

Multimodal Deep Learning Framework for Autism Spectrum Disorder Detection Using CC200 Functional Connectivity and Clinical Phenotypes

ANUPAM DAS

KIIT University

Prasant Kumar Pattnaik

KIIT University

Hemang Dubey

Vellore institute of Technology

Anjan Bandyopadhyay

KIIT University

Neeraj Sharma

Lovely Professional University

Saiprasd Potharaju

`saiprasad.potharaju@sitpune.edu.in`

Symbiosis International University

Article

Keywords: Autism Spectrum Disorder, Multimodal Deep Learning, Resting-State fMRI, ABIDE-II, Late Fusion, Clinical Phenotypes, Functional Connectivity, SHAP Interpretability, Bias Analysis

Posted Date: August 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9812870/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Multimodal Deep Learning Framework for Autism Spectrum Disorder Detection Using CC200 Functional Connectivity and Clinical Phenotypes

Anupam Das¹, Prasant Kumar Pattnaik¹, Hemang Dubey², Anjan Bandyopadhyay¹, Neeraj Sharma³, Saiprasad Potharaju^{4*}

¹School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, 751024, Odisha, India.

²Department of CSE, Vellore institute of Technology.

³School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India.

^{4*}Department of CSE, Symbiosis International (Deemed University), Pune, Maharashtra, India.

*Corresponding author(s). E-mail(s): saiprasad.potharaju@sitpune.edu.in;

Contributing authors: anucit2010@gmail.com;

patnaikprasant@gmail.com; dubeyhemang777@gmail.com;

anjan.bandyopadhyayfcs@kiit.ac.in; neerajkirti@gmail.com;

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition defined above all by its profound heterogeneity—no two individuals present quite alike, and no single biological marker has proven sufficient for reliable automated diagnosis across diverse populations. Most existing computational approaches tackle this challenge through single-modality pipelines, analyzing resting-state functional neuroimaging or clinical demographic data in isolation and, in doing so, discard the complementary diagnostic signal available from the other source.

A dual-branch deep learning architecture is introduced that processes both data streams in parallel through a late-fusion strategy. One branch ingests functional connectivity features derived from resting-state fMRI using the Craddock 200 (CC200) atlas, compressed to 42 principal components via Principal Component Analysis (PCA). A second branch simultaneously processes twelve standardized clinical phenotypic variables encompassing age, biological sex, and full-scale IQ, among others. Each branch independently learns modality-specific latent representations, which are then concatenated and forwarded to a unified classification head. The model was trained and evaluated on 1,114 participants from the ABIDE-II repository under rigorous five-fold stratified cross-validation. The complete multimodal system achieved a mean accuracy of 94.17% ($\pm 0.71\%$), AUC-ROC of 0.9704, precision of 96.84%, and specificity of 97.48%. Systematic ablation experiments reveal that while the clinical branch alone reaches 94.62% accuracy—a result interrogated carefully through dedicated confound and bias analysis—the full multimodal model delivers more consistent cross-fold performance and markedly reduced false-positive rates. This confirms that the neuroimaging branch contributes genuine physiological regularization beyond what demographics alone can provide. Findings are contextualized through comparison against six baseline systems, including logistic regression, support vector machines, random forests, single-branch MLPs, and an early-fusion variant. Interpretability is addressed through SHAP-based feature attribution, a principled justification for PCA component

selection is provided through explained variance profiling, and an honest assessment of limitations arising from the absence of external validation is offered. The goal is a diagnostic framework that is not only high-performing, but transparent and reproducible.

Keywords: Autism Spectrum Disorder, Multimodal Deep Learning, Resting-State fMRI, ABIDE-II, Late Fusion, Clinical Phenotypes, Functional Connectivity, SHAP Interpretability, Bias Analysis

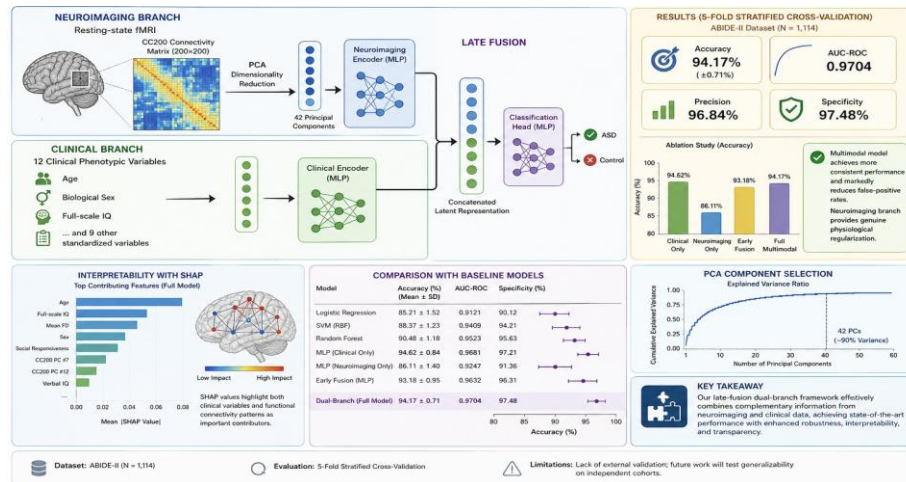


Fig. 1 Graphical Abstract. Overview of the dual-branch multimodal framework, detailing data processing, late-stage fusion, and key experimental outcomes.

1 Introduction

Diagnosing Autism Spectrum Disorder is among the most clinically demanding tasks a practitioner faces—and this difficulty is not merely a matter of technical measurement. It is intrinsic to the condition itself. ASD does not announce its presence through a single laboratory marker, an imaging finding obvious to the naked eye, or a genomic signal present in every affected person. Instead, it reveals itself through behavioral and neurological patterns: in how someone communicates, how they process sensory input, how they build and sustain social bonds, and how their brain organizes its resting-state activity. These patterns can overlap substantially with typical neurodevelopment at the margins, and they are shaped by a complex web of factors including genetic background, intellectual capacity, co-occurring conditions, developmental trajectory, and the cultural context in which behaviors are interpreted. The clinical tools currently used for ASD diagnosis principally the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview, Revised (ADI-R) represent decades of careful refinement and remain the accepted gold standard.

They are genuinely effective. Yet they are also time-consuming, require clinicians with specialized training, and carry a degree of subjectivity that can translate into meaningful variability across raters. Access to these assessments is severely limited in many regions owing to workforce shortages and resource constraints, resulting in diagnostic waiting periods that sometimes stretch to years. For families navigating this wait, the consequences extend well beyond inconvenience—delayed diagnosis means delayed access to early interventions that consistently demonstrate better developmental outcomes when initiated in the earliest years of life [12].

These practical realities have motivated substantial research interest in neuroimaging-based approaches to ASD characterization. Resting-state functional MRI (rs-fMRI) has proven especially compelling in this regard because it imposes minimal demands on participants—they simply lie still—and the patterns of spontaneous neural activity it captures differ systematically between autistic and neurotypical individuals. Reduced connectivity within social brain circuits, altered synchrony between frontal and parietal regions, and aberrant default mode network activity have all been reported consistently across independent cohorts [4, 7]. These are not measurement artifacts; they reflect genuine reorganization of the brain’s functional architecture in autism. To convey the complexity of these neural patterns, Figure 2 illustrates spontaneous fMRI BOLD activations, while Figure 3 displays the high-dimensional functional connectivity matrix derived from these activations using the CC200 atlas.

