

NeuroASD-Net: A Deep Learning-Based Approach for Autism Detection from Structural MRI

Harendra Sharma

Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, NCR Campus, Delhi-NCR Campus, Delhi-Meerut Road, Modinagar, Ghaziabad, U.P., India

E-mail: harendra.pth@gmail.com

ORCID iD: <https://orcid.org/0000-0002-1158-6559>

Oshin Sharma*

Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, NCR Campus, Delhi-NCR Campus, Delhi-Meerut Road, Modinagar, Ghaziabad, U.P., India

E-mail: oshins1@srmist.edu.in

ORCID iD: <https://orcid.org/0000-0003-1416-1596>

*Corresponding author

Received: May 9, 2026; Revised: June 20, 2026; Accepted: August 1, 2026; Published: October 8, 2026

Abstract: Autism spectrum disorder (ASD), a complex neurodevelopmental disorder, is typified by social interaction challenges, communication difficulties, and repetitive activities. Effective intervention requires an early and precise diagnosis. This paper introduces NeuroASD-Net, a novel framework for ASD detection from structural Magnetic Resonance Imaging (MRI). The approach integrates advanced preprocessing techniques such as histogram matching and Contrast Limited Adaptive Histogram Equalization (CLAHE) to standardize input data. The next stage is the segmentation of key brain regions, like cerebrospinal fluid, white matter, and gray matter, using k-Means clustering. This segmentation step isolates critical regions for subsequent analysis. Following segmentation, a Two-Level Feature Extraction (TLFE) model is applied. In the first level, spatial and textural patterns are extracted using Local Binary Patterns (LBP), Gabor filters, and the Gray Level Co-occurrence Matrix (GLCM), capturing essential morphological characteristics of the brain. Feature selection is then optimized using the Self-Adaptive Black-Winged Kite Algorithm (SA-BKA). In the second level, high-level features are extracted using EffiDSR-Net, in which the EfficientNet model is enhanced with Dual Scale Residual (DSR) Blocks and a Convolutional Block Attention Module (CBAM). The features from both levels are fused to form a comprehensive feature set. Finally, classification is performed using a Softmax layer, achieving precise ASD detection. The proposed framework demonstrates enhanced diagnostic accuracy, achieving an Accuracy of 99.28%, Sensitivity of 99.31%, Specificity of 99.24%, F1-score of 99.31%, and an Area Under the ROC Curve (AUC) of 0.995, indicating its effectiveness for clinical ASD detection.

Index Terms: Autism Spectrum Disorder, NeuroASD-Net, k-Means Clustering, Self-Adaptive Black-Winged Kite Algorithm, EffiDSR-Net.

1. Introduction

ASD is a neurodevelopmental disorder marked by impairments in social interaction, restricted interests, and repetitive behavior [1]. It is a heterogeneous neurodevelopmental disorder with a strong genetic basis, which presents itself in several ways. Repeated behavior, problems with verbal and non-verbal communication skills and impairments in social interaction are some of the characteristics of ASD [2]. Individuals with ASD exhibit diverse patterns in learning, behavior, and attention, with significant variability in their abilities and support needs. While some people can lead independent lives, others need constant care and support throughout their lifetime [3]. ASD symptoms are commonly noticed within the first three years of life and most often linger throughout a lifetime. In addition to intellectual disability, ASD contributes to severe challenges in a person's speech, feelings, actions, self-control, education, and memory [4]. ASD is a developmental disorder affecting social interaction, conduct, and performance.

Early screening and diagnosis are important for therapy, family support, educational preparation, and timely medical care for children with ASD to ensure that appropriate support is provided and care is extended to children with such disorders [5].

MRI diagnosis of ASD is time-consuming and difficult. To help professionals diagnose ASD, artificial intelligence (AI) has been developed [6]. The traditional diagnostic processes take a long time; it usually takes several months to confirm with many neurology and speech professionals [7]. The process of determining ASD in children is complex and multifaceted, and current methods usually rely on a single modality with limited data [8]. Facial expression analysis provides a solution because early identification is challenging due to brain function issues [9]. Very little study has been done on immunological anomalies in the very early stages of life, such as altered neonatal cytokine/chemokine profiles in relation to an aberrant developmental trajectory. Real-world performance is accurately reflected by conventional evaluations, which often lack objectivity [10].

The Convolution Neural Network (CNN) design, which is suited for high-dimensional data interpretation, combines layers that are fully linked, convolutional, pooling, batch-normalization, and dropout [11]. [12] used longitudinal T1-weighted structural MRI to identify illnesses across sessions. [13] proposed a framework for automatic detection based on machine learning and these features. Using the pre-trained VGG16 model, the transfer learning technique is employed to benchmark and validate deep learning's (DL) potential in neuroimaging data analysis [13]. The CNN and Transfer Learning Techniques were assessed in conjunction with data from the Autism Brain Imaging Data Exchange's ABIDE I and ABIDE II datasets. The creation of diagnostic biomarkers for brain disease is made possible by the four-dimensional nature of rs-fMRI data [14]. Because convolution neural networks (CNN) are more accurate than standard statistical methods, researchers combine genetic algorithms with artificial intelligence (AI) techniques for training [15].

This paper focuses on the recognition of ASD through DL and statistical feature extraction methods combined with the use of MRI data. Initial steps include preprocessing and a segmentation step in MRI scans to isolate the regions of interest in the brain. Advanced statistical feature extraction using LBP, Gabor filters, and GLCM captures substantial abnormalities in the brain texture and spatial features. The most advanced DL model, named EfficientNet, uses superior feature extraction capabilities with parameter efficiency and is applied to classification purposes. Statistical analysis supports deep learning as a strong framework for ASD diagnosis with better classification accuracy. The following are the paper's primary contributions:

- Two-Level Feature Extraction (TLFE) is performed. First Level extracts spatial and textural features using Gabor filters, Local Binary Patterns (LBP), and the Gray Level Co-occurrence Matrix (GLCM) to capture essential morphological and structural patterns of the brain.
- The second stage employs an advanced EfficientNet model equipped with Dual Scale Residual (DSR) Blocks and a Convolutional Block Attention Module (CBAM), providing effective feature extraction at a higher level.
- The proposed framework makes use of a Self-Adaptive Black-Winged Kite Algorithm (SA-BKA) to provide efficient feature selection.
- In addition, the main contributions of this work are explicitly summarized as follows:
- A novel hybrid multi-level feature extraction approach that combines texture features with DL-based features to better classify ASD.
- An efficient feature selection technique based on the Self-Adaptive Black-Winged Kite Algorithm (SA-BKA) that improves discrimination ability.
- A modified version of EfficientNet model called EffiDSR-Net, which uses DSR and CBAM attentions.
- Extensive experiments conducted using multisite MRI data that show improved results against other machine learning and DL approaches.

The paper is organized as follows: A review of relevant work is presented in Section 2. In Section 3, the proposed methodology is introduced. Section 4 outlines the experimental setup, datasets used, and presents the results of the proposed framework. Lastly, Section 5 concludes the paper.

2. Literature Review

Jain et al. [16] suggested merging a deep CNN, the Dwarf Mongoose optimized Residual Network (DM-ResNet), to classify autism illnesses from MRI brain scans. The process includes feature extraction, segmentation, preprocessing, and hyperparameter optimization using the DM optimization algorithm. The Gaussian Mixture Model (GMM) and Fuzzy C Means (FCM) are used in the feature extraction procedure, and the DM-ResNet categorizes functional connectivity features based on Region of Interest.

Devika et al. [17] proposed the use of an outlier identification technique for the ASD diagnosis utilizing structural magnetic resonance imaging (sMRI). Using only the sMRI images of healthy subjects, a GAN is trained to discover spatiotemporal patterns of structural brain connections. The ASD sMRI scan reconstructions are then recognized as outliers by the GAN discriminator when the GAN generator reconstructs the following three slices on a stack of three slices that is provided as input. Two more baselines are used to compare this model: a more straightforward UNet and a

more intricate Self-Attention GAN.

Nogay and Adeli [18] proposed a two-phase automatic autism diagnostic framework using sMRI. In the first phase, the photos are preprocessed, unclear ones are removed, Canny edge detection (CED) is utilized to identify the edges and cropped to the right size, and then data augmentation is applied five times to enlarge the dataset. This strategy prevents manipulation of data and ensures discrimination both for normal development and autism spectrum disorders (ASD). The hyperparameters of the Deep CNN (DCNN) are optimized by employing the grid search optimization technique in the second step. Superiority of the model is demonstrated through comparison with previous literature and pre-trained models, and dependability is demonstrated through five-fold cross-validation.

Yakolli et al. [19] utilized the Autism Brain Imaging Data Exchange (ABIDE) repository data in the development and implementation of CNN and ML models in order to categorize people with ASD. The graphical user interface, Gradio, is used to visualize the data analysis. The paper analyses machine learning models with clinical data to screen toddlers for ASD, using methodologies of DT, RF, and LR. The proposed methodology will enable early detection and diagnosis as it allows a more efficient diagnosis of even minute characteristics. In this regard, DL models undergo training, using fMRI and s-MRI pictures to differentiate between ASD and non-ASD cases. Certain performance indicators, including accuracy, ROC curve, and AUC, are applied in measuring the capacity of a system concerning classification. A doctor can easily use the automated method that would determine whether a picture is normal or has ASD.

Belhaouari et al. [20] presented an innovative approach of time-frequency decomposition called Progressive Fourier Transform (PFT), integrated into a CNN, for the enhancement of the ASD diagnosis. Since PFT holds the potential of tracing features in the fields of time and frequency, the network may extract richer, more discriminative patterns to be classified. With this reliable substitute (CNN+PFT), the novel approach outperforms methods traditionally used in finding ASD regarding its precision and efficiency. This method uses the complementary strengths of CNN and PFT in order to amplify the precision of diagnosis while keeping computational economy intact.

Koc et al. [21] aimed at improving the diagnostic sensitivity of ASD by using multiple neuroimaging modalities. A combination of resting-state functional MRI and sMRI data from the ABIDE database is used for categorizing participants with ASD and healthy controls according to ML approaches. Brain connection features are identified that characterize the disorder. The data are denoised and transformed into 3D s-MRI pictures for the removal of confounding variables. Hybrid convolutional recurrent neural networks perform better than convolutional or recurrent neural networks. Overall, the model is diagnostic, where people are classified as either autistic or not from the indicators of functional and anatomical connections based on the standard of the ADOS.

Oruganti et al. [22] proposed an outlier-rediscovery-based method utilizing a structural magnetic resonance imaging (sMRI)-based approach to identify ASD, based on the idea that a GAN can be trained on sMRI images of healthy patients to find out spatial-temporal patterns in structural brain connections. Reconstructions of ASD sMRI scans are identified as outliers by the GAN generator, which also reconstructs nearby slices. Two baselines are used to compare the model: a more straightforward UNet and a more intricate Self-Attention GAN. Numerous tests demonstrate that, with minimal training data, the ASD detection system outperforms the state-of-the-art.

