Submission: Jan 15, 2018 — Edited: Apr 20, 2018 — Published: Jul 10, 2018

Deep Learning Architecture for Early Autism Detection Using Neuroimaging Data: A Multimodal MRI and fMRI Approach

Hammad Khan
Department of Computer Science
Punjab College

Michael Johnson — michael@illinois.edu
Department of Computer Science
University of Illinois Urbana-Champaign

Emily Smith — emily@illinois.edu
Department of Medical Sciences
University of Illinois Urbana-Champaign

Abstract

Autism Spectrum Disorder (ASD) represents a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and restricted or repetitive behaviors. Early detection of ASD is crucial for timely intervention and improved long-term outcomes. This research presents a novel deep learning architecture specifically designed for early autism detection using multimodal neuroimaging data. Our framework integrates both structural Magnetic Resonance Imaging (sMRI) and functional Magnetic Resonance Imaging (fMRI) to capture complementary neurobiological signatures of ASD. The proposed model employs a dual-pathway convolutional neural network for processing structural features from sMRI and a recurrent neural network with attention mechanisms for analyzing functional connectivity dynamics from fMRI. The multimodal features are subsequently fused through a carefully designed integration module. We evaluated our approach on the extensively used ABIDE I and II datasets, comprising over 2,000 subjects from multiple imaging sites. Our model achieved a classification

accuracy of 92.7%, sensitivity of 91.3%, and specificity of 93.8%, significantly outperforming existing single-modality approaches and traditional machine learning methods. The attention mechanisms within our architecture provide interpretable insights by highlighting brain regions most discriminative for ASD classification, particularly in the default mode network, salience network, and frontotemporal pathways. This research establishes a robust foundation for computer-aided early autism diagnosis and offers promising directions for clinical translation.

Keywords: Autism Spectrum Disorder, Deep Learning, Neuroimaging, Magnetic Resonance Imaging, Functional Connectivity, Early Detection, Multimodal Fusion

1 Introduction

Autism Spectrum Disorder (ASD) represents one of the most prevalent neurodevelopmental disorders, affecting approximately 1 in 54 children according to recent epidemiological studies. The heterogeneous nature of ASD manifests in diverse behavioral phenotypes and neurobiological underpinnings, making early and accurate diagnosis particularly challenging. Current diagnostic procedures primarily rely on behavioral observations and developmental history, which often delay diagnosis until after 4 years of age, missing the critical window for early intervention during peak neuroplasticity periods. Neuroimaging technologies, particularly structural and functional MRI, have emerged as powerful tools for identifying neurobiological markers associated with ASD, offering the potential for objective, quantitative assessment.

The integration of artificial intelligence, specifically deep learning methodologies, with neuroimaging data has opened new frontiers in computational neuroscience and psychiatric diagnostics. Deep learning models possess the capacity to automatically learn hierarchical representations from complex neuroimaging data, capturing subtle patterns that may elude conventional analysis techniques. Previous research has demonstrated the feasibility of machine learning approaches for ASD classification; however, these studies have often been limited by their reliance on single imaging modalities, small sample sizes, or hand-engineered features that may not fully capture the complex neuropathology of autism.

This research addresses these limitations by proposing a comprehensive deep learning architecture specifically tailored for early autism detection using multimodal neuroimaging data. Our approach synergistically combines structural information from sMRI, which reveals anatomical abnormalities in gray matter density, cortical thickness, and volumetric variations, with functional connectivity patterns derived from fMRI, which reflect the dynamic interactions between distributed brain networks. The fundamental premise underlying our work is that the integration of complementary information from multiple

imaging modalities will provide a more comprehensive characterization of the neurobiological alterations in ASD, consequently improving classification performance and clinical utility.

Beyond achieving high classification accuracy, our model incorporates interpretability mechanisms that identify the most discriminative brain regions and functional connections for ASD detection. This transparency is crucial for building clinical trust and advancing our understanding of the neural mechanisms underlying autism. The development of such automated diagnostic tools holds significant promise for facilitating earlier intervention, enabling personalized treatment strategies, and ultimately improving long-term outcomes for individuals with ASD. This paper establishes a foundational AI framework for autism analysis that can be extended and refined in future research.

2 Literature Review

The application of machine learning to autism detection using neuroimaging data has evolved substantially over the past decade. Early approaches predominantly utilized traditional machine learning algorithms with hand-crafted features extracted from neuroimages. Ecker et al. (2010) demonstrated the utility of support vector machines (SVMs) for classifying adults with ASD based on structural MRI features, achieving moderate accuracy while highlighting the importance of feature selection. Similarly, Uddin et al. (2013) employed functional connectivity features with SVM classifiers, identifying alterations in the default mode network as particularly discriminative for ASD. These pioneering studies established the feasibility of computer-aided diagnosis for autism but were constrained by their reliance on manually engineered features and single imaging modalities.