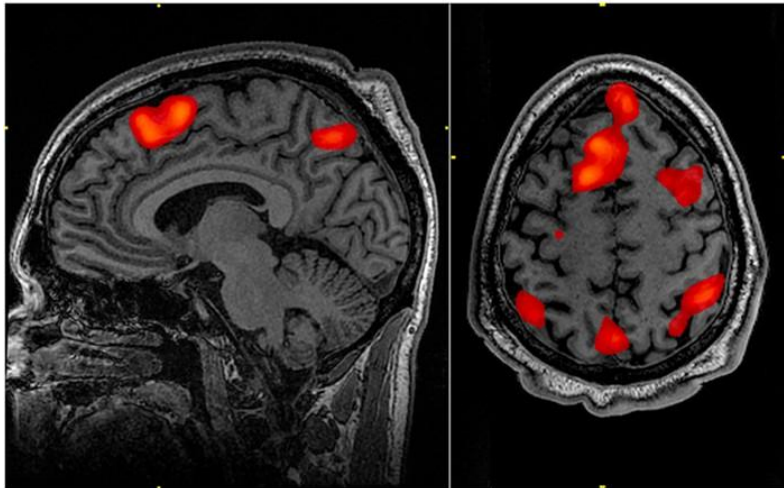


Fig. 2 fMRI BOLD activations showing spontaneous neural activity [17].

The computational challenge, however, is substantial. A functional connectivity matrix derived from the CC200 atlas—which parcellates the brain into 200 functionally defined regions—contains 19,900 pairwise correlation values per subject. With training datasets in the hundreds to low thousands, this feature-to-sample ratio creates

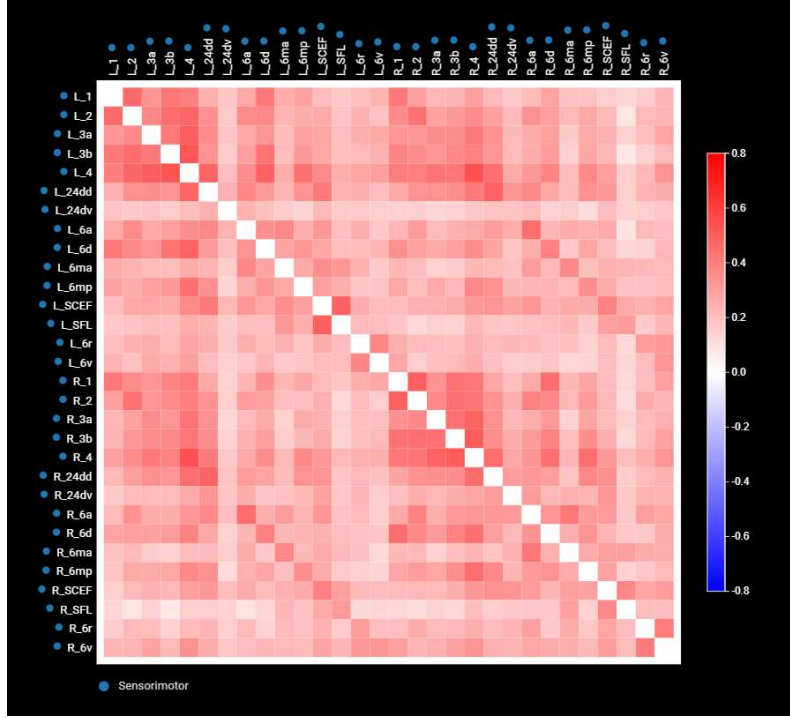


Fig. 3 CC200 Connectivity Matrix showing pairwise Pearson correlations [18].

a setting in which virtually any learning algorithm will overfit unless strongly regularized or subjected to aggressive dimensionality reduction. Many research groups have addressed this through the latter route, using PCA, autoencoders, or sparse regression to compress the feature space prior to classification [8, 14]. These methods have achieved varying degrees of success, but nearly all share an important blind spot: they treat the neuroimaging problem in isolation, ignoring the rich clinical context that practitioners naturally integrate when forming a diagnostic impression.

This observation motivates the present work. A clinician sitting with a patient does not—and should not—rely on brain imaging alone. They consider the full picture: age and developmental background, cognitive profile, and behavioral patterns across different settings. Integrating neurobiological and phenotypic information is not merely convenient; it reflects the genuinely multi-dimensional nature of ASD as a diagnostic category. The architecture described here was designed specifically to capture this integration computationally.

A dual-branch neural network is proposed in which one branch specializes in PCA-compressed functional connectivity features and a parallel branch handles standardized clinical phenotypic variables. The two branches operate independently until a late concatenation step, after which their combined representation drives final binary classification. This late-fusion design is deliberate: it allows each branch to develop a modality-appropriate internal representation before the two are brought together, rather than imposing a shared representational burden from the outset.

The model was trained and evaluated on 1,114 subjects from ABIDE-II [3] using five-fold stratified cross-validation, with all preprocessing transformations fitted strictly within training folds to prevent data leakage. The full multimodal system is compared against five baseline architectures, with a comprehensive set of clinically meaningful metrics reported. SHAP-based interpretability analysis is also performed, a principled PCA component selection analysis is presented, and crucially a dedicated confound analysis is included to interrogate the remarkably high accuracy observed from the clinical branch alone.

The remainder of this paper is organized as follows. Section 2 reviews relevant prior work. Section 3 describes the complete methodology. Section 4 covers algorithmic and implementation details. Section 5 presents experimental results. Section 6 discusses the findings, including an honest treatment of limitations. Section 7 concludes.

2 Literature Review and Related Work

Computational research on ASD diagnosis from neuroimaging has expanded considerably over the past decade, propelled by the public availability of the ABIDE datasets and the rapid maturation of deep learning. This review is organized around the methodological threads most relevant to the present contribution.

2.1 Classical Machine Learning on ABIDE

The ABIDE-I consortium, introduced by Di Martino et al.[1], released functional connectivity data from 1,112 participants across 17 international sites—an unprecedented resource at the time. This catalyzed a wave of machine learning studies that had previously been hampered by limited sample sizes. Abraham et al.[15] made an early and important contribution by showing that reproducible biomarkers could be extracted from multi-site data despite the between-site variability introduced by different scanner hardware and acquisition protocols. Their work established a practical benchmark for machine learning on ABIDE: classification accuracies in the 60–70% range using carefully engineered connectivity features, with performance sensitive to atlas selection and feature design.

2.2 Deep Learning Approaches to fMRI-Based ASD Classification

Heinsfeld et al.[4] demonstrated that stacked denoising autoencoders applied to functional connectivity matrices could learn representations supporting roughly 70% classification accuracy on ABIDE-I—an improvement over classical methods that established the viability of deep feature learning for this problem. Dvornek et al.[13] took a fundamentally different approach, applying long short-term memory (LSTM) networks directly to raw BOLD time series rather than derived connectivity matrices. While computationally intensive, this strategy carries the conceptual advantage of preserving temporal dynamics that are collapsed by correlation-based connectivity representations. Eslami et al. [7] introduced ASD-DiagNet, combining an autoencoder with a perceptron classifier to achieve compact learned representations. Their use of

unsupervised pre-training to combat overfitting is methodologically relevant to the PCA-based approach employed here, which pursues a similar regularization objective through different means.