Wang et al. [23] proposed using fMRI to identify ASD by applying the method known as Multi-site Clustering and Nested Feature Extraction (MC-NFE). After dividing multi-site training data into groups of healthy controls and those with ASD, it learns latent representations utilizing a linear reconstruction model that is motivated by similarity. Using a nested singular value decomposition (SVD) approach, inter-site heterogeneity is decreased and FC features are extracted. A linear Support Vector Machine (SVM) is also used in the method to identify ASD.

Sharif and Khan [24] presented an ML-based method for automatically diagnosing ASD, employing characteristics extracted from the cerebral brain volume and the corpus callosum. By selecting characteristics that are most crucial for discriminatory abilities for ASD classification, the recommended method has not only produced high identification accuracy, but it has also made ML model training easier. Further, the presented findings from this study demonstrate the value of the transfer learning approach in evaluating and validating deep learning's ability to analyze neuroimaging data. To achieve this, we have classified ASD using the pre-trained VGG16 model.

Alkahtani et al. [25] recognized autistic children using facial landmark detection by utilizing various transfer learning strategies found in deep CNNs. To increase the CNN model's prediction accuracy, an empirical investigation is performed to determine the optimal optimizer and hyperparameter setups. A transfer learning method, such as hybrid VGG19 and MobileNetV2, uses many ML methods like random forest (RF), logistic regression (LR), linear SVM, decision trees (DT), gradient boosting, MLP Classifier, and K-nearest neighbors (KNN). The proposed study's results show that MobileNetV2 transfer learning techniques outperform those developed in existing systems. Table 1 compares the existing papers related to ASD.

Table 1. Comparison of the literature review papers

Authors' Names	Aim	Methods	Advantages	Disadvantages
Jain et al. [16]	To develop an efficient and accurate DL framework to use MRI brain scans for the early identification of ASD.	FCM, GMM, DM-ResNet, CNN	applicability for real-world scenarios.	High Computational Cost
Devika et al. [17]	To propose a reliable, efficient, and interpretable method for ASD detection	GAN	Efficient Outlier Detection, Scalability	Interpretability Limitations, High Model

				Complexity
Nogay, and Adeli, [18]	To develop an automatic diagnostic model for ASD based on sMRI.	GSO, DCNN, CED	Generalizability, Automated Feature Extraction	Single-Modality Limitation, No Real-Time Application
Yakolli et al. [19]	To develop and implement ML and DL models for the early diagnosis and ASD classification.	DT, LR, RF, CNN	Scalability and Extensibility	No Support for Large Datasets
Belhaouari et al. [20]	To develop an advanced and reliable diagnostic system for ASD	CNN, PFT	High Classification Accuracy	Limited Real-time Applicability
Koc et al. [21]	To improve the precision of diagnosing ASD by utilizing multiple neuroimaging modalities	CNN, RNN	Large and Diverse Dataset	Potential Overfitting, Complexity of Implementation
Oruganti et al. [22]	To create a powerful and efficient system framework for ASD diagnosis using sMRI	UNet, GAN	Large and Diverse Dataset	Complexity of Implementation, Overfitting
Wang et al. [23]	To develop a robust and interpretable framework, MC-NFE, for diagnosing ASD	SVD, SVM	Handles High-dimensional Data	Dependence on Quality of Clustering
Sharif, and Khan [24]	To develop an efficient ML-based system for automatically identifying ASD using neuroimaging data.	VGG16	Feature Selection and Reduced Complexity	require large datasets for effective model training
Alkahtani et al. [25]	To create and assess a DL-based system for ASD early detection using facial landmark analysis of children.	MobileNetV2, VGG19, SVM, RF, KNN	Cost-effective, consistency and reliability	Limited Dataset

Despite significant advancements, existing ASD detection methods exhibit several limitations. Most techniques adopt single-level feature extraction, which cannot adequately account for the structural differences in brain MRI images. Moreover, deep learning frameworks may not be sufficiently interpretable and robust for heterogeneous and multisite imaging datasets. In contrast, conventional manual feature extraction methods, though interpretable, have limited expressive power. Lastly, most of the literature does not consider feature selection optimization techniques, resulting in redundancy and poor classification accuracy.

Research Gap

The highly commendable research on ASD classification is still somewhat restrictive in its use of one dataset, which may not necessarily apply to all children with ASD who are of various ages, races, and socioeconomic classes. The model would still need more development in order to attain certain standards in terms of efficiency and accuracy, which can be deemed clinically acceptable. There may be some other factors pertaining to ASD that have been overlooked simply because the facial landmarks being studied do not take into account such parameters, thereby leading to inaccurate results. It does not investigate the capability of the model to differentiate between ASD and other developmental disorders. It also considers the possibility of bias in the dataset. Large validation datasets and diverse data are very important for generalizability. In view of these gaps, the goal of this paper is to formulate an effective ASD diagnosis system that makes full use of both types of features, handcrafted features and deep learning-based features. The main motivation behind this research paper is to incorporate multi-level feature extraction, improve the generalizability of the dataset, and optimize feature selection to enhance classification performance. This motivation drives the design of the NeuroASD-Net framework.

3. Proposed Methodology

The proposed NeuroASD-Net technique uses structural MRI for detecting ASD in an organized way. The complex pre-processing approach begins with the histogram matching and CLAHE to enhance image quality and normalize the input data properly. After the completion of pre-processing, the k-means clustering algorithm is used for region segmentation, such as gray matter, white matter, and cerebrospinal fluid. Lastly, a TLFE model is employed. At the first level, LBP, Gabor filter, and GLCM algorithms are utilized for spatial and texture feature extraction in particular morphological patterns. In selecting the most optimal features, SA-BKA is employed. In the second level, EffiDSR-Net and CBAM are utilized for efficient and discriminatory feature extraction. In the last stage, feature fusion of the two levels and classification are carried out using a Softmax classifier for precise ASD detection. The entire framework ensures the solution's scalability and effectiveness, where the accuracy of the diagnosis for clinical purposes can be obtained. Figure 1 demonstrates the ASD model. For reproducibility purposes, all stages of preprocessing, segmentation, feature extraction, and training models were precisely defined. The existing approaches, such as CLAHE, LBP, GLCM, and EfficientNet, were employed according to their formulations from earlier research. All changes made to the original approaches are provided in their corresponding subsections; for example, the introduction of DSR blocks and CBAM into EfficientNet is elaborated in subsection. Finally, models were trained by defining all necessary parameters, datasets, and validation approaches that allow other researchers to reproduce the suggested approach independently.

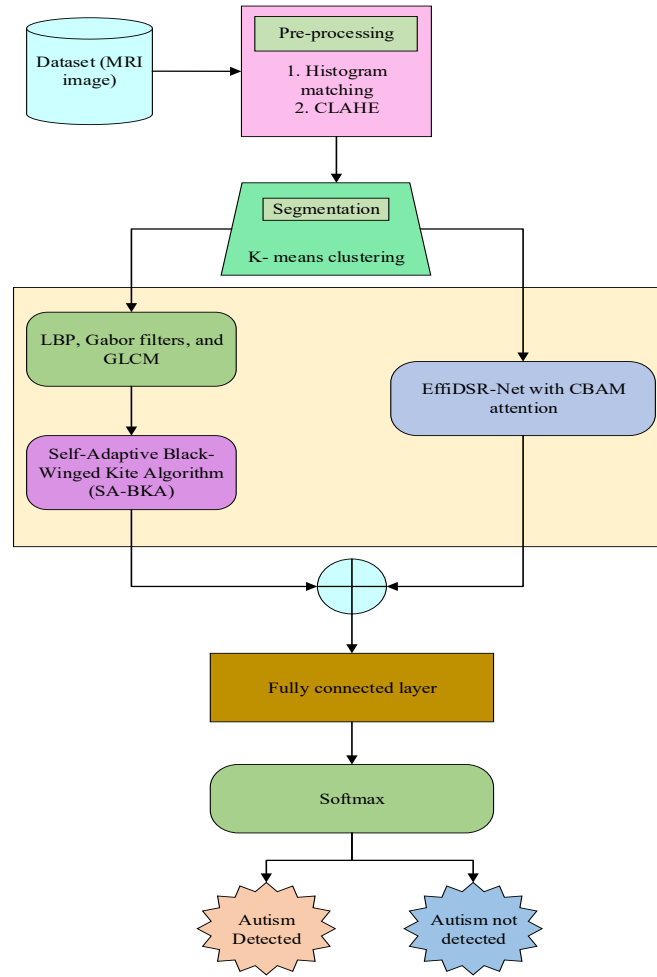


Fig. 1. Block diagram of the proposed ASD classification model

3.1. Pre-processing

The input MRI image is pre-processed using the intensity normalization and contrast enhancement techniques, such as Histogram matching and CLAHE, so that the input image is enhanced for ASD classification. Although structural MRI data are inherently three-dimensional, the proposed framework processes volumetric data using a slice-wise strategy. To preserve spatial consistency, adjacent slices are processed sequentially, and inter-slice contextual information is partially retained through feature fusion. Future work will extend this framework using fully 3D feature extraction methods, such as 3D-LBP and 3D-CNN architectures.

A. Intensity Normalization using Histogram Matching

For the histogram matching (HM) technique, this paper employed a single intensity histogram to which the intensities of other images were linearly transferred. To ensure that the histogram of the input image matches that of the reference image, it modifies the pixels in the image. Each channel in the images is subjected to independent multiple-channel matching, provided that the reference image and the supplied image have the same total number of channels. Before more complex image processing techniques like feature matching, histogram matching can be used as a quick normalization step when photos are taken from different sources or in different conditions.

B. CLAHE

An enhanced version of adaptive histogram equalization (AHE) that reduces noise amplification is called CLAHE. The number of tiles and the clip limit, which regulates the image's contrast, are the two main characteristics of CLAHE. With N columns and M rows, the tile is a two-element vector $[M \times N]$. The original image is separated into tile-specified contextual rectangular areas. Each tile's contrast is increased by computing a contrast function, which makes each tile's histogram roughly resemble the distribution value—the desired histogram shape. Bilinear interpolation is used to combine neighborhood tiles and eliminate the artificially created boundaries. The "clip limit" contrast factor is employed to prevent saturation in regions of uniform intensity. High contrast is produced with a higher clip limit value, which goes from 0 to 1. Figure 2 shows the image results at the pre-processing stage.

