Journal of
Sensor and Actuator Networks

MDPI

Article

A Feature Fusion Framework for Improved Autism Spectrum Disorder Prediction Using sMRI and Phenotype Information

Bhagya Lakshmi Polavarapu [1], V. Dinesh Reddy [2,*], Mahesh Kumar Morampudi [1], Md Muzakkir Hussain [1] and Ashu Abdul [1]

[1] Department of Computer Science and Engineering, SRM University-AP, Amaravati 522240, Andhra Pradesh, India
[2] Symbiosis Institute of Technology, Hyderabad Campus, Symbiosis International (Deemed University), Pune 412115, Maharashtra, India
* Correspondence: dineshvemula@gmail.com

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by a wide range of symptoms and severity, posing significant challenges for accurate diagnosis. Approaches that rely on a single data source, or unimodal data, often fail to capture the disorder's inherent heterogeneity. A multimodal approach, which integrates diverse data types, can create a more holistic and precise understanding of ASD. This paper introduces the Multimodal ASD (MMASD) framework, a novel predictive model for ASD. The MMASD framework is built upon two distinct input modalities: structural magnetic resonance imaging (sMRI) and corresponding phenotype data. The sMRI data provides detailed neuroanatomical metrics, including brain tissue segmentation, volumetric measurements, and cortical thickness. Complementing this, the phenotype data encompasses the clinical and behavioral characteristics of each individual. In the proposed framework, latent features are independently extracted from both modalities and then fused to generate a comprehensive representation of the multimodal information. These fused features are then used to predict ASD by leveraging the outputs of various classifiers. A majority voting ensemble is employed to determine the final prediction. The MMASD framework achieves a high accuracy of 97.27%, surpassing the performance of current state-of-the-art approaches and demonstrating the efficacy of integrating neuroimaging and clinical data for ASD prediction.

Keywords: autism; sMRI; phenotype data; multimodal data; bagging and boosting; majority voting

Check for updates

Academic Editors: Lei Shu and Chengwen Luo

Received: 16 October 2025
Revised: 6 January 2026
Accepted: 2 February 2026
Published: 15 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Autism is a disorder that starts in childhood. It is identified by a limited range of interests, activities, and communication development [1]. It is characterized by social interaction problems, repetitive behaviors, and communication with others. According to the diagnostic and statistical manual of mental disorders [DSM-5] [2], people affected by autism vary in differing degrees, and a specific set of behaviors expressed identifies the condition. Autism was first identified by Leo Kanner in the year of 1943, trying to determine how sociodemographic characteristics like parent education, age, race, sex, and social class relate to autism [3]. According to an estimate conducted in 2021 by the World Health Organization (WHO), 0.63% of children suffer from autism spectrum disorder (ASD)

worldwide [4]. The female-to-male reporting ratio of ASD is 1:4. However, to date, no accurate treatment for ASD has been developed.

Furthermore, the accurate diagnosis of ASD is difficult because the symptoms vary from one to another. The lengthy diagnosis process is subjective and interview-based, so a physician must examine a child's behavior and developmental history. It may cause a delay of almost 3.5 years from the time parents first contact a doctor and the confirmation of an ASD diagnosis [5]. So, to ensure proper treatment, a diagnosis is essential. The predictive models can be crucial in diagnosing ASD because of the shortage of healthcare experts, especially specialists like developmental pediatricians or child psychologists. The diagnosis process of ASD involves neuroimaging methods and psychological evaluations. Neuroimaging analysis is useful in identifying patterns in the anatomical structure, the function of the brain, and abnormalities in specific brain areas that connect with symptoms of autism [6]. Furthermore, artificial intelligence (AI) and machine learning (ML) techniques are applied to these data modalities to diagnose ASD. ML algorithms can quickly process vast amounts of data, allowing for efficient screening and analysis. This speed is essential in healthcare settings where time is critical [7].

Unimodal data refers to data collected from a single modality. In the context of ASD prediction, unimodal datasets include behavioral data [8], genetic data [9], neuroimaging data [10], EEG data [11], eye-tracking data [12], and phenotype data [13]. These sources typically focus on isolated types of information, each offering a limited perspective that may not identify complex patterns and boundaries of the disorder [14]. For instance, if only brain imaging data is used, the model might identify neurological markers but miss behavioral and environmental factors crucial for a comprehensive diagnosis. Even though using unimodal data can reach high accuracy in ASD prediction, it lacks the robustness and depth needed to capture the full complexity of the disorder. Unimodal approaches often miss the nuanced interactions between social, behavioral, and physiological factors that multimodal data captures, making it less effective at providing comprehensive and personalized assessments.

Multimodal data refers to data collected from multiple modalities or types, integrating various sources to provide a comprehensive understanding of ASD. Key combinations of multimodal data include behavioral and neuroimaging data [15], which correlate observational assessments with brain structure and function; genetic and behavioral data, which explore how genetic predispositions influence specific behaviors; and EEG and eye-tracking data [12], which capture both neural processing and visual attention dynamics. Speech and language data combined with behavioral assessments [16] can provide insights into communication challenges in social interactions. By merging these diverse sources, multimodal approaches enable personalized profiling, capturing unique combinations of behaviors and symptoms, enhancing predictive accuracy, and leading to better diagnostic and therapeutic outcomes for ASD patients. Multimodal data offers significant advantages over unimodal approaches in ASD assessment by providing a more comprehensive, personalized, and balanced view of the disorder. By combining diverse data sources such as video, audio, and physiological sensors, multimodal approaches capture a fuller range of behavioral and neurological markers that single-modality methods often overlook. This integration allows for more balanced assessments by including objective measurements and subjective observations, reducing potential biases inherent in relying on a single data source like questionnaires. Additionally, multimodal approaches offer a more individualized screening experience, better accommodating the variability and unique presentations seen in autism, unlike unimodal methods that may miss atypical or subtle behaviors. They also enhance scalability and contextual relevance by using wearable and remote monitoring technologies that capture natural behaviors in real-world settings rather than controlled

clinic environments [17]. To address the limitations of unimodal data and take advantage of multimodal data, we propose a framework using multimodal data such as structural magnetic resonance imaging (sMRI) and phenotype data for ASD prediction that utilizes deep learning for feature extraction and majority voting for classification. sMRI offers high spatial resolution, capturing subtle, stable structural abnormalities that serve as consistent biomarkers essential for accurate ASD prediction and can be effectively combined with phenotypic data. While functional MRI (fMRI) offers insights into brain function and connectivity, its higher variability, complexity, and participant burden often make sMRI a preferred choice for studies focusing on structural aspects and their direct relationship with behavioral traits. Paired with phenotype data reflecting behavioral, cognitive, and social traits, the proposed multimodal approach provides a comprehensive view that links brain structure to observable ASD characteristics. This improves diagnostic accuracy compared to existing approaches and provides insights into individual differences.

Addressing the extensive phenotypic and neuroanatomical heterogeneity of ASD necessitates the integration of multimodal data as unimodal systems frequently lack the requisite resilience for accurate diagnosis. By synthesizing structural imaging features with clinical behavioral indicators, the proposed framework effectively circumvents the limitations inherent to isolated data sources. To fully exploit this fused, high-dimensional feature space, we implement a consensus-driven decision strategy utilizing a majority voting ensemble. Rather than relying on a single algorithm, this approach incorporates a diverse array of classifiers spanning both bagging and boosting paradigms, each contributing unique inductive biases. This architectural diversity is crucial for detecting subtle structural patterns that individual models may fail to identify. Ultimately, by aggregating these varied predictive insights, the majority voting mechanism mitigates model variance and reduces the risk of overfitting, ensuring the diagnostic stability essential for clinical deployment. To investigate how multimodal data and majority voting contribute to accurate predictions of ASD, the following research question has been formulated: How does the integration of multimodal data (sMRI and phenotype) through a majority voting ensemble of classifiers improve the accuracy and reliability of ASD prediction compared to unimodal approaches?

The contributions of the paper are as follows:

- We employ a framework for ASD prediction with multimodal data.
- Preprocessing and extracting slice information from the axial plane of 3D sMRI scans.
- Extracting the salient features from phenotype and sMRI data using dimensionality reduction and feature extraction techniques, achieving high accuracy even in limited data scenarios.
- Proposed "majority voting" that aggregates the predictions from multiple classifiers (bagging or boosting) to determine the final prediction, ensuring a more robust and accurate classification outcome.
- With extensive experiments, we comprehensively compared the proposed framework with other methods, providing clear evidence of its performance benefits.

The remainder of the paper is organized into sections as follows: Section 2 explains related work, Section 3 presents methodology, Section 4 discusses performance evaluation, and the conclusion is presented in Section 5.

2. Related Work

A broad range of symptoms can be seen in individuals with ASD compared to those without, and these symptoms can be used to predict the development of ASD. This section presents various approaches used in the literature on different modalities of data to detect/diagnose ASD. The summary of all the literature based on different techniques and modalities of data used for ASD classification is presented in Table 1.

Table 1. Literature survey on predicting ASD using different modalities.

S. No.	Authors/Reference	Method	Dataset	Accuracy
1	Muhammad Shoaib Farooq et al. [8]	Federated learning is applied to share the models to the server without sharing the data. LR is used for the classification at the server.	Screening data	98%
2	Tania Akter et al. [18]	Authors applied feature transformation methods and various classification techniques to ASD datasets for different age groups. Among all the SVM achieves better accuracy.	Screening data	100%
3	Suman Raj et al. [19]	This study explores the use of ML techniques, including Naïve Bayes, SVM, LR, KNN, neural networks, and CNNs. The CNN-based approach performs better.	Screening data	98.3%
4	Vakadkar et al. [20]	ML models such as SVM, random forest, LR, and KNN are applied. LR achieves better accuracy.	Screening data	97.5%
5	Hussain et al. [21]	The "relief F" feature selection method is applied to extract the features, and a multilayer perceptron is applied for the classification.	Screening data	100%
6	Sherkatghanad et al. [22]	CNN with varying filter sizes to capture connections between brain regions, with max-pooling, dropout, and cross-validation.	fMRI	70.22%
7	Kun Gao et al. [23]	Segmentation and parcellation maps were generated from sMRI, and a CNN paired with a Siamese network was employed to classify ASD.	sMRI	70%
8	Ingalhalika et al. [24]	ComBat harmonization technique with an ANN is applied for ASD classification.	fMRI	70%
9	Ke et al. [10]	By training a 3D CNN, the researchers visualized key brain regions involved in ASD classification.	fMRI	70%
10	Heinsfeld et al. [25]	Two stacked denoising autoencoders were used to extract the features, and an MLP was used for classification.	fMRI	72%
11	Lakmini et al. [26]	InceptionV3 was applied to fMRI data.	fMRI	98.06%
12	Lakmini et al. [27]	All TL techniques were applied to fMRI data with different learning rates.	fMRI	98.38%
13	Mostafa et al. [28]	Convolutional autoencoder is applied to sMRI data.	sMRI	96.2%
14	Lakshmi et al. [29]	TL-based techniques were applied to fMRI data. Among all VGG16 achieves high accuracy.	fMRI	95.12%
15	Manoj et al. [30]	Random forest (RF), an SVM, and an MLP are applied to evaluate ASD vs. TD classification performance.	sMRI	95.27%
16	Nogay et al. [31]	Optimal CNN models were developed via grid search optimization.	sMRI	85.42%
17	Sachdev et al. [32]	An MLP is applied to fMRI matrices for ASD prediction.	fMRI	72.46%
18	Kong et al. [33]	Features were extracted using LeNet5, and an MLP was used for classification.	fMRI	83.5%
19	Jiannan Kang et al. [34]	The study uses the minimum redundancy maximum relevance (MRMR) feature selection method and SVM classifiers for autism classification.	EEG and Eye Tracking	85%

Table 1. *Cont.*

S. No.	Authors/Reference	Method	Dataset	Accuracy
20	Han et al. [12]	An ANN is applied to features extracted from EEG and eye tracking data.	EEG and Eye Tracking	95.56%
21	Wang et al. [13]	A federated learning approach was applied to fMRI and non-imaging demographic data.	fMRI and Phenotype	73.52%
22	Wang et al. [15]	A deep CNN is applied to the features extracted from fMRI and non-imaging demographic data.	fMRI and Phenotype	77.27%
23	Moon et al. [16]	Random forest is applied to video recording and VR-game-based data.	Video and Audio	87.8%
24	Liu et al. [35]	Features are extracted from fMRI and non-imaging data, and DeepGCNs are applied.	fMRI and Phenotype	91.62%
25	Dong et al. [36]	A GCN, SVM, and other machine learning techniques are applied to ABIDE data.	sMRI and fMRI	71.35%
26	Chen et al. [37]	RNN and GRU feature extraction is employed, fused via the MKLF algorithm and classified using neural networks.	sMRI and fMRI	81.48%
27	Dong et al. [38]	sMRI and the phenotype are fused and given to the classifier.	sMRI and Phenotype	81.35%

2.1. Analysis of Screening Data

Screening for autism involves various tools and assessments that help identify behavioral patterns associated with ASD. The modified checklist for autism in toddlers (M-CHAT) is one tool for assessing ASD based on questions and behaviors of children.