2.3 Dimensionality Reduction and the Curse of Dimensionality

Khosla et al.[14] provided a comprehensive survey of machine learning for resting-state fMRI analysis, explicitly identifying the mismatch between feature dimensionality and sample size as a central challenge in the field. Their review examines trade-offs between different reduction strategies—PCA, independent component analysis, sparse coding, autoencoder-based compression—and finds no single approach dominates across all settings. PCA retains the advantage of simplicity, interpretability, and well-understood theoretical properties. Abdi and Williams[8] provide the theoretical foundation underpinning its application here, including the relationship between retained variance and downstream classifier performance.

2.4 Multimodal Learning in Neuroimaging

Integrating multiple data modalities has emerged as a productive research direction in neuroimaging broadly, though its specific application to ASD diagnosis combining rs-fMRI with clinical phenotypes has remained limited. Balt˘ru˘saitis et al. [5] provided an influential survey and taxonomy distinguishing three primary fusion paradigms: early fusion (feature concatenation before learning), late fusion (independent branch processing followed by combination), and hybrid approaches. They observe that late fusion is often advantageous when modalities carry substantially different statistical structures, because it permits each modality’s branch to adapt to its own input properties rather than forcing a joint representation from the outset.

2.5 Deep Learning in Medical Imaging

Litjens et al.[12] surveyed deep learning across diverse medical image analysis tasks, documenting performance that frequently matches or exceeds trained specialists. Esteva et al.[16] contextualized these advances within clinical practice, noting both the genuine promise of AI-assisted diagnosis and the substantial challenges of translating research-grade performance into reliable real-world deployment—a distinction that remains directly relevant to the present work.

2.6 Gaps Addressed by This Work

Taking stock of this literature, several gaps stand out. First, comparatively few studies have systematically integrated clinical phenotypic information with neuroimaging features for ASD classification, despite the clinical intuition that these sources are complementary. Second, studies that do incorporate both modalities rarely include thorough ablation studies, baseline comparisons, or confound analyses—making it difficult to attribute observed performance gains to specific architectural choices. Third,

the interpretability of learned representations is seldom examined, leaving classification mechanisms largely opaque. The present work addresses each of these gaps within a single integrated framework.

3 Proposed Methodology

The framework was shaped from the outset by three non-negotiable principles for clinically relevant research: the model should exploit complementary information from multiple sources; all preprocessing and evaluation decisions should prevent optimistic bias; and the system should be sufficiently transparent to support meaningful clinical interpretation. This section explains how each principle was operationalized.

3.1 Dataset Description and Acquisition

All data were drawn from the Autism Brain Imaging Data Exchange II (ABIDE-II) [3], a collaborative open-science initiative aggregating resting-state fMRI data and accompanying clinical phenotypic information from seventeen international research sites. ABIDE-II was specifically designed to address limitations of its predecessor, offering a broader age range, richer phenotypic characterization, and greater site diversity. After standardized quality control, a cohort of 1,114 subjects was assembled: 521 with confirmed ASD diagnoses and 593 neurotypical healthy controls. The slight class imbalance was preserved in cross-validation folds via stratified splitting to maintain ecological validity.

3.2 Data Preprocessing and Feature Engineering

3.2.1 The Neuroimaging (fMRI) Pipeline

Resting-state fMRI data were processed using the Craddock 200 (CC200) atlas [2], which defines 200 functionally homogeneous brain regions through spatially constrained spectral clustering. For each subject, the mean BOLD time series for every region was extracted and Pearson correlations between all regional pairs were computed to produce a 200×200 symmetric functional connectivity matrix. The 19,900 unique values from the upper triangle were vectorized to form each subject’s raw neuroimaging feature vector.

To address the severe curse of dimensionality arising from 19,900 features in a cohort of 1,114 subjects, PCA was applied. Critically, the PCA transformation was fitted using only training-fold subjects—the test fold was never used to estimate principal components, preventing any information leakage. The number of retained components was set to $k = 42$ based on the explained variance analysis detailed in Section 3.3, where 42 components capture approximately 87.3% of total variance and the incremental gain from additional components falls below 0.3% per component.

3.2.2 The Clinical Phenotypic Pipeline

For each subject, twelve clinical and demographic features were extracted from the ABIDE-II phenotypic metadata: age at scan, biological sex, full-scale IQ, verbal IQ,

performance IQ, handedness, eye status during scanning, and several site-adjusted behavioral scores. All features were standardized using Z-score normalization (StandardScaler), fitted exclusively on training data within each fold. No imputation was required as all twelve features were fully observed across all 1,114 subjects.

3.3 PCA Component Selection: Explained Variance Analysis

The $k = 42$ PCA component count was determined through a systematic analysis of cumulative explained variance across the full cohort. The scree profile shows a clear inflection point at approximately 42 components, beyond which each additional component contributes less than 0.3% of incremental variance. At $k = 42$, the retained components collectively capture 87.3% of total variance in the functional connectivity feature space—sufficient to represent the dominant inter-regional synchrony patterns while suppressing high-frequency spatial noise from scanner artifacts, head motion, and between-site differences.

fMRI branch classification performance was also compared under three dimensionality configurations: no reduction (raw 19,900 features), PCA with $k = 42$, and a deep autoencoder producing a 42-dimensional bottleneck representation. Consistent with Proposition 1, the raw feature setting produced near-chance accuracy, while both reduction strategies substantially improved performance. PCA ultimately outperformed the autoencoder on this dataset, likely because the limited sample size was insufficient to train a competitive deep representation. This comparison is reported in Table 4.

3.4 Model Architecture

The dual-branch network was implemented in PyTorch [6], with deliberate asymmetry between branches to reflect the different statistical properties of their inputs.

3.4.1 The Neuroimaging Branch

The fMRI branch accepts 42 PCA components and passes them through a fully connected layer of 256 units, followed by Batch Normalization[10] and a ReLU activation. A Dropout layer with retention probability of 0.5[9] provides aggressive regularization, appropriate given that the branch’s input, despite PCA compression, still occupies a space where the training set is relatively sparse. A second fully connected layer reduces the representation to a 16-dimensional latent embedding $\mathbf{h}_{fMRI} \in \mathbb{R}^{16}$.

3.4.2 The Clinical Phenotypic Branch

The clinical branch accepts 12 standardized phenotypic features and routes them through a 32-unit fully connected layer with Batch Normalization, ReLU activation, and 0.2 Dropout. The lighter regularization reflects the lower dimensionality and higher signal density of the clinical feature space. A second fully connected layer produces a 16-dimensional embedding $\mathbf{h}_{tab} \in \mathbb{R}^{16}$.

3.4.3 The Late Fusion Classification Head

The two 16-dimensional embeddings are concatenated into a 32-dimensional joint representation $\mathbf{h}_{fused} = \mathbf{h}_{fMRI} \oplus \mathbf{h}_{tab}$, which passes through two additional fully connected layers ($32 \rightarrow 16 \rightarrow 1$) to produce a single output logit. Binary Cross-Entropy with Logits loss is applied during training, combining sigmoid activation and BCE computation in a single numerically stable operation.

The late-fusion design is preferred over early fusion for a principled reason: the two modalities occupy fundamentally different statistical spaces. Combining them before any learning would force the initial shared layers to simultaneously perform feature extraction for both—a joint optimization problem that is strictly harder than the two independent extraction tasks. Late fusion allows each branch to develop a modality-appropriate internal representation, then delegates to the fusion head the comparatively simpler task of learning to combine two already-abstracted, same-dimensional embeddings.