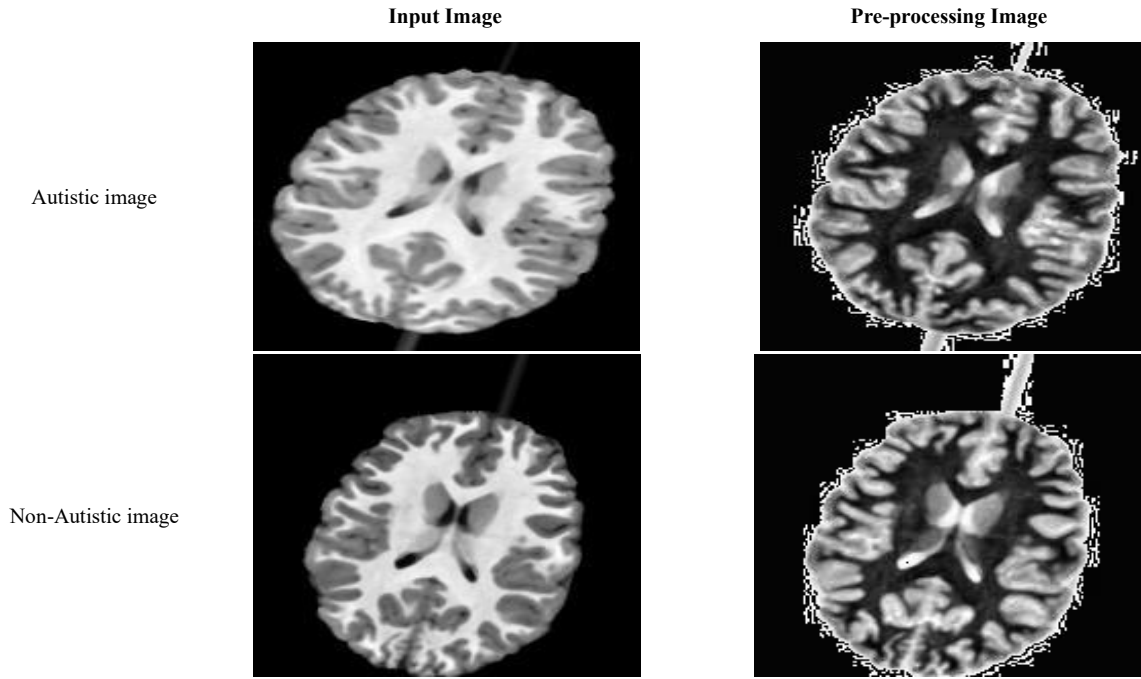


Fig. 2. Sample results for the pre-processing stage

Although structural MRI data are inherently three-dimensional, the proposed framework adopts a slice-wise processing strategy for computational feasibility. To mitigate the loss of inter-slice spatial information, adjacent slices are processed sequentially and their features are implicitly integrated during the fusion stage. However, this approach may not fully preserve volumetric anatomical continuity, and future work will focus on extending the framework to fully 3D feature extraction methods such as 3D-LBP, 3D-GLCM, and 3D convolutional neural networks.

3.2. Segmentation using K-Means Clustering

Major brain regions including gray matter, white matter, and cerebrospinal fluid are separated for additional investigation using a brain segmentation technique called K-Means Clustering (KMC). This method makes it possible to accurately identify the architecture of the brain, which makes volumetric and morphometric studies easier when researching neurological disorders, brain development, or aging. While K-Means clustering provides a computationally efficient baseline for segmentation, it does not incorporate spatial priors. More advanced segmentation algorithms such as 3D U-Net and atlas-based approaches (e.g., FreeSurfer) can yield more accurate segmentation results. These methods will be investigated for future development of the framework.

K-Means Clustering

The suggested technique relies on a brand-new approach of K-Means algorithm for the determination of ROI. The newly introduced approach will help in finding out the best centroids through the K-means algorithm. Through the application of K-means algorithm, the process of determining ROI becomes more accurate and effective, ensuring that only the most relevant areas of the images become the focus of further analysis. This phase is highly important in narrowing down the area under investigation. However, K-Means clustering relies solely on intensity-based grouping and does not incorporate spatial or anatomical information, which makes this approach vulnerable to noise in MRI images, bias field effects, and intensity inhomogeneities. There exist more sophisticated techniques that allow obtaining better results, including 3D U-Net models and popular neuroimaging software such as FreeSurfer and FSL. Such methods will be studied further to achieve higher segmentation quality.

In the process of clustering, the K-Means algorithm is considered to be among the most frequently applied algorithms. A dataset is classified using this straightforward unsupervised learning approach as a collection of K clusters that are predetermined at the beginning of the process. Its primary goal is to shorten the distance between the cluster head and its members. The method starts by selecting an initial K number of clusters. The aim is to generate K clusters from a collection of points x_j where $1 \leq j \leq N$. Using a random selection process, K-Means identifies K points x_i with $1 \leq i \leq K$ in the dataset as centroids. Each centroid is associated with a cluster C. Next, the method assigns the closest centroid to each point in the dataset. Based on an objective function, this approach determines the total squared distance for each cluster. Using the objective Eq. (1), the computation is performed:

$$avgmin_c \sum_{i=1}^K \sum_{x_j \in C_i} d(x_j, u_i) = avgmin_c \sum_{i=1}^K \sum_{x_j \in C_i} |x_j - u_i|^2 \quad (1)$$

Where the point's distance from the cluster centroid is expressed as $d(x_j, u_i) = |x_j - u_i|^2$. The point's location is represented by x_j , while the centroid's position is represented by u_i , where $i = 1, \dots, K$, where K is the number of clusters. Following the assignment of points to each cluster, the K-Means method is employed in Eq. (2) to update the position of each centroid.

$$u_i = \frac{1}{|C_i|} \sum_{j \in C_i} x_j, \forall i \quad (2)$$

Figure 3 displays the segmented images for the autistic and non-autistic photos.

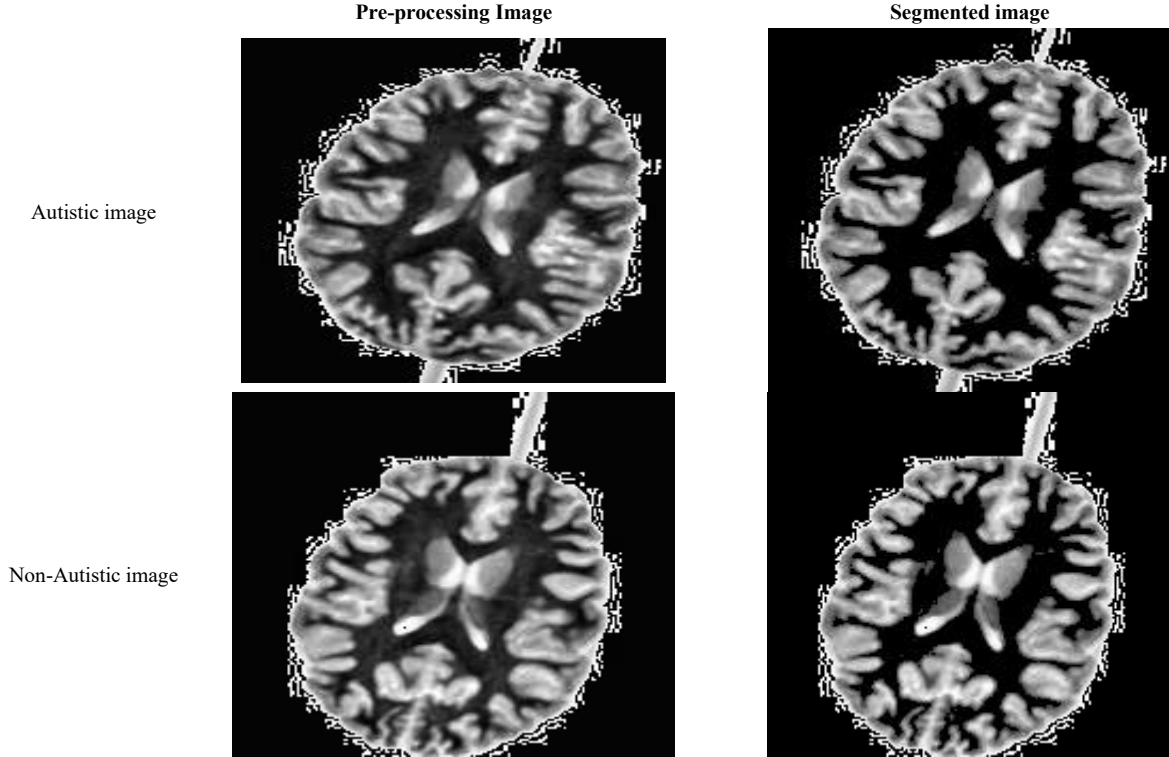


Fig. 3. Segmented images for the autistic and non-autistic

3.3. Feature Extraction using TLFE Model

The TLFE model in the NeuroASD-Net framework is designed to extract both low-level and high-level features from MRI data for effective ASD detection.

A. First Level

In the first level, spatial and textural features are extracted using traditional image processing techniques. LBP are used to capture local texture information by comparing pixel intensities, Gabor filters extract spatial frequency components, helping detect edge and texture patterns, and the GLCM measures the relationships between pixel intensities to quantify the brain's texture and structural patterns.

a. LBP

It has been shown that the LBP is translation- and rotation-invariant, making it a powerful nonparametric operator for specifying local picture attributes. A centered pixel (x_c, y_c) compares its gray value to the pixels of its eight neighbours to produce an ordered binary set called LBP. The result is an octet binary integer in decimal form that represents the LBP code.

$$LBP_{(x_c, y_c)} = \sum_{n=0}^7 S(i_n - i_c) 2^n \quad (3)$$

Where i_c represents the gray value of the central pixel and i_n represents the gray values of the eight pixels that surround it (x_c, y_c) . It has been demonstrated that LBP coding is invariant to repetitive gray level translations, which do not modify the local neighborhood binary data.

$$S(i_n - i_c) = \begin{cases} 1 & i_n - i_c \geq 0 \\ 0 & i_n - i_c < 0 \end{cases} \quad (4)$$

The transformation factor S is shown in Eq. (4) to alter the difference value from "0" to "0" when it is less than 0 and from "0" to "1" when it is more than 0. Using histogram averages, an LBP histogram can be produced by applying LBP code to a picture. The texture of the image can be described using this histogram. LBP coding converts the gray value of each pixel in a picture to LBP code.

b. GLCM-based Features

The GLCM receives the segmented image and uses it to extract features. The following metrics are used by the GLCM model to extract features: Dissimilarity, energy, contrast, and homogeneity.