Akter et al. [18] employed feature transformation techniques such as log, Z-score, and sine functions to Q-chart datasets. Further, different classification methods are implemented on these datasets. Among all, support vector machines (SVM) outperforms in terms of accuracy for classifying ASD. Raj et al. [19] applied machine learning techniques and a convolution neural network (CNN) on three distinct datasets, namely, children, adults, and adolescents. Vakadkar et al. [20] used logistic regression on screening data to classify patients with and without ASD. Hussain et al. [21] use the "relief F" feature selection technique to extract features from the autism questionnaire (AQ) dataset and apply classification algorithms. They showed that the multi-layer perceptron achieves high accuracy. Farooq et al. [8] proposed the federated learning (FL) technique by training two different ML classifiers, namely, logistic regression and a support vector machine. These classifiers are locally trained on various datasets, and the results are sent to a central server to decide the most accurate classifier.

2.2. Analysis of Neuroimaging Data

Magnetic resonance imaging (MRI) gives precise pictures of the anatomy and physiology of the brain, highlighting possible variations in the areas of the brain linked to communication, social interaction, and sensory processing. Sherkatghanad et al. [22] proposed a CNN model to classify ASD using the craddock400 functional parcellation atlas of the brain from fMRI images. Ingalhalikar et al. [24] used harmonization using ComBat on an fMRI dataset to extract insights into discriminative connectivity patterns. Further, to classify ASD and typical development (TD), an artificial neural network (ANN) was utilized. Ke et al. [10] used fMRI and CNN to classify autism. Heinsfeld et al. [25] used autoencoders to extract lower-dimensional features and applied unsupervised learning to the features. The results showed an anti-correlation of brain activity between anterior

and posterior areas, corroborating the evidence of anterior–posterior disturbance in brain connectivity for ASD patients. Lakmini et al. [26] transformed 4D fMRI data into 2D images and applied inception V3 for classification. Lakmini et al. [27] conducted a survey on TL techniques for ASD prediction using fMRI images and optimized hyperparameters like learning rate and epochs with an accuracy of 98.38%. Mostafa et al. [28] proposed an approach that utilizes T1-weighted MRI images for predicting ASD through the application of a convolutional autoencoder (CAE) with an accuracy of 96.2%. Lakshmi et al. [29] applied different TL techniques on fMRI data; among all methods, VGG16 achieved the highest accuracy.

Manoj et al. [30] introduced morphological distance-related features (MDRFs), which were then used as input to three classifiers, RF, SVM, and MLP, to assess classification performance in distinguishing ASD from TD. Nogay et al. [31] conducted two quadruple and one octal classification task using deep learning on preprocessed sMRI scans. The images were cropped with Canny Edge Detection and augmented fivefold, and optimal CNN models were identified through grid search optimization. Sachdev et al. [32] employed fMRI-derived connectivity matrices to extract features, which were analyzed using logistic regression and refined with an AND operation to retain only statistically significant features. These selected features were then classified with an MLP for distinguishing ASD from TD. Kong et al. [33] transformed multi-site fMRI data into glass brain images, extracted features using LeNet-5, and constructed partial correlation matrices. Feature selection and classification were carried out with an MLP, while dataset heterogeneity was mitigated using the Split–Merge–Split (SMS) partitioning method.

ASD is a highly heterogeneous condition with complex neural underpinnings. Neuroimaging data, on its own, may struggle to capture subtle brain structure or function abnormalities specific to ASD, leading to diagnostic limitations. ASD has highly individualized brain features with no single biomarker for all individuals. This makes it difficult for deep learning models to capture the diverse patterns associated with ASD, leading to mixed results in accuracy. Incorporating subjective measurement (like behavior ratings) alongside the neuroimaging data will create a balanced assessment.

2.3. Analysis of Multimodal Images

Multimodal data refers to the data from different modalities such as image, text and video data. Dcouto et al. [39] conducted a survey on existing multimodal data for ASD prediction. Wang et al. [15] proposes a multimodal ASD diagnosis method based on a deep convolutional neural network (DeepGCN). It integrates whole-brain functional connectivity from fMRI and non-imaging demographic data to construct a population graph, achieving 77.27% accuracy. Han et al. [12] introduced a multimodal diagnosis framework that integrates both EEG and eye-tracking data. A stacked denoising autoencoder is used for feature extraction, and an ANN is used for ASD classification with an accuracy of 95.56%. Wang et al. [13] designed an FL-based framework that addresses privacy concerns and utilizes a hypergraph for multimodal feature fusion. The framework is applied to multi-site ABIDE data, including fMRI and non-imaging demographic data for ASD prediction. Moon et al. [16] used audio and video recordings of autistic adolescents in the VR environment. A random forest algorithm is applied for the final classification of the combined data. However, the authors focused on dynamic behavioral features, and the absence of missing neurobiological insight leads to lower accuracy (87.8%). Liu et al. [35] used multimodal data, and features were extracted using functional connectivity matrices and multi-scale DeepGCNs. These features, combined with non-imaging data, were used for ASD diagnosis via node classification, achieving an accuracy of 91.62%. Dong et al. [36] trained graph convolutional networks (GCNs), edge-variational graph convolutional

networks (EV-GCNs), fully connected networks (FCNs), autoencoders followed by fully connected networks (AE-FCNs), and an SVM on ABIDE functional connectivity, structural, and phenotypic data to classify ASD versus TD participants. Dong et al. [38] extracted sMRI features using FreeSurfer and combined them with phenotypic variables. Feature selection was applied to sMRI, while phenotypic features were concatenated directly, and classifiers were evaluated using a standardized cross-validated pipeline. Chen et al. [37] proposed an MKLF-RAG framework that integrates rs-fMRI and sMRI features from ABIDE using RNN-based (DRNN) and GRU-based (DGRU) feature selection. Selected functional and structural features are fused via a multi-kernel learning fusion (MKLF) strategy and classified using fully connected layers, enabling both ASD diagnosis and pathogenic ROI identification. Lou et al. [40] introduced SDR-Former, a framework designed specifically to address the scalability issues of multi-phase liver lesion classification. Unlike traditional feature-fusion methods, SDR-Former employs a Siamese Neural Network (SNN) architecture that allows for weight sharing across varying numbers of input phases.

Existing ASD research includes a wide array of prediction methods, with a significant portion relying on unimodal data from sources such as functional MRI (fMRI), diffusion tensor imaging (DTI), or behavioral assessments alone, while a growing number of studies incorporate multimodal approaches. While these unimodal studies have been foundational, they are often limited in their ability to capture the full spectrum of the disorder. For example, neuroimaging alone may not correlate strongly with the severity of behavioral symptoms, and behavioral data can lack objective biological grounding. The proposed framework distinguishes itself by focusing on a particular combination of sMRI and detailed phenotype data, providing a comprehensive analysis that bridges both neurobiological and clinical dimensions. Specifically, the proposed approach aims to understand the intricate relationship between brain structure and ASD symptom patterns, a connection that is less explored in the literature and adds significant depth to multimodal diagnostics beyond simple performance metrics.

3. Proposed Methodology

To address the challenges of diagnosing ASD using heterogeneous data, we proposed a Multimodal ASD (MMASD) model. The high-level end-to-end workflow is illustrated in Figure 1, which depicts the progression of the pipeline. The corresponding detailed architecture and modules are presented in Figure 2. sMRI slice extraction and preprocessing are applied to the imaging data, while phenotype data undergo preprocessing and encoding, as described in Section 3.1 and illustrated in Figure 2a. To learn latent representations, feature extraction is performed on the preprocessed sMRI data using a deep network, and dimensionality reduction is applied to the phenotype data, as discussed in Section 3.2 and shown in Figure 2b,c. The extracted sMRI and phenotype features are then combined through multimodal feature fusion, as described in Section 3.3 and depicted in Figure 2d. Finally, ensemble-based classification using bootstrap aggregation and boosting techniques is applied to the fused features, as discussed in Section 3.4 and shown in Figure 2e. A voting classifier is employed to leverage the diversity of multiple models and generate the final prediction of ASD or TD. During the training stage, all components of the proposed architecture, including the ResNet-50 feature extractor, the autoencoder, and the ensemble classifiers, are optimized using 80% of the available data. The remaining 10% of the data is used for validation and hyperparameter tuning, while the final 10% is reserved for testing [41]. Algorithm 1 explains the step-by-step process for the proposed method.

In this work phenotype data ($i \in [1, n]$) are represented as $Dp = [D_{p1}, D_{p2}, D_{p3}, ...D_{pn}]$ and sMRI data as $Ds = [D_{s1}, D_{s2}, D_{s3}, ..D_{sn}]$. D_{p_i}, D_{si} represents the phenotype and sMRI data of the i^{th} person.

Algorithm 1 Algorithm for MMASD.

Input: Input training dataset $D = x, y$
Output: Prediction of ASD

Preprocess the datasets D_p and D_s
Apply appropriate feature selection techniques on D_p and D_s
Apply Fusion operation (F) on the selected features from D_p and D_s datasets as follows F = $D_p \oplus D_s$
Set all the model parameters according to the input data.
Train all the base classifiers
for $n \leftarrow 1$ **to** N **do**
    Learn output based on data D
end for
Calculate the accuracies for the training models.
for $i \leftarrow 1$ **to** $Classifiers$ **do**
    Apply majority voting as discussed in Section Majority Voting
end for
Make decision based on Majority Voting

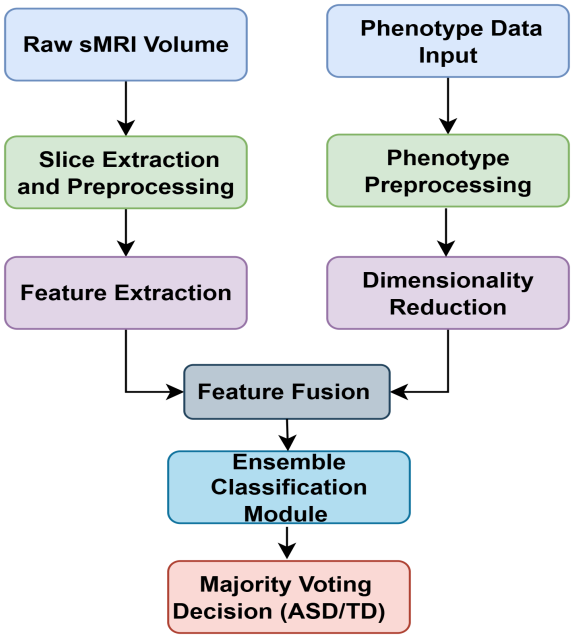


Figure 1. Workflow of proposed approach.

3.1. Data Extraction and Preprocessing

This section describes the procedures used for slice extraction and preprocessing of the sMRI data, along with the preparation of the associated phenotype information.

3.1.1. sMRI Slice Extraction

sMRI is the most used neuroimaging technique for ASD research. It gives more precise details about the brain anatomy, including the brain regions' shape, size, and volume. sMRI analysis in ASD discloses structural abnormalities in brain regions such as the hippocampus, amygdala, and prefrontal cortex, which are related to emotion processing, executive functioning, and social cognition, which are often impaired in individuals with ASD [42]. In the ABIDE dataset, the data files are available in the Neuroimaging Informatics Technology Initiative (NIfTI) format (.nii) that stores neuroimage data. In sMRI scans, data is presented in three distinct planes depicting various aspects of the brain anatomy: axial, coronal, and sagittal, as shown in Figure 3. Each of these planes offers unique perspectives, allowing for comprehensive examination and visualization of brain structures. In this work,

we use the axial plane because it does not suffer from drastic anatomical variances and offers a representative image of the brain [43,44].