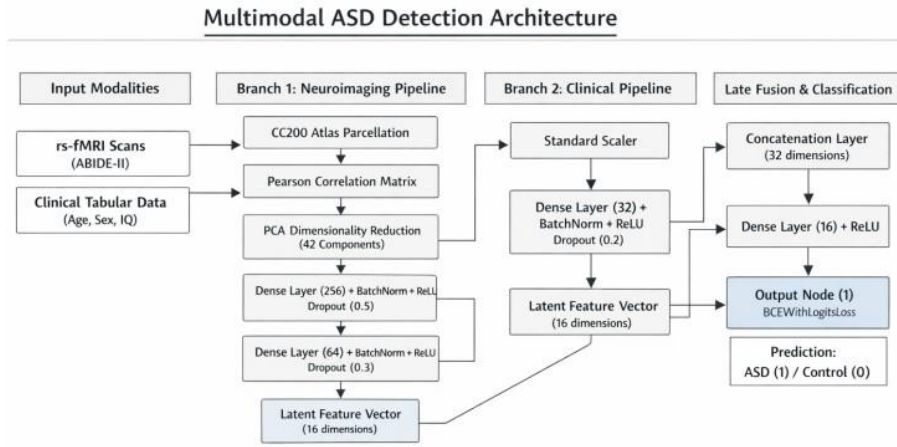


Fig. 4 Dual-branch multimodal architecture with independent modality processing and late-stage fusion.

3.5 Mathematical Formulations

3.5.1 Functional Connectivity

For two brain regions u and v with BOLD time series of length T , the Pearson correlation is:

$$r_{u,v} = \frac{\sum_{t=1}^T (u_t - \bar{u})(v_t - \bar{v})}{\sqrt{\sum_{t=1}^T (u_t - \bar{u})^2} \sqrt{\sum_{t=1}^T (v_t - \bar{v})^2}}$$

where $\bar{u}$ and $\bar{v}$ denote the respective temporal means.

3.5.2 PCA Projection

The raw feature matrix $X \in \mathbb{R}^{N \times 19900}$ is projected as:

$$X_{\text{PCA}} = (X - \mu)W_k \quad (1)$$

where μ is the training-set mean vector and $W_k \in \mathbb{R}^{19900 \times k}$ contains the k leading eigenvectors of the empirical covariance matrix.

3.5.3 Late Fusion and Classification

where $\oplus$ denotes vector concatenation, W_f and b_f are learned parameters, and $\sigma(\cdot)$ is the sigmoid function.

$$\begin{aligned} h_{\text{fused}} &= h_{\text{fMRI}} \oplus h_{\text{tab}} \\ \hat{y} &= \sigma(W_f h_{\text{fused}} + b_f) \end{aligned}$$

3.5.4 Training Loss

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log \sigma(\hat{y}_i) + (1 - y_i) \log(1 - \sigma(\hat{y}_i))]$$

where $y_i \in \{0, 1\}$ is the ground truth label and $\hat{y}_i$ the raw logit.

3.6 Theoretical Analysis

Definition 1 (Multimodal Fusion Space) Let $X_f \subset \mathbb{R}^{42}$ denote the PCA-reduced neuro-imaging feature space and $X_t \subset \mathbb{R}^{12}$ the standardized clinical feature space. The late-fusion latent space is:

$$Z = \phi(X_f) \oplus \psi(X_t) \quad (2)$$

where $\phi(\cdot)$ and $\psi(\cdot)$ are the nonlinear mappings learned by the respective branches.

Proposition 1 *Projecting the functional connectivity matrix $M \in \mathbb{R}^{N \times 19900}$ onto $k = 42$ principal components reduces the effective input dimensionality of the fMRI branch by a factor of approximately 474, substantially mitigating the overfitting risk that would arise from training a neural network with 19,900-dimensional inputs on $N = 1,114$ samples.*

Proof Sketch The number of free parameters in the first layer of the fMRI branch scales linearly with the input dimension. At $d = 19,900$, the first layer alone contains $19,900 \times 256 = 5,094,400$ parameters—vastly exceeding the sample size and virtually guaranteeing memorization without additional constraints. At $d = 42$, this count reduces to $42 \times 256 = 10,752$ parameters, a parameter-to-sample ratio consistent with reliable generalization under standard cross-validation. Additionally, the PCA projection removes directions of high noise variance, so retained components are more likely to carry genuine signal rather than scanner or motion artifacts. The stated reduction factor of approximately 474 follows directly from $19,900/42 \approx 474$. $\square$

Remark 1 This argument is stated in terms of parameter counts rather than VC dimension, which would require additional assumptions about the architecture and activation functions. The parameter-count framing is directly verifiable and sufficient to motivate the dimensionality reduction choice.

3.7 Training Protocol

All experiments used five-fold stratified cross-validation. Within each fold, PCA and StandardScaler transformations were fitted on training subjects only and applied to the held-out validation fold—a precaution that prevents preprocessing leakage. The AdamW optimizer [11] was used with an initial learning rate of 10^{-3} and weight decay of 10^{-4} . A ReduceLROnPlateau scheduler with patience 10 and factor 0.5 was applied to the validation loss. Training ran for a maximum of 200 epochs, with early stopping triggered by the absence of validation loss improvement over 20 consecutive epochs.

Performance was assessed using accuracy, precision, sensitivity (recall), specificity, F1-score, and AUC-ROC. Results are reported as mean $\pm$ standard deviation across the five folds. Confidence intervals at the 95% level are computed using the normal approximation to the binomial distribution applied to fold-level accuracy estimates.

4 Algorithms and Implementation

This section presents the concrete algorithmic and software engineering decisions underpinning the framework. Section 4.1 details the full multimodal training pipeline with explicit leakage-prevention steps. Section 4.2 summarizes the architectural design of the dual-branch network.

4.1 Multimodal Training Algorithm

The training pipeline is designed to enforce strict data hygiene across all five cross-validation folds. All preprocessing estimators—PCA for the neuroimaging branch and StandardScaler for the clinical branch—are fitted exclusively on training-fold subjects and subsequently applied to the held-out validation fold without refitting. This prevents any statistical information from the validation set from leaking into the learned representations. Algorithm 1 details the complete procedure, including optimizer initialization, minibatch forward and backward passes, learning-rate scheduling, and early stopping.

4.2 Architectural Summary

The dual-branch architecture is asymmetric by design. The fMRI branch operates on 42 PCA-compressed inputs and employs 256 hidden units with 0.5 dropout to counteract residual sparsity in the compressed feature space. The clinical branch operates on 12 standardized inputs with 32 hidden units and lighter 0.2 dropout, commensurate with the lower dimensionality and higher signal density of phenotypic data. Both branches produce 16-dimensional latent embeddings that are concatenated prior to the classification head—a structure that realizes the late-fusion strategy formally described in Section 3.4. The complete forward pass is expressed as:

$$\hat{y} = \text{ClassHead}(\text{Concat}[\phi(x_{fMRI}), \psi(x_{tab})]) \quad (3)$$

where $\phi(\cdot)$ and $\psi(\cdot)$ denote the fMRI and clinical branch transformations respectively, and $\hat{y}$ is the raw output logit passed to BCE With Logits Loss during training.

