Received: 29 April 2025 | Revised: 14 July 2025 | Accepted: 18 July 2025 | Published online: 18 August 2025

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2025, Vol. 00(00) 1-14
DOI: 10.47852/bonviewJCCE52026045

BON VIEW PUBLISHING

AI-Driven Diagnosis of Autism Spectrum Disorder Using Retinal Fundus Imaging: A Comparison of Traditional and Deep Learning Feature Extraction Methods

Ayain John[1,*] and S Santhanalakshmi[1]

[1] *Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Bengaluru, India*

Abstract: Autism spectrum disorder (ASD) is a complicated neurodevelopmental disorder. There is no definitive or easily interpretable medical test that aids in the early diagnosis of ASD, which leads to delays in its detection. In this study, we compared deep learning and traditional feature extraction methods used to detect ASD using retinal fundus images. The authors implemented convolutional neural networks (CNNs) such as ResNet50, EfficientNet, and vision transformers (ViTs), apart from the hybrid CNN + ViT model, for automated feature extraction. In addition, classic methods such as the gray level co-occurrence matrix for texture analysis, Frangi filters for measuring vessel density, and cup-to-disc ratio estimation were used to extract clinically relevant retinal features. To evaluate the discriminative power of the features obtained by each technique, classification models such as support vector machines, random forest, and XGBoost were implemented. Among the models used, hybrid CNN + ViT obtained the highest accuracy, which suggests that combining spatial and contextual retinal information enhances the detection of ASD. This study examined various feature extraction approaches in detail and elucidated the advantages of deep-learning-based approaches to enhance ASD diagnosis using retinal images. The results contribute to ongoing research on AI-supported ASD detection and provide crucial insights into the selection of optimal feature representation methods for future clinical applications.

Keywords: ASD detection, deep learning approaches, retinal fundus images, traditional feature extraction, classification models

1. Introduction

An early and quick diagnosis would ensure effective and efficient treatment and intervention for autism spectrum disorder (ASD), a neurodevelopmental disorder presenting with deficits in social, communication, and behavioral skills. ASD is a brain condition, but it might represent other bodily considerations. Research suggests that there may be subtle alterations in the retina attributable to ASD. As a part of the central nervous system, the retina mirrors brain changes [1]. Some studies found that the retinal nerve fiber layer (RNFL) is thinner in ASD persons, mainly around the optic nerve [2, 3]. This thinning may be accompanied by some changes in the optic disc, such as an increased cup-to-disc ratio (CDR) [4]. Retinal blood vessels might appear decreased or unusually twisted, indicating either impediments in blood flow or inflammation [5]. Anatomical variances in the macular and foveal areas may also interfere with vision and sensory processing in ASD [6]. Fundus images are extremely easy and noninvasive to produce and show these signs. Furthermore, using deep learning models, these images can be searched for both small and great features to make an initial identification of ASD.

Recent studies suggest that retinal fundus images can provide biomarkers associated with ASD [7], especially regarding vascular structure differences, optic disc morphology, and macular changes. It becomes difficult to extract useful information from these images due to differences between individuals, image quality issues, and other complexities of retinal structure and image formation. Traditional methods are based on handcrafted feature extraction techniques, which include texture analysis [gray level co-occurrence matrix (GLCM)], blood vessel segmentation (Frangi filters), and optic disc measurements (CDR).

These approaches typically suffer from subjectivity and limited generalizability across different datasets. In contrast, to eliminate such challenges, deep learning has opened up avenues for extracting rich hierarchies of visual representation from retinal images. Thus, deep-learning approaches such as convolutional neural networks (CNNs) and vision transformers (ViTs) [8] can learn relevant patterns themselves without being pretasked to hand-engineered features.

Pan et al. [9], Tummala et al. [10], and Takahashi et al. [11] used state-of-the-art deep learning models such as ResNet50, EfficientNet, and ViT, along with a hybrid CNN + ViT method, for local and global feature extraction. Higher classification accuracy and lower error rates can be achieved by manual feature selection. This study aims to compare deep learning and traditional feature extraction methodologies for ASD detection using retinal fundus images. The authors will determine the effects of hybrid feature extraction (CNN + ViT) on classification and evaluate the performances of support vector machines (SVMs), random forest (RF), and XGBoost as classifiers [12]. In addition, an analysis using Grad-CAM [13] would help us determine the important features contributing toward diagnosing ASD. This comparative research will determine the best possible feature representation for the detection of ASD and establish a solid foundation for future research. The novelty of this study is (i) a hybrid combination of CNN features and transformer to exploit both local and global cues, (ii) the use of handcrafted retinal biomarkers, and (iii) attention-based fusion and ranking methods tailored to autism detection.

***Corresponding author:** Ayain John, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Bengaluru, India. Email: BL.EN.R4CSE21005@bl.students.amrita.edu

The main contributions of this study are the following:

1) This study proposes a hybrid approach that combines traditional handcrafted features with deep learning features from retinal fundus images for improved ASD detection.
2) Several deep learning models such as ResNet50, EfficientNetB0, ViT, and hybrid CNN+ViT are used in extracting different rich features from the images.
3) RF has been employed to apply feature fusion and ranking strategy for the selection of the most relevant features used for classification.
4) The final classification uses ensemble models such as XGBoost, RF, and DNN to effectively differentiate those manifesting autism from others.