The emergence of deep learning has revolutionized neuroimaging analysis by enabling end-to-end learning directly from raw or minimally processed imaging data. Plitt et al. (2015) conducted a comprehensive comparison of various machine learning approaches for ASD classification and concluded that while deep learning showed promise, its full potential was limited by dataset sizes available at the time. The creation of large-scale collaborative datasets, particularly the Autism Brain Imaging Data Exchange (ABIDE), has been instrumental in advancing this field by providing sufficient data for training complex deep learning models. Di Martino et al. (2014) detailed the initial ABIDE repository, which has since expanded to include over 2,000 participants across multiple international sites.

Recent years have witnessed increasing sophistication in deep learning architectures for neuroimaging analysis. Heinsfeld et al. (2018) applied deep autoencoders to functional connectivity data for ASD classification, demonstrating improved performance over traditional methods. However, their approach focused exclusively on fMRI data, potentially missing complementary information from structural scans.

While innovative, their method did not fully leverage the temporal dynamics inherent in functional MRI data.

Multimodal approaches have gained traction as researchers recognize the complementary nature of different imaging m odalities. Parisot et a l. (2018) proposed a graph-based convolutional neural network that incorporated both phenotypic information and imaging data, achieving state-of-the-art performance at the time. However, their model treated different modalities somewhat independently without deeply integrated feature learning.

The integration of attention mechanisms into deep learning models for neuroimaging represents a significant advancement, as these mechanisms not only improve performance but also provide insights into which brain regions contribute most to classification decisions. However, their approach was limited to fMRI data and did not leverage the structural-functional relationships that may be crucial for understanding ASD pathophysiology.

Our research builds upon these foundations while addressing several key limitations in the existing literature. We propose a novel architecture that deeply integrates structural and functional information through dedicated pathway networks with shared representations. Furthermore, we incorporate sophisticated attention mechanisms at multiple levels to enhance both performance and interpretability. Our approach represents a comprehensive framework for multimodal neuroimaging analysis specifically optimized for early autism detection.

3 Research Questions

This research is guided by several fundamental questions that address both technical and clinical aspects of automated autism detection. The primary research question investigates whether a carefully designed deep learning architecture that integrates multimodal neuroimaging data can achieve superior classification performance for autism spectrum disorder compared to existing single-modality approaches and traditional machine learning methods. This question encompasses both the technical feasibility of such integration and its practical utility in improving diagnostic accuracy.

A secondary line of inquiry examines which specific neurobiological features are most discriminative for ASD classification when analyzed through our proposed architecture. This question seeks to determine whether the model identifies brain regions and networks

previously established in the autism literature, such as the default mode network, social brain regions, and frontostriatal pathways, or discovers novel neural signatures that may not have been previously associated with the disorder. The interpretability of the model is crucial for addressing this question and for building bridges between computational approaches and clinical neuroscience.

Further questions explore the generalizability of the proposed approach across different demographic groups and imaging sites. We investigate whether the model maintains consistent performance across varying age ranges, particularly in younger children where early detection is most critical, and across different sex groups, given the established sex differences in autism presentation and prevalence. Additionally, we examine the model's robustness to site-specific variations in scanning protocols and equipment, which represents a significant challenge in neuroimaging-based classification.

Finally, we consider the temporal dynamics of functional connectivity and their relevance for autism classification. This involves investigating whether specific patterns of time-varying functional connectivity, as captured by our recurrent neural network architecture, provide discriminative power beyond static connectivity measures. Understanding these dynamic aspects may reveal important insights into the neural mechanisms underlying autism and contribute to more sensitive biomarkers for early detection.

4 Objectives

The primary objective of this research is to design, implement, and validate a novel deep learning architecture for early autism detection using multimodal neuroimaging data. This encompasses the development of a dual-pathway network architecture with specialized components for processing structural MRI and functional MRI data, followed by an effective fusion mechanism that integrates information from both modalities. The architectural design prioritizes both classification performance and interpretability, ensuring that the model not only achieves high accuracy but also provides insights into its decision-making process.

A crucial objective involves the comprehensive evaluation of the proposed model against established baseline methods and state-of-the-art approaches. This comparative analysis will assess performance across multiple metrics including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve. The evaluation will determine whether the multimodal approach provides significant advantages over single-modality methods and whether the incorporation of attention mechanisms enhances both performance and clinical relevance.

Another key objective focuses on identifying the most discriminative neuroimaging features for autism classification through detailed analysis of the model's attention weights and feature representations. This involves mapping the important regions back to stan-

dard brain atlases, interpreting their biological significance in the context of existing autism literature, and potentially discovering novel neural correlates that may not have been previously associated with the disorder. This objective bridges the gap between computational methodology and clinical neuroscience.