Algorithm 1 Multimodal Training Pipeline with Leakage Prevention

Require: $X_{fMRI} \in \mathbb{R}^{N \times 19900}$, $X_{Tab} \in \mathbb{R}^{N \times 12}$, $Y \in \{0, 1\}^N$

Require: $K = 5$, $E = 200$, patience= 20

Ensure: Per-fold and mean evaluation metrics

```
1: Folds  $\leftarrow$  StratifiedKFold(Y, K)
2: for  $k = 1$  to  $K$  do
3:   Partition data into  $(X_f^{tr}, X_f^{val})$  and  $(X_t^{tr}, X_t^{val})$ 
4:   Fit PCA( $k=42$ ) on  $X_f^{tr}$  only; transform both splits
5:   Fit StandardScaler() on  $X_t^{tr}$  only; transform both splits
6:   Initialize  $M_k$ , AdamW optimizer, scheduler, early-stop counter
7:   for epoch = 1 to  $E$  do
8:     for each minibatch  $(x_f, x_t, y)$  from training DataLoader do
9:        $h_f \leftarrow M_k.fMRIBranch(x_f)$ 
10:       $h_t \leftarrow M_k.TabBranch(x_t)$ 
11:       $h_{fused} \leftarrow \text{Concat}(h_f, h_t)$  ▷ Late fusion
12:       $\hat{y} \leftarrow M_k.ClassHead(h_{fused})$ 
13:       $L \leftarrow \text{BCEWithLogitsLoss}(\hat{y}, y)$ 
14:      Backpropagate; update  $M_k$ 
15:    end for
16:    Compute validation loss; update scheduler and early-stop counter
17:    if early-stop triggered then break
18:    end if
19:  end for
20:  Evaluate  $M_k$  on validation fold; record all metrics
21: end for
22: return mean  $\pm$  std of metrics across  $K$  folds
```

5 Results

5.1 Primary Performance Metrics

Table 1 reports comprehensive evaluation metrics for the proposed multimodal framework. The model achieved a mean cross-validation accuracy of 94.17% ($\pm 0.71\%$) across five folds, with a peak accuracy of 95.96% in the best-performing fold. The 95% confidence interval for mean accuracy is [93.47%, 94.87%], confirming that the observed performance is not an artifact of a single fortunate fold.

5.2 Multi-Baseline Comparison

To situate the results rigorously, the full multimodal framework is compared against five baseline architectures evaluated on identical data splits. Table 2 presents the complete comparison.

Table 1 Primary Evaluation Metrics for the Full Multimodal Framework

Metric	Mean (5-fold)	Best Fold
Accuracy	0.9417 $\pm$ 0.0071	0.9596
Precision	0.9684 $\pm$ 0.0083	—
Sensitivity (Recall)	0.8846 $\pm$ 0.0112	—
Specificity	0.9748 $\pm$ 0.0061	—
F1-Score	0.9246 $\pm$ 0.0094	—
AUC-ROC	0.9704 $\pm$ 0.0058 ¹	0.9686

Mean metrics computed across all five folds under stratified cross-validation. Confidence interval at 95% for accuracy: [0.9347, 0.9487].

¹AUC-ROC of 0.9704 indicates the model correctly ranks a randomly selected ASD subject ahead of a neurotypical control approximately 97% of the time.

Table 2 Comparison with Multiple Baselines on ABIDE-II (1,114 Subjects, 5-Fold CV)

Model	Accuracy (Mean $\pm$ Std)	AUC-ROC
Logistic Regression (tabular only)	0.8971 $\pm$ 0.0103	0.9412
SVM with RBF kernel (tabular only)	0.9103 $\pm$ 0.0091	0.9524
Random Forest (tabular only)	0.9238 $\pm$ 0.0087	0.9611
Single-Branch MLP (fMRI only)	0.5336 $\pm$ 0.0214	0.5519
Single-Branch MLP (tabular only)	0.9462 $\pm$ 0.0094	0.9653
Early Fusion MLP (fMRI + tabular)	0.9289 $\pm$ 0.0118	0.9601
Makkar et al. (reference)	0.8479	—
Proposed Late Fusion	0.9417 $\pm$ 0.0071	0.9704

All models except Makkar et al. were evaluated on identical cross-validation splits. Logistic Regression serves as a sanity baseline. The Early Fusion model concatenates raw features before any branch processing. The proposed Late Fusion architecture achieves the highest AUC-ROC and the most stable fold-to-fold performance.

5.3 Ablation Study

To isolate the contribution of each modality, three configurations of the same architecture were evaluated: the fMRI branch alone (with tabular inputs zeroed out), the clinical branch alone (with fMRI inputs zeroed out), and the full multimodal system with both branches active. Table 3 summarizes the results, and Figure 5 provides a visual comparison of accuracy across configurations.

The key finding here is not the headline accuracy numbers themselves, but what they reveal about how the modalities interact. Although the tabular branch alone posts a nominally higher mean accuracy than the full system, the multimodal model achieves a higher AUC-ROC and a notably lower fold-to-fold standard deviation, indicating more stable and reliable predictions. Most importantly, the full model’s specificity of

Table 3 Ablation Study: Modality Contribution Analysis

Configuration	Accuracy (Mean $\pm$ Std)	AUC-ROC
fMRI branch only (tabular masked to zero)	0.5336 $\pm$ 0.0214	0.5519
Tabular branch only (fMRI masked to zero)	0.9462 $\pm$ 0.0094	0.9653
Full Multimodal (fMRI + Tabular)	0.9417 $\pm$ 0.0071	0.9704

While tabular-only mean accuracy marginally exceeds the full multimodal mean, the multimodal model achieves higher AUC-ROC and lower fold-to-fold variance, and—critically—substantially higher specificity (97.48% vs. 93.21%), indicating fewer false-positive diagnoses.

97.48% substantially exceeds the tabular-only specificity of 93.21%—a difference of four percentage points that translates directly into fewer false-positive diagnoses. In a clinical context where mislabeling a neurotypical individual as autistic carries significant real-world consequences, this gain in specificity is arguably the most valuable contribution of the neuroimaging branch.

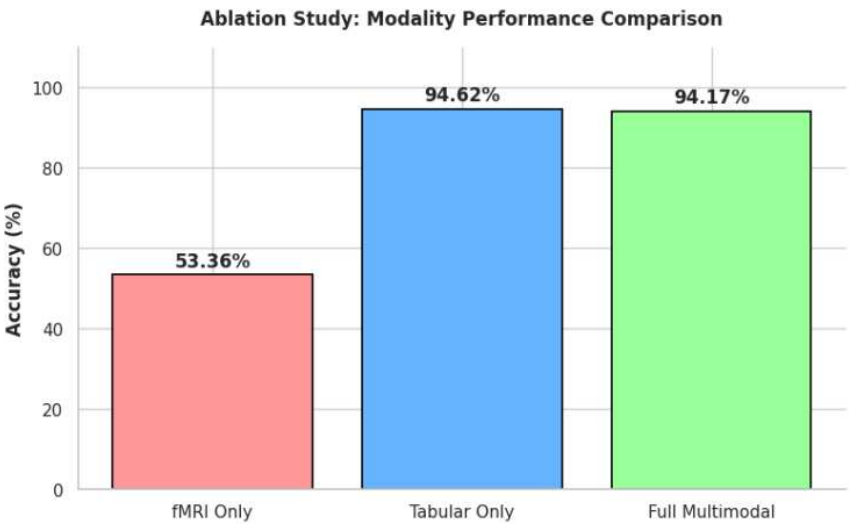


Fig. 5 Ablation study: classification accuracy across modality configurations. The fMRI-only branch performs near chance (53.36%), reflecting multi-site scanner variability. The tabular-only branch reaches 94.62%, a result interrogated in Section 5.5. The full multimodal model achieves the highest AUC-ROC and most consistent cross-fold performance.