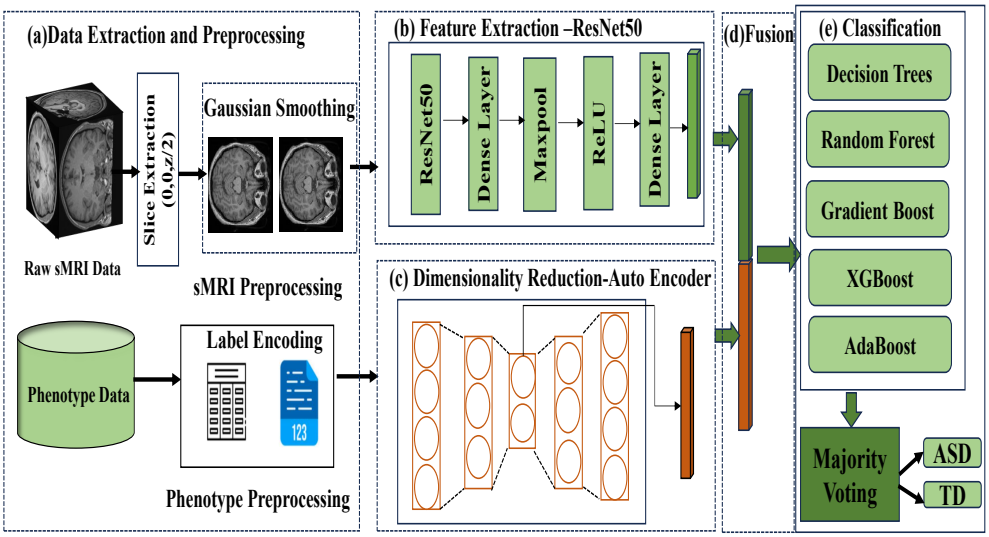


Figure 2. Proposed MMASD architecture for multimodal ASD prediction. The framework follows a linear pipeline consisting of (**a**) data extraction and preprocessing of phenotype and sMRI inputs, (**b**) feature transformation using an ResNet50 (**c**) dimensionality reduction using autoencoders, (**d**) early multimodal feature fusion via concatenation in a shared latent space, and (**e**) ensemble-based classification using bagging and boosting models with a final majority voting decision for ASD/TD prediction.

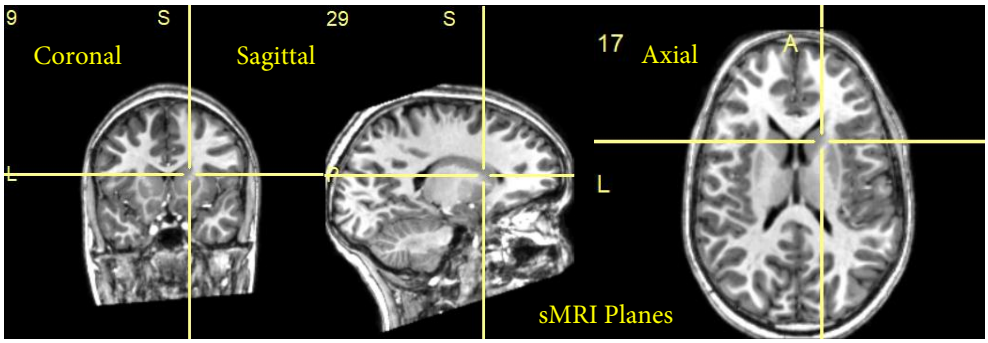


Figure 3. sMRI Three planes: coronal, sagittal, and axial. Orientation markers indicate A = Anterior, L = Left, and S = Superior.

In brain imaging analysis, it is customary to begin by extracting the central slice along the Z-dimension (axial plane). The axial plane, also known as the transverse plane, provides cross-sectional views of the brain from top to bottom. These images are useful for examining structures such as the ventricles, basal ganglia, and white matter tracts. We accessed the image at coordinates (x, y, and z/2) to retrieve the middle slice of the axial plane, where z is the number of slices along the z-axis. These extracted middle slices are stored in Portable Network Graphics (PNG) format. The brain's middle axial plane consists of structures implicated in ASD. The selected middle axial slice intersects several key structures implicated in ASD, including periventricular regions, basal ganglia, thalamic areas, and cortical gray–white matter interfaces. sMRI study will elucidate the structural changes in brain anatomy [45].

3.1.2. sMRI Preprocessing

Preprocessing of sMRI images is crucial for enhancing quality, dependability, and interpretability. It helps to improve the model's performance and is also essential to

determining the structure and functioning of the brain. sMRI preprocessing consists of image resizing, smoothing, and enhancement. After acquiring the image from the NIfTI, the generated images are resized to 224 × 224. Additionally, image smoothing and enhancement are applied. Image smoothing, often called blurring, is a process used to decrease an image's noise or detail. Nearby pixels go through a mathematical procedure to achieve a softer or more uniform appearance. The equation of the Gaussian smoothing technique for the proposed approach is given in Equation (1). Furthermore, contrast enhancement is applied to enhance the contrast between an image's light and dark areas.

$$G(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

here, x is the distance, and σ is the distribution.

3.1.3. Phenotype Preprocessing

The problem with raw data is that it contains missing values or noisy data and categorical variables. To achieve high accuracy, the preprocessing is performed on the phenotype dataset.

Handling missing values: In the medical context, substituting missing values with the mean or median is deemed inappropriate [46]. Specifically for phenotype data, we choose to eliminate features when the count of missing values closely approximates γ, a hyperparameter. Let C represent the set of columns, while R denotes the number of rows, and M signifies the set of missing values. C_d is the dataset after handling the missing values. The drop operation is mathematically expressed in Equation (2).

$$C_{\mathrm{d}} = \{c \in C \mid \frac{|M_c|}{R} < \gamma\} \tag{2}$$

here, $|M_c|$ is the count of missing values in column c, and γ is a constant representing the threshold for considering the ratio of missing values to the number of rows. Columns with a ratio below this threshold are retained in C_{d}.

Handling categorical variables: The feature set Dp comprises categorical features. To handle these features, the label encoding algorithm is employed.

Let S be the set of categories and f be a categorical feature in Dp. The label encoding function $L_f : S \to \mathbb{N}$ maps each category to a unique natural number as discussed in Equation (3).

$$L_f(s_i) = \{i, \quad \text{for } i = 1, 2, \ldots, |S|\} \tag{3}$$

here, s_i represents the i-th category in S, and $|S|$ is the total number of unique categories. The label encoding assigns a unique numerical label to each category, facilitating the representation of categorical data with numerical values.

3.2. Feature Extraction and Dimensionality Reduction

MMASD considers feature extraction on sMRI data and dimensionality reduction on phenotype data to obtain salient features and reduce the computational time.

We used RestNet [47] for feature extraction from sMRI data. Further, we performed a comparative analysis with a CNN [48] and VGG [49] as presented in Table 2. From Table 2, we can infer that the ResNet50 with decision trees gives better accuracy. The reason for ResNet's success is that, compared to the CNN and VGG, it can handle the challenges involved in training very deep neural networks, leading to better performance.

It skips some intermediate layers and allows activations on one layer to relate to activations on succeeding layers [47]. ResNet-50 is composed of multiple convolutional layers organized into residual blocks, which facilitate the training of deep networks by alleviating

the vanishing gradient problem. Each residual block learns a residual mapping by adding the input of the block to the output of a sequence of convolutional operations, thereby preserving low-level information while enabling the learning of high-level representations. Mathematically, given an input feature map **x**, a residual block computes:

$$y = \mathcal{F}(x, W) + x, \tag{4}$$

where $\mathcal{F}(\cdot)$ denotes the residual function consisting of convolution, batch normalization, and ReLU activation, and W represents the learnable parameters of the block. In the proposed approach, the original classification head of ResNet-50 is removed, and the network outputs a latent embedding vector representing high-level anatomical features extracted from the sMRI data. During training, the parameters of the ResNet-50 network are optimized using the cross-entropy loss function, defined as

$$\mathcal{L}_{\text{ResNet}} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i), \tag{5}$$

where y_i and $\hat{y}_i$ denote the ground-truth label and the predicted probability for the i-th sample, respectively. This loss function guides the network to learn feature representations that are discriminative for ASD and TD classification. The feature extraction process based on the ResNet-50 architecture is illustrated in Figure 4. A detailed comparison between the original ResNet-50 architecture and the modified ResNet-50 employed in the proposed framework is provided in Supplementary Section S1, where the architectural differences and corresponding modifications are explicitly described. The features extracted from ResNet50 are further fused with the phenotype data.

Table 2. Results for feature extraction techniques on sMRI data.

FE Technique	Model	Accuracy	Precesion	Recall	F1-Score
CNN	XGBoost	0.81	0.79	0.82	0.80
	Decision Tree	**0.84**	**0.81**	**0.86**	**0.84**
	Gradient Boost	0.77	0.76	0.78	0.77
	AdaBoost	0.74	0.72	0.74	0.73
	Random Forest	0.69	0.67	0.71	0.69
VGG16	XGBoost	0.79	0.80	0.74	0.77
	Decision Tree	**0.85**	**0.85**	**0.85**	**0.85**
	Gradient Boost	0.79	0.78	0.77	0.78
	AdaBoost	0.73	0.71	0.75	0.73
	Random Forest	0.6	0.60	0.5	0.54
ResNet50	XGBoost	0.82	0.80	0.84	0.82
	Decision Tree	**0.87**	**0.86**	**0.88**	**0.87**
	Gradient Boost	0.84	0.83	0.83	0.83
	AdaBoost	0.78	0.76	0.78	0.77
	Random Forest	0.70	0.73	0.62	0.67

Note: Bold value indicates the model with the highest accuracy.

The dimensionality reduction techniques include PCA [50], t-SNE [51], UMAP, TriMap, PacMAP, and autoencoders [52]. All the results for these models are compared to determine the most effective approach. This comparison aims to identify the best approach to dimen-

sionality reduction while maintaining significant details and improving performance. The results presented in Table 3 show that the autoencoder outperforms the other approaches. It is a parametric model with learned weights and biases, enabling fast and stable inference for new patients via a simple forward pass. In contrast, UMAP and TriMAP are graph-based methods optimized for fixed datasets, where mapping unseen samples is computationally complex and less stable. So, we chose the autoencoder as the dimensionality reduction method for phenotype data in the proposed approach.

Table 3. Performance comparison of different dimensionality reduction techniques on ABIDE phenotype data.

DR Technique	Classifier	Accuracy	Precision	Recall	F1-Score
Autoencoder	**XGBoost**	**0.92**	**0.91**	**0.93**	**0.92**
	Decision Tree	0.90	0.88	0.93	0.90
	Gradient Boost	0.90	0.90	0.90	0.90
	AdaBoost	0.83	0.82	0.83	0.82
	Random Forest	0.81	0.81	0.81	0.81
PCA	XGBoost	0.77	0.73	0.75	0.76
	Decision Tree	**0.81**	**0.80**	**0.80**	**0.80**
	Gradient Boost	0.80	0.82	0.75	0.78
	AdaBoost	0.71	0.72	0.66	0.68
	Random Forest	0.61	0.61	0.53	0.57
t-SNE	**XGBoost**	**0.87**	**0.86**	**0.87**	**0.87**
	Decision Tree	0.86	0.87	0.84	0.86
	Gradient Boost	0.83	0.82	0.83	0.83
	AdaBoost	0.80	0.79	0.81	0.80
	Random Forest	0.79	0.78	0.78	0.78
UMAP	**XGBoost**	**0.87**	**0.86**	**0.87**	**0.87**
	Decision Tree	0.88	0.90	0.85	0.87
	Gradient Boost	0.83	0.81	0.84	0.83
	AdaBoost	0.80	0.79	0.81	0.80
	Random Forest	0.79	0.77	0.78	0.79
TriMAP	**XGBoost**	**0.88**	**0.87**	**0.89**	**0.88**
	Decision Tree	0.87	0.86	0.88	0.87
	Gradient Boost	0.86	0.85	0.87	0.86
	AdaBoost	0.81	0.80	0.82	0.81
	Random Forest	0.80	0.79	0.81	0.80
PaCMAP	**XGBoost**	**0.89**	**0.88**	**0.90**	**0.89**
	Decision Tree	0.88	0.87	0.89	0.88
	Gradient Boost	0.87	0.86	0.88	0.87
	AdaBoost	0.82	0.81	0.83	0.82
	Random Forest	0.80	0.79	0.81	0.80

Note: Bold value indicates the model with the highest accuracy.