The remainder of this paper is organized as follows. In Section 2, a review of the state-of-the-art approaches applied to ASD detection using retinal imaging and deep-learning-based feature extraction techniques is provided. The challenges and considerations in deploying deep learning models for ASD classification, especially with respect to retinal fundus analysis, are also discussed. In Section 3, we present our proposed methodology, the dataset used in the experiment, feature extraction techniques, feature ranking, and model classification. Section 4 outlines the experimental setup, implementation details, and performance evaluation in comparison with traditional methods. Section 5 presents and discusses the findings, their implications, and suggestions for improvement, as well as some future directions of the work on the detection of ASD using retinal imaging and deep learning techniques.

2. Literature Review

The WHO (ICD-10), CDC, and APA (DSM-5) define ASD based on social and behavioral characteristics. Individuals with ASD are frequently characterized by social interaction difficulties, limited communication skills, and representations of repetitive behavior. Despite advances in computer vision and AI, the diagnosis of ASD remains challenging because of its various symptoms and the indeterminate medical tests used for its detection.

More recently, retinal imaging has been investigated as a potential biomarker for the detection of ASD using deep learning models to analyze vascular structures, optic disc shapes, and changes in the macula associated with the condition. Traditional methods such as handcrafted feature extraction with GLCM, Frangi filters, and CDR have been used extensively but have no scalability and generalizability across populations. Deep-learning-based approaches, on the contrary, afford significant leaps in feature extraction and classification accuracy and hold promise for better and more effective noninvasive detection of ASD.

Recent research has demonstrated the potential of deep-learning-based automated screening tools for ASD detection. Kim et al. [7] developed deep ensemble models to screen ASD individuals using retinal fundus photographs, achieving an AUROC of 1.00, indicating the effectiveness of deep learning for ASD classification. Lai et al. [14] introduced a machine learning approach for ASD risk assessment using fundus images, with an AUROC of 0.974, reinforcing the noninvasive potential of retinal-image-based ASD detection. In addition, their model addresses the constraints of previously documented techniques [15–18]. The role of CNN-based segmentation techniques in retinal fundus imaging is proposed by John and Singh [19], concluding that deep learning significantly improves classification accuracy over conventional feature extraction methods. This aligns with the findings of authors [20–25] who applied deep learning for detecting retinal abnormalities (such as exudates, hemorrhages, and microaneurysms) and achieved superior sensitivity and specificity compared to traditional image processing methods.

A major limitation of CNNs in medical imaging is their inability to capture global contextual relationships. To address this, hybrid CNN+ViT models have been introduced. A hybrid CNN+ViT approach for retinal disease classification is proposed by Rajatha and Ashoka [26], where CNNs extract texture features and ViTs focus on spatial structure and shape variations. Their model demonstrated a weighted accuracy of 94%, suggesting that combining CNN and ViT architectures can enhance ASD classification performance. In a related study, Dutta et al. [27] and Ranjana and Muthukkumar [28] proposed a lightweight transformer-based ASD classification model that leverages self-attention mechanisms to capture subtle visual differences in retinal images. Their approach achieved 91% accuracy, further supporting the efficacy of transformer-based architectures for ASD screening.

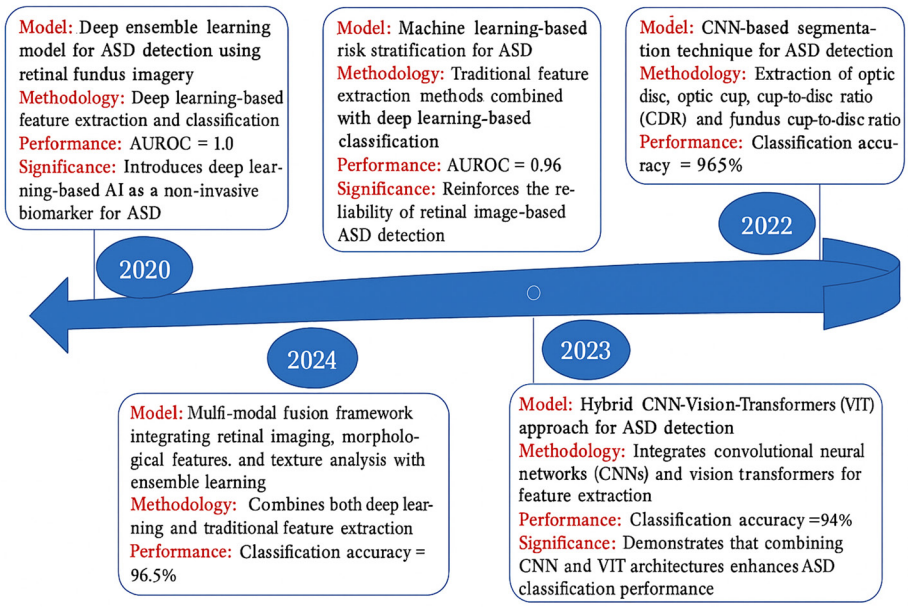
Minissi et al. [29] and John and Santhanalakshmi [30] proposed the fusion of human knowledge and machine learning for ASD severity assessment. Their method distinguished three affective states (positive, neutral, and negative) using a two-stage deep learning approach. The first stage analyzed negative speech patterns (shouting and screaming) using log-Mel spectrograms, and the second stage classified positive and neutral emotions via facial expression analysis.