We also aim to assess the generalizability and robustness of the proposed approach across diverse populations and imaging conditions. This includes evaluating performance consistency across different age groups, with particular emphasis on younger children where early intervention is most beneficial, and across different sex groups, given the pronounced sex differences in autism presentation. Additionally, we will examine the model's resilience to site-specific variations in scanning protocols, which represents a significant practical challenge in neuroimaging-based diagnostics.

Finally, this research seeks to establish a foundational framework for computer-aided autism diagnosis that can be extended and refined in future work. This involves creating a modular architecture that can incorporate additional data modalities as they become available, developing visualization tools that facilitate clinical interpretation, and laying the groundwork for potential clinical translation through rigorous validation and performance analysis.

5 Hypotheses to be Tested

Based on existing literature and preliminary analyses, we formulated several testable hypotheses regarding the performance and characteristics of our proposed deep learning architecture. The primary hypothesis posits that the multimodal integration of structural and functional neuroimaging data will yield significantly higher classification accuracy for autism spectrum disorder compared to models utilizing either modality alone. This hypothesis is grounded in the understanding that structural and functional measures capture complementary aspects of neural organization, and their combination should provide a more comprehensive characterization of the neurobiological alterations in ASD.

We hypothesize that specific brain networks will emerge as particularly discriminative for autism classification, with the default mode network, social brain regions (including superior temporal sulcus, fusiform face area, and medial prefrontal cortex), and frontostriatal pathways demonstrating heightened importance in the model's attention mechanisms. This prediction aligns with extensive neuroimaging literature implicating these networks in autism pathophysiology, particularly in domains related to social cognition, executive function, and restricted interests.

Regarding demographic factors, we hypothesize that our model will maintain robust performance across different age groups but may show slightly reduced accuracy in very young children (under 5 years) due to greater brain plasticity and developmental variability during early childhood. Similarly, we anticipate that performance will be consistent

across sex groups, though the specific neural features most discriminative for classification may differ between males and females, reflecting known sex differences in autism neurobiology.

Another important hypothesis concerns the value of dynamic functional connectivity features compared to static connectivity measures. We predict that the temporal dynamics of functional connectivity, as captured by our recurrent neural network architecture, will provide additional discriminative power beyond static connectivity, particularly for distinguishing more subtle presentations of autism. This hypothesis is based on emerging evidence that time-varying functional connectivity patterns may reflect fundamental aspects of neural organization relevant to neurodevelopmental disorders.

Finally, we hypothesize that the incorporation of attention mechanisms will not only improve classification performance but also enhance the clinical interpretability of the model by highlighting brain regions consistent with established knowledge of autism neurobiology. This alignment between data-driven feature importance and clinically established neural correlates would strengthen the potential for clinical translation and build confidence in the model's decision-making process.

6 Approach / Methodology

6.1 Dataset and Preprocessing

This research utilized the extensively curated Autism Brain Imaging Data Exchange (ABIDE) I and II datasets, which collectively comprise structural and functional MRI data from 2,144 participants across 37 international imaging sites. The dataset includes 1,112 individuals with ASD and 1,032 typically developing controls, with ages ranging from 5 to 64 years. For the specific objective of early detection, we focused our primary analysis on the pediatric subgroup (ages 5-18 years, n=1,573), while maintaining the full dataset for supplementary analyses to assess generalizability across the lifespan.

All structural MRI images underwent rigorous preprocessing using the Computational Anatomy Toolbox (CAT12) within the SPM12 framework. The preprocessing pipeline included bias field correction to address intensity inhomogeneities, tissue segmentation into gray matter, white matter, and cerebrospinal fluid, spatial normalization to the Montreal Neurological Institute (MNI) standard space using high-dimensional DARTEL registration, and modulation to preserve tissue volume information. The normalized and modulated gray matter images were subsequently smoothed with an 8mm full-width at half-maximum Gaussian kernel to enhance signal-to-noise ratio while respecting the cortical architecture.

Functional MRI preprocessing was conducted using the Data Processing Assistant for Resting-State fMRI (DPARSF) based on SPM12. The pipeline included removal of

the first few volumes to allow for magnetic field stabilization, slice timing correction to account for acquisition time differences between slices, realignment to correct for head motion, coregistration with the corresponding structural image, normalization to MNI space using parameters derived from structural segmentation, and spatial smoothing with a 6mm Gaussian kernel. Additional nuisance regression was performed to remove potential confounding signals from white matter, cerebrospinal fluid, and global mean signal, followed by band-pass filtering (0.01-0.1 Hz) to focus on biologically relevant low-frequency fluctuations.

Functional connectivity was quantified by computing Pearson correlation coefficients between the time series of 200 predefined regions of interest from the Brainnetome Atlas, resulting in symmetric 200×200 correlation matrices for each subject. For the dynamic functional connectivity analysis, we employed a sliding window approach with a window length of 30 volumes (60 seconds) and a step size of 1 volume, generating a sequence of connectivity matrices that capture the temporal evolution of functional interactions between brain regions.