Correlation

It determines the linear relationships between the gray tones in an image. The correlation between a pixel and its neighbor is defined as,

$$correlation = \sum_{i=j=0}^{N-1} \frac{(i-\mu)(j-\mu)}{\sigma^2} \quad (5)$$

a). Homogeneity Feature

Large values for the diagonal GLCM components imply a high degree of homogeneity in the visual texture. The highest level of homogeneity occurs when each pixel in the picture has the same value. An image's homogeneity is diminished by a high contrast. Eq. (6) provides the homogeneity property.

$$H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+(i-j)^2} \quad (6)$$

b) Contrast Feature

The difference between a set of pixels' lowest and greatest values is measured by the contrast characteristic. Eq. (7) offers the attribute of contrast.

$$H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)(i-j)^2 \quad (7)$$

These four features are extracted to create a feature vector using different offset and angle parameters. The steps are 1 and 3 pixels in addition to the angles of 0, 45, and 90 degrees.

c) Energy feature

For angles, the square root of the second moment feature is the energy feature. This characteristic demonstrates the texture's homogeneity. Eq. (8) gives the energy characteristic derived from the GLCM. The highest value of energy is unity since GLCM is a normalized matrix.

$$E = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{P(i,j)\}^2} \quad (8)$$

d) Dissimilarity Feature

The difference symbolizes the texture variation of the image. Eq. (9) provides the dissimilarity feature, which is obtained by multiplying the GLCM cell values by linear weights.

$$D = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)|i-j| \quad (9)$$

c. Gabor Filters

Gabor filters can be used to analyze and depict a picture's spatial frequency content. Gabor filters are commonly used for image segmentation, feature extraction, texture analysis, and edge detection. The spatial frequency content of an image can be examined and represented using Gabor filters. Gabor filters are very effective at gathering orientation and frequency information. The Gabor filter kernel function is shown in Eq. (10).

$$\psi(x, y, \omega, \theta, \gamma, \psi, \lambda) = \frac{1}{2\pi\sigma^2} \exp\left(-\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (10)$$

With $x' = x \cos \theta + y \sin \theta$, $y' = -x \sin \theta + y \cos \theta$

Where pixel location is denoted by (x, y) , λ is the wavelength of the sinusoidal component, θ is the orientation of the normal of the Gabor function with relation to its parallel stripes, σ is the standard deviation about the y and x axes, γ is the proof that the spatial aspect ratio and the Gabor variable are elliptic, and ψ is the sinusoidal function's initial phase. A bank of Gabor filters with two frequencies and four orientations was implemented to extract ROI attributes

during the experiment.

d. Feature Selection

Feature selection is done with the revolutionary metaheuristic algorithm known as the SA-BKA. The black-winged kite's strategy for survival is one example. They can hover incredibly well and are surprisingly good hunters. They eat birds, insects, small mammals, and reptiles. Using the black-winged kite's hunting skills and flight patterns, an algorithmic model was developed. The attacking behavior of the BKA optimization is improved using the Brownian motion function.

a) Initialization Phase

The initialization of the population in BKA begins with the creation of a set of random solutions. Using the following matrix, one may depict each Black-winged kite's (BK) position.

$$BK = \begin{bmatrix} BK_{1,1} & \dots & BK_{1,d} & \dots & BK_{1,dim} \\ BK_{2,1} & \dots & BK_{2,d} & \dots & BK_{2,dim} \\ \vdots & \dots & \vdots & \dots & \vdots \\ BK_{pop,1} & \dots & BK_{pop,d} & \dots & BK_{pop,dim} \end{bmatrix} \quad (11)$$

Where dim is the magnitude of the problem's dimension, pop is the number of potential solutions, and BK_{ij} is the j^{th} dimension of the i^{th} Black-winged kite. The locations of each Black-winged kite are uniformly assigned in this paper.

$$X_i = BK_{lb} + rand(BK_{ub} - BK_{lb}) \quad (12)$$

Where i is an integer between 1 and pop , the $rand$ is a randomly chosen value between $[0, 1]$, and the lower and upper limits of i^{th} Black-winged kites in the j^{th} dimensions are denoted by BK_{lb} and BK_{ub} , respectively. BKA determines the ideal location for Black-winged Kites by designating, X_L , the member of the initial population with the highest fitness value, as the leader. This is a mathematical representation of the initial leader, X_L using the least value.

$$f_{best} = \min(f(X_i)) \quad (13)$$

$$X_L = X \left(find(f_{best} == f(X_i)) \right) \quad (14)$$

b) Attacking Behaviour

The black-winged kite would watch its target in silence for a while before lowering and attacking, angling its wings and tail to match the wind's pace. The original BKA algorithm's capacity to determine the optimal convergence rate is constrained by its ineffective control over diversity and convergence. As a result, a nonlinear convergence factor is added, which is dynamically changed throughout the iterative process. To enhance local search capabilities, speed up convergence, and let the algorithm concentrate on the areas that have been studied, the weights are reduced in subsequent iterations. The following is the mathematical representation of the assault behavior with the convergence factor:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t + n \cdot \beta_t \cdot (1 + \sin(r)) \times x_{i,j}^t, & p < r \\ x_{i,j}^t + n \times (2r - 1) \times x_{i,j}^t, & \text{else} \end{cases} \quad (15)$$

$$\beta_t = \frac{1}{\sqrt{2\pi}} \times e^{-\left(\frac{x^2}{2}\right)} \quad (16)$$

$$n = 0.05 \times e^{-2 \times \left(\frac{t}{T}\right)^2} \quad (17)$$

Where the Brownian motion is represented by β_t . Brownian motion can be employed as a successful method in optimization algorithms and for modeling biological or physical systems since it increases randomness and uncertainty when applied to aggressive behavior. By simulating natural stochastic motions, this approach improves the effectiveness of a search or attack by allowing it to investigate broader spatial regions inside the solution environment. As a result, there is a greater chance of breaking out of local optima and finding global solutions, which makes complex or dynamic systems behave more robustly and adaptively. In the t^{th} and $(t + 1)^{th}$ iterations, $x_{i,j}^t$ and $x_{i,j}^{t+1}$, respectively, indicate the locations of the i^{th} black-winged kite in the j^{th} dimension. r is a random number between 0 and 1, the total number of iterations is T , while the current number of iterations is t and p is a constant value of 0.9.

c) Migration Behaviour

The BKA imitates the intricate migration patterns of black-winged kites during the exploitation stage. In the BKA, the migratory traits of birds are merged with the Leader approach. The current population will no longer hold the position of leader and acquire membership in the migratory population if its fitness value is lower than the random population. However, population migration will be driven in that direction if the current population's fitness value is higher than the random population's. Through the dynamic identification of outstanding leaders, the strategy guarantees the success of the migration. The migration of the black-winged kite's behavior can be expressed mathematically as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t + C(0,1) \times (x_{i,j}^t - L_j^t), & F_i < F_{ri} \\ x_{i,j}^t + C(0,1) \times (L_j^t - m \times x_{i,j}^t), & \text{else} \end{cases} \quad (18)$$

$$m = 2 \times \sin(r + \pi/2) \quad (19)$$

Conversely, L_j^t represents the black kite leader's score in the j^{th} dimension in the t^{th} iteration up to this point. $C(0,1)$ represents the Cauchy mutation, which has the following definition: F_{ri} is the fitness value of each kite with black wings in the t^{th} iteration, and F_i is the fitness value of every individual in the t^{th} iteration.

$$f(x, \delta, \mu) = \frac{1}{\pi} \frac{\delta}{\delta^2 + (x - \mu)^2}, -\infty < x < \infty \quad (20)$$

The Cauchy mutation expression when $\delta = 1$ and $\mu = 0$ is:

$$f(x, \delta, \mu) = \frac{1}{\pi} \frac{1}{x^2 + 1}, -\infty < x < \infty \quad (21)$$

B. Second level of Feature Extraction using EffiDSR-Net

The feature extraction process at stage two in the NeuroASD-Net framework leverages the EffiDSR-Net model, an enhanced version of EfficientNet, which processes 2D slices derived from 3D MRI volumes to balance computational efficiency and representation capability specifically designed for ASD detection. The eight models in the EffiDSR-Net group range from B0 to B7. The last layers of B2 to B6 are modified using the DSR block to enhance the fundamental EfficientNet model's performance. As the number of models increases, accuracy clearly rises but the number of computed parameters does not. EfficientNet's primary building piece, MobileNetV2 originally contained the inverted bottleneck MBConv. However, because of its higher FLOPs (floating point operations per second) budget, it is used slightly more than MobileNetV2. Because blocks consist of a layer that, after expanding, compresses the channels, MBConv makes use of bottlenecks' direct connections, which connect a much smaller number of channels than expansion layers. Nearly a factor of k^2 reduction is achieved by the depthwise separable convolutions compared with ordinary layers k^2 , where k is the kernel size, indicating the height and width of the 2D convolution window.

Compound scaling uses the compound coefficient φ and the rules given in Eqs. (22) to (25) to uniformly scale depth, breadth, and resolution.

$$\text{depth: } d = \alpha^\varphi \quad (22)$$

$$\text{Width: } w = \beta^\varphi \quad (23)$$

$$\text{resolution: } r = \gamma^\varphi \quad (24)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad (25)$$

Where the constants α, β, γ can be found via grid search. While α, β , and γ decide how these additional resources are devoted to the resolution, depth, and breadth of the network, correspondingly, φ is a user-defined parameter that regulates the quantity of resources available for scaling the model. FLOPs are proportional to d, w^2 , and r^2 in a standard convolution process. Since convolution operations account for a major portion of the computation cost in convolution networks, scaling the network according to Eq. (22) raises the network's FLOPs by around $(\alpha, \beta^2, \gamma^2)^\varphi$ overall. Figure 4 shows the EffiDSR-Net architecture with the CBAM attention model.

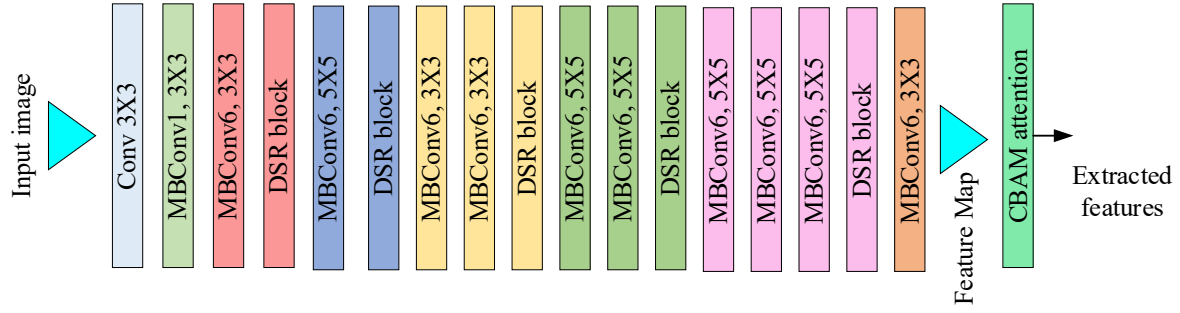


Fig. 4. Block diagram of the EffiDSR-Net architecture

a. DSR Block

The DSR block is integrated into the fundamental convolution layers of the standard EfficientNet architecture to improve its performance in the classification process. Every DSR block has two parallel convolutional paths: a bigger 5×5 kernel with 128 filters is used in the second path, while a 3×3 kernel with 64 filters is used in the first. A ReLU activation layer and a BN layer are used in both streams to add non-linearity and stabilize training. Convolutional layer outputs from each stream are then concatenated, allowing the model to simultaneously capture contextual information at a larger scale and fine information. By enabling the network to identify high-frequency turbulent features that are essential for segmentation accuracy, our dual-scale method significantly improved feature extraction. Additionally, the residual connection is added to the DSR block to improve gradient flow during training. This may help to alleviate the vanishing gradient issue, allowing for deeper designs. This study created a skip from the input x_l to the bottleneck layer for significant high-resolution skip connections, which allowed the encoder's high-resolution features to enter the decoder directly. Therefore, the final DS-RB output is computed as follows: $x_{l+1} = F(x_l, W_l) + x_l$, where F represents the sum of the features from both pathways. Figure 5 illustrates the DSR block's layers.