In the autoencoder, the number of features is reduced to 20 because it yields better accuracy performance when compared with all the features, as shown in Table 4. To select the number of features to reduce from the phenotype data, we conducted experiments with different sets of features; the accuracy with 20 features was almost similar to all features. So, we reduced the number of features to 20. The distribution of phenotype data before and after applying dimensionality reduction is presented in Figure 5. The autoencoder is used to convert the data into lower-dimension space. It reduces the number of features by using a feed-forward neural network model. The autoencoder operation on phenotype data D_p is given in mathematical Equations (6) (encoder) and Equation (7) (decoder), respectively

$$D_{penc} = \phi_{enc}(v) = Tanh(W_{enc} * V + B_{enc}) \quad (6)$$

where W_{enc} and B_{enc} are the weight matrix and bias for the encoder. *Tanh* is the hyperbolic tangent function. V is the reduced data after applying autoencoders

$$V = \phi_{dec}(D_{penc}) = (W_{dec} * D_{enc}) + B_{dec} \quad (7)$$

On the decoder side, D_{penc} is phenotype encoded data, W_{dec} is the weight matrix, and B_{dec} is bias for the decoder.

$$\mathcal{L}_{rec} = \frac{1}{N}\sum_{i=1}^{N}\left\|V_i - \hat{V}_i\right\|_2^2 \quad (8)$$

where $\mathcal{L}_{rec}$ denotes the reconstruction loss used to train the autoencoder. The term V_i represents the original phenotype feature vector of the i-th sample, and $\hat{V}_i$ denotes the reconstructed output generated by the decoder for the same sample. The operator $\|\cdot\|_2^2$ corresponds to the squared Euclidean (L2) norm, which measures the reconstruction error between the input data and its reconstruction. The parameter N denotes the total number of training samples. By minimizing $\mathcal{L}_{rec}$, the autoencoder is encouraged to learn a compact and informative latent representation that preserves the essential characteristics of the phenotype data.

Table 4. Evaluating feature performance based on autoencoder.

DR Technique	No. of Features	Accuracy
Auto Encoders	27	0.90
	25	0.91
	23	0.91
	21	0.91
	20	**0.92**

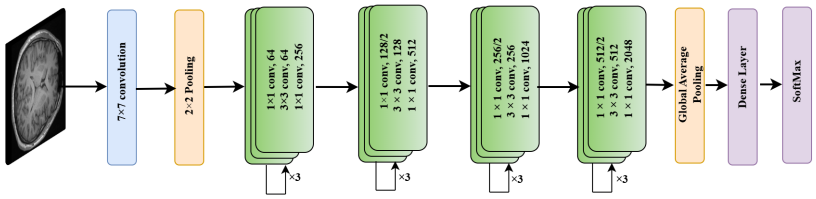


Figure 4. Feature extraction using ResNet50.

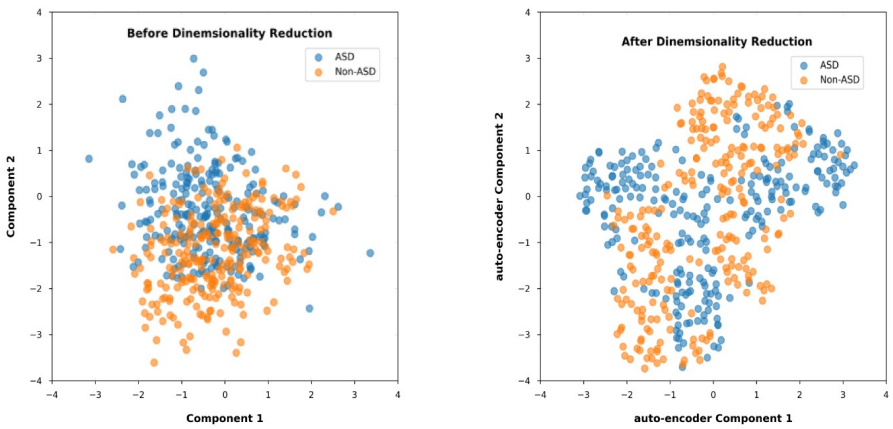


Figure 5. PCA-based assessment of the autoencoder module's impact on feature separability.

In the proposed autoencoder, the latent representation is explicitly used as the output feature vector for multimodal fusion. The decoder is employed only during training to enforce a reconstruction constraint and is discarded during inference (as discussed in Supplementary Section S2).

3.3. Feature Fusion

The proposed framework adopts an early fusion strategy, selected primarily to enable direct interaction between sMRI-derived structural features and phenotype variables at the feature level. This approach is particularly critical for ASD prediction, where clinical attributes (e.g., age, sex, and behavioral scores) are closely coupled with brain morphology. By combining modalities before classification, early fusion allows the model to learn joint representations that capture complementary and correlated information across modalities.

Let $D_p \in \mathbb{R}^{n \times d_p}$ and $D_s \in \mathbb{R}^{n \times d_s}$ denote the phenotype and sMRI feature matrices, respectively, where n represents the number of subjects, and d_p and d_s are the dimensionalities of each modality. A unified multimodal representation is obtained by projecting each modality-specific feature space into a common latent embedding space and subsequently performing a concatenation-based fusion. Formally, let $\phi_p : \mathbb{R}^{d_p} \to \mathbb{R}^{d'_p}$ and $\phi_s : \mathbb{R}^{d_s} \to \mathbb{R}^{d'_s}$ represent learnable or fixed feature transformation functions (e.g., linear projections or neural embeddings) applied to each modality:

$$\tilde{D}_p = \phi_p(D_p), \quad \tilde{D}_s = \phi_s(D_s), \tag{9}$$

resulting in modality-specific embeddings $\tilde{D}_p \in \mathbb{R}^{n \times d'_p}$ and $\tilde{D}_s \in \mathbb{R}^{n \times d'_s}$. The final fused representation $F \in \mathbb{R}^{n \times (d'_p + d'_s)}$ is obtained via concatenation in a shared latent space:

$$F = \tilde{D}_p \oplus \tilde{D}_s, \tag{10}$$

where $\oplus$ denotes vector concatenation along the feature dimension. This formulation preserves modality-specific information while providing a unified multimodal representation for downstream classification.

3.4. Classification Module

In the MMASD approach, bagging and boosting techniques such as decision trees, random forest, XGBoost, AdaBoost, and gradient boost are used for classification.

Decision Tree Classifier: A decision tree is a tree-like model of decisions and their possible consequences [53]. It is a set of recursive rules, based on these rules, the class is identified. Decision trees for left and right are shown in Equation (11) and Equation (12), respectively.

$$D(n) = \begin{cases} D(L_c), & \text{if } F > x \\ D(R_c), & \text{otherwise} \end{cases} \tag{11}$$

$$D(n) = \begin{cases} D(R_c), & \text{if } F \leq x \\ D(L_c), & \text{otherwise} \end{cases} \tag{12}$$

here, D is the decision tree, L_c is the left child, R_c is the right child, n is the node, F is the feature or attribute, x is a threshold value, and CL is the class label, If n is a leaf node then D(n) = CL.

Random Forest Classifier: Multiple decision trees are trained on various data subsets to create an ensemble of decision trees, and the final decision is obtained by aggregating the decisions of these individual trees [54]. The prediction process involves the calculation of gini, entropy, and information gain as presented in Equations (13)–(16).

$$\text{Gini}(rt) = 1 - \sum_{i=1}^{l} x_i^2 \tag{13}$$

$$\tau(t) = -\sum_{i=1}^{l} x_i \log_2(x_i) \tag{14}$$

$$I_g = \tau(P) - \sum_{i=1}^{k} \frac{S_{C_i}}{S_P} \times \tau(C_i) \tag{15}$$

$$\hat{y} = \text{mode}(y_1, y_2, \ldots, y_T) \tag{16}$$

here, $\tau(t)$ is the entropy, I_g is information gain, x_i is the proportion of the samples, S_{C_i} is samples in children, S_p represents samples in the parent node, and y_i is the prediction class.

AdaBoost Classifier: AdaBoost operates by training learners on a training set iteratively [55]. Initially, all the weights of the model are equal. As each iteration progresses, the weights of samples increase. This strategy compels the learner to pay attention to the challenging samples within the training set. After every iteration, the weak learner is merged with the learners to create a stronger learner. The final prediction made by the AdaBoost model is determined through a sum of the predictions made by these learners presented in Equation (17) and Equation (18), respectively. w_i is weak classifier weights updated by using

$$w_i^{(l+1)} = \frac{w_i^{(t)} \cdot \exp(-\alpha_t \cdot y_i \cdot h_t(x_i))}{\sum_{j=1}^{N} w_j^{(t)}} \tag{17}$$

for final classifier

$$H(x) = \sin\left(\sum_{t=1}^{T} \alpha_t \cdot h_t(x)\right) \tag{18}$$

here, $Df(x_1, y_1)$ is considered as initial samples of the dataset and $Df(x_n, y_n)$ as n^{th} instances. Then it takes the weights α_t of all the samples $w(i) = \frac{1}{n}$, here i = 1, …, n. $h_t(x_i)$ is the weights for weak classifier, and T is the total weak classifiers in the model.

Gradient Boost Classifier: The gradient boosting approach is an ensemble technique wherein the training data pass through the model, then the weak learners are passed to the next decision tree, and the process is repeated until it correctly classifies the training data and gradually reduces the error of the model [56]. The gradient boost is represented in Equations (19) and (20).

$$E_m(i) = E_{m-1}(i) + \eta \cdot h_m(i) \tag{19}$$

$$F(i) = \sum_{m=1}^{M} \eta \cdot h_m(i) \tag{20}$$

where $E_m(i)$ is the current ensemble model, $E_{m-1}(i)$ the ensemble model from the previous round, η is the learning rate, and $h_m(i)$ is the weak learner.

XGBoost Classifier: XGBoost is an ensemble-based algorithm. In the XGBoost approach, weights are allocated to the input data, and then they are fit to the decision tree to get the predictions. The approach uses extreme gradients to enhance the weak classifier. It is a step-by-step additive method. The incorrect predictions are then passed to the next decision tree. The process continues until the model has low bias and low variance. The mathematical expression for XGBoost is presented in Equation (21).

The initial model is generated by fitting data to a weak classifier [57].

$$O(\theta) = \sum_{i=1}^{n} L(y_{act}, y_{pred}) + \sum_{p=1}^{P} \Omega(f_p) \tag{21}$$

here, $O(\theta)$ is the objective function, y_{act} and y_{pred} are the actual and predicted values of the classification problem, and $\Omega(f_p)$ is the penalty term for regularization.

Majority Voting

The majority voting algorithm is used to stack a group of classifiers and make decision based on the majority voting of each classifier. The algorithm is represented as $y_a^{(b)}$, where a denotes the instance and b represents the classifier index. The final predication ϕ_a is as expressed in Equation (22).

It formalizes the decision aggregation strategy employed by the ensemble framework. Let ϕ_a denote the final predicted class label for a specific data instance a. The ensemble consists of M distinct base classifiers (in this study, $M = 5$), where the variable b serves as the iterator representing the b-th classifier in the ensemble. The term $y_a^{(b)}$ represents the predicted class label output by the b-th classifier for the input instance a. The core of the voting mechanism relies on the indicator function $V(\cdot)$, which effectively acts as a binary vote counter. It is defined such that

$$V(y_a^{(b)} = c) = \begin{cases} 1 & \text{if the } b\text{-th classifier predicts class } c \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

The summation $\sum_{b=1}^{M} V(y_a^{(b)} = c)$ calculates the total number of votes (or "vote tally") received for a specific class candidate c across all M classifiers. Finally, the argmax operation evaluates this sum for all possible classes (e.g., $c \in \{\text{ASD}, \text{TD}\}$) and selects the class c that maximizes the vote count. This ensures that the final decision ϕ_a reflects the consensus of the majority, thereby minimizing the variance and bias associated with any single classifier's prediction.