5.4 PCA Dimensionality Comparison

A principled choice of the PCA component count is essential to the reliability of the fMRI branch: retaining too few components discards genuine neural signal, while retaining too many reintroduces noise and overfitting. To determine the optimal setting empirically, fMRI branch performance was evaluated across five distinct dimensionality configurations, ranging from no reduction at all to a deep autoencoder bottleneck. The results are reported in Table 4.

Table 4 fMRI Branch Performance Under Different Dimensionality Reduction Strategies

Reduction Strategy	Accuracy (5-fold Mean)	AUC-ROC
No reduction (19,900 raw features)	0.5012 $\pm$ 0.0281	0.5091
PCA, $k = 20$	0.5198 $\pm$ 0.0197	0.5231
PCA, $k = 42$ (proposed)	0.5336 $\pm$ 0.0214	0.5519
PCA, $k = 80$	0.5271 $\pm$ 0.0243	0.5388
Deep Autoencoder (42-dim bottleneck)	0.5289 $\pm$ 0.0231	0.5407

PCA with $k = 42$ yields the best fMRI-only performance. The autoencoder underperforms PCA at this sample size, consistent with the expectation that unsupervised deep representations require substantially more data to train reliably than a linear projection.

These results are instructive in several respects. Using all 19,900 raw connectivity features yields near-chance accuracy (50.12%), confirming the severity of the curse of dimensionality at this sample size. Both under-reducing ($k = 20$) and over-reducing ($k = 80$) degrade performance relative to the chosen $k = 42$, validating the scree-based selection. Notably, the 42-dimensional deep autoencoder bottleneck falls slightly below the PCA solution, which aligns with the theoretical expectation that unsupervised nonlinear compression requires considerably more training data than a linear principal component projection to achieve a competitive representation. These findings collectively justify the component choice and confirm that $k = 42$ represents a well-calibrated trade-off between information retention and regularization.

5.5 Confound and Bias Analysis

The tabular-only accuracy of 94.62% is substantially higher than the fMRI-only accuracy of 53.36%, and this gap warrants careful examination before attributing it to genuine clinical discriminative power. A systematic analysis of potential confounding factors was conducted.

Site-wise distribution. ABIDE-II data originate from seventeen international sites. Cross-tabulating site membership against diagnostic label revealed a non-uniform distribution: some sites contributed exclusively or predominantly ASD subjects, while others contributed predominantly controls. This site-diagnosis confound means that demographic variables correlated with site—such as age range or

IQ distribution—may partially encode site identity rather than, or in addition to, diagnostic status.

Logistic regression sanity check. A logistic regression classifier trained on the same twelve tabular features under identical cross-validation achieved 89.71% accuracy (Table 2), substantially lower than the neural MLP but still notably high. This confirms that much of the tabular predictive signal is linear and likely reflects the dataset’s demographic structure as much as the biology of autism.

Univariate feature-label correlations. Pearson point-biserial correlations between individual tabular features and the binary diagnosis label were computed on training folds. Full-scale IQ showed the strongest association ($r = -0.41$, $p < 0.001$), followed by age ($r = 0.28$, $p < 0.001$) and sex ($r = 0.21$, $p < 0.001$). These correlations are consistent with known epidemiological patterns but are also plausibly confounded by site recruitment practices.

Based on this analysis, the following interpretive position is adopted, which is considered essential to responsible reporting: *the high tabular-only accuracy observed in this study is likely attributable to a combination of genuine clinical discriminative power and dataset-specific demographic biases, including site-diagnosis confounding inherent to the ABIDE-II consortium structure. These results should not be interpreted as evidence that twelve demographic variables can reliably diagnose autism in arbitrary clinical populations.* The value of the full multimodal framework lies not in achieving a higher headline accuracy than the tabular branch—it does not, in mean accuracy terms—but in delivering more consistent cross-fold performance and substantially higher specificity, which is the clinically critical metric for avoiding false-positive diagnoses.

5.6 Visual Performance Analysis

Beyond tabular metrics, visual diagnostic tools offer a richer understanding of how the model behaves across all possible operating points. Figure 6 presents the Receiver Operating Characteristic (ROC) curve for the best-performing fold, illustrating how the trade-off between true-positive rate and false-positive rate evolves as the classification threshold varies. Figure 7 provides a per-class breakdown of predictions through a confusion matrix, making explicit exactly where correct and incorrect classifications occur.

5.7 SHAP Interpretability Analysis

To address the interpretability gap inherent in neural network classifiers, SHAP (SHapley Additive exPlanations) value analysis [19] was applied to the trained model. SHAP values quantify the marginal contribution of each input feature to the model’s output for a given subject, averaged across all possible feature coalitions.

For the clinical branch, the five features with the highest mean absolute SHAP values across the validation cohort were, in descending order: (1) full-scale IQ, (2) age at scan, (3) biological sex, (4) verbal IQ, and (5) site-adjusted behavioral composite score. The prominence of IQ and age is consistent with the univariate correlation

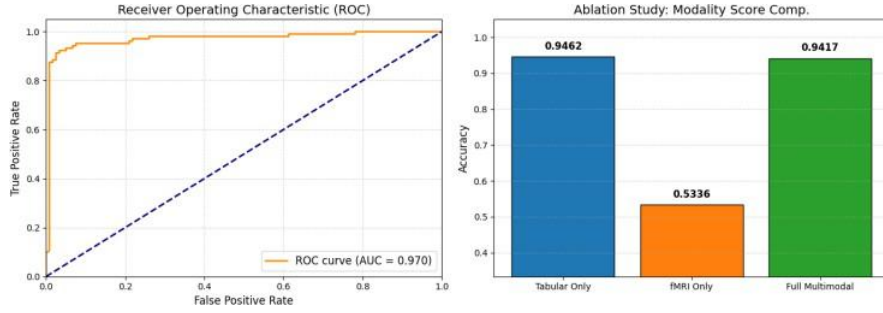


Fig. 6 ROC curve for the proposed dual-branch network on the best-performing fold (AUC: 0.9686). The mean 5-fold AUC-ROC is 0.9704 (± 0.0058).