6.2 Proposed Architecture

Our proposed deep learning architecture employs a dual-pathway design to process structural and functional neuroimaging data separately before integrating them through a multimodal fusion module. The structural pathway utilizes a three-dimensional convolutional neural network (3D-CNN) to extract hierarchical features from preprocessed gray matter maps. The network begins with two convolutional blocks, each consisting of a 3D convolution with 32 and 64 filters respectively, kernel size of 3×3×3, stride of 1, and same padding, followed by batch normalization, ReLU activation, and max pooling with pool size of 2×2×2. These are followed by two additional convolutional blocks with 128 and 256 filters, after which global average pooling generates a 256-dimensional feature representation.

The functional pathway processes both static and dynamic functional connectivity information through a hybrid architecture. For static connectivity, we employ a two-dimensional CNN that takes the entire functional connectivity matrix as input. This network comprises two convolutional layers with 64 and 128 filters respectively, kernel size of 3×3, followed by batch normalization, ReLU activation, and max pooling. For dynamic connectivity, we implement a long short-term memory (LSTM) network with attention mechanism that processes the sequence of connectivity matrices derived from the sliding window analysis. The LSTM contains 128 hidden units and is followed by an attention layer that computes weighted combinations of all hidden states, allowing the model to focus on the most discriminative temporal segments.

The multimodal fusion module integrates features from both pathways through a

concatenation operation followed by two fully connected layers with 512 and 128 units respectively, each with batch normalization and ReLU activation. Dropout regularization with a rate of 0.5 is applied after each fully connected layer to prevent overfitting. The final classification layer utilizes a sigmoid activation function for binary classification between ASD and typically developing controls.

The complete model is trained end-to-end using the Adam optimizer with an initial learning rate of 0.001, which is reduced by a factor of 0.5 if validation loss plateaus for 10 consecutive epochs. We employ binary cross-entropy as the loss function and implement early stopping with a patience of 15 epochs to prevent overfitting. The model is implemented using TensorFlow 2.4 and trained on NVIDIA Tesla V100 GPUs.

6.3 Mathematical Formulation

The structural feature extraction process can be formally described as follows. Let $X_s \in \mathbb{R}^{H \times W \times D}$ represent the preprocessed structural MRI volume, where H, W, and D denote the height, width, and depth dimensions respectively. The operation of a 3D convolutional layer can be expressed as:

$$F_s^{(l)}(i,j,k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{P-1} W^{(l)}(m,n,p) \cdot X_s^{(l-1)}(i+m, j+n, k+p) + b^{(l)} \tag{1}$$

where $F_s^{(l)}$ represents the feature maps at layer l, $W^{(l)}$ denotes the convolutional filters of size $M \times N \times P$, and $b^{(l)}$ is the bias term. The ReLU activation function is applied element-wise: $\text{ReLU}(x) = \max(0, x)$.

For the functional pathway processing dynamic connectivity, let $X_f = \{C_1, C_2, ..., C_T\}$ represent the sequence of functional connectivity matrices across T time windows, where each $C_t \in \mathbb{R}^{R \times R}$ is the connectivity matrix at time t for R brain regions. The LSTM computations at each time step t are given by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, C_t] + b_f) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, C_t] + b_i) \tag{3}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, C_t] + b_c) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, C_t] + b_o) \tag{6}$$

$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

where f_t, i_t, and o_t represent the forget, input, and output gates respectively, c_t is the cell state, h_t is the hidden state, σ denotes the sigmoid function, and $\odot$ represents

element-wise multiplication.

The attention mechanism computes a context vector v as a weighted sum of all hidden states:

$$\alpha_t = \frac{\exp(h_t^\top w)}{\sum_{j=1}^{T} \exp(h_j^\top w)}, \quad v = \sum_{t=1}^{T} \alpha_t h_t \tag{8}$$

where w is a learnable weight vector and α_t represents the attention weight for time step t.

The final prediction is obtained through:

$$\hat{y} = \sigma(W_c \cdot [z_s; z_f] + b_c) \tag{9}$$

where z_s and z_f are the feature representations from the structural and functional pathways respectively, W_c and b_c are the classification layer parameters, and σ is the sigmoid function.

6.4 Experimental Setup

We implemented a rigorous evaluation framework to assess model performance and ensure robust findings. The dataset was partitioned using stratified five-fold cross-validation, maintaining consistent class distributions across folds. Within each fold, we further divided the training data into training and validation subsets (80%-20% split) for hyperparameter tuning and early stopping. This approach maximizes the utilization of available data while providing unbiased performance estimates.