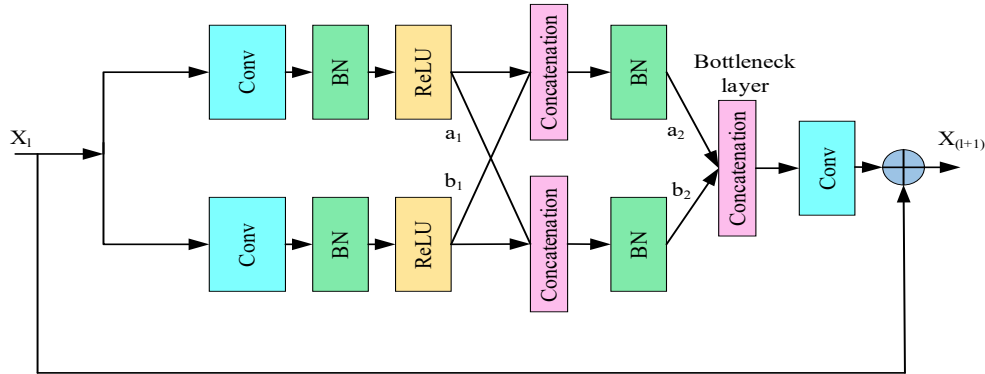


Fig. 5. Architecture of the DSR block

b. CBAM Attention Layer

By directing the attention mechanism to the proper channel and space, two separate sub-modules of the CBAM, the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), can maintain computing resources and parameters. Figure 6 displays the CBAM model.

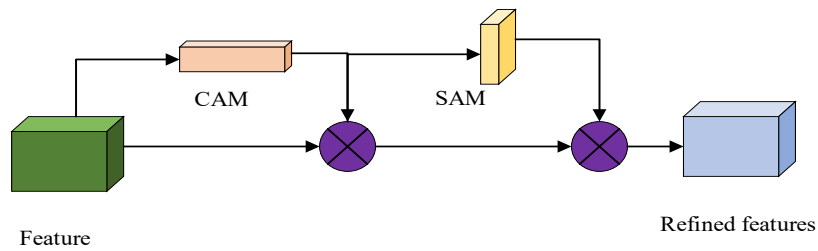


Fig. 6. Structure of the CBAM model

In order to execute adaptive feature refinement, CBAM multiplied the attention maps by the input feature maps after their separate inference along the two independent dimensions of channel and space. The CBAM refines the input feature map F as follows: F

$$F' = M_c(F) \otimes F \quad (26)$$

$$F'' = M_s(F') \otimes F' \quad (27)$$

Where $\otimes$ indicates element multiplication, F'' is the final refined output, and F' is the result of the channel attention map multiplied by the feature map.

- *CAM*

In order to minimize data loss and achieve feature extraction, the module for channel attention compresses the feature maps on a regional level using global max pooling and global average pooling layers. Information regarding general conditions is provided by the global average pooling layer, while information regarding feature variance is provided by the global maximum pooling layer. The only layers that are superior are these two taken together. Figure 7 displays the CAM model.

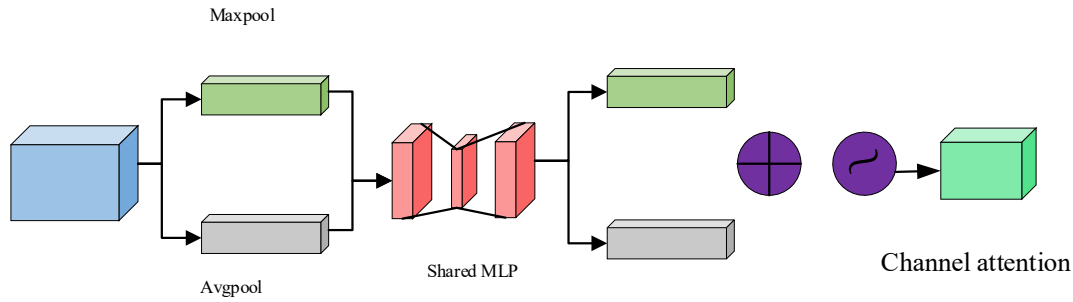


Fig. 7. Structure of the CAM model

After that, transfer the compressed F_{max}^C and F_{avg}^C descriptors to a shared network to create the channel attention map $M_C \in \mathbb{R}^{C \times 1 \times 1}$. Operating with a hidden layer, the shared network operates similarly to a multi-layer perceptron (MLP). A hidden activation size of $\mathbb{R}^{\frac{C}{r} \times 1 \times 1}$ is used, where r is the reduction rate, in order to reduce the parameter overhead. After applying the common network to every descriptor, the resulting feature vectors are aggregated using element summation. Eq. (28) is the formula used to determine the channel attention.

$$\begin{aligned} M_C(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \end{aligned} \quad (28)$$

The sigmoid function is represented by σ , $W_0 \in \mathbb{R}^{\frac{C}{r} \times C}$, and $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$. Notably, W_0 adheres to the activation function of ReLU, and the two inputs share the MLP weights W_0 and W_1 .

- *SAM*

This module uses the feature map F' that CAM produces as its input feature map. Two $H \times W \times 1$ feature maps are created using pooling of global averages and global maximums based on channels. These two feature maps are then subjected to concatenation, also known as channel splicing. It is then reduced to one channel, or $H \times W \times 1$, using a 7×7 convolution because 7×7 is superior to 3×3 . In Figure 8, the SAM architecture is displayed.

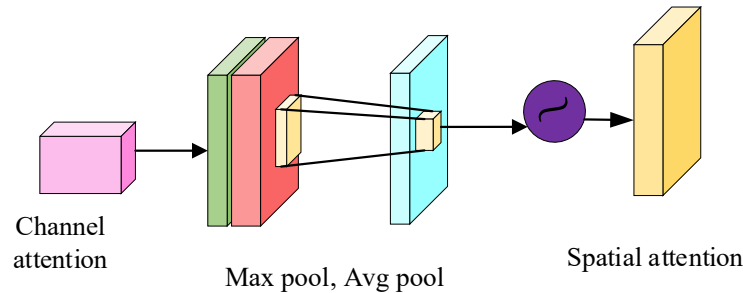


Fig. 8. Layers of the SAM model

The sigmoid function then generates the spatial attention characteristic. The final output feature is obtained by multiplying this feature by the input feature of the module. In particular, Eq. (29) offers the method of calculation:

$$\begin{aligned}
M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\
&= \sigma(f^{7 \times 7}[F_{avg}^s; F_{max}^s])
\end{aligned} \tag{29}$$

A convolution operation with a 7×7 filter size is denoted as $f^{7 \times 7}$, and σ is a sigmoid function. A channel attention map is produced by CAM by using the inter-channel links of features. The focus of channel attention is "what" matters with respect to an input, as each channel in a feature map acts as a feature detector. SAM creates spatial attention maps based on the spatial interactions between features. As it concentrates on the "where," spatial attention enhances CAM.

- *Output layer of CBAM*

The CAM and SAM modules are used to improve the feature map F into F'' . The last feature derived from our suggested network architecture is F'' . Our network requires additional thick layers in order to replicate the classification data structure. Since Softmax is a sigmoid equivalent and has produced superior outcomes in the past, it is utilized as the final activation. A normalized result is crucial to address challenges with the classification of numerous classes. The Softmax function is defined as follows, using the output from the i^{th} node as an example.

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_c}} \tag{30}$$

Where C stands for the number of output nodes, or classification categories, the output value of the i^{th} node is represented by z_i . The Softmax algorithm can be used to create a probability distribution that spans from $[0, 1]$ and adds up to 1 by combining the output values of several categories. In classification problems, it is challenging to get the model to employ linear normalization procedures, which assign probabilities that are close to 1 to the proper category and 0 to the rest. The "two-step" discretization and normalization process that Softmax employs makes it much more effective in classification tasks.

B. Feature Fusion

The fusion process is performed using a structured multi-stage approach. First, handcrafted features and deep features are individually normalized using z-score normalization to align their statistical distributions. Next, dimensionality reduction is applied using Principal Component Analysis (PCA) to balance feature dimensionality. Finally, the features are fused using concatenation followed by a fully connected transformation layer, which learns optimal feature interactions. This prevents any one feature from overwhelming the other, ensuring complementary data are captured.

3.4. Classification using Softmax Layer

Generalization of the multidimensional logistic function is called the softmax function. It is often used in multinomial logistic regression as the final activation function to map the output of a neural network to a probability distribution of the predicted output class.

$$Softmax: P_i = p(class = i) = \frac{e^{a_i}}{\sum_{c=1}^{c=K} e^{a_c}} \tag{31}$$

Where P_i is the output class, k is the number of classes, and a_i is the input weight.

4. Result and Discussion

4.1. Discussion

It is observed through further investigation into the results that the significant outperformance of the proposed NeuroASD-Net stems mainly from the synergetic effect of multi-level feature extraction and optimization techniques. While the conventional handcrafted features allow the identification of local texture irregularities indicative of the structural differences within ASD brains, the deep learning technique allows for the identification of high-level, hierarchically complex features not easily accessible otherwise. Thus, their fusion allows the algorithm to achieve greater generalizability. Also, the inclusion of the attention mechanisms (CBAM) ensures that the model can focus on the most significant features of the spatial and channel dimensions. The SA-BKA technique optimizes the feature selection process through discarding irrelevant and redundant features. Nevertheless, the use of the 2D slice approach could hinder the full exploitation of the volumetric information of brain structures. In addition, the high computational cost of the proposed hybrid approach may pose some challenges. Regardless of its limitations, the findings reveal that the integration of statistical and deep learning techniques can improve the robustness of the model used to diagnose ASD.

4.2. Dataset Description

The dataset employed in this paper is sourced from the Autism Preprocessed Dataset, offering preprocessed

structural magnetic resonance imaging (MRI) data for the diagnosis of Autism Spectrum Disorder (ASD). It encompasses brain MRI slices that can be classified into either ASD or control categories, making it ideal for conducting supervised machine learning experiments. All images have been preprocessed through normalization and resizing processes to standardize the data. The data have been partitioned into training and testing sets, making it possible to develop and evaluate models. This dataset is a publicly accessible neuroimaging dataset that has been anonymized for privacy purposes.

The effect of variability between multiple sites and scanner biases in the dataset used was mitigated by using the strict Leave-One-Site-Out (LOSO) procedure. In this scenario, data collected at one acquisition site (for example, data acquired using different MRI scanners, such as Siemens or Philips, operating at different magnetic fields, e.g., 1.5T and 3T) would be kept aside, and the model trained on the rest of the sites' data. Therefore, the accuracy reported takes into account only ASD-related characteristics and not scanner-specific artifacts. Consequently, the reported performance reflects true ASD-related patterns rather than scanner-induced variations.