The majority voting classifier combines predictions from diverse base classifiers, such as decision trees and random forests as shown in Figure 6. Each base classifier independently predicts whether a child is likely to have autism based on their behavioral and developmental features. The system then aggregates these predictions and selects the most common prediction as the outcome. Ensemble classifiers, such as majority voting, play a crucial role in enhancing predictive accuracy for autism diagnosis by combining the results of multiple algorithms. This approach has proven effective in various clinical settings, where integrating the outputs of different classifiers reduces the biases and limitations associated with individual models [58]. In the context of ASD prediction, ensemble methods have been shown to improve both reliability and robustness, offering a more balanced and accurate diagnostic outcome. By leveraging the strengths of multiple algorithms, majority voting classifiers help mitigate errors and provide a more comprehensive analysis, making them a valuable tool in supporting clinical decisions for ASD diagnosis [59]. The advantage of using the majority voting classifier in autism prediction lies in its ability to integrate insights from multiple base classifiers, potentially improving prediction accuracy and robustness. By considering the collective wisdom of diverse models, the process can help mitigate individual classifier biases and errors, leading to more reliable predictions. In our proposed approach, we encountered issues with overfitting and underfitting due to model complexity and data limitations. To address overfitting, we applied hyperparameter tuning to control model complexity, such as adjusting maximum depth and minimum samples per

split in individual models. For boosting methods like XGBoost and gradient boosting, we tuned parameters such as learning rate and number of estimators to enhance performance while avoiding noise. To counter underfitting, we optimized the number of estimators in the ensemble models, ensuring the model effectively captured relevant patterns in the data.

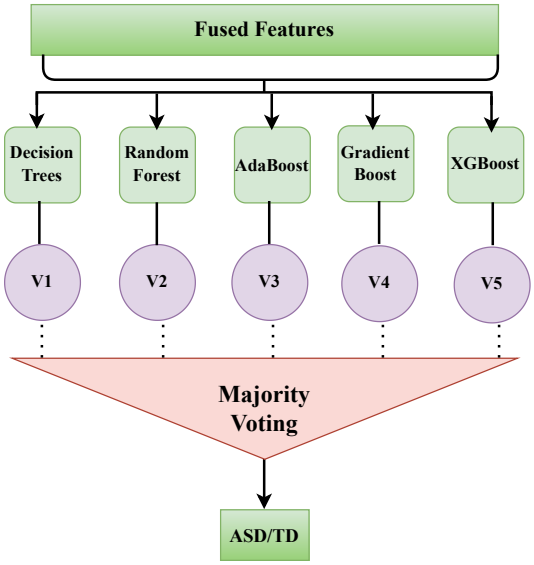


Figure 6. Overview of majority voting approach.

4. Performance Evaluation

4.1. Dataset Description

The proposed method considers the ABIDE-I dataset comprising phenotype, fMRI, and sMRI data of the same patients. The dataset consists of a total of 1100 data samples, with 570 representing typically developing (*TD*) individuals and 530 representing individuals with ASD, collected from 17 different universities. TD includes children with normal behavior, growth, and developmental patterns for their age group. The dataset is divided into train, test and validation splits in the ratio of 80, 10 and 10, respectively [41]. The ABIDE dataset project is generated through the connectome computation system pipeline (https://fcon_1000.projects.nitrc.org/indi/abide, Accessed on 23 October 2024). The phenotypic variables included in this study were age at scan, sex, full-scale IQ (FIQ), eye status at scan, PCP quality assessment protocol values, and manual quality assessment annotations. Age, sex, and FIQ were used to account for demographic and cognitive variability, while eye status and quality assessment measures were incorporated for data quality control and reliability. The diagnostic group served as the classification label. Representative axial slices of the preprocessed sMRI data, stratified by diagnostic group, are visualized in Figure 7. To demonstrate the heterogeneity of the multi-site cohort, the demographic distributions are further detailed in the subsequent figures. Specifically, Figure 8 illustrates the gender distribution across the various acquisition sites, while Figure 9 depicts the site-wise class balance between ASD subjects and TD controls.

4.2. MRI Acquisition Parameters

sMRI data were obtained from the ABIDE dataset, which comprises T1-weighted brain images acquired across multiple imaging sites using different MRI scanners and acquisition protocols. These parameters are not uniform across all subjects. In general, T1-weighted structural images were acquired using standard anatomical sequences such as Magnetization Prepared Rapid Gradient Echo (MPRAGE) or Spoiled Gradient Recalled Echo (SPGR), depending on the imaging site. Scans were collected on MRI systems from differ-

ent vendors, including Siemens, GE, and Philips, with magnetic field strengths primarily of 1.5 Tesla and 3 Tesla. Voxel resolutions also varied across sites, with most scans acquired at approximately 1 mm isotropic resolution or near-isotropic voxel sizes. Due to the multi-site nature of the ABIDE dataset, detailed acquisition parameters such as repetition time (TR), echo time (TE), flip angle, and exact voxel dimensions differ between centers and are not consistently available for all subjects. As a result, site-specific acquisition settings were not explicitly modeled in this study.

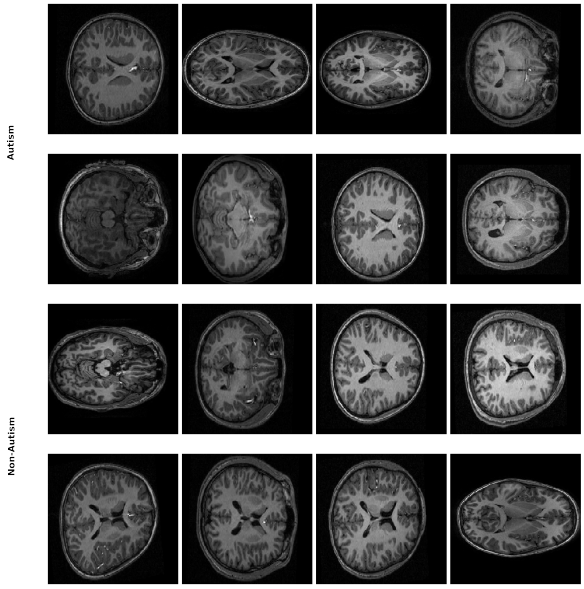


Figure 7. ASD and NonASD data.

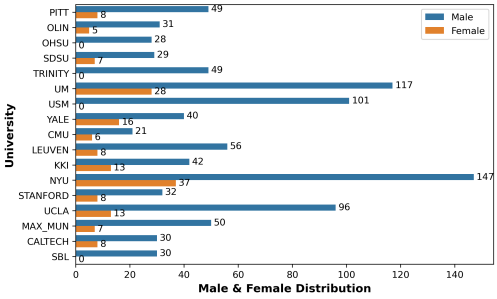


Figure 8. Gender distribution.

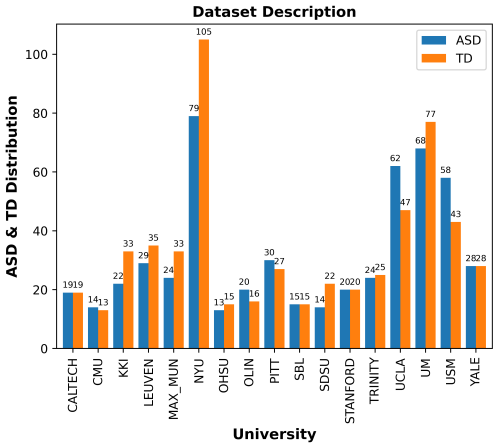


Figure 9. University-wise data description.

4.3. Environmental Setup and Evaluation Metrics

The proposed model was executed on an NVIDIA DGX-1 server at the School of Engineering and Sciences, SRM University-AP, India. This high-performance server features eight Tesla V100 GPUs, offering a substantial computational capacity with 40,960 CUDA cores and 64 GB of memory per GPU. This setup supports the efficient processing of large datasets and complex models during training and evaluation. The model was implemented using TensorFlow, a popular deep-learning framework for neural network development. Data manipulation was handled with NumPy (version 1.26.4) and Pandas (version 2.2.2), while Matplotlib (version 3.9.0) and Seaborn (version 0.13.2) were used for visualization.

The experimental analysis involved testing phenotype, sMRI, and fusion data. Dimensionality reduction techniques for the phenotype data include PCA, t-SNE, and autoencoders. The number of components was set to 20 in all approaches after conducting regressive experiments. Extracting features from sMRI data involved the utilization of a CNN, VGG, and ResNet. The CNN consists of 3 convolution layers, pooling layers, and batch normalization. In the case of VGG and ResNet, dense layer and pooling layers were added, and the final layer was omitted to extract the features. The ResNet architecture used in the MMASD approach consists of three dense layers and two max-pooling layers after the base ResNet. The dense layers (2048, 1024, and 512) with ReLU activations and max-pooling layers (2×2) downsample feature maps to enhance feature learning. The hyperparameters were tuned with regressive experiments in the bootstrap aggregation and boosting algorithms. To identify the optimal ensemble classifier configuration, we applied a systematic grid search over the hyperparameter ranges defined in Table 5. The configurations were evaluated on a validation partition to select the settings that maximized predictive performance. All hyperparameter optimization was conducted exclusively on the training data, with no access to the independent test set. Additionally, all data-dependent transformations, including feature scaling and autoencoder training, were fitted solely within the training folds. This strict separation prevents data leakage and ensures that the reported performance accurately reflects the model's generalization ability on unseen data.

Table 5. Hyperparameters for Bagging and Boosting approaches.

Method	Hyperparameters	Selected Parameters
Decision Trees	Depth: {3–7}, Criterion: {Entropy, Information Gain}	Depth 4, Criterion: Entropy
Random Forest	n_estimators: {100–150}, Criterion: {Entropy, Information Gain}, Depth: {4–6}	n_estimators 100, Criterion: Entropy, Depth 5
AdaBoost	n_estimators: {50, 70, 100} Learning rate: {0.1, 0.2, 0.3} Depth: {4–6}	n_estimators 50, Learning rate 0.3, Depth 5
Gradient Boost	n_estimators: {100, 120, 200} Learning rate {0.25, 0.20, 015} Depth: {4–6}	n_estimators 200, Learning rate 0.25, Depth 6
XGBoost	n_estimators: {100, 120, 150} Learning rate: {0.25, 0.20, 0.15} Depth: {4–6}	n_estimators 100, Learning rate 0.1, Depth 6

To evaluate the performance of the proposed approach, performance metrics such as accuracy (Acc), precision (P_r), recall (R_e), F1-score (F_1), and specificity (S_p) [60] are considered. The performance metrics are mathematically expressed as shown in Table 6. Where T_{ps} signifies the successful prediction of the positive class (ASD) distribution, while F_{ps} represents instances where the model incorrectly predicts the positive class distribution. On the other hand, T_{ns} denotes the correct prediction of the negative class (TD) distribution, and F_{ps} corresponds to instances where the model incorrectly predicts the negative class distribution. The model's accuracy, denoted as Acc, reflects its overall performance. It quantifies how well the model predicts the correct class labels by comparing the predicted outputs with the actual ground truth labels. The accuracy is computed by summing the number of correctly classified instances, which include true positives and true negatives, and dividing this sum by the total number of evaluated instances. Precision P_r gauges the ratio of correctly predicted positive instances T_P among all positive predictions, while recall R_e assesses the ratio of actual positives that were accurately classified. The F_1 score is the harmonic mean of precision and recall, and S_p measures the proportion of actual negatives that are correctly identified by the model and provides a comprehensive measure of the model's performance. To rigorously assess the framework's ability to generalize across varying scanner protocols and demographic distributions, we adopted a Leave-One-Site-Out Cross-Validation (LOSO-CV) strategy. In this scheme, data from one acquisition site are held out as an independent test set, while the model is trained using data from all remaining sites. This procedure is iteratively repeated so that each site serves as the test set exactly once. As a result, the reported performance metrics provide a robust evaluation of the model's resilience to site-specific heterogeneity and its potential for real-world deployment.

Table 6. Performance metrics for evaluating MMASD approach.