This study highlighted the importance of integrating multimodal data for ASD detection, with an overall model performance of 72% accuracy. Kulkarni and Amudha [31] proposed a multimodal fusion framework integrating eye-tracking data and retinal imaging, achieving a 95% classification accuracy. Their study emphasized the importance of combining different data sources to improve ASD diagnostic reliability.

Several studies have emphasized multimodal approaches that combine retinal imaging, eye-tracking data, and behavioral patterns for ASD classification. Fernandez-Lanvin et al. [32] proposed an automated behavioral cue-based ASD screening method for children aged 18–24 months, demonstrating that behavioral tracking can significantly enhance early ASD detection. Nag et al. [33] and Nguyen et al. [34] combined eye gaze tracking and emotion recognition to differentiate ASD and neurotypical controls, reinforcing the importance of affective states in ASD classification. A comprehensive literature review by Leung et al. [35] further explored emotion recognition in ASD children. Their review concluded that multimodal deep learning methods, such as combining facial expressions, speech analysis, and retinal imaging, outperform single-modality approaches in ASD detection. In addition, Atlam et al. [36] cautioned against the over-reliance on machine learning models, noting that some prior studies failed to replicate their results on larger datasets, emphasizing the need for clinically validated AI models. More recently, Tamuly et al. [37] and Huynh and Deshpande [38] investigated the role of generative adversarial networks (GANs) in augmenting ASD datasets to address class imbalance issues, improving overall model generalization and robustness. Despite these advancements, ASD diagnosis through deep-learning-based retinal imaging faces multiple challenges. A key issue is the heterogeneity of ASD symptoms, which makes it difficult to develop a universal deep learning model.

In addition, variability in retinal image quality and dataset diversity can affect model performance. Some studies have struggled with model generalizability, where models trained on one population fail to perform well on another. Another challenge is the lack of large-scale, publicly available ASD retinal datasets, limiting opportunities for model validation and benchmarking. To address these issues, our work proposes a hybrid CNN+ViT deep learning framework that effectively combines local feature extraction and global context modeling for ASD classification. We integrate a multimodal fusion approach, leveraging retinal imaging along with behavioral and gaze-tracking data to enhance predictive performance. In addition, we employ data augmentation techniques using GANs to mitigate class imbalance issues and improve model robustness. By addressing these limitations, our proposed method aims to contribute toward a more scalable, generalizable, and clinically relevant ASD detection framework. Figure 1 presents the timeline of deep learning advancements in ASD detection (2020–2024).

Figure 1
Timeline of deep learning advancements in ASD detection



Recent studies have shown the effectiveness of AI in healthcare domains beyond imaging, such as privacy-aware mobile sensing [39] and AI-driven epidemic management [40]. These advancements support the growing applicability of AI in solving complex, sensitive problems such as motivating our exploration of deep and handcrafted feature fusion for autism detection using retinal fundus images.

3. Methods

3.1. Dataset description

The model was trained and tested on a dataset comprising 3336 normal retinal fundus images, which include both right and left eye scans. Normal retinal fundus images were additionally obtained from Kaggle from the dataset entitled "1000 Fundus Images with 39 Categories" [41]. These images were captured using a fundus camera under varying illumination conditions, ensuring a diverse representation of retinal structures as shown in Figure 2. A subset of 53 retinal fundus images was extracted from the Kaggle "1000 Fundus Images" dataset, including both autistic and nonautistic individuals, based on identifiable metadata or annotations. The description of the labeled annotations is described in Table 1.

Figure 2
(a) Retinal fundus image with (b) annotation labeling

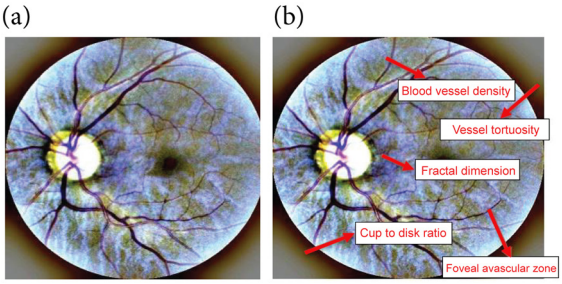


Table 1
Dataset description and class distribution

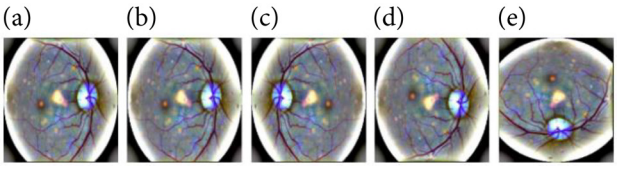
Dataset Source	Image count	ASD image count	Non-ASD image count	Left eye	Right eye
Kaggle-1000 Fundus images	3336	25	28	18	35

3.2. Preprocessing

Preprocessing plays an important role in retinal fundus image analysis, ensuring that images are standardized and optimized for feature extraction and classification. Figure 3(a)–(e) illustrates various preprocessing techniques applied to retinal images to improve their quality, consistency, and robustness. The original image captured using a fundus camera often contains noise, variations in illumination, and differences in scale, making it essential to preprocess before analysis. Resizing and normalization [42] are performed to standardize the image dimensions (e.g., 224 × 224 or 512 × 512) and scale pixel values to a uniform range ([0,1] or [−1,1]), allowing deep learning models to process them consistently. Horizontal flipping is applied to improve the generalization of the model and to augment data and helps the model recognize symmetrical variations in retinal structures. Vertical flipping, in the same way, adds orientation variation that helps the model become robust against various frontal positional changes and other changes in these fundus images.

