ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Multimodal Explainable Automated Diagnosis of Autistic Spectrum Disorder

Meryem Ben Yahia[1], Moncef Garouani[2*] and Julien Aligon[2]

1- Université Jean Monnet, Saint-Etienne, France
2- IRIT, UMR 5505 CNRS, Université Toulouse Capitole, Toulouse, France

Abstract.
Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by symptoms that affect social interaction, communication, and behavior, the diagnosis being complicated by significant individual variability and the absence of definitive biomarkers. Current artificial intelligence methods have improved diagnostic accuracy, but their reliance on subjective assessments or single-modal data, coupled with their "black-box" nature, limits consistency and clinical applicability. Addressing current limitations, this paper introduces a multimodal ASD detection framework using deep neural networks (DNN) with explainable AI (xAI) to enhance model transparency. Our model achieves a mean 5-fold cross-validation accuracy of 98.64% ($\pm$ 0.86%), surpassing existing methods and demonstrating potential for clinical dependability of ASD diagnoses. The source code is available at : Multimodal-Explainable-Diagnosis-of-ASD.git

1 Introduction

Autism Spectrum Disorder encompasses a wide range of related conditions, including unique symptoms and traits [1]. Rather than a singular disorder, it is a syndrome with distinct subgroups, leading to varied presentations across individuals [2]. The etiology of ASD is multifaceted, involving genetic and neurological factors [3].

Over recent years, advanced neuroimaging techniques, notably functional magnetic resonance imaging (fMRI) [4, 5] and magnetoencephalography (MEG) [6], have been extensively employed to investigate the structural and functional brain characteristics associated with neurodevelopmental disorders. Studies leveraging fMRI have demonstrated that individuals with ASD often exhibit atypical neural oscillations and disrupted functional connectivity throughout various stages of development [4, 5, 7]. To capture these distinctions, researchers have extracted numerous fMRI-based metrics, such as connectivity patterns, neural activation markers, and nonlinear dynamic indicators, providing quantitative insights into the unique neural profiles of children with ASD [1, 8].

Despite the utility of fMRI, recent research highlights that relying solely on neural data may be insufficient to capture the complexity of ASD, which affects individuals on multiple levels, from cellular to behavioral[3]. Phenotypic patterns, specifically cognitive indices like Full-Scale IQ (FIQ), Performance IQ (PIQ), and Verbal IQ (VIQ), offer additional non-invasive means of understanding ASD's impact without intrusive methods that may alter the observed behaviors or introduce biases.

*Corresponding author. E-mail addresse : moncef.garouani@irit.fr .

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

However, to date, these data modalities have primarily been applied independently in ASD studies to identify biomarkers and build diagnostic models using advanced ML techniques [1]. Most of studies focus on single-modality data analysis, which may not be sufficient given ASD heterogeneity and its range of abnormal manifestations levels [8]. The multimodal diagnostic framework outlined in this work integrates fMRI and phenotypic data to address this constraint.

To increase clinicians confidence in the outcomes of AI-based ASD diagnostic systems, this study emphasizes model interpretability [9, 10]. Specifically, we utilize SHapley Additive exPlanations (SHAP) [11], a robust game-theoretic approach to explainability, to provide insights into the model decision-making process. To our knowledge, this is the first work to integrate internal neurophysiological data with external non-invasive phenotypic data for explainable ASD detection. Our key contributions are as follows :

- **Exploration of non-invasive modalities**, integrating neurophysiological imaging data with cognitive assessments for explainable ASD detection.
- **Development of a multimodal method** combining fMRI and phenotypic data using feature selection and a deep neural network.
- **An interpretable framework** utilizing SHAP values to reveal significant brain connectivity patterns related to ASD diagnosis.

2 Methodology

The proposed multimodal identification framework for ASD is illustrated in in Figure 1, where fMRI and phenotypic data are concatenated as one input vector. It mainly consists of three sequential steps : data acquisition, feature extraction and selection, and multimodal combination. The details of each part are described in the following subsections.

2.1 Data acquisition and preprocessing

We utilized the Autism Brain Imaging Data Exchange (ABIDE I), which aggregates fMRI imaging and phenotypic data from multiple research sites. The dataset includes 1,112 subjects (539 ASD, 573 controls), aged 7 to 64 years.