Performance Metrics	Equation
Acc	$\text{Acc} = \frac{T_{ps}+T_{ns}}{T_{ps}+T_{ns}+F_{ps}+F_{ns}}$
P_r	$P_r = \frac{T_{ps}}{T_{ps}+F_{ps}}$
R_e	$R_e = \frac{T_{ps}}{T_{ps}+F_{ns}}$
F_1 Score	$F_1 = \frac{2 \cdot P_r \cdot R_e}{P_r+R_e}$
S_e	$S_e = \frac{T_{ns}}{T_{ns}+F_{ps}}$

4.4. Comparative Analysis

By comparing the performance of models trained on individual feature sets with the model trained on fused features, we demonstrate the advantages of feature fusion. In this section, we present a baseline comparison of the proposed approach in Section 4.4.1, a comprehensive comparison with the multimodal approaches in Section 4.4.2, and a comparison with the unimodal approaches in Section 4.4.3.

4.4.1. Ablation Study

To ensure methodological rigor and address potential concerns regarding dataset variability, all classifiers and ensemble approaches in this study were evaluated on the exact same dataset, comprising the same set of subjects, identical preprocessing steps, and consistent training–test splits. This guarantees that performance differences are attributed solely to modeling and fusion strategies, rather than inconsistencies in data sources. Furthermore, a structured baseline comparison framework was established to provide a clear reference for evaluating the effectiveness of the proposed multimodal fusion and ensemble methods. Specifically, three primary baselines were defined: (1) phenotype-only classifiers,

(2) sMRI-only classifiers, and (3) multimodal fusion classifiers. Each baseline was trained and validated under identical conditions, ensuring a fair comparison across modalities and methods. Additionally, all reported metrics were computed uniformly, and statistical hypothesis testing was applied to confirm the significance of observed differences. These steps were taken to create a robust experimental design and eliminate ambiguity in performance interpretation.

Table 7 summarizes the comparative results across the different models and ensembles. XGBoost emerged as the top-performing classifier for phenotype data with an accuracy of 0.92, while the decision tree achieved the best performance for sMRI data with 0.87. Gradient Boost demonstrated superior results on fusion data, achieving an accuracy of 0.96. Notably, the multimodal fusion classifiers consistently outperformed their unimodal counterparts, highlighting the advantage of leveraging complementary information from both modalities. Furthermore, ensemble methods such as majority voting and stacking were systematically compared across all baselines, demonstrating stacking as the most effective strategy, achieving accuracies of 0.93, 0.88, and 0.97 for phenotype, sMRI, and fusion datasets, respectively. The use of hypothesis testing to validate these improvements underscores the statistical robustness of our findings, reinforcing that ensemble approaches provide meaningful performance gains over single-model classifiers. The p-value [61] is a crucial metric in this analysis, allowing us to assess the significance of our results. Based on results presented in Table 7, the obtained p-values are all below the standard significance level of 0.05; this statistical analysis confirms that our model's performance is both accurate and statistically significant, enhancing the reliability of our ASD prediction results. Table 8 summarizes the bootstrap (n = 1000) performance of the proposed majority voting ensemble when utilizing the fused multimodal features, presenting both point estimates and 95% confidence intervals. The framework achieved an exceptional sensitivity (recall) of 0.99 (95% CI: 0.9710–1.0000), demonstrating the model's critical ability to minimize false negatives and correctly identify ASD cases. Additionally, the narrow confidence intervals across accuracy (0.97) and F1-score (0.97) indicate that these results are statistically stable and not artifacts of data sampling. Overall, these metrics confirm that the fusion-based ensemble delivers the high reliability and precision necessary for clinical decision support systems.

Table 7. Results for individual and majority classifier.

Techniques	Data	Acc	(P_r)	(R_e)	(F_1)	(p-Value)
XGBoost	Phenotype	0.92	0.91	0.93	0.92	0.034
Decision Tree	sMRI	0.87	0.86	0.88	0.87	0.025
Gradient Boost	Fusion	0.96	0.93	0.99	0.96	0.00154
Majority Voting	Phenotype	0.93	0.99	0.96	0.96	-
	sMRI	0.88	0.87	0.88	0.87	-
	Fusion	0.97	0.95	0.99	0.97	-

Table 8. Performance metrics with 95% confidence intervals for majority voting (fusion).

Metric	Value	95% Confidence Interval	
		Lower	Upper
Accuracy	0.97	0.9400	1.0000
Sensitivity (Recall)	0.99	0.9710	1.0000
Precision	0.95	0.9092	0.9908
F1-score	0.97	0.9400	1.0000

Majority voting, where the predictions from multiple classifiers are aggregated, proved more effective in obtaining accuracy than any single algorithm alone. This illustrates the importance of combining the knowledge of several algorithms to improve overall classification accuracy. The confusion matrix is used to understand the performance of the model. The confusion matrix for phenotype data, sMRI data, and fusion data is illustrated in Figure 10. As per the results of the confusion matrix, it is easy to understand that the model performed better with fusion data compared to alone sMRI and phenotype data. Additionally, from Figure 10, we can infer that the model achieved 110 true positives and 105 true negatives, indicating a strong capability to correctly classify both positive and negative instances with only 1 false positive and 4 false negatives, the model demonstrates a high degree of accuracy, minimizing both incorrect alarms and missed positive cases. The area under the receiver operating characteristic curve (AUC-ROC) measure the model's ability to distinguish between the classes. The AUC-ROC curve for the MMASD approach is presented in Figure 11. Figure 12 illustrates the ROC analysis, where the proposed MMASD framework achieved a mean AUC of 0.972 ± 0.02. The narrow 95% confidence interval demonstrates minimal variance across bootstrap folds, indicating high statistical stability in the model's predictive performance. These results confirm that the ensemble approach maintains robust diagnostic sensitivity even when accounting for data variability.

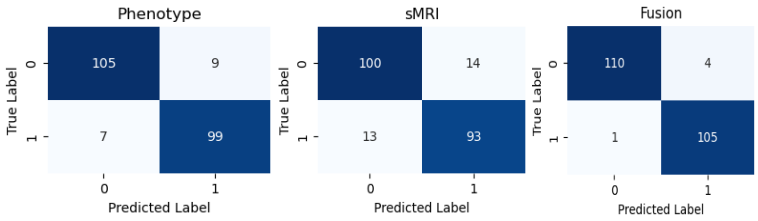


Figure 10. Confusion matrix for MMASD approach.

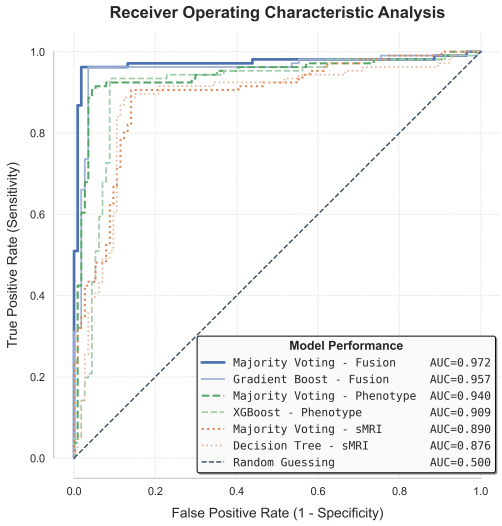


Figure 11. AUC-ROC for MMASD approach.

4.4.2. Comparison with Multimodal Approaches

This section provides a comparative analysis of our proposed approach with existing multimodal methods (Han et al. [12], Wang et al. [13], Wang et al. [15], Moon et al. [16], Liu et al. [35], Alharthi et al. [43], and Chen et al. [37]) for ASD prediction, summarized in Table 9. The authors of [12] applied a denoising autoencoder on EEG and eye-tracking data, effectively extracting noise-resistant features from the multimodal input. However, their single-method approach may fail to capture the full spectrum of insights available in multimodal data. The FL framework of Wang et al. [13] achieved a moderate accuracy of

73.52% and was likely limited by data heterogeneity across sources. In a separate study, the authors of [15] employed DeepGCN with fMRI and phenotype data, leveraging graph networks to capture relationships between brain activity and phenotype indicators. However, in our proposed work, accuracy is considerably higher (97.27% vs. 77.27%), suggesting that while GCNs are suitable for mapping network features, they may not capture the same level of detail from raw multimodal data as MMASD ensemble strategy. A random forest classifier [16] was applied to audio, video, and VR-generated data, achieving an accuracy of 87.8%. However, a limitation of using only a random forest classifier is its potential to overlook complex interactions between multimodal features. The authors also used DeepGCN [35] on fMRI and phenotype data. Although effective in capturing brain–phenotype relationships, this approach relies heavily on graph-based representations, which may not fully capture complementary information from non-graph-structured data. The ViT model was applied to the ABIDE NYU fMRI and sMRI dataset for ASD classification [43], but a limited dataset was used. The proposed approach leverages a more diverse dataset, enhancing robustness and generalizability in ASD classification. Chen et al. [37] utilized fMRI and sMRI data and employed a recurrent neural network (RNN) and gated recurrent unit (GRU) for feature extraction, fused via the MKLF algorithm and classified using neural networks. However, this approach is susceptible to the vanishing gradient problem, which may prevent convergence to a global minimum. Our proposed approach combines majority voting with sMRI and phenotype data, enabling more robust handling of data variability and a richer integration of multimodal features, thereby enhancing both accuracy and generalizability in ASD prediction.

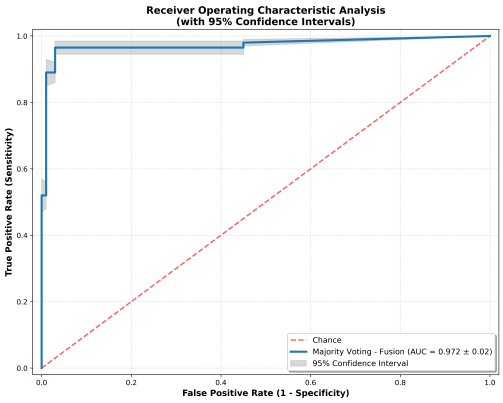


Figure 12. ROC for MMASD approach with 95% CI.

Table 9. Comparative analysis with multimodal approaches.

S. No.	Author	Data	Method	Accuracy
1	Han et al. [12]	EEG and Eye Tracking	Stacked autoencoder	95.56%
2	Wang et al. [13]	fMRI and Phenotype	Federated Learning	73.52%
3	Wang et al. [15]	fMRI and Phenotype	DeepGCN	77.27%
4	Moon et al. [16]	Audio and Video	Random Forest	87.8%
5	Liu et al. [35]	fMRI and Phenotype	DeepGCN	91.62%
6	Alharthi et al. [43]	fMRI and sMRI	ViT	87.12%
7	Chen et al. [37]	fMRI and sMRI	GNN,LSTM	81.48%
8	MMASD	sMRI and Phenotype	Resnet50 and majority voting	97.27%

4.4.3. Comparison with Unimodal Approaches

This section presents comparison of proposed work (fusion+majority voting) with existing approaches (Alharthi et al. [43], Devika et al. [44], Gao et al. [62], Mishra et al. [63], Kong et al. [64], and Chen et al. [37]). The comparative analysis with existing approaches is presented in Table 10. Devika et al. [44] employed spatiotemporal patterns extracted from sMRI images using a generative adversarial network (GAN) achieving an accuracy of 95%. While GANs capture complex spatiotemporal dependencies, they may struggle with modeling finer structural or temporal variances, which are very important for diagnosing ASD patients. Gao et al. [62] utilized morphological brain networks to extract deep features from an sMRI dataset and applied a basic CNN for feature extraction with an accuracy of 70%. Our proposed approach employs TL techniques and captures more refined and relevant features from sMRI data, leading to a significant enhancement in model performance. Deep convolutional neural networks (DCNNs) [63] on the sMRI dataset achieved an accuracy of 81.35%, whereas the proposed approach has an accuracy of 97.27%. The lack of hyperparameter tuning, high computational complexity, and single modality limit the performance and scalability of their model. Deep neural networks (DNNs) with individual brain networks as feature representation on sMRI data [64] achieved an accuracy of 90.39%. However, they relied on a small dataset, which limits the model's generalizability. The authors of [65] proposed a multi-model ensemble classifier (MMEC) on distinct planes of fMRI data and achieved an accuracy of 97.87%. However, they used unimodal data, which can result in bias by capturing only a limited aspect of brain function or structure and potentially overlooking critical information needed for a holistic understanding of ASD, while our approach utilizes multimodel data, capturing high-resolution anatomical features enhanced with individual characteristics, to build a comprehensive prediction model.

Table 10. Comparative analysis with unimodal approaches.