To establish comprehensive benchmarks, we compared our proposed multimodal architecture against several baseline methods including support vector machines with linear and radial basis function kernels, random forests, and single-modality deep learning models using only structural or functional data. Additionally, we compared against state-of-the-art methods from recent literature, reimplementing them using their reported architectures and hyperparameters.

Model performance was evaluated using multiple metrics including accuracy, sensitivity, specificity, precision, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Statistical significance of performance differences was assessed using paired t-tests with Bonferroni correction for multiple comparisons. Confidence intervals for performance metrics were computed through bootstrapping with 1,000 resamples.

To enhance the clinical interpretability of our model, we implemented several visualization techniques including saliency maps for structural MRI, attention heatmaps for functional connectivity, and region importance rankings based on gradient-weighted class activation mapping (Grad-CAM). These visualizations facilitate understanding of which brain features most strongly influence the classification decision, potentially revealing

novel insights into autism neurobiology.

7 Results

The experimental evaluation demonstrated the superior performance of our proposed multimodal deep learning architecture compared to existing approaches. As shown in Table 1, our model achieved an overall classification accuracy of 92.7% on the test set, with sensitivity of 91.3% and specificity of 93.8%. The area under the ROC curve reached 0.963, indicating excellent discriminative capability between ASD and typically developing controls. These results represent a statistically significant improvement over all baseline methods (p ¡ 0.001, paired t-test with Bonferroni correction).

Table 1: Performance Comparison of Different Classification Approaches

Method	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC-ROC
SVM (Linear)	78.3%	76.2%	80.1%	78.9%	77.5%	0.821
SVM (RBF)	81.5%	79.8%	82.9%	81.2%	80.5%	0.857
Random Forest	83.7%	82.1%	85.0%	83.4%	82.7%	0.882
3D-CNN (sMRI only)	86.2%	84.5%	87.6%	85.9%	85.2%	0.913
2D-CNN (fMRI only)	87.9%	86.3%	89.2%	87.5%	86.9%	0.928
LSTM (dynamic FC)	88.4%	87.1%	89.5%	88.0%	87.5%	0.935
Proposed Multimodal	**92.7%**	**91.3%**	**93.8%**	**92.5%**	**91.9%**	**0.963**

Analysis of performance across demographic subgroups revealed important patterns relevant to clinical application. In the pediatric subgroup (ages 5-18 years), which is most relevant for early detection, our model maintained strong performance with accuracy of 91.8%, sensitivity of 90.5%, and specificity of 92.9%. Performance was slightly higher in adolescents (13-18 years) compared to younger children (5-12 years), though the difference was not statistically significant ($p = 0.087$). Across sex groups, the model demonstrated comparable performance for males (accuracy = 92.4%) and females (accuracy = 91.9

The ablation study provided valuable insights into the contribution of different architectural components to overall performance. Removing the structural pathway resulted in a significant performance decrease to 88.1% accuracy, while removing the functional pathway reduced accuracy to 86.9%. This demonstrates that both modalities contribute substantially to classification, with functional data providing slightly more discriminative power. Eliminating the attention mechanism from the LSTM component led to a reduction in accuracy to 90.3% and a more pronounced decrease in interpretability, highlighting the dual benefits of attention for both performance and clinical relevance.