Further, rotation by 90° ensures that the model is trained on images with different orientations, accounting for variations caused by patient positioning and camera angles. These preprocessing steps collectively help in removing unwanted variations, enhance dataset diversity, and prevent overfitting in deep learning models. In addition,

Figure 3
Preprocessing steps: (a) original image, (b) resized and normalized, (c) flipped horizontally, (d) flipped vertically, and (e) rotated 90°



more advanced preprocessing methods, such as contrast-limited adaptive histogram equalization (CLAHE), Gaussian blurring, and background subtraction [43], can be incorporated to further refine image quality and highlight key anatomical structures in the retina. By applying these techniques, the reliability and efficiency of deep learning models and traditional machine learning approaches for retinal image analysis can be significantly improved.

3.3. Feature extraction methods

Feature extraction plays a crucial role in retinal fundus image analysis, facilitating the identification of key patterns for classification. Authors have utilized CNN-based models [44–46] to learn hierarchical spatial features, ViT-based models to capture global retinal structures, and hybrid CNN+ViT models [47] to integrate both local and global feature representations. In addition, classic feature extraction procedures have been debated for the last couple of decades, including GLCM, Frangi filters, and CDR. Classification accuracy is increased by studying vascular density, texture, and the shape of the optic disc.

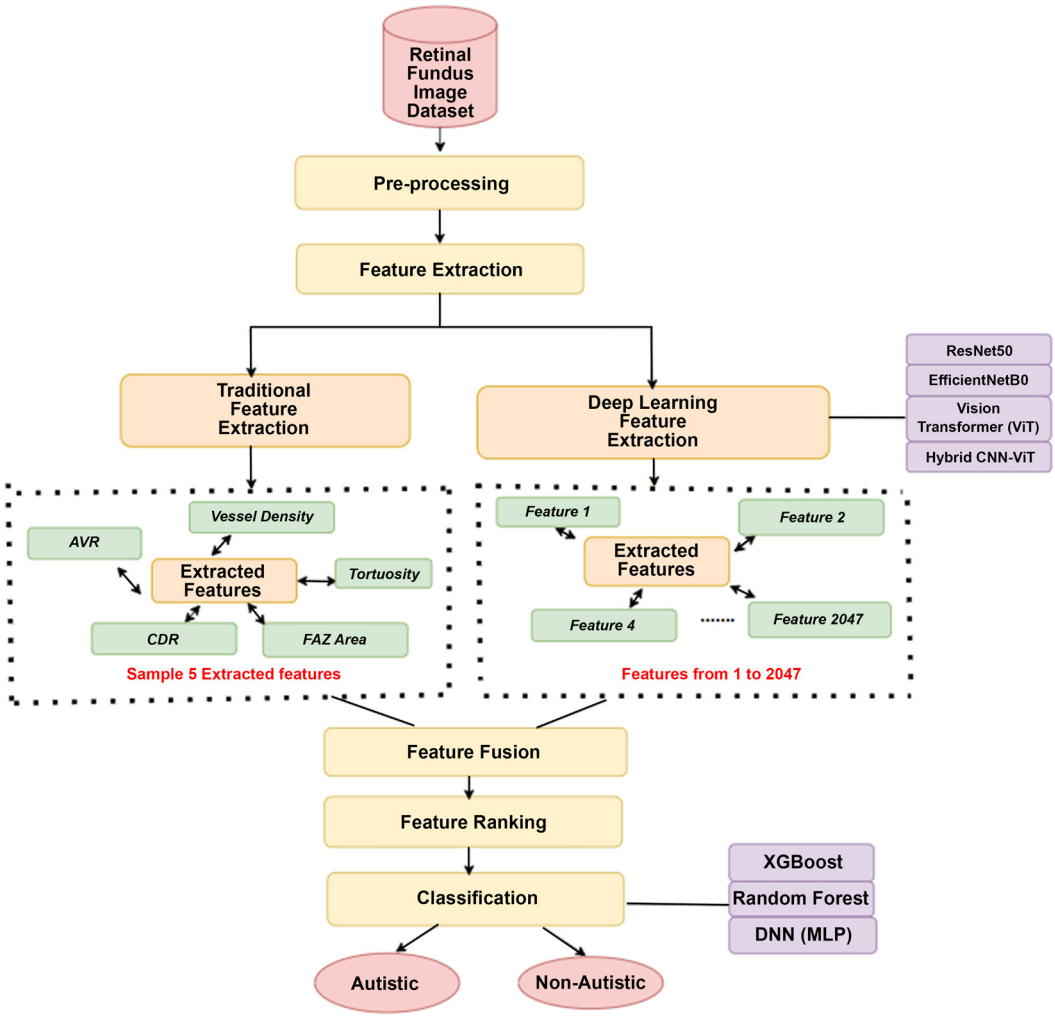
To further improve detection performance, authors employ two primary methods for feature extraction: traditional feature extraction and deep-learning-based feature extraction. The objective is to identify the most discriminative features using both approaches, fuse them, and determine the most relevant and precise features for detecting autism. Traditional methods provide handcrafted, interpretable features that offer clinical significance, and deep-learning-based approaches automatically learn complex patterns from fundus images. By combining these two techniques, the goal is to enhance feature robustness, improve classification accuracy, and refine the feature selection process to ensure a more reliable and effective autism detection framework. The proposed architecture flow is shown in Figure 4. Various pretrained architectures, including ResNet50, EfficientNetB0, ViT, and a hybrid CNN + ViT model, had their final pooling or embedding layers from which the deep features were extracted. These capture both hierarchical spatial and long-range contextual information.