2.1.1 Phenotypic Data processing

To ensure dataset integrity, we excluded records with corrupted or missing fMRI scans. Given that traditional imputation methods may introduce biases and compromise the validity of analyses for biological data [12], we framed the missing data challenge as an optimization problem. The goal was to maximize the feature-to-example ratio while minimizing the missing-example-to-feature one. The process retained 5 features, summarized in Table 1.

2.1.2 fMRI Data preparation

We utilized the BASC122 brain atlas[1], which defines 122 distinct networks, to delineate specific regions of interest (ROIs) from the imaging data. Functional

[1]A brain atlas is a detailed map that categorizes different regions and structures of the brain, often used as a reference in neuroscience for studying brain anatomy and function.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

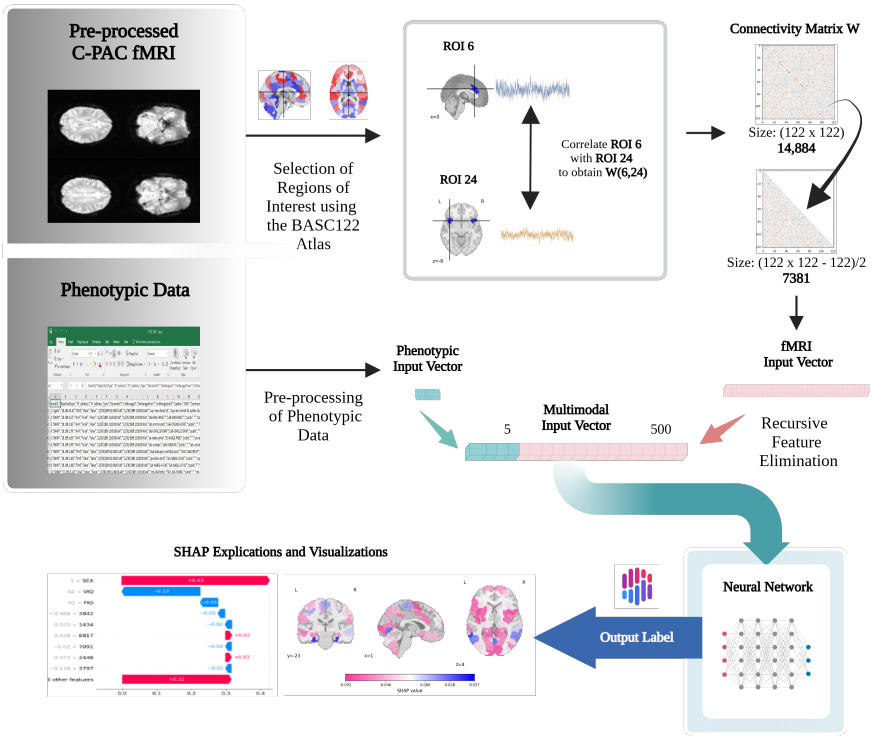

Fig. 1: Proposed multimodal identification framework of ASD diagnosis.

Table 1: Retained phenotypic features.

Feature	Description	Type
Age	Age at the time of the scan	Numeric
Sex	Biological sex	Categorical
FIQ	Full-Scale Intelligence Quotient	Numeric
VIQ	Verbal Intelligence Quotient	Numeric
PIQ	Performance Intelligence Quotient	Numeric

connectomes[2] were constructed using the tangent embedding of the Ledoit-Wolf regularized covariance estimator. The 4D fMRI data were converted into 2D time-series representations by applying 3D masks at each time point.

From this time-series data, a symmetric tangent connectivity matrix was generated and simplified by retaining only the lower triangular values, resulting in a 1D feature vector of size 7,381. Each element represents the interaction between a pair of ROIs. Due to high dimensionality, Recursive Feature Elimination (RFE) was applied to select the 500 most essential fMRI features. These were combined with the 5 phenotypic features for model input, totaling 505 features.