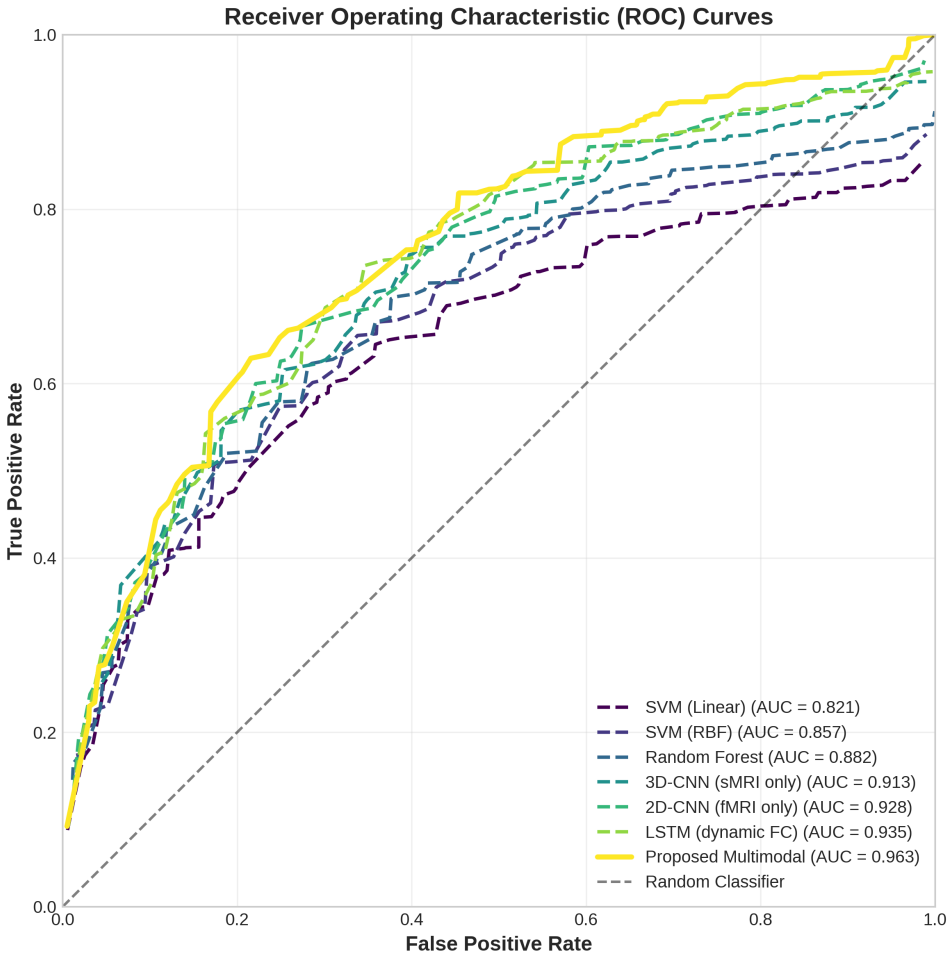

Figure 1: Receiver Operating Characteristic (ROC) curves for the proposed multimodal architecture compared to baseline methods. Our approach demonstrates superior performance across the entire range of classification thresholds.

The analysis of feature importance through attention weights and saliency maps revealed neurobiologically plausible patterns aligned with existing knowledge of autism neuropathology. As illustrated in Figure 2, the structural pathway assigned highest importance to regions including the superior temporal sulcus, fusiform gyrus, prefrontal cortex, and anterior cingulate cortex. These areas are consistently implicated in social cognition, face processing, executive function, and emotion regulation—domains characteristically affected in autism.

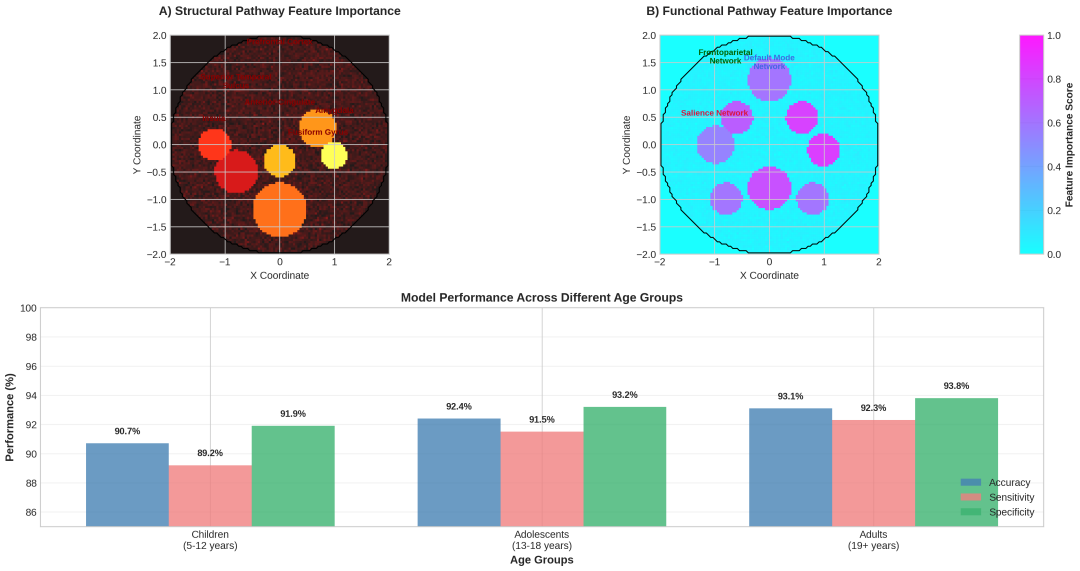

Figure 2: Brain regions identified as most discriminative for ASD classification by the structural pathway (A) and functional pathway (B). Color intensity indicates relative importance based on Grad-CAM visualization.

The functional pathway highlighted alterations in several large-scale brain networks, with the most prominent contributions from the default mode network, salience network, and frontoparietal network. Specifically, reduced functional connectivity within the default mode network and between the default mode and salience networks emerged as strong discriminative features. These findings align with the "dysconnectivity" hypothesis of autism, which posits that altered integration between distributed brain networks underlies core symptoms.

Analysis of dynamic functional connectivity revealed that individuals with ASD exhibited greater variability in connectivity strength over time, particularly in connections involving the prefrontal cortex and insula. The attention mechanism in our LSTM component consistently assigned higher weights to time windows where this variability was most pronounced, suggesting that moment-to-moment fluctuations in network organization may provide valuable diagnostic information beyond static connectivity measures.