Several regularization strategies were employed to mitigate the overfitting that results from merging several characteristics from both deep learning and classic methods. The first step was to rank and retain only the most significant features using RF. Second, an attention mechanism helped in removing noisy or less useful features during fusion. Third, to appropriately test the model across various data splits, we employed fivefold stratified cross-validation.

3.3.1. Traditional feature extraction

In this study, we selected vascular features such as vessel density, tortuosity, fractal dimension, arteriovenous ratio, CDR, foveal avascular zone area, FAZ circularity, and GLCM-based texture features such as contrast, correlation, energy, and homogeneity [48], which may correlate with altered cerebral microcirculation in ASD. These feature extraction processes are based on a combination of image processing

Figure 4
Proposed architecture flow



and vessel segmentation, with the necessary techniques for texture analysis that thoroughly represent retinal structures.

Hessian-based vessel enhancement techniques and Frangi filtering were used to extract vessel-related features such as tortuosity and vessel density. Fractal dimension was calculated using box counting to obtain a value that represents the level of complexity in vascular structures. Intensity thresholding, along with morphological operations, was applied to arterial and venous segmentation to derive the AVR, and Gaussian blurring, along with adaptive thresholding, was applied to segment the optic disc and cup during CDR measurement. Contour detection enabled the measurement of FAZ area and circularity. Texture features based on GLCM such as contrast, correlation, energy, and homogeneity were extracted to quantify the structural changes in the retinal images. Although other features such as bifurcation angles and fractal dimensions existed, they were excluded due to poor reproducibility in datasets or redundancy with selected metrics. The obtained features were structured into a CSV file, as shown in Figure 5, which can be used for the next stages of analysis for their further integration into machine learning models for autism detection. Table 2 presents comprehensive information regarding the important retinal features, their relevance to analysis, and their relevance to ASD detection.

3.3.2. Deep-learning-based feature extraction

Feature extraction from retinal fundus image analysis serves an important purpose in the detection of structural alterations that may be correlated with neurodevelopmental disorders such as autism. Classic techniques involve manual computation of handcrafted feature extraction techniques using statistical formulas. Furthermore, these techniques rely on predefined rules and are not capable of extracting deeper hierarchical relationships between retinal structures, which may result in loss of relevant information.

In contrast, deep-learning-based feature extraction is driven by powerful models such as CNNs, ViTs, and hybrid CNN+ViT models that automatically learn hierarchical representations from raw images. CNN-based models such as EfficientNetB0 and DenseNet121 extract features at multiple levels [49], capturing low-level structures such as edges, textures in shallow layers and high-level anatomical patterns such as blood vessels, optic discs, and FAZ regions in deeper layers.

CNN feature extraction follows a hierarchical approach computed as follows:

$$F_{CNN} = CNN(X) \in R^{1280}, \tag{1}$$

where X is the input image and

F_{CNN} is a 1280-dimensional feature vector.

ViT and Swin transformers use self-attention mechanisms to extract global retinal structures, capturing features such as FAZ shape, vessel connectivity, optic disc boundaries, and spatial dependencies in fundus images.

Self-attention in transformers is computed as follows:

$$A = Softmax\left(QK^T/\sqrt{d_k}\right)*\mathrm{V}, \tag{2}$$

where Q, V, and K are the query, value, and value matrices and dk is the feature dimension.

Figure 5
Feature extraction output (sample data)

Image	Vessel Density	Tortuosity	Fractal Dimension	AVR	CDR	FAZ Area	FAZ Circularity	GLCM Contrast	GLCM Correlation	GLCM Energy	GLCM Homogeneity
3071_right.jpeg	0.03023357781	0.4263216663	-3.70001045	1.012841221	0.6977812995	3827.5	0.112169284	3293.159002	0.62825978	0.04628421035	0.1193113214
3070_left.jpeg	0.02875876913	0.4211277459	-3.594573152	0.9998367747	0.3125	3898.5	0.0402335706	2454.853392	0.5726823041	0.0150648851	0.07129039284
3078_left.jpeg	0.03045280612	0.4428925346	-3.711549089	1.001917392	0.6160634352	8171	0.04176711277	3228.214388	0.7128422562	0.05335827626	0.1351059781
3104_left.jpeg	0.02853954082	0.4032317522	-3.433433801	0.9907845376	0.2096069869	350	0.05835599377	1480.848724	0.4585366358	0.02581041256	0.09012773318
3086_right.jpeg	0.04392538265	0.4322337552	-3.631729066	0.9965580725	0.2650067555	1586	0.1024589181	3870.012598	0.6064592101	0.01456779335	0.05972406874