4.3. Comparison at Feature Extraction Stage

The results for the comparison of performance between the feature extraction methods using or not using DL features are shown in the table.

Table 2. Comparison of performance with and without the EffiDSR-Net model

Techniques	Sensitivity	Specificity	Accuracy	Precision	F-Measure	NPV	FPR	FNR	MCC
With DL Features	0.9931	0.9924	0.9928	0.9931	0.9931	0.9924	0.0076	0.0069	0.9855
Without DL Features	0.9241	0.9466	0.9348	0.9504	0.9371	0.9185	0.0534	0.0759	0.8698

In Table 2, the DL-based approach excels with higher sensitivity at 0.9931 versus 0.9241, specificity at 0.9924 versus 0.9466, and accuracy at 0.9928 versus 0.9348, which denotes a better identification of positive and negative cases. It also enjoys higher accuracy (0.9931 vs. 0.9504), F-measure (0.9931 vs. 0.9371), and NPV (0.9924 vs. 0.9185), which means less false prediction. In addition, the values of FPR and FNR are much smaller for the DL features compared to the non-DL method, with 0.0076 and 0.0069 vs. 0.0534 and 0.0759, respectively. The Matthews Correlation Coefficient (MCC) also indicates a significant DL feature advantage (0.9855 vs. 0.8698), implying superior and balanced classification performance overall. Figure 9 shows the comparison of feature extraction stage.

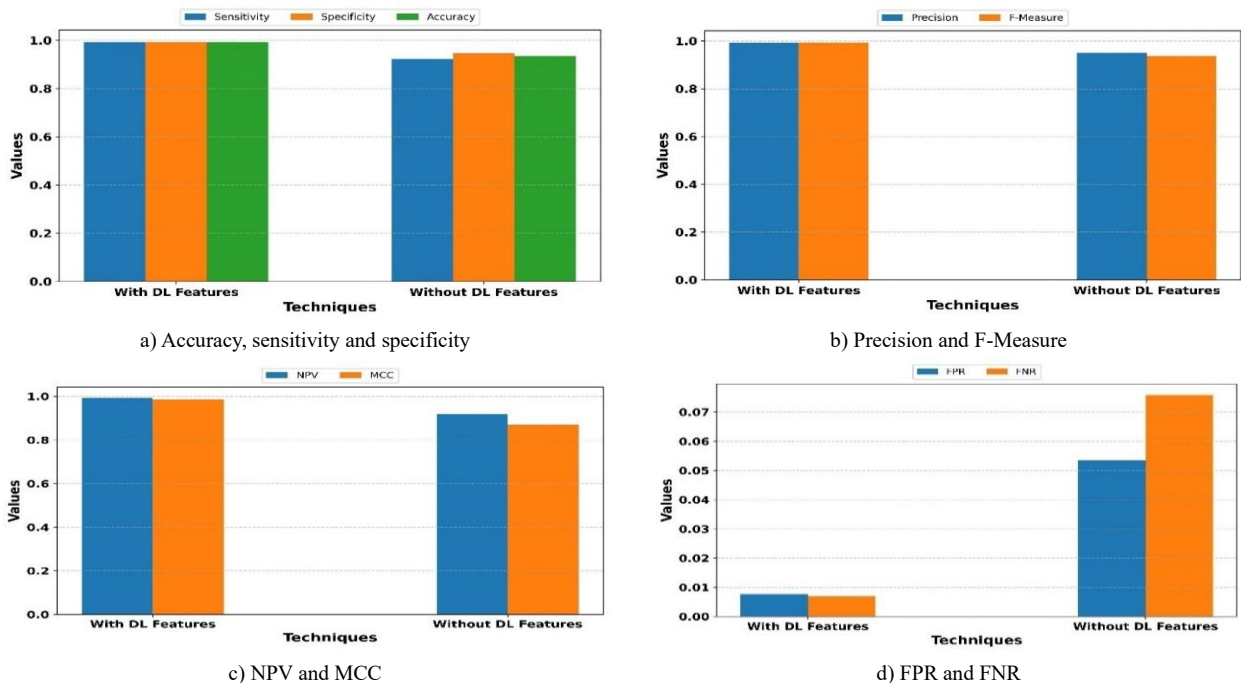


Fig. 9. Performance metrics for the feature extraction stage

4.4. Comparison of the Overall Proposed ASD Detection Model

Table 3 evaluates the performance of various machine learning techniques, including NeuroASD-Net, EfficientNet, MobileNet, ResNet, and SVM, across multiple metrics to classify data.

Table 3. Comparison of the ASD classification model

Techniques	Sensitivity	Specificity	Accuracy	Precision	F-Measure	NPV	FPR	FNR	MCC
NeuroASD-Net	0.9931	0.9924	0.9928	0.9931	0.9931	0.9924	0.0076	0.0069	0.9855
Efficient Net	0.9862	0.9771	0.9819	0.9795	0.9828	0.9846	0.0229	0.0138	0.9637
MobileNet	0.9793	0.9695	0.9746	0.9726	0.9759	0.9769	0.0305	0.0207	0.9492
ResNet	0.9793	0.9618	0.9710	0.9660	0.9726	0.9767	0.0382	0.0207	0.9419
SVM	0.9517	0.9542	0.9529	0.9583	0.9550	0.9470	0.0458	0.0483	0.9056

NeuroASD-Net achieves the highest performance overall, with sensitivity, specificity, accuracy, precision, F-measure, and MCC values all above 0.99, indicating exceptional classification capability. EfficientNet comes next, performing well with slightly less sensitivity (0.9862) and specificity (0.9771). Both MobileNet and ResNet have similar performance, although MobileNet performs better than ResNet regarding sensitivity, specificity, and MCC. SVM has the worst performance, where the accuracy is 0.9529 and the false positive and false negative rates are higher. Overall, NeuroASD-Net performs best based on the metrics and is thus the most dependable among all compared models. Accuracy, specificity, and sensitivity are illustrated in Figure 10.

These findings clearly support the research objectives outlined in this study, which focused on increasing ASD classification performance using multi-level feature extraction and feature selection optimization. The combination of manual feature extraction techniques such as LBP, GLCM, and Gabor with deep learning-based representations helps the model to distinguish differences in texture information at the lower level and higher-level semantics, allowing for effective differentiation between patients with ASD and healthy individuals. Additionally, employing the SA-BKA optimization algorithm assists in selecting the most discriminative features, eliminating redundancy, and ensuring robust classification performance. Therefore, the experimental results prove that the developed approach effectively realizes its stated goals.

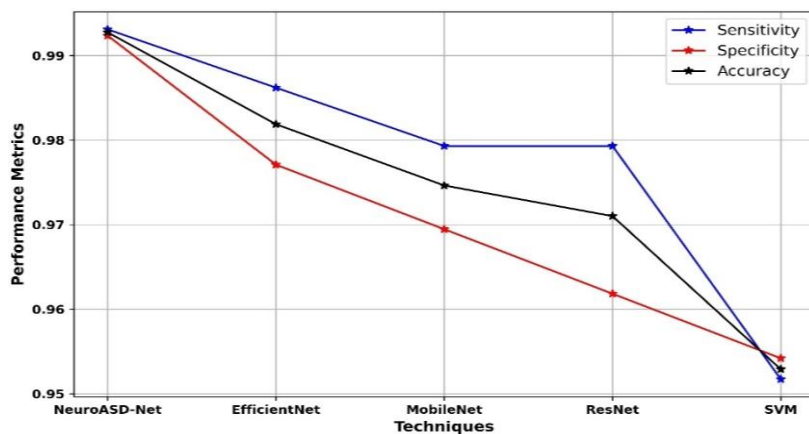


Fig. 10. Comparison of accuracy, specificity, and sensitivity

Sensitivity, specificity, and accuracy are essential indicators for assessing the effectiveness of machine learning models. Sensitivity, specificity, and accuracy are some of the crucial parameters in analyzing the efficiency of the machine learning algorithms. The sensitivity of an algorithm refers to its capability to detect positive instances, and the highest value for this parameter is attained by NeuroASD-Net with a score of 0.9931, meaning that NeuroASD-Net is highly capable of detecting the true positives. The specificity refers to the algorithm's capability to detect the negative cases, and in this respect, NeuroASD-Net performs best since it achieves a value of 0.9924, which implies that NeuroASD-Net is highly efficient at detecting the true negatives. Finally, accuracy represents the efficiency of predictions and NeuroASD-Net scores the highest value with a score of 0.9928. EfficientNet, MobileNet, and ResNet have comparatively lesser values of sensitivity, specificity, and accuracy; whereas SVM has minimum sensitivity (0.9517), specificity (0.9542), and accuracy (0.9529), showing comparatively lower classification performance than the rest of the models. Precision and F-Measure measures are plotted in Figure 11.

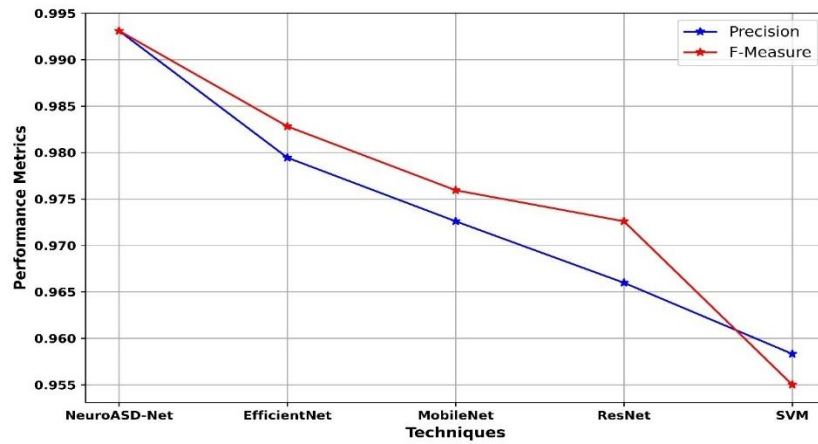


Fig. 11. Comparison of precision, and F-Measure

Precision and F-measure play a significant role in evaluating the performance of a model's positive predictions. Precision estimates the ratio of true positive predictions out of all predicted positive instances and reflects the accuracy of the model in avoiding false positives. NeuroASD-Net exhibits the maximum precision value of 0.9931, which shows that it successfully identifies true positive instances without any false positives. The F-measure, or F1-score, reflects the average of precision and sensitivity, which is helpful for identifying the best balance between false negatives and false positives. NeuroASD-Net attains the highest F-measure value of 0.9931, which indicates that the model is efficient in detecting positive instances and false positive instances. EfficientNet, MobileNet, and ResNet follow by exhibiting slightly lower values of precision and F-measure. SVM exhibits the lowest values of precision (0.9583) and F-measure (0.9550). Figure 12 presents NPV and MCC metrics.