[2]Functional connectomes are representations of the functional connections in the brain, illustrating how different regions interact with one another during rest or task performance.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

2.1.3 Deep Neural Network Model

Our DNN architecture features two hidden layers with ReLU activation functions and dropout to prevent overfitting [13]. The model was trained using standard feature scaling, L2 regularization, and a sigmoid function in the output layer. Hyperparameters were fine-tuned using stratified 5-fold cross-validation. The optimal configuration included 96 units, L2 regularization of 0.002, a dropout rate of 0.3, and a learning rate of 0.003.

3 Experiments and Results

To assess the effectiveness of our proposed multimodal framework, we benchmarked its performance against state-of-the-art unimodal methods. The following subsections present the results of our performance evaluation and provide an interpretation of the model's outputs using explainable AI techniques.

3.1 Model Performance Evaluation

The multimodal approach provides a robust and highly accurate model, effectively integrating diverse features to enhance performance stability. The model achieves a mean accuracy of 98.64%, with a standard deviation of 0.86, indicating consistently high performance across different folds. Both precision and recall values are near-perfect for ASD and Non-ASD classes, underscoring the model reliability in distinguishing between the two.

Comparing with previous studies (Table 2), our Multimodal model, leveraging both fMRI and phenotypic data, outperforms existing methods by achieving an AUC of 1.00, along with 99% accuracy and recall. These results set a new benchmark in classification performance, and highlights the efficacy of our approach in capturing complex patterns within ASD data.

Table 2: Research publications on the ABIDE dataset to ASD detection.

Authors	**Data modality**	**AUC**	**Accuracy**	**Recall**
[5]	fMRI	-	0.95	0.97
[4]	fMRI	0.96	0.87	0.87
[7]	fMRI	0.91	0.89	0.93
[14]	fMRI	0.78	0.75	0.77
[15]	fMRI	-	0.70	0.74
Our Method	fMRI + Phenotypic	**1.00**	**0.99**	**0.99**

3.2 Model Results Interpretation

SHAP values quantify each variable's contribution to the ML model's predictions. We illustrate SHAP explanations using two complementary examples presented in Figures 2 and 3, which pertain to the same two patients (patient 50606 (a), predicted ASD, and 50572 (b), predicted non-ASD).

In Figure 2, blue denotes a contribution toward non-ASD and red toward ASD. Phenotypic variables emerge as some of the most influential features for both patients. The remaining fMRI features collectively contribute to the model's predictions, though their individual impacts might be minor.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

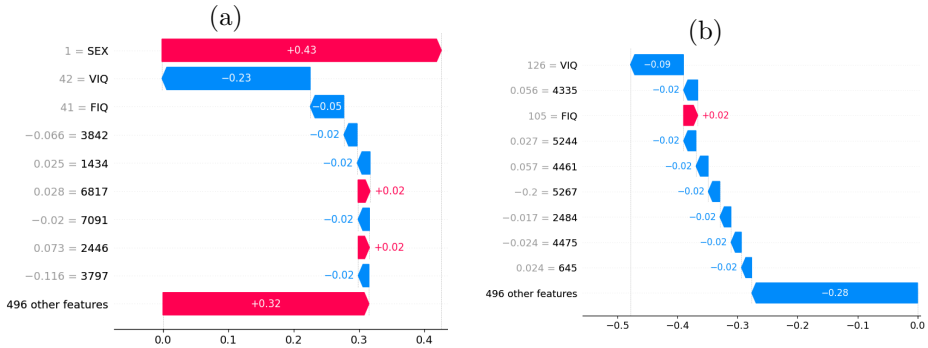

Fig. 2: SHAP values for patients with (a) and without ASD (b).

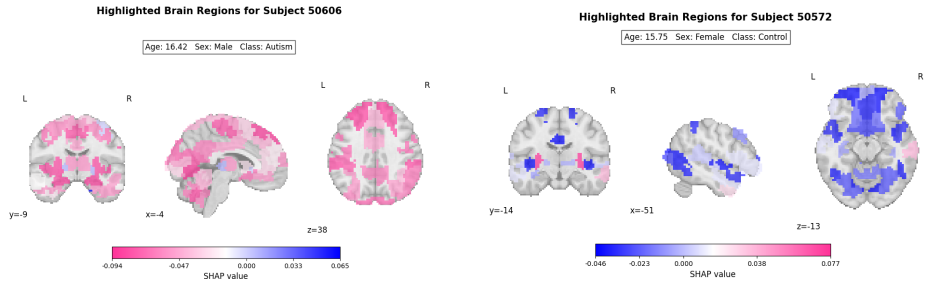

Fig. 3: Inter-regions connectivity for patients with (a) and without ASD (b).