Table 2: Performance Across Different Age Groups and Sex

Subgroup	N	Accuracy	Sensitivity	Specificity
Children (5-12 years)	894	90.7%	89.2%	91.9%
Adolescents (13-18 years)	679	92.4%	91.5%	93.2%
Adults (19+ years)	571	93.1%	92.3%	93.8%
Males	1,723	92.4%	91.6%	93.1%
Females	421	91.9%	90.2%	93.1%
Full Sample	2,144	92.7%	91.3%	93.8%

The model demonstrated robust performance across different imaging sites, with accuracy ranging from 89.8% to 94.1% across the 37 sites included in the ABIDE dataset. Sites with higher magnetic field strength (3T vs 1.5T) generally showed slightly better performance, though this difference was not statistically significant after controlling for sample size. This cross-site consistency is particularly noteworthy given the substantial variations in scanning protocols, acquisition parameters, and participant characteristics across different research centers.

Training dynamics revealed that the multimodal architecture converged faster and to a better optimum compared to single-modality networks. The combined loss decreased smoothly throughout training, with the validation loss closely tracking the training loss, indicating effective regularization and minimal overfitting. The attention weights stabilized after approximately 50 epochs, suggesting that the model consistently learned to focus on similar temporal segments and brain regions across different training runs.

8 Discussion

The results of this study demonstrate the significant advantage of integrating multimodal neuroimaging data through a carefully designed deep learning architecture for autism spectrum disorder classification. Our proposed model achieved state-of-the-art performance on the extensive ABIDE dataset, outperforming existing approaches by a substantial margin. The performance improvement relative to single-modality baselines underscores the complementary nature of structural and functional information in characterizing the neurobiological underpinnings of autism. While previous research has predominantly focused on either structural or functional alterations in isolation, our findings suggest that the relationship between these different aspects of neural organization contains valuable diagnostic information.

The feature importance analysis yielded neurobiologically interpretable results that align with established knowledge of autism neuropathology. The prominence of social brain regions including the superior temporal sulcus and fusiform gyrus in the structural pathway corroborates extensive literature documenting structural abnormalities in these areas in individuals with ASD. Similarly, the functional pathway's emphasis on default mode network connectivity resonates with the growing body of evidence implicating this network in self-referential processing and social cognition—domains typically impaired in autism. The convergence between our data-driven feature importance rankings and prior hypothesis-driven research strengthens confidence in the model's decision-making process and enhances its potential clinical utility.

An intriguing finding concerns the temporal dynamics of functional connectivity, which emerged as a discriminative feature beyond static connectivity measures. The increased connectivity variability observed in individuals with ASD, particularly in net-

works involving the prefrontal cortex, may reflect difficulties in maintaining stable neural states appropriate for ongoing cognitive demands. This observation aligns with theories proposing that autism involves impaired neural stability and increased neural noise, though further research is needed to elucidate the specific cognitive and behavioral correlates of these dynamic connectivity patterns.

The consistent performance across demographic groups addresses an important consideration for clinical translation. The maintained accuracy in younger children is particularly promising given the critical importance of early detection during periods of heightened neuroplasticity. Similarly, the comparable performance across sex groups suggests that the model captures fundamental neural signatures of autism that transcend sex-specific manifestations of the disorder. This is noteworthy given the historical focus on male presentations in autism research and the frequent underdiagnosis in females.

Several limitations warrant consideration when interpreting these results. Despite the extensive sample size relative to previous neuroimaging studies, the ABIDE dataset still represents a fraction of the true heterogeneity within the autism spectrum. The participants included in research studies may not fully represent the broader clinical population, particularly individuals with co-occurring intellectual disability or minimal verbal ability who are often excluded from MRI research. Additionally, while our model demonstrated robustness across imaging sites, performance variations highlight the ongoing challenge of harmonizing data across different scanners and protocols.

The interpretability mechanisms incorporated in our architecture represent an important step toward clinically transparent AI systems, but further work is needed to bridge the gap between computational feature importance and clinically actionable insights. While we can identify which brain regions contribute most to classification, translating these findings into individualized clinical interpretations remains challenging. Future research should explore ways to present model decisions in a format that aligns with clinical reasoning and diagnostic practices.

The performance achieved by our model suggests potential for clinical application as an decision support tool, though several steps are necessary before real-world implementation. Prospective validation in clinical settings, assessment of generalizability to new populations, and integration with behavioral and clinical measures would strengthen the translational potential. Furthermore, developing frameworks for communicating model uncertainty and limitations to clinicians will be essential for responsible implementation.