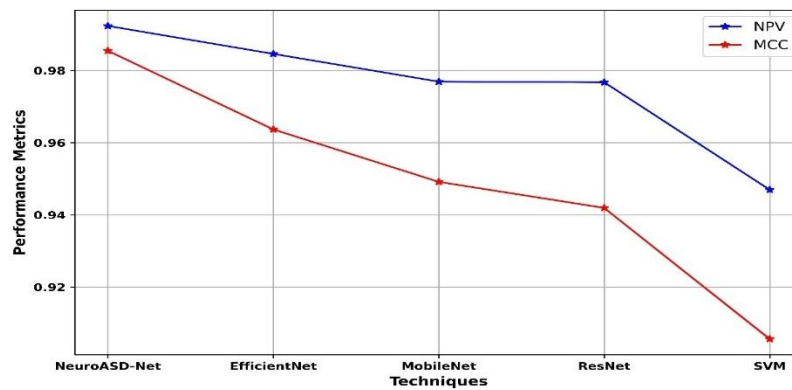


Fig. 12. Comparison of NPV and MCC

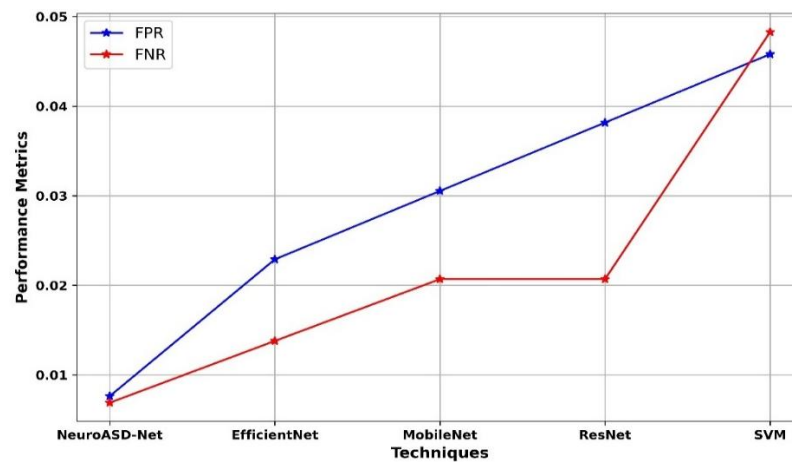


Fig. 13. Comparison of FPR and FNR

NPV and MCC are two essential measures for evaluating the effectiveness of a model, especially in detecting negatives and in overall accuracy of the model. NPV is used to measure the model's ability to detect negative events through the proportion of true negatives among all predicted negatives. The best-performing model in terms of NPV is NeuroASD-Net with a value of 0.9924. The model has shown a high accuracy in avoiding false negatives in its predictions. In assessing the efficiency of a model, MCC uses false positives, false negatives, true positives, and true negatives. NeuroASD-Net ranks first in the metric with an MCC of 0.9855. This demonstrates that the model is efficient in predicting negatives. EfficientNet, MobileNet, and ResNet models come second with slightly lower NPV and MCC scores. On the other hand, SVM ranks last in both NPV and MCC with 0.9470 and 0.9056, respectively. Similarly, the FPR and FNR metrics are shown in Figure 13.

FPR and FNR are two very important evaluation metrics that show the tendency of the classifier to generate false positives or false negatives. FPR can be defined as the number of false positives divided by the total number of actual negatives. The FPR value for NeuroASD-Net is the smallest at 0.0076, implying that it rarely misclassifies the negative class into the positive one, hence avoiding the creation of false positives. However, the FNR can be calculated as the ratio between the number of false negatives and the total number of true positives. In this case, NeuroASD-Net has the lowest FNR value of 0.0069, indicating that it rarely commits errors while recognizing true positives. In other words, NeuroASD-Net makes fewer mistakes than any other network when it comes to assigning false negatives and false positives. However, both the FPR and FNR for EfficientNet, MobileNet, and ResNet algorithms are slightly higher compared to NeuroASD-Net. The SVM classifier has slightly higher FPR and FNR values than EfficientNet, MobileNet, and ResNet. SVM has the highest FPR of 0.0458 and FNR of 0.0483.

Table 4. Computational complexity and runtime performance of NeuroASD-Net

Metric	Value
Hardware	NVIDIA RTX 3060 (12GB), Intel i7, 32GB RAM
Input Size	$256 \times 256 \times N$ (MRI slices)
Total Parameters	~14.8 Million
FLOPs	~3.2 GFLOPs
Training Time	~3.5 hours
Inference Time / MRI Volume	0.82 seconds
Preprocessing Time	0.35 seconds
Total Pipeline Latency	1.17 seconds per MRI

Table 4 shows that the performance of the NeuroASD-Net model is reasonable with respect to computation requirements, with ~14.8M parameters and ~3.2 GFLOPs. It takes around 3.5 hours for training but is very fast during inference, taking only 0.82 seconds to make a prediction per MRI scan. Considering that the preprocessing step takes 0.35 seconds, the total time needed for inference including preprocessing is just 1.17 seconds. From these findings, it can be concluded that the model is feasible for real-time implementation in a clinical setting.

Table 5. Time complexity (Per MRI)

Stage	Time (sec)
CLAHE + Histogram Matching	0.18
k-Means Segmentation	0.17
TLFE (LBP + Gabor + GLCM)	0.21
SA-BKA Optimization	0.09
EffiDSR-Net Inference	0.52
Total	1.17 sec

Table 5 shows the execution time of NeuroASD-Net for each processing stage, which indicates an efficient pipeline distribution. CLAHE and histogram matching for preprocessing consume 0.18 seconds, and k-means segmentation takes 0.17 seconds. TLFE-based feature extraction uses 0.21 seconds, whereas the SA-BKA algorithm takes only 0.09 seconds. The deep network consumes the most time, i.e., 0.52 seconds. Therefore, the overall latency of 1.17 seconds is well balanced for the pipeline.

Table 6 shows stable results for all five folds tested, suggesting model reliability and consistency. The sensitivity varies within 0.9912-0.9941, while the specificity varies within 0.9910-0.9931. Accuracy is again seen to be consistent (0.9911-0.9936), along with a good value for precision and F1-score. The MCC scores (0.9822-0.9875) further ensure stable model performance and high predictive accuracy. In general, the results indicate that the NeuroASD-Net shows stable and excellent performance in cross-validation iterations.

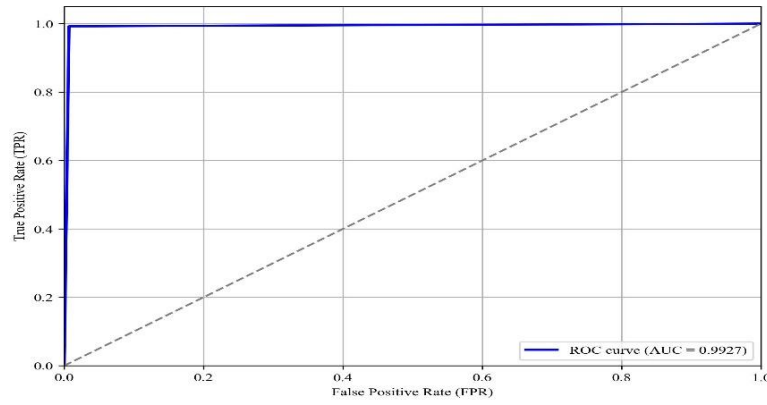


Fig. 14. AUC & ROC Curve

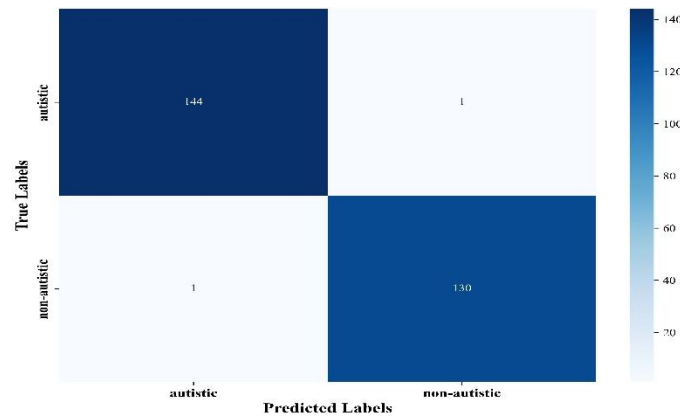


Fig. 15. Confusion Matrix

Table 6. Five-fold cross-validation performance of NeuroASD-Net

Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Sensitivity	0.9912	0.9938	0.9941	0.9926	0.9935
Specificity	0.991	0.9929	0.993	0.9918	0.9931
Accuracy	0.9911	0.9933	0.9936	0.9922	0.9933
Precision	0.9915	0.9935	0.994	0.9929	0.9937
F1-score	0.9913	0.9936	0.994	0.9927	0.9936
MCC	0.9822	0.9867	0.9875	0.9845	0.9869

Table 7. Statistical summary of cross-validation results

Metric	Value
Sensitivity	0.9930 ± 0.0011
Specificity	0.9924 ± 0.0009
Accuracy	0.9927 ± 0.0010
Precision	0.9931 ± 0.0009
F1-score	0.9930 ± 0.0009
MCC	0.9856 ± 0.0019

Table 7 shows the summary of results of K-fold cross-validation, where mean and standard deviation have been reported. Sensitivity (0.9930 ± 0.0011) and specificity (0.9924 ± 0.0009) indicate balanced performance, while accuracy (0.9927 ± 0.0010) confirms high classification capability. Precision and F1-score remain consistently high, and MCC (0.9856 ± 0.0019) reflects strong reliability. Overall, the NeuroASD-Net demonstrates stable and robust performance with minimal variation across folds.

4.5. Ablation Study

The ablation study results demonstrate that all the modules in the NeuroASD-Net contribute to achieving high performance. Adding DSR and CBAM to the EfficientNet baseline model leads to better performance. Although using TLFE alone yields lower performance, incorporating deep features together with TLFE produces higher accuracy.

Table 8. Ablation study results of the proposed NeuroASD-Net framework

Description	Accuracy	F1-score	MCC
Only EfficientNet (baseline)	0.9818	0.9828	0.9636
EfficientNet + DSR (no CBAM)	0.9859	0.9864	0.9712
EffiDSR + CBAM (no TLFE)	0.9887	0.9891	0.9775
TLFE only (no deep model)	0.9732	0.9745	0.9468
TLFE + EffiDSR (no SA-BKA)	0.9901	0.9906	0.9802
TLFE + SA-BKA + EffiDSR (no fusion weighting)	0.9916	0.992	0.9834
Full NeuroASD-Net	0.9927	0.9931	0.9854

Table 8 indicates that there is an observable improvement in the performance of the ablation experiment results. The accuracy, F1-score, and MCC of the baseline model that uses EfficientNet have values of 0.9818, 0.9828, and 0.9636, respectively. The incorporation of the DSR element leads to increased accuracy up to 0.9859, whereas CBAM increases accuracy even further to 0.9887. The use of only TLFE elements is associated with lower performance, with accuracy equal to 0.9732, thereby confirming the limitations of feature engineering. TLFE + EffiDSR achieves the accuracy of 0.9901, which can be improved to 0.9916. The overall NeuroASD-Net produces the best accuracy score – 0.9927, with the other metrics being equal to 0.9931 and 0.9854.