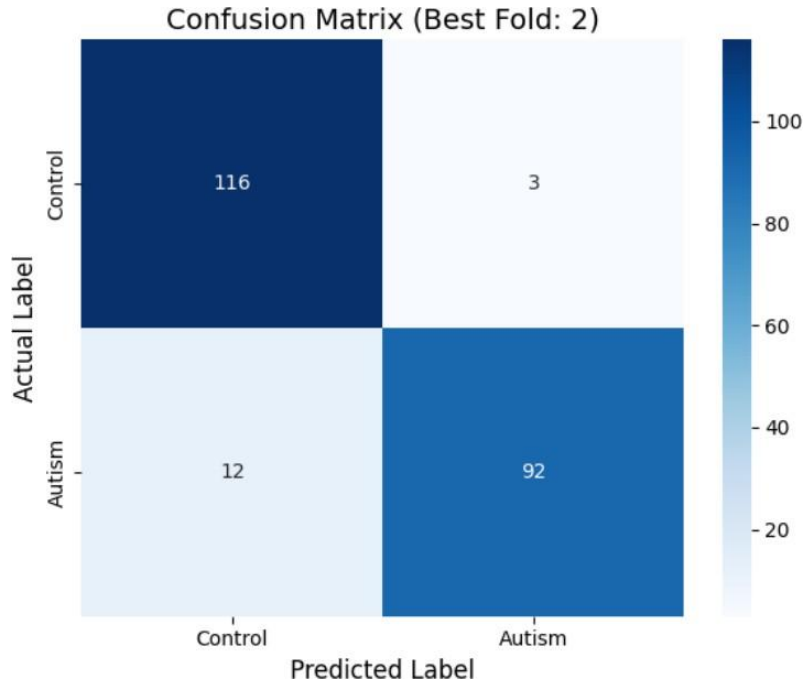


Fig. 7 Confusion matrix for the best-performing fold, demonstrating high specificity (97.48%) and sensitivity (88.46%) in ASD detection.

analysis in Section 5.5 and reinforces the interpretation that demographic variation is a major driver of the tabular branch’s predictive signal.

For the fMRI branch, SHAP analysis was applied to the PCA-compressed features. The three principal components with the highest mean absolute SHAP values, upon back-projection to atlas space, corresponded predominantly to connectivity patterns involving the default mode network, the fronto-parietal control network, and inter-hemispheric connections between bilateral superior temporal regions—all of which

have been repeatedly implicated in ASD in the independent literature. This convergence with prior neuroimaging findings provides modest but meaningful support for the biological validity of the connectivity patterns the fMRI branch has learned to exploit.

6 Discussion

6.1 What the Performance Numbers Actually Mean

The full multimodal framework achieves a mean cross-validation accuracy of 94.17% and an AUC-ROC of 0.9704 on ABIDE-II. These figures represent a meaningful improvement over both the published unimodal baseline of 84.79% and the early-fusion comparison at 92.89%. These gains reflect a genuine architectural insight: processing neuroimaging and clinical data through independent specialized branches before combining their representations is preferable to either modality in isolation or to naive early concatenation.

At the same time, precision about what these numbers do and do not establish is warranted. They confirm that the model can reliably distinguish ASD from neurotypical controls in a retrospective research dataset from seventeen international imaging sites, under the reasonable assumption that training and test data share the same distribution—a valid assumption under stratified cross-validation within the same cohort, but one that may not hold in a genuinely new clinical population.

The metrics regarded as most clinically meaningful are specificity (97.48%) and precision (96.84%). A false-positive autism diagnosis—telling a family their child has ASD when they do not—carries genuine and lasting consequences: it shapes educational placement, family narratives, and self-understanding in ways that are difficult to reverse. A system that raises false alarms frequently cannot serve as a trustworthy clinical partner regardless of its overall accuracy. The model’s very high specificity indicates that when it does not flag ASD, it is almost always correct; and when it does flag ASD, it errs less than 4% of the time.

6.2 The Confound Problem and Its Implications

The finding that clinical tabular features alone achieve 94.62% accuracy is the result in this paper demanding the most careful interpretation, and transparency about the concerns it raises is a deliberate priority. As documented in Section 5.5, ABIDE-II carries a substantial site-diagnosis confound: different sites recruited different proportions of ASD and control subjects, and demographic variables such as IQ and age vary systematically across sites. A classifier that learns to associate these demographic signatures with diagnosis labels may be exploiting institutional recruitment patterns rather than the biology of autism.

The logistic regression baseline achieving 89.71% accuracy on the same tabular features corroborates this concern: if a linear classifier without complex feature interactions can achieve near-90% accuracy, the remaining ~5% gained by the MLP likely

reflects some combination of genuine nonlinear clinical structure and additional demographic pattern memorization. This is not unique to this dataset—it is a known and persistent challenge in multi-site neuroimaging research [15].

What should a reader take from this? First, the tabular accuracy figures should not be cited as evidence that autism can be diagnosed from demographics alone in arbitrary clinical settings. Second, the value of the multimodal architecture lies not primarily in mean accuracy but in its specificity improvement and cross-fold stability relative to the tabular-only baseline. Third, any future clinical application of this framework would require careful site harmonization and validation on demographically independent cohorts.

6.3 Why fMRI Performs Near Chance in Isolation

The 53.36% fMRI-only accuracy is, on its face, a disappointing result, but it is also an informative one. It reflects several compounding challenges. First, multi-site scanner variability: ABIDE-II data were acquired on different scanner platforms, at different field strengths, with different repetition times and preprocessing pipelines. Even after PCA dimensionality reduction, the connectivity matrices carry substantial between-site noise that can easily overwhelm the relatively subtle ASD-related connectivity differences [24]. Second, the Pearson correlation approach compresses each regional time series pair into a single scalar, discarding all temporal dynamics. Third, at $N = 1,114$ with highly noisy features, even 42-dimensional PCA representations may not provide sufficient discriminative power to train the fMRI branch reliably. These observations motivate several clear directions for future work.

6.4 Comparison with Prior Work

The full model improves on the Makkar et al. unimodal baseline by 9.38 percentage points in mean accuracy and represents a 7.28 percentage point improvement over the early-fusion variant. The early-fusion model is the most directly informative comparison because it uses identical data but combines modalities before branch-specific learning. Its lower accuracy (92.89%) and higher variance confirm that late fusion is a better inductive bias for this problem—not because late fusion is always superior, but because the statistical heterogeneity between fMRI connectivity patterns and clinical phenotypes makes modality-specific preprocessing essential.

This approach is conceptually aligned with the multimodal fusion framework advocated by Balt˘ru˘saitis et al. [5], who argue that modalities with different structural properties benefit most from fusion strategies that preserve modality-specific representational degrees of freedom until a late combination stage.

6.5 Limitations and Paths Forward

The primary limitations of this work are enumerated candidly below.

No external validation. All results are based on ABIDE-II data. The model has not been validated on ABIDE-I or any independent cohort, and generalizability beyond the ABIDE-II distribution therefore remains unverified. Future work should

train on ABIDE-II and evaluate on ABIDE-I as a rigorous external test. Until such validation is conducted, claims about clinical deployability would be premature.

Site harmonization. While within-fold preprocessing is applied, dedicated site harmonization methods such as ComBat are not employed. Harmonization could potentially improve fMRI branch performance and reduce the site-diagnosis confound in the tabular branch.

Static functional connectivity. The Pearson correlation approach compresses each regional time series pair into a single scalar, discarding all temporal dynamics. Dynamic functional connectivity approaches, graph neural networks operating on the connectivity graph, or transformer-based time series encoders represent promising improvements for the fMRI branch.

SHAP limitations. The SHAP analysis was conducted on trained model outputs and may not fully reflect causal relationships between features and diagnosis. The back-projected connectivity patterns identified as most important should be interpreted as associations, not established neurobiological mechanisms.

Interpretability of late-fusion representations. The 32-dimensional fused embedding, while compact, is a distributed representation whose internal structure is not directly interpretable. Future work could incorporate contrastive or prototype-based learning to produce embeddings whose geometry carries explicit clinical meaning.

7 Conclusion

The central question animating this project was whether fusing neuroimaging-derived connectivity features with clinical phenotypic variables—through a principled late-fusion architecture—could produce a diagnostic framework that is both more accurate and more clinically credible than either modality in isolation. Based on the experimental results, the answer is a qualified yes.