With 5 phenotypic features and 500 fMRI features, our findings suggest that the phenotypic features act as high-level, summarized indicators of the activation patterns in the fMRI. We hypothesize that phenotypic data encapsulates connectivity measurements between ROIs in the fMRI data. Thus, when combined with fMRI data, phenotypic variables offer additional interpretability as simplified proxies of the broader neural activity.

Figure 3 presents the explanations related to the fMRI data, into a brain scan representation. The magnitude of each region's contribution reveals that some regions exert significantly greater influence, yet there is a clear cohesiveness among them: the majority align with the decision. For patient with ASD (a), pink regions dominate, emphasizing regions associated with a positive ASD prediction. In non-ASD case (b), blue regions are more prevalent, reflecting connectivity patterns that actively oppose an ASD classification.

4 Conclusion and Perspectives

This study presents a multimodal diagnostic framework for ASD, achieving 98.64% of accuracy with minimal variability by integrating resting-state fMRI and phenotypic data through a deep NN. Using explainable AI components, the framework addresses obstacles in ASD diagnosis, such as individual variability and the complexity of interpreting high-dimensional data. Our findings show

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

that phenotypic features complement and summarize the complex patterns in fMRI data, offering potential pathways to uncover meaningful biological discoveries. Future advancements could integrate more data modalities and enhance real-time application, boosting clinical utility and trust in AI diagnostics.

References

[1] K. K. Hyde et al. "Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review". In: *Review Journal of Autism and Developmental Disorders* 6.2 (2019), pp. 128–146.

[2] J. Maser et al. "Spectrum concepts in major mental disorders". In: *The Psychiatric Clinics of North America* 25.4 (2002), pp. xi–xiii.

[3] J. Han et al. "A Multimodal Approach for Identifying Autism Spectrum Disorders in Children". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30 (2022), pp. 2003–2011.

[4] F. Z. Subah et al. "A Deep Learning Approach to Predict Autism Spectrum Disorder Using Multisite Resting-State fMRI". In: *Applied Sciences* 11.8 (2021).

[5] M. I. Al-Hiyali et al. "Principal Subspace of Dynamic Functional Connectivity for Diagnosis of Autism Spectrum Disorder". In: *Applied Sciences* 12.18 (2022).

[6] M. Kikuchi et al. "Magnetoencephalography in the study of children with autism spectrum disorder". In: *Psychiatry and Clinical Neurosciences* (2015).

[7] C. P. Chen et al. "Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism". In: *NeuroImage: Clinical* 8 (2015), pp. 238–245.

[8] N. Silva. "Autism Spectrum Disorder: History, Concept and Future Perspective". In: *J of Neur. Research Reviews & Reports* (2023), pp. 1–8.

[9] M. Garouani et al. "Investigating the Duality of Interpretability and Explainability in Machine Learning". In: *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*. 2024, pp. 861–867.

[10] M. Garouani et al. "Unlocking the Black Box: Towards Interactive Explainable Automated Machine Learning". In: *Intelligent Data Engineering and Automated Learning – IDEAL 2023*. Cham: Springer Nature Switzerland, 2023, pp. 458–469.

[11] S. M. Lundberg et al. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017.

[12] J. A. C. Sterne et al. "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls". In: *BMJ* 338 (2009).

[13] M. Garouani et al. "Autoencoder-kNN meta-model based data characterization approach for an automated selection of AI algorithms". In: *Journal of Big Data* 10.1 (2023).

[14] B. Yamagata et al. "Machine learning approach to identify a resting-state functional connectivity pattern serving as an endophenotype of autism spectrum disorder". In: *Brain imaging and behavior* 13 (2019).

[15] C. L. Alves et al. "Diagnosis of autism spectrum disorder based on functional brain networks and machine learning". In: *Scientific Reports* (2023).