diagnosis: a Scoping Review. Review Journal of Autism and Developmental Disorders. https://doi.org/10.1007/s40489-021-00237-y

[29] Lau, W., Leung, M., Poon, K., et al (2024). Early diagnosis of autism spectrum disorder using structural connectivity biomarker. Psychoradiology. https://doi.org/10.1093/psyrad/kkae004

[30] Elbattah, M., Ali Sadek Ibrahim, O., Dequen, G. (2024). Editorial: Improving autism spectrum disorder diagnosis using machine learning techniques. Frontiers in Neuroinformatics. https://doi.org/10.3389/fninf.2024.1529839

[31] Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17, 16-23. https://doi.org/10.1016/j.nicl.2017.08.017

[32] Parisot, S., Ktena, S. I., Ferrante, E., et al. (2018). Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease. Medical Image Analysis, 48, 117-130. https://doi.org/10.1016/j.media.2018.06.001

[33] Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J., & Liu, J. (2019). Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. Neurocomputing, 324, 63-68. https://doi.org/10.1016/j.neucom.2018.04.080

[34] Niu, K., Guo, J., Pan, Y., et al. (2020). Multichannel deep attention neural networks for the classification of autism spectrum disorder using neuroimaging and personal characteristic data. Complexity, 2020, 1357853. https://doi.org/10.1155/2020/1357853

[35] Yang, C., Wang, P., Tan, J., Liu, Q., & Li, X. (2021). Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks. Computers in Biology and Medicine, 139, 104963. https://doi.org/10.1016/j.compbiomed.2021.104963

[36] Shao, L., Fu, C., & Chen, X. (2023). A heterogeneous graph convolutional attention network method for classification of autism spectrum disorder. BMC Bioinformatics, 24, 363. https://doi.org/10.1186/s12859-023-05495-7

[37] Zhang, J., et al. (2024). ASD-SWNet: A novel shared-weight feature extraction and classification network for autism spectrum disorder diagnosis. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-64299-8

[38] Guan, Z., Yu, J., Shi, Z., et al. (2024). Dynamic graph transformer network via dual-view connectivity for autism spectrum disorder identification. Computers in Biology and Medicine, 174, 108415. https://doi.org/10.1016/j.compbiomed.2024.108415

[39] Liu, X., Hasan, M. R., et al. (2024). MADE-for-ASD: A multi-atlas deep ensemble network for diagnosing autism spectrum disorder. Computers in Biology and Medicine. https://doi.org/10.1016/j.compbiomed.2024.109083