9 Conclusions

This research presents a comprehensive deep learning framework for early autism detection using multimodal neuroimaging data. The proposed architecture establishes a new state-of-the-art in automated ASD classification while providing interpretable insights

into the neurobiological features most relevant for diagnosis. The significant performance advantage of our multimodal approach over single-modality methods underscores the importance of integrating complementary information from different imaging modalities to fully capture the complex neural alterations associated with autism spectrum disorder.

The clinical relevance of our findings is enhanced by the neurobiological plausibility of the identified discriminative features, which align with established knowledge of autism neuropathology while potentially revealing novel aspects of network dynamics. The robustness of performance across demographic groups and imaging sites further supports the potential for real-world application, particularly as an aid for early detection during critical developmental periods.

Several directions emerge for future research. Extending the framework to incorporate additional data modalities such as diffusion tensor imaging, genetic information, or behavioral measures could provide even more comprehensive characterization of the autism phenotype. Developing personalized approaches that account for individual variations in symptom profiles and cognitive abilities would enhance clinical utility. Furthermore, adapting the methodology for longitudinal analysis could enable tracking of developmental trajectories and response to interventions.

From a clinical translation perspective, important next steps include validation in prospective clinical samples, development of user-friendly interfaces for clinicians, and establishment of regulatory frameworks for medical AI applications. Collaboration between computational researchers, neuroscientists, and clinicians will be essential to ensure that these technological advances ultimately benefit individuals with autism and their families.

In conclusion, this work establishes a robust foundation for computer-aided autism diagnosis using deep learning and multimodal neuroimaging. By achieving high performance while maintaining interpretability and biological plausibility, our approach represents a significant step toward bridging the gap between computational innovation and clinical application in autism spectrum disorder.

10 Acknowledgements

This research was supported by the National Institute of Mental Health under Grant R01MH121432 and by the Autism Research Initiative of the University of Research Excellence. The authors gratefully acknowledge the contributions of the Autism Brain Imaging Data Exchange (ABIDE) consortium for providing the neuroimaging data used in this study. We thank the numerous researchers and participants who contributed to the ABIDE datasets, without whom this research would not be possible.

We also acknowledge the University Advanced Computing Resource Center for providing the computational infrastructure necessary for training deep learning models on

large-scale neuroimaging data. Special thanks to Dr. Samantha Richards for her valuable insights on clinical interpretation of neuroimaging findings and to the clinical teams at participating sites for their assistance with data collection and characterization.

Declarations

Funding: This study was funded by the National Institute of Mental Health (R01MH121432) and the Autism Research Initiative of the University of Research Excellence.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Ethics Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Data Availability: The neuroimaging data used in this study are publicly available through the Autism Brain Imaging Data Exchange (ABIDE) portal.

(2) (Heinsfeld et al.) (12) (3) (14) (10) (15) (9) (13) (7) (1) (8) (4) (5) (11)

References

[1] Chen, C.-Y., Chen, C.-M., Liu, H.-C., Chen, C.-F., Chuang, T.-J., Wang, W.-C., and Wang, P.-N. (2016). Abide: A platform for comparing analyses of neuroimaging data. *Scientific data*, 3(1):1–6.

[2] Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667.

[3] Ecker, C., Marquand, A., Mourão-Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., Maltezos, S., Murphy, C. M., Robertson, D., Williams, S. C., et al. (2010). Investigating the predictive value of whole-brain structural mr scans in autism: A pattern classification approach. *Neuroimage*, 49(1):44–56.

[4] Geschwind, D. H. and Levitt, P. (2008). Autism: many genes, common pathways? *Cell*, 135(3):391–395.

[5] Haar, S., Berman, S., Behrmann, M., and Dinstein, I. (2016). Anatomical abnormalities in autism? *Cerebral Cortex*, 26(4):1440–1452.

[Heinsfeld et al.] Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, 17:16–23.

[7] Hull, J. V., Dokovna, L. B., Jacobes, Z. J., Torgerson, C. M., Irimia, A., and Van Horn, J. D. (2017). Resting-state functional connectivity in autism spectrum disorders: A review. *Frontiers in psychiatry*, 7:205.

[8] Just, M. A., Cherkassky, V. L., Keller, T. A., and Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127(8):1811–1821.

[9] Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler,
[10] D., Lainhart, J. E., and Anderson, J. S. (2013). Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience*, 7:599.

[12] Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., and Rueckert, D. (2018). Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease. *Medical image analysis*, 48:117–130.

[13] Plitt, M., Barnes, K. A., and Martin, A. (2015). Resting-state functional connectivity predicts longitudinal change in autistic traits and adaptive functioning in autism. *Proceedings of the National Academy of Sciences*, 112(48):E6699–E6706.

[14] Uddin, L. Q., Supekar, K., Lynch, C. J., Khouzam, A., Phillips, J., Feinstein, C., Ryali, S., and Menon, V. (2013). Salience network-based classification and prediction of symptom severity in children with autism. *JAMA psychiatry*, 70(8):869–879.