S. No.	Reference	Study	Acc	P_r	R_e	F_1
1	Devika et al. [44]	SGAN	95%	-	-	-
2	Gao et al. [62]	CNN-ResNet	70%	68.75%	81.25%	68.68%
3	Mishra et al. [63]	DCNN	81.35%	80.85%	79.95%	80.40%
4	Kong et al. [64]	DNN	90.39%	95.88%	84.37%	-
5	Herath et al. [65]	Ensemble learning	97.89%	97.47%	98.33%	97.76%
6	MMASD	Majority Voting	97.27%	95.45%	99.05%	97.22%

In the MMASD approach, we extract deep features from both modalities, namely, phenotype and sMRI data. Phenotype data offers crucial information that enables a more detailed assessment of a person's ASD behaviors, essential for adopting personalized treatments [66]. sMRI data is a feasible and reliable tool for ASD diagnosis because it is faster and safer due to low radiation and provides high anatomical accuracy. Phenotype and sMRI data together improve the precision and efficacy of ASD prediction models, resulting in more individualized and focused interventions for ASD patients. We adopted an early fusion mechanism to integrate information from both datasets, aiming to capture intra-modality correlations effectively. We employed a majority voting mechanism to predict autism, leveraging the majority rank from individual classifiers to make the final decision. This strategy allows for comprehensive utilization of information from both modalities, enhancing the robustness of the predictive model with an accuracy of 97.27%.

4.5. Discussion

We have proposed MMASD, a multimodal, majority-voting-based approach for ASD prediction that leverages both phenotype and sMRI data to enhance model accuracy and reliability. Phenotype data offers detailed clinical and behavioral information, including ADOS scores, IQ levels, demographics, and social responsiveness, collectively reflecting the diverse behavioral characteristics associated with ASD. Complementing this, sMRI data provides insights into neurobiological markers, such as cortical thickness, gray matter volume, and structural abnormalities in brain regions associated with social cognition and language processing. Our evaluation confirms that when used independently, phenotype and sMRI data yielded lower prediction accuracy than the combined multimodal approach, as detailed in Section 4.4.1. The use of a majority voting ensemble allows the model to aggregate predictions across multiple classifiers, balancing the unique strengths of each while minimizing individual biases. This ensemble method, combined with the rich multimodal data, results in a robust, sensitive, and specific ASD classification model that is more reliable and clinically applicable. The performance of the MMASD framework compares favorably to the current state of the art in the field. Many existing studies have focused on unimodal approaches, achieving notable success but also facing inherent limitations. For instance, models based solely on neuroimaging have demonstrated the potential of biomarkers but can be sensitive to variations in scanner protocols and preprocessing pipelines. Conversely, approaches relying exclusively on behavioral or clinical data may not capture the underlying neurodevelopmental patterns in individuals who present with atypical symptoms. Our work aligns with a growing consensus that multimodal integration is essential for tackling the heterogeneity of ASD. By successfully fusing neuroanatomical and phenotypic data, the MMASD framework achieves an accuracy of 97.27%, representing a significant advancement over methods that do not combine these complementary information streams.

The LOSO-CV results are detailed in Table 11, demonstrating consistently high performance when each acquisition site was held out for testing. Model accuracy remained robust, ranging from 0.95 to 0.98 across the validation folds. A key observation is the model's high recall, which approached unity in several folds (e.g., KKI, NYU, SBL, and UM), indicating an exceptional capacity for correctly identifying individuals with ASD. Aggregating across all sites, the MMASD framework achieved a final accuracy of 97.27%, with a precision of 95.45%, a recall of 99.06%, and an F1-score of 97.22%. The stability of these metrics provides strong evidence that the model is not overfitted to any single site's data characteristics and underscores its potential for robust performance in real-world clinical applications. The use of a single axial sMRI slice provides a computationally efficient and robust representation that enables stable training and evaluation on heterogeneous, multi-site datasets such as ABIDE. Despite its simplicity, a carefully selected axial slice can capture key anatomical patterns that are informative for ASD classification, allowing the model to achieve strong predictive performance while maintaining manageable computational and memory requirements. At the same time, extending the framework to multi-slice or full 3D volumetric representations is expected to further enhance performance by incorporating richer spatial context and three-dimensional anatomical relationships. Such extensions represent a promising direction for future work and are likely to improve accuracy as larger datasets and computational resources become available. To ensure the framework's direct applicability in clinical settings, we addressed critical practical challenges regarding data heterogeneity and acquisition variability. Missing values within the phenotypic records were managed through rigorous preprocessing and standardization protocols, ensuring consistent input representations without compromising sample size. Furthermore, the

inherent variability introduced by differing scanner hardware and research-versus-clinical protocols was mitigated by our standardized image normalization pipeline.

Table 11. Site-wise majority voting performance (Leave-One-Site-Out Cross-Validation).

Site	Accuracy	Precision	Recall	F1-Score
Caltech	0.96	0.94	0.98	0.96
CMU	0.97	0.95	0.99	0.97
KKI	0.98	0.96	1.00	0.98
Leuven	0.95	0.93	0.98	0.95
MaxMun	0.97	0.95	0.99	0.97
NYU	0.98	0.96	1.00	0.98
OHSU	0.97	0.95	0.99	0.97
Olin	0.96	0.94	0.98	0.96
Pitt	0.97	0.95	0.99	0.97
SBL	0.98	0.96	1.00	0.98
Stanford	0.96	0.94	0.99	0.96
Trinity	0.95	0.93	0.98	0.95
UCLA	0.97	0.95	0.99	0.97
UM	0.98	0.96	1.00	0.98
USM	0.97	0.95	0.99	0.97
Yale	0.96	0.94	0.98	0.96
SDSU	0.97	0.95	0.99	0.97
MMASD	0.97	0.95	0.99	0.97

For real-time implementation of the proposed model, we are working with the medical experts from SRM Hospital, Chennai, India. We are integrating the proposed MMASD model directly with the hospital's Electronic Health Record (EHR) system through a secure, API-based interface capable of ingesting multimodal patient data as inputs. This integration supports automated data preprocessing pipelines that standardize, align, and format heterogeneous data sources, enabling seamless and efficient communication between the EHR and the model. The system functions as a clinical decision-support tool, delivering insightful analyses to healthcare professionals while supporting informed decision-making without adding manual workload. For scalability and flexibility, we plan to deploy the model using either on-premise or cloud-based infrastructure, depending on the hospital's available resources. The observed improvements may be partly influenced by circularity; therefore, we explicitly acknowledge this limitation and interpret the findings with appropriate caution to ensure transparency and rigor in the analysis. Although the training phase involves complex optimization, the deployed model performs inference in milliseconds, making it fully compatible with standard clinical workstations. This computational investment is justified by the significant reduction in false negatives and the automation of manual feature extraction, which streamlines the diagnostic workflow.

5. Conclusions

Autism is a neurodevelopmental disorder, and the symptoms vary from one person to another. Early prediction of ASD significantly raises the quality of living among individuals with autism and their families. It can deal with age-related issues and enhance social interactions and regular life abilities. The existing unimodal approaches predict ASD with high accuracy. However, they lack a holistic picture of the disorder and are prone to biases in a single dataset. Multimodal data in ASD prediction is essential due to the diverse and varying symptoms exhibited by individuals. So, in this paper, we propose a multimodal approach by considering both the phenotype and sMRI data. Features from these modalities are extracted and then fused. This process involves capturing relevant information from each modality to integrate insights from multiple perspectives. In the

MMASD approach, majority voting is used for the final prediction of ASD to improve prediction accuracy and reliability by leveraging the collective wisdom of multiple models, making it valuable for early detection and intervention in children at risk of ASD.

Future Work

Our future work will focus on age-specific studies to create support systems and interventions targeted to different developmental stages. By focusing on distinct age groups, we aim to identify unique biomarkers and risk factors that emerge from infancy through adulthood. Furthermore, we aim to collect real-time data from medical and educational institutions to improve the reliability of our prediction models. Integrating real-time data streams will allow us to continuously monitor behavioral and environmental factors, providing a dynamic and comprehensive understanding of autism's early signs and progression. To facilitate clinical adoption and build trust in our model, we will also incorporate Explainable Artificial Intelligence (XAI) techniques. This will enable the framework to provide clinicians with clear insights into which specific neurobiological or behavioral features are driving its predictions for an individual. Further, we will extend this study by collecting and integrating larger and more diverse multimodal ASD datasets, incorporating additional structural MRI scans and enriched phenotypic information from multiple clinical sites. With increased sample size and sufficient computational resources, LLM-based and multimodal transformer models are expected to more effectively capture complex cross-modal relationships between neuroanatomical features and clinical attributes. We also plan to investigate strategies to improve modality alignment and reduce hallucination-related effects to enhance reliability in clinical settings. In addition, we will explore OverLoCK [67]-inspired architectural principles such as hierarchical context modeling and dynamic feature modulation within lightweight designs tailored for medical imaging. As larger, multimodal ASD datasets become available, these advanced architectures may enable improved representation learning, stronger cross-modal integration and enhanced generalization, thereby increasing their clinical applicability for ASD prediction.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/jsan15010021/s1, Section S1: ResNet50 Architecture Comparison; Section S2: Autoencoder Architecture Comparison.

Author Contributions: B.L.P.: Numerical experimentation, data analysis, writing—original draft. V.D.R.: Conceptualization, methodology development, supervision, writing—review and editing. M.K.M.: Methodology development, formal analysis, validation. M.M.H.: Software implementation, visualization, investigation. A.A.: Resources, validation, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this study is publicly available as part of the Autism Brain Imaging Data Exchange (ABIDE) initiative. The ABIDE datasets can be accessed at https://fcon_1000.projects.nitrc.org/indi/abide/, accessed on 14 October 2025.

Conflicts of Interest: The authors declare no competing interests.

References

1. International Classification of Diseases. ICD-10-CM Diagnosis. ICD10Data. 2024. Available online: https://www.cdc.gov/nchs/icd/icd-10-cm/index.html (accessed on 11 July 2025).
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)*; DSM-5 Report; American Psychiatric Association: Washington, DC, USA, 2021.
3. Kanner, L. Early infantile autism. *J. Pediatr.* **1944**, *25*, 211–217. [CrossRef]