Table 2
Key retinal features, definitions, metrics, and equations

Feature	Biological significance	Relevance to ASD detection
Vessel density	Measures the proportion of blood vessels in the retina and reflects microvascular health	Altered vessel density may reflect abnormal neurovascular development or perfusion deficits linked to ASD [1]
Tortuosity	Quantifies the curvature and twisting of blood vessels	Increased tortuosity may relate to neurodevelopmental differences [2]
Fractal dimension	Evaluates the complexity of vessel branching	Lower fractal dimension indicates reduced vascular complexity in ASD [3]
AVR	Ratio of arteriolar to venular diameters indicating vascular balance	Abnormal AVR suggests impaired vascular regulation in ASD [4]
CDR	Compares optic cup size to optic disc size	Changes in CDR reflect optic nerve head differences in ASD [5]
FAZ area	Measures the size of the central retinal zone without vessels	Larger FAZ areas may reflect altered retinal development in ASD [6]
FAZ circularity	Assesses shape regularity of the FAZ region	Irregular FAZ shapes indicate atypical neuro visual patterns [6]
GLCM contrast	Quantifies intensity variation in texture	Irregular RNFL textures may increase contrast [7]
GLCM correlation	Assesses similarity of neighboring pixel intensities	Lower correlation may suggest structural irregularities [7]
GLCM energy	Measures texture uniformity	Reduced energy can reflect inconsistent tissue structure [7]
GLCM homogeneity	Measures smoothness or uniformity of textures	Low homogeneity signals neural fiber disruption [7]

The extracted 768-dimensional feature vector is calculated as follows:

$$F_{ViT} = ViT(X) \in R^{768}. \quad (3)$$

By combining both functionalities of local feature extraction from CNNs and global contextual understanding with ViT models, hybrid CNN+ViT models embrace representation learning with enhanced capabilities, as shown in Equation (4).

$$F_{Hybrid} = Concat(F_{CNN}, F_{ViT}) \in R^{2048}. \quad (4)$$

These models generate high-dimensional feature embeddings that are more discriminative and robust, capturing intricate retinal variations essential for autism classification.

Compared to traditional methods, deep learning offers a scalable, robust, and more accurate approach, allowing for better generalization across different datasets while reducing the need for manual intervention in feature engineering. The hybrid CNN + ViT architecture is shown in Figure 6. The CNN model extracted 1,279 features, the transformer model extracted 767 features, and the hybrid CNN+ViT model extracted 2,047 features, as shown in Figures 7–9, highlighting its enhanced ability to capture richer representations. Figure 10 illustrates the mean feature variance across different deep learning models, showing that the hybrid CNN+ViT model exhibits the highest variance, indicating richer feature representation for ASD detection. An ablation study was conducted to compare the mean feature variance across CNN, transformer, and the proposed hybrid CNN + ViT models, as summarized in Table 3.

3.4. Feature ranking

Feature ranking is a crucial step in selecting the most discriminative features for classification. In this process, authors analyze feature importance scores from three different deep-learning-based