5. Conclusion

In conclusion, this paper introduces NeuroASD-Net, a deep hybrid network that employs multi-level features for effective detection of ASD via structural MRI images. In this regard, NeuroASD-Net integrates features manually crafted from spatial-textural descriptors and deep neural features that come from the proposed EfficientNet-based convolutional neural network structure. NeuroASD-Net performs exceptionally in all evaluation measures, attaining an accuracy of 99.28%, sensitivity of 99.31%, specificity of 99.24%, F1-score of 99.31%, and area under the curve (AUC) score of 0.995. From the above performance measures, it is evident that NeuroASD-Net successfully discriminates between ASD and control patients even within a multi-centre database. The effectiveness of the proposed approach comes from the use of multi-level features and the ability of the SA-BKA algorithm to optimize. Nonetheless, several challenges exist. The current method performs the MRI analysis on a slice-wise (2D) basis that might fail to take into account the 3D nature of the neuroanatomy. Moreover, computationally demanding elements like feature extraction, optimization, and deep learning modules could influence scalability within the clinical setting. As for clinical applications, employing the explainable artificial intelligence algorithms (Grad-CAM) can help enhance comprehension of how this model generates a solution and increase the chances of its application within medicine. As for future advancements in this case, we will consider the shift toward purely 3D-based technologies, the growth of the database, and efficiency enhancement during analysis. The technique of domain adaptation should be considered here so that the model can become more versatile and flexible in different MRI facilities. Overall, NeuroASD-Net is a rather promising and scalable platform for the automatic diagnosis of autism spectrum disorder.

All the Declarations and Statements

Author Contributions Statement

Harendra Sharma: Proposed research ideas, constructed the overall framework, performed experiments and drafted the initial manuscript.

Oshin Sharma: Reviewed and edited the manuscript, ensured clarity and coherence, and helped coordinate project milestones and deadlines.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Funding Declaration

On behalf of all authors, the corresponding author states that no funds were received for this project.

Data Availability Statement

This study analysed publicly available datasets. The results obtained and datasets can be found here: “<https://www.kaggle.com/dswainsonsojana/autism-preprocessed>”, accessed on “December 2025”.

Ethical Declarations

No experiments involving human participants or animals were conducted in this study.

Acknowledgments

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

The authors declare that no generative AI or AI-assisted technologies were used in the preparation, writing, or editing of this manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

ASD - Autism Spectrum Disorder
 MRI - Magnetic Resonance Imaging
 CLAHE - Contrast Limited Adaptive Histogram Equalization
 GLCM - Gray Level Co-occurrence Matrix
 TLFE - Two-Level Feature Extraction
 SA-BKA - Self-Adaptive Black-Winged Kite Algorithm
 DSR - Dual Scale Residual
 CBAM - Convolutional Block Attention Module
 CNN - Convolution Neural Network
 DL - Deep Learning
 LBP - Local Binary Patterns
 GMM - Gaussian Mixture Model
 FCM - Fuzzy C Means
 CED - Canny Edge Detection
 DCNN - Deep CNN
 ABIDE - Autism Brain Imaging Data Exchange
 PFT - Progressive Fourier Transform
 MC-NFE - Multi-site Clustering and Nested Feature Extraction
 SVD - Singular Value Decomposition
 RF - Random Forest
 LR - Logistic Regression
 DT - Decision Trees
 KNN - K-Nearest Neighbors
 HM - Histogram Matching
 FLOPs - Floating Point Operations Per Second
 CAM - Channel Attention Module
 SAM - Spatial Attention Module
 LOSO - Leave-One-Site-Out

Appendix

Not applicable (no additional materials beyond figures/tables in main text).

References

- [1] Bahathiq, R.A., Banjar, H., Bamaga, A.K. and Jarraya, S.K., 2022. Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging. *Frontiers in Neuroinformatics*, 16, p.949926.
- [2] Ali, M.T., ElNakieb, Y., Elnakib, A., Shalaby, A., Mahmoud, A., Ghazal, M., Yousaf, J., Abu Khalifeh, H., Casanova, M., Barnes, G. and El-Baz, A., 2022. The role of structure MRI in diagnosing autism. *Diagnostics*, 12(1), p.165.
- [3] Saputra, D.C.E., Maulana, Y., Win, T.A., Phann, R. and Caesarendra, W., 2023. Implementation of machine learning and deep learning models based on structural MRI for identification autism spectrum disorder. *Journal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(2), pp.307-318.
- [4] Duan, Y., Zhao, W., Luo, C., Liu, X., Jiang, H., Tang, Y., Liu, C. and Yao, D., 2022. Identifying and predicting autism spectrum disorder based on multi-site structural MRI with machine learning. *Frontiers in human neuroscience*, 15, p.765517.
- [5] Elshoky, B.R.G., Younis, E.M., Ali, A.A. and Ibrahim, O.A.S., 2022. Comparing automated and non-automated machine learning for autism spectrum disorders classification using facial images. *Etri Journal*, 44(4), pp.613-623.
- [6] Moridian, P., Ghassemi, N., Jafari, M., Salloum-Asfar, S., Sadeghi, D., Khodatars, M., Shoeibi, A., Khosravi, A., Ling, S.H., Subasi, A. and Alizadehsani, R., 2022. Automatic autism spectrum disorder detection using artificial intelligence methods with

- MRI neuroimaging: A review. *Frontiers in Molecular Neuroscience*, 15, p.999605.
- [7] Kadi, I., Abbas, M. and Eddine, M.M.C., 2024. Intelligent models for early Autism detection from MRI images. *Studies in Engineering and Exact Sciences*, 5(2), pp.e8464-e8464.
 - [8] Han, J., Jiang, G., Ouyang, G. and Li, X., 2022. A multimodal approach for identifying autism spectrum disorders in children. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30, pp.2003-2011.
 - [9] Talaat, F.M., Ali, Z.H., Mostafa, R.R. and El-Rashidy, N., 2024. Real-time facial emotion recognition model based on kernel autoencoder and convolutional neural network for autism children. *Soft Computing*, pp.1-14.
 - [10] Alcañiz, M., Chicchi-Giglioli, I.A., Carrasco Ribelles, L.A., MarínMorales, J., Minissi, M.E., Teruel García, G., Sirera, M. and Abad, L., 2022. Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis. *Autism Research*, 15(1), pp.131-145.
 - [11] Feng, M. and Xu, J., 2023. Detection of ASD children through deep-learning application of fMRI. *Children*, 10(10), p.1654.
 - [12] Kuttala, D., Mahapatra, D., Subramanian, R. and Oruganti, V.R.M., 2022. Dense attentive GAN-based one-class model for detection of autism and ADHD. *Journal of King Saud University-Computer and Information Sciences*, 34(10), pp.10444-10458.
 - [13] Durkut, M., Blok, E., Suleri, A. and White, T., 2022. The longitudinal bidirectional relationship between autistic traits and brain morphology from childhood to adolescence: a population-based cohort study. *Molecular Autism*, 13(1), p.31.
 - [14] Ashraf, A., Qingjie, Z., Bangyal, W.H.K. and Iqbal, M., 2023. Analysis of brain imaging data for the detection of early age autism spectrum disorder using transfer learning approaches for internet of things. *IEEE Transactions on Consumer Electronics*, 70(1), pp.4478-4489.
 - [15] Nagesh, N., Patil, P., Patil, S. and Kokatanur, M., 2022. An architectural framework for automatic detection of autism using deep convolution networks and genetic algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, 12(2), pp.1768-1775.
 - [16] Jain, S., Tripathy, H.K., Mallik, S., Qin, H., Shaalan, Y. and Shaalan, K., 2023. Autism Detection of MRI Brain Images Using Hybrid Deep CNN With DM-Resnet Classifier. *IEEE Access*.
 - [17] Devika, K., Mahapatra, D., Subramanian, R. and Oruganti, V.R.M., 2022. Outlier-based autism detection using longitudinal structural MRI. *IEEE Access*, 10, pp.27794-27808.
 - [18] Nogay, H.S. and Adeli, H., 2023. Diagnostic of autism spectrum disorder based on structural brain MRI images using grid search optimization and convolutional neural networks. *Biomedical Signal Processing and Control*, 79, p.104234.
 - [19] Yakolli, N., Anusha, V., Khan, A.A., Shubhashree, A. and Chatterjee, S., 2023. Enhancing the diagnosis of autism spectrum disorder using phenotypic, structural, and functional MRI data. *Innovations in Systems and Software Engineering*, pp.1-12.
 - [20] Belhaouari, S.B., Talbi, A., Hassan, S., Al-Thani, D. and Qaraqe, M., 2023. PFT: A novel time-frequency decomposition of bold fMRI signals for autism spectrum disorder detection. *Sustainability*, 15(5), p.4094.
 - [21] Koc, E., Kalkan, H. and Bilgen, S., 2023. Autism Spectrum Disorder Detection by Hybrid Convolutional Recurrent Neural Networks from Structural and Resting State Functional MRI Images. *Autism Research and Treatment*, 2023(1), p.4136087.
 - [22] Oruganti, V.R.M., Mahapatra, D. and Subramanian, R., 2022. Outlier-based autism detection using longitudinal structural mri. *arXiv preprint arXiv:2202.09988*.
 - [23] Wang, N., Yao, D., Ma, L. and Liu, M., 2022. Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with resting-state fMRI. *Medical image analysis*, 75, p.102279.
 - [24] Sharif, H. and Khan, R.A., 2022. A novel machine learning based framework for detection of autism spectrum disorder (ASD). *Applied Artificial Intelligence*, 36(1), p.2004655.
 - [25] Alkahtani, H., Aldhyani, T.H. and Alzahrani, M.Y., 2023. Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks. *Applied Sciences*, 13(8), p.4855.

Authors' Profiles



Harendra Sharma is a Ph.D. research scholar and is currently working in the Department of Computer Science & Engineering of SRM Institute of Science and Technology, Delhi-NCR Campus Ghaziabad, UP. His research interests include medical image processing, machine learning, autism spectrum disorder. He has published more than 7 research papers in Scopus indexed journals and conferences.



Dr. Oshin Sharma is currently working as an Associate Professor in the Department of Computer Science & Engineering of SRM Institute of Science and Technology, Delhi-NCR Campus Ghaziabad, UP. She has 7.5 years of teaching experience from various reputed Universities. Her research interests include Energy efficient Cloud/Fog computing, networking, industry 4.0, Machine learning towards health care and Internet of things. She has published more than 40 research papers in SCI and Scopus indexed journals. She is also a reviewer of various Journals and Conferences such as: *Int. J. of Ad Hoc and Ubiquitous Computing*, *Journal of information science and engineering*, *Parallel and distributed computing*.

How to cite this paper:

Harendra Sharma, Oshin Sharma, "NeuroASD-Net: A Deep Learning-Based Approach for Autism Detection from Structural MRI", International Journal of Intelligent Systems and Applications (IJISA), Vol.18, No.5, pp.191-212, 2026. DOI:10.5815/ijisa.2026.05.10