4. Autism Speaks. Autism and Health: A Special Report. 2021. Available online: https://www.autismspeaks.org/what-autism/health/autism-health-special-report/ (accessed on 15 July 2025).
5. McMorris, C.A.; Cox, E.; Hudson, M.; Liu, X.; Bebko, J.M. The diagnostic process of children with autism spectrum disorder: Implications for early identification and intervention. *J. Dev. Disabil.* **2013**, *19*, 42.
6. Muhle, R.A.; Reed, H.E.; Stratigos, K.A.; Veenstra-VanderWeele, J. The emerging clinical neuroscience of autism spectrum disorder: A review. *JAMA Psychiatry* **2018**, *75*, 514–523. [CrossRef] [PubMed]
7. Kavitha, N.; Kasthuri, N. An efficient automatic diabetic retinopathy grading using a two-way cascaded convolution neural network. *Discov. Comput.* **2025**, *28*, 7. [CrossRef]
8. Farooq, M.S.; Tehseen, R.; Sabir, M.; Atal, Z. Detection of autism spectrum disorder (ASD) in children and adults using machine learning. *Sci. Rep.* **2023**, *13*, 9605. [CrossRef]
9. Abdelwahab, M.M.; Al-Karawi, K.A.; Hasanin, E.; Semary, H. Autism spectrum disorder prediction in children using machine learning. *J. Disabil. Res.* **2024**, *3*, 20230064. [CrossRef]
10. Ke, F.; Choi, S.; Kang, Y.H.; Cheon, K.A.; Lee, S.W. Exploring the structural and strategic bases of autism spectrum disorders with deep learning. *IEEE Access* **2020**, *8*, 153341–153352. [CrossRef]
11. Bosl, W.J.; Tager-Flusberg, H.; Nelson, C.A. EEG analytics for early detection of autism spectrum disorder: A data-driven approach. *Sci. Rep.* **2018**, *8*, 6828. [CrossRef]
12. Han, J.; Jiang, G.; Ouyang, G.; Li, X. A multimodal approach for identifying autism spectrum disorders in children. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2003–2011. [CrossRef]
13. Wang, H.; Jing, H.; Yang, J.; Liu, C.; Hu, L.; Tao, G.; Zhao, Z.; Shen, N. Identifying autism spectrum disorder from multi-modal data with privacy-preserving. *npj Ment. Health Res.* **2024**, *3*, 15. [CrossRef]
14. Lord, C.; Elsabbagh, M.; Baird, G.; Veenstra-Vanderweele, J. Autism spectrum disorder. *Lancet* **2018**, *392*, 508–520. [CrossRef] [PubMed]
15. Wang, M.; Guo, J.; Wang, Y.; Yu, M.; Guo, J. Multimodal autism spectrum disorder diagnosis method based on DeepGCN. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 3664–3674. [CrossRef] [PubMed]
16. Moon, J.; Ke, F.; Sokolikj, Z.; Chakraborty, S. Applying multimodal data fusion to track autistic adolescents' representational flexibility development during virtual reality-based training. *Comput. Educ. X Real.* **2024**, *4*, 100063. [CrossRef]
17. Rani, P.; Kumar, P.; Babulal, K.S.; Kumar, S. Interpolation of Erythrocytes and Leukocytes Microscopic image using MsR-CNN with Yolo v9 Model. *Discov. Comput.* **2025**, *28*, 18. [CrossRef]
18. Akter, T.; Satu, M.S.; Khan, M.I.; Ali, M.H.; Uddin, S.; Lio, P.; Quinn, J.M.; Moni, M.A. Machine learning-based models for early stage detection of autism spectrum disorders. *IEEE Access* **2019**, *7*, 166509–166527. [CrossRef]
19. Raj, S.; Masood, S. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Comput. Sci.* **2020**, *167*, 994–1004. [CrossRef]
20. Vakadkar, K.; Purkayastha, D.; Krishnan, D. Detection of autism spectrum disorder in children using machine learning techniques. *SN Comput. Sci.* **2021**, *2*, 386. [CrossRef]
21. Hossain, M.D.; Kabir, M.A.; Anwar, A.; Islam, M.Z. Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets. *Health Inf. Sci. Syst.* **2021**, *9*, 17. [CrossRef]
22. Sherkatghanad, Z.; Akhondzadeh, M.; Salari, S.; Zomorodi-Moghadam, M.; Abdar, M.; Acharya, U.R.; Khosrowabadi, R.; Salari, V. Automated detection of autism spectrum disorder using a convolutional neural network. *Front. Neurosci.* **2020**, *13*, 1325. [CrossRef]
23. Gao, K.; Sun, Y.; Niu, S.; Wang, L. Unified framework for early stage status prediction of autism based on infant structural magnetic resonance imaging. *Autism Res.* **2021**, *14*, 2512–2523. [CrossRef]
24. Ingalhalikar, M.; Shinde, S.; Karmarkar, A.; Rajan, A.; Rangaprakash, D.; Deshpande, G. Functional connectivity-based prediction of Autism on site harmonized ABIDE dataset. *IEEE Trans. Biomed. Eng.* **2021**, *68*, 3628–3637. [CrossRef] [PubMed]
25. Heinsfeld, A.S.; Franco, A.R.; Craddock, R.C.; Buchweitz, A.; Meneguzzi, F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *Neuroimage Clin.* **2018**, *17*, 16–23. [CrossRef] [PubMed]
26. Herath, L.; Meedeniya, D.; Marasingha, M.; Weerasinghe, V. Autism spectrum disorder diagnosis support model using Inception V3. In *Proceedings of the 2021 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 16 September 2021*; IEEE: New York, NY, USA, 2021; Volume 4, pp. 1–7.
27. Herath, L.; Meedeniya, D.; Marasingha, J.; Weerasinghe, V. Optimize transfer learning for autism spectrum disorder classification with neuroimaging: A comparative study. In *Proceedings of the 2022 2nd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 23–24 February 2022*; IEEE: New York, NY, USA, 2022; pp. 171–176.
28. Mostafa, S.; Wu, F.X. Diagnosis of autism spectrum disorder with convolutional autoencoder and structural MRI images. In *Neural Engineering Techniques for Autism Spectrum Disorder*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 23–38.

29. Lakshmi, P.B.; Reddy, V.D.; Ghosh, S.; Sengar, S.S. Classification of Autism Spectrum Disorder Based on Brain Image Data Using Deep Neural Networks. In *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 209–218.
30. Manoj, G.; Gupta, V.; Bhattacharya, A.; Aleem, S.G.A.; Vedantham, D.; Prince A, A.; Agastinose Ronickom, J.F. Diagnostic classification of autism spectrum disorder using sMRI improves with the morphological distance-related features compared to morphological features. *Multimed. Tools Appl.* **2025**, *84*, 4979–5000. [CrossRef]
31. Nogay, H.S.; Adeli, H. Multiple classification of brain MRI autism spectrum disorder by age and gender using deep learning. *J. Med. Syst.* **2024**, *48*, 15. [CrossRef]
32. Sachdeva, J.; Mittal, R.; Mehta, J.; Jain, R.; Ranjan, A. Resolving autism spectrum disorder (ASD) through brain topologies using fMRI dataset with multi-layer perceptron (MLP). *Psychiatry Res. Neuroimaging* **2024**, *343*, 111858. [CrossRef]
33. Kang, L.; Chen, M.; Huang, J.; Xu, J. Identifying autism spectrum disorder based on machine learning for multi-site fMRI. *J. Neurosci. Methods* **2025**, *416*, 110379. [CrossRef]
34. Kang, J.; Han, X.; Song, J.; Niu, Z.; Li, X. The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Comput. Biol. Med.* **2020**, *120*, 103722. [CrossRef]
35. Liu, S.; Wang, S.; Sun, C.; Li, B.; Wang, S.; Li, F. DeepGCN based on variable multi-graph and multimodal data for ASD diagnosis. *CAAI Trans. Intell. Technol.* **2024**, *9*, 879–893. [CrossRef]
36. Dong, Y.; Batalle, D.; Deprez, M. A framework for comparison and interpretation of machine learning classifiers to predict autism on the ABIDE dataset. *Hum. Brain Mapp.* **2025**, *46*, e70190. [CrossRef]
37. Chen, J.; Zhang, H.; Zou, Q.; Liao, B.; Bi, X.A. Multi-kernel Learning Fusion Algorithm Based on RNN and GRU for ASD Diagnosis and Pathogenic Brain Region Extraction. *Interdiscip. Sci. Comput. Life Sci.* **2024**, *16*, 755–768. [CrossRef]
38. Dong, Y.; Batalle, D.; Deprez, M. Reproducible comparison and interpretation of machine learning classifiers to predict autism on the ABIDE multimodal dataset. *medRxiv* **2024**. [CrossRef]
39. Dcouto, S.S.; Pradeepkandhasamy, J. Multimodal Deep Learning in Early Autism Detection—Recent Advances and Challenges. *Eng. Proc.* **2024**, *59*, 205.
40. Lou, M.; Ying, H.; Liu, X.; Zhou, H.Y.; Zhang, Y.; Yu, Y. Sdr-former: A siamese dual-resolution transformer for liver lesion classification using 3d multi-phase imaging. *Neural Netw.* **2025**, *185*, 107228. [CrossRef] [PubMed]
41. Joseph, V.R. Optimal ratio for data splitting. *Stat. Anal. Data Mining ASA Data Sci. J.* **2022**, *15*, 531–538. [CrossRef]
42. Ecker, C.; Suckling, J.; Deoni, S.C.; Lombardo, M.V.; Bullmore, E.T.; Baron-Cohen, S.; Catani, M.; Jezzard, P.; Barnes, A.; Bailey, A.J.; et al. Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: A multicenter magnetic resonance imaging study. *Arch. Gen. Psychiatry* **2012**, *69*, 195–209. [CrossRef] [PubMed]
43. Alharthi, A.G.; Alzahrani, S.M. Multi-Slice Generation sMRI and fMRI for Autism Spectrum Disorder Diagnosis Using 3D-CNN and Vision Transformers. *Brain Sci.* **2023**, *13*, 1578. [CrossRef]
44. Devika, K.; Mahapatra, D.; Subramanian, R.; Oruganti, V.R.M. Outlier-based autism detection using longitudinal structural MRI. *IEEE Access* **2022**, *10*, 27794–27808. [CrossRef]
45. Piven, J.; Bailey, J.; Ranson, B.J.; Arndt, S. An MRI study of the corpus callosum in autism. *Am. J. Psychiatry* **1997**, *154*, 1051–1056. [CrossRef]
46. Sterne, J.A.; White, I.R.; Carlin, J.B.; Spratt, M.; Royston, P.; Kenward, M.G.; Wood, A.M.; Carpenter, J.R. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ* **2009**, *338*, b2393. [CrossRef]
47. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
48. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; IEEE: New York, NY, USA, 2016; pp. 770–778.
50. Labrín, C.; Urdinez, F. Principal component analysis. In *R for Political Data Science*; Chapman and Hall/CRC: London, UK, 2020; pp. 375–393.
51. Belkina, A.C.; Ciccolella, C.O.; Anno, R.; Halpert, R.; Spidlen, J.; Snyder-Cappione, J.E. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nat. Commun.* **2019**, *10*, 5415. [CrossRef] [PubMed]
52. Wang, Y.; Yao, H.; Zhao, S. Auto-encoder based dimensionality reduction. *Neurocomputing* **2016**, *184*, 232–242. [CrossRef]
53. Song, Y.Y.; Ying, L. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatry* **2015**, *27*, 130.
54. Jackins, V.; Vimal, S.; Kaliappan, M.; Lee, M.Y. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *J. Supercomput.* **2021**, *77*, 5198–5219. [CrossRef]
55. Schapire, R.E. Explaining adaboost. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.

56. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
57. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
58. Kline, A.; Wang, H.; Li, Y.; Dennis, S.; Hutch, M.; Xu, Z.; Wang, F.; Cheng, F.; Luo, Y. Multimodal machine learning in precision health: A scoping review. *NPJ Digit. Med.* **2022**, *5*, 171. [CrossRef]
59. Tanveen, N.; Trisha, J.; Sravani, A.D. Advancing autism spectrum disorder diagnosis through ensemble learning. *i-Manag. J. Artif. Intell. Mach. Learn.* **2024**, *2*, 29–38.
60. Canbek, G.; Sagiroglu, S.; Temizel, T.T.; Baykal, N. Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. In *Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 5–8 October 2017*; IEEE: New York, NY, USA, 2017; pp. 821–826.
61. García-Berthou, E.; Alcaraz, C. Incongruence between test statistics and P values in medical papers. *BMC Med. Res. Methodol.* **2004**, *4*, 13. [CrossRef]
62. Gao, J.; Chen, M.; Li, Y.; Gao, Y.; Li, Y.; Cai, S.; Wang, J. Multisite autism spectrum disorder classification using convolutional neural network classifier and individual morphological brain networks. *Front. Neurosci.* **2021**, *14*, 629630. [CrossRef]
63. Mishra, M.; Pati, U.C. A classification framework for Autism Spectrum Disorder detection using sMRI: Optimizer based ensemble of deep convolution neural network with on-the-fly data augmentation. *Biomed. Signal Process. Control* **2023**, *84*, 104686. [CrossRef]
64. Kong, Y.; Gao, J.; Xu, Y.; Pan, Y.; Wang, J.; Liu, J. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. *Neurocomputing* **2019**, *324*, 63–68. [CrossRef]
65. Herath, L.; Meedeniya, D.; Marasinghe, J.; Weerasinghe, V.; Tan, T. Autism spectrum disorder identification using multi-model deep ensemble classifier with transfer learning. *Expert Syst.* **2022**, *42*, e13623. [CrossRef]
66. Moridian, P.; Ghassemi, N.; Jafari, M.; Salloum-Asfar, S.; Sadeghi, D.; Khodatars, M.; Shoeibi, A.; Khosravi, A.; Ling, S.H.; Subasi, A.; et al. Automatic autism spectrum disorder detection using artificial intelligence methods with MRI neuroimaging: A review. *Front. Mol. Neurosci.* **2022**, *15*, 999605. [CrossRef]
67. Lou, M.; Yu, Y. OverLoCK: An Overview-first-Look-Closely-next ConvNet with Context-Mixing Dynamic Kernels. In *Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 10–17 June 2025*; IEEE: New York, NY, USA, 2015; pp. 128–138.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.