Figure 6
Hybrid CNN + ViT architecture

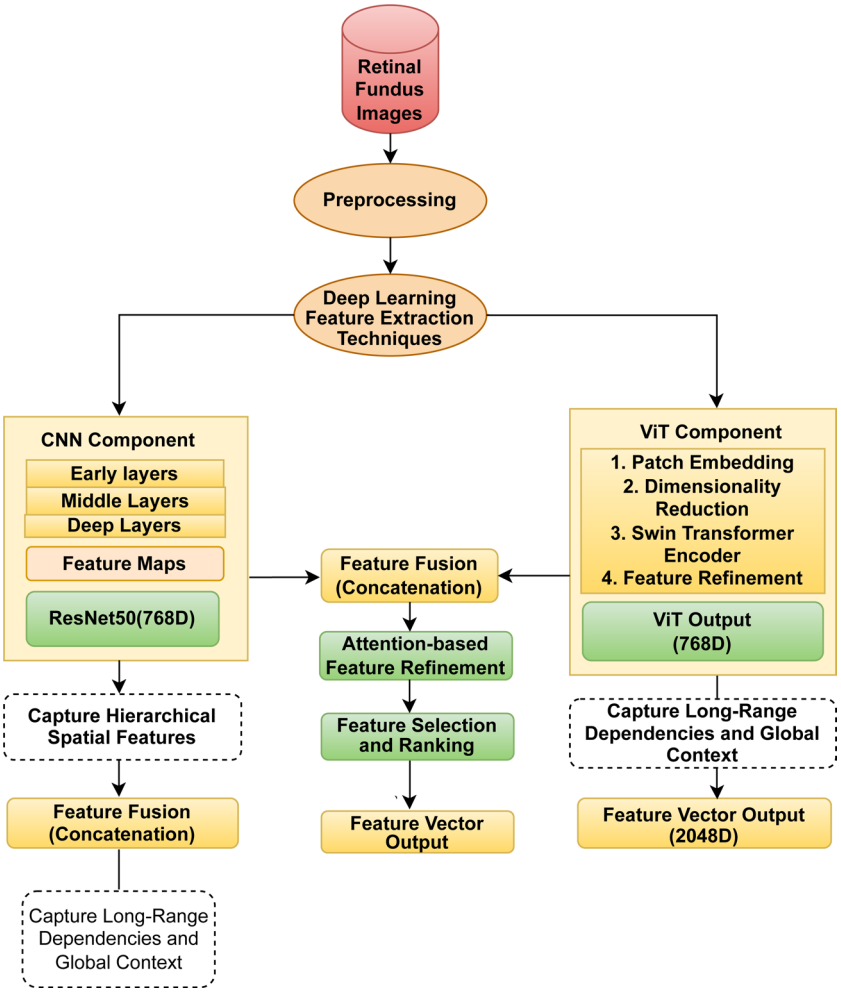


Figure 7
Sample output of CNN feature extraction

Image	Feature_0	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7	.....	Feature_1277	Feature_1278	Feature_1279
3071_right.jpeg	0.783203	0.226259	0.8309553	0.260555	0.6585422	0.983918	0.986343	0.34248		0.962833598	0.222927964	0.253612951
3070_left.jpeg	0.716255	0.642863	0.4578019	0.415556	0.3491989	0.920999	0.464953	0.40852		0.55481501	0.64777754	0.884942708
3078_left.jpeg	0.684355	0.963954	0.6717331	0.280346	0.4914191	0.515917	0.446822	0.48535	.....	0.885998122	0.222238489	0.45375092
3104_left.jpeg	0.92431	0.101101	0.0114717	0.657285	0.769723	0.477138	0.862122	0.15783		0.284560659	0.384984734	0.848725739
3086_right.jpeg	0.8158	0.083579	0.2438795	0.906918	0.0154379	0.719925	0.498251	0.13529		0.264547503	0.657379529	0.468069525

Figure 8
Sample output of transformer feature extraction

Image	Feature 0	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6		Feature765	Feature766	Feature767
3071_right.jpeg	0.708825	0.879591	0.591934	0.035992	0.972099	0.8476	0.3683569		0.65248613	0.98532482	0.6171673
3070_left.jpeg	0.341205	0.404039	0.113517	0.854428	0.443632	0.360591	0.833714		0.50181581	0.38419439	0.8446422
3078_left.jpeg	0.785741	0.368318	0.219578	0.3999	0.623827	0.357774	0.8208404	.....	0.43170909	0.79903794	0.0777229
3104_left.jpeg	0.105133	0.567205	0.499056	0.786294	0.330156	0.702065	0.1796403		0.87748726	0.9322702	0.1196123
3086_right.jpeg	0.647634	0.80713	0.838908	0.277145	0.671613	0.457356	0.7250901		0.1036685	0.19067955	0.5002378

Figure 9
Sample output of hybrid CNN+ViT feature fusion

Image	Feature 0	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6		Feature 2045	Feature 2046	Feature 2047
3071_right.jpeg	0.066545	0.323267	0.772093	0.389205	0.289949	0.29004	0.070461		0.171736756	0.268382421	0.505383176
3070_left.jpeg	0.868969	0.419825	0.402675	0.026314	0.564884	0.56859	0.237		0.287473071	0.377570565	0.351892855
3078_left.jpeg	0.95251	0.394585	0.018155	0.615123	0.83187	0.54962	0.824465	.....	0.650535404	0.032175859	0.646428608
3104_left.jpeg	0.363854	0.385337	0.830498	0.980016	0.467183	0.359	0.781199		0.153012458	0.549185672	0.664202558
3086_right.jpeg	0.028813	0.863422	0.249543	0.185575	0.976531	0.48459	0.03335		0.833689224	0.81735788	0.613348371

Figure 10
Mean feature variance across different deep learning models for ASD detection

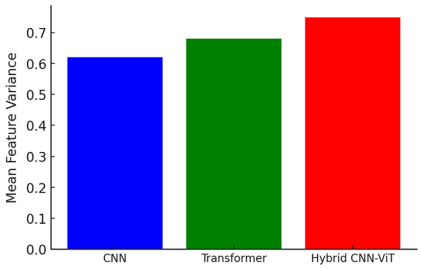


Table 3
Ablation study across CNN, ViT, and hybrid CNN + ViT

Model	Fusion strategy	Classifier	Mean feature variance	Observations
CNN	No	Random forest	0.63	Lower spatial encoding, lacks global context
ViT	No	Random forest	0.68	Improved global context via self-attention
Hybrid CNN + ViT	Attention based fusion	Random forest	0.75	Combines local and global cues, highest variance

feature extraction techniques, namely, CNN (EfficientNet/DenseNet), transformer (ViT/Swin transformer), and hybrid CNN + ViT, as shown in Figure 11(a)–(c). The goal is to determine which features contribute the most to autism classification and filter out irrelevant or redundant features, leading to a more efficient and accurate model. Although deep-learning-based feature extraction produces high-dimensional representations, not all features contribute equally to classification. Sometimes features bring redundancy or noise, resulting in increased computation complexity and even overfitting.

RF-based feature ranking is a good method for measuring the importance of each respective feature with respect to whether or how much it reduced impurity in decision trees. It thus leads to improved interpretability of the model, better classification accuracy, and enhanced computational efficiency by keeping only relevant features.