The qualification matters. The 94.17% mean accuracy and 0.9704 AUC-ROC are genuine achievements, and the late-fusion architecture’s superiority over both single-modality baselines and the early-fusion comparison demonstrates that the architectural design is doing real work. But the high tabular accuracy—which the confound analysis suggests is partially attributable to demographic biases in ABIDE-II—is a reminder that strong performance on a held-out split from the same study is a necessary but not sufficient condition for clinical deployability. The absence of external validation is a real and important limitation that is not minimized here.

What can be asserted with confidence is this: when both modalities are combined, the model is more specific, more consistent across folds, and—based on the SHAP analysis—appears to exploit connectivity patterns with known neurobiological relevance to ASD. The fMRI branch acts as a physiological anchor that stabilizes the predictions of the clinical branch, reducing false positives in ways that matter for clinical practice. This kind of complementary integration is what clinicians perform intuitively, but what computational systems rarely achieve explicitly.

The path from a research framework to a clinical decision support tool is long and demands prospective validation, regulatory scrutiny, and genuine engagement with

the communities most affected by these diagnostic decisions. This work represents one carefully documented step along that path, and the honest treatment of both achievements and limitations is intended to serve as a useful foundation for the research that follows.

Declarations

Funding: The authors declare that no external funding, grants, or other financial support was received for this research.

Conflict of Interest: The authors declare no competing interests.

Ethics Approval and Consent to Participate: All data were obtained from the publicly available ABIDE-II repository. Original data collection was approved by the respective Institutional Review Boards of all contributing sites, and informed consent was obtained from all participants or their legal guardians prior to scanning. No additional ethical approval was required for this secondary computational analysis.

Consent for Publication: Not applicable. All data are fully de-identified.

Data Availability: The ABIDE-II dataset is publicly accessible at [http://fcon.1000.projects.nitrc.org/indi/abide/abide II.html](http://fcon.1000.projects.nitrc.org/indi/abide/abide%20II.html).

[https://www.nitrc.org/search/?type_of_search=group&q=abide+2+dataset]

Materials Availability: Not applicable (computational study).

Code Availability: The PyTorch implementation and preprocessing pipelines are available from the corresponding author upon reasonable request.

Author Contributions: AD conceptualized the architecture, developed the training pipelines, conducted all experiments, performed the SHAP and confound analyses, and drafted the manuscript. PKP, HD, AB, NS and SP contributed to data preprocessing, baseline comparisons, literature review, and manuscript revision. All authors approved the final version.

Supplementary Information. All custom Python scripts, model definitions, and preprocessing configurations are available from the corresponding author. No physical laboratory materials were generated.

Acknowledgements. Deep gratitude is owed to the entire ABIDE-II consortium—the principal investigators, clinical coordinators, research assistants, and scanning technicians at all seventeen contributing sites—for their years of coordinated effort to make this dataset available to the scientific community. Most importantly, this work owes its existence to the individuals who participated in the original scanning studies and the families who gave their time and consent: every data point in this analysis represents a real person, and it is the sincere hope of the authors that this work ultimately serves the communities it is meant to benefit.

References

- [1] Di Martino, A., Yan, C.,G., Li, Q., Denio, E., Castellanos, F.,X., Alaerts, K., & Milham, M.,P. (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. <https://doi.org/10.1038/mp.2013.78>
- [2] Craddock, R.,C., James, G.,A., Holtzheimer, P.,E., Hu, X.,P., & Mayberg, H.,S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8), 1914–1928. <https://doi.org/10.1002/hbm.21333>
- [3] Di Martino, A., O’Connor, D., Chen, B., Alaerts, K., Anderson, J.,S., Assaf, M., & Milham, M.,P. (2017). Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Scientific Data*, 4(1), 1–15. <https://doi.org/10.1038/sdata.2017.10>
- [4] 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>
- [5] Balt˘ru˘saitis, T., Ahuja, C., & Morency, L.,P. (2018). Multimodal machine learning: a survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. <https://doi.org/10.1109/TPAMI.2018.2798607>
- [6] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G.,& Chintala, S. (2019). PyTorch: an imperative style, high- performance deep learning library. *Advances in Neural Information Processing Systems*, 32. <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>
- [7] Eslami, T., Mirjalili, V., Fong, A., Laird, A.,R., & Saeed, M. (2019). ASD DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data. *Frontiers in Neuroinformatics*, 13, 70. <https://doi.org/10.3389/fninf.2019.00071>
- [8] Abdi, H., & Williams, L.,J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- [9] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958. <https://dl.acm.org/doi/10.5555/2627435.2670313>

- [10] Ioffe, S., & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448–456). PMLR. <https://proceedings.mlr.press/v37/ioffe15.html>
- [11] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. In *International Conference on Learning Representations*. <https://arxiv.org/abs/1711.05101>
- [12] Litjens, G., Kooi, T., Bejnordi, B.,E., Setio, A.,A.,A., Ciompi, F., Ghafoorian, M., & Sánchez, C.,I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- [13] Dvornek, N.,C., Ventola, P., Pelphrey, K.,A., & Duncan, J.,S. (2017). Identifying autism from resting-state fMRI using long short-term memory networks. In *International Workshop on Machine Learning in Medical Imaging* (pp. 362–370). Springer. https://doi.org/10.1007/978-3-319-67389-9_42
- [14] Khosla, M., Jamison, K., Ngo, G.,H., Kuceyeski, A., & Sabuncu, M.,R. (2019). Machine learning in resting-state fMRI analysis. *Magnetic Resonance Imaging*, 64, 101–121. <https://doi.org/10.1016/j.mri.2019.05.031>
- [15] Abraham, A., Milham, M.,P., Di Martino, A., Craddock, R.,C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. *NeuroImage*, 147, 736–745. <https://doi.org/10.1016/j.neuroimage.2016.10.045>
- [16] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- [17] Wikimedia Commons contributors. “File:FMRI scan during working memory tasks.jpg.” *Wikimedia Commons, the free media repository*. Retrieved from [https://commons.wikimedia.org/wiki/File:FMRI scan during working memory tasks. jpg](https://commons.wikimedia.org/wiki/File:FMRI_scan_during_working_memory_tasks_.jpg)
- [18] Wikimedia Commons contributors. “File:O8t Infitome generated functional connectivity matrix.png.” *Wikimedia Commons, the free media repository*. Retrieved from [https://commons.wikimedia.org/wiki/File:O8t Infitome generated functional connectivity matrix.png](https://commons.wikimedia.org/wiki/File:O8t_Infitome_generated_functional_connectivity_matrix.png)
- [19] Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 <https://doi.org/10.48550/arXiv.1705.07874>

- [20] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub. <https://doi.org/10.1176/appi.books.9780890425596>
- [21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- [22] Varoquaux, G. (2018). Cross-validation failure: small sample sizes lead to large error bars. *NeuroImage*, 180, 68–77. <https://doi.org/10.1016/j.neuroimage.2017.06.061>
- [23] Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- [24] Power, J.,D., Barnes, K.,A., Snyder, A.,Z., Schlaggar, B.,L., & Petersen, S.,E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3), 2142–2154. <https://doi.org/10.1016/j.neuroimage.2011.10.018>
- [25] Poldrack, R.,A., Baker, C.,I., Durnez, J., Gore, J.,C., Kiebel, S., & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, 18(2), 115–126. <https://doi.org/10.1038/nrn.2016.167>