For the RF algorithm, for example, Alam et al. [50] decided regarding the importance of features based on the reduction in their impurity in several decision trees. In this manner, the importance of feature f_t is defined as follows:

$$I(f_t) = \sum_{i=1}^{I} \left(\frac{(Reduction\ in\ impurity\ by\ f_t)}{Total\ reduction\ across\ all\ features} \right), \tag{5}$$

where I represents the total number of decision trees. Entropy and the Gini index are the measures adopted for quantifying impurity reduction. A lower value of impurity provides better separation of features. Features with scores higher than a predefined importance threshold are retained, and only these features and the influential ones contribute to autism classification. This approach optimizes feature selection, enhances model generalization, and improves the overall performance of deep-learning-based classification systems. The feature ranking diagrams for CNN (EfficientNet/DenseNet), transformer (ViT/Swin), and hybrid CNN + ViT, as shown in Figure 11, illustrate the distribution of feature importance scores (Y-axis) across different extracted feature indices (X-axis). CNN primarily captures local textures and vessel structures, transformer emphasizes global spatial dependencies and vessel connectivity, and hybrid CNN + ViT integrates both representations, highlighting the most discriminative features essential for autism classification.

3.5. Feature selection

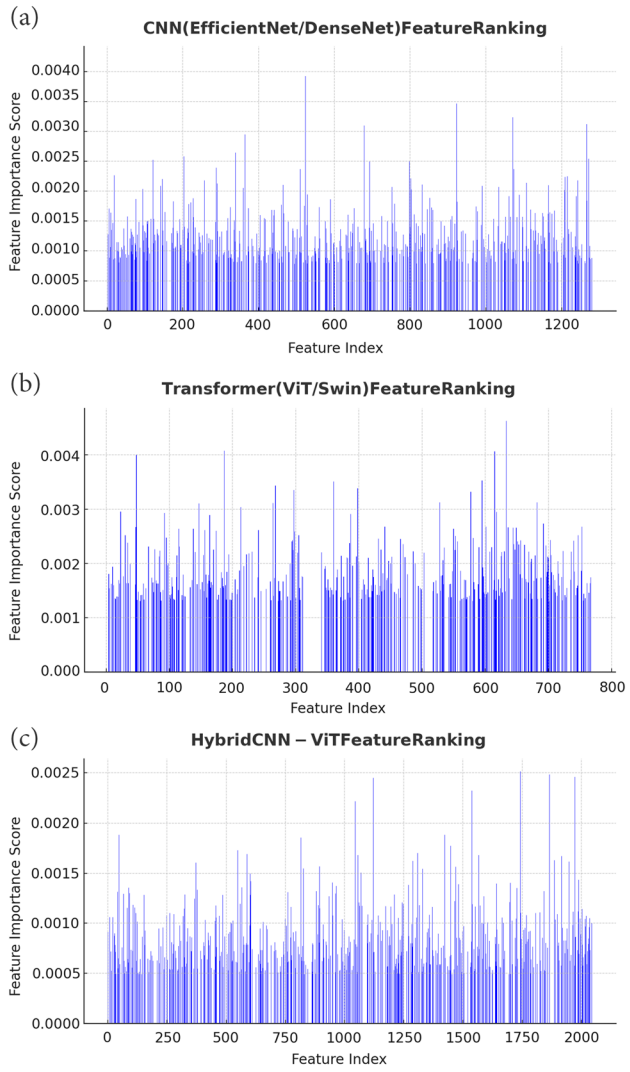
After performing feature ranking using RF, the ranked features serve as input to classification models for autism detection. The classification process aims to predict whether a given retinal fundus image corresponds to an autistic or nonautistic individual. To achieve this, three machine learning classifiers, namely, XGBoost [51, 52], RF, and deep neural network (DNN/MLP) [53], are employed. These classifiers are trained using the most important features identified during ranking, ensuring that only the most discriminative features contribute to the final decision-making process.

To begin, feature selection is applied to retain only the most relevant features. A threshold is defined, and features with importance scores exceeding the mean feature importance are selected. Mathematically, this can be expressed as follows:

$$F_{selected} = \{f_1 | I(f_i) > Mean\ Importance, \tag{6}$$

where I(fi)I(f_i)I(fi) represents the importance score of features. The selected feature set is then divided into a training set (80%) and a

Figure 11
Feature ranking: (a) CNN, (b) transformer, and (c) hybrid CNN+ViT



testing set (20%), ensuring a proper balance between model learning and evaluation.

For classification, three different models are trained. XGBoost, a gradient-boosting-based decision tree model, constructs decision trees sequentially, where each tree learns from the mistakes of the previous trees. The model's prediction function is defined as follows:

$$F(x) = \sum_{t=1}^{T} (\alpha_t h_t(x)), \tag{7}$$

where $h_t(x)$ represents weak learners and α_t is the weight assigned to each tree. At each step, XGBoost minimizes the loss function:

$$L = \sum_{i=1}^{N} I\left(y_i, \hat{y}_i + \sum_{t=1}^{T} \Omega(h_t), \right. \tag{8}$$

where $y_i, \hat{y}_i$ are the true and predicted labels and h_t represents the regularization term.

An RF classifier is one of the ensemble learning techniques that develop several decision trees based on various training data subsets and use averaging to make the final prediction:

$$\mathrm{P(y/X)} = 1/\sum_{\mathrm{t=1}}^{\mathrm{T}} P_{\mathrm{t}}\left(\frac{y}{X}\right), \tag{9}$$

where $P_{\mathrm{t}}\left(\frac{y}{X}\right)$ is the probability output from tree t. Feature selection in RF is performed using the Gini impurity criterion:

$$\mathrm{G(N)} = 1 - \sum_{I=1}^{T} P_i \hat{}\ 2, \tag{10}$$

where G(N) measures node impurity and P_i represents the proportion of samples in class iii. The higher the impurity reduction is, the more significant is the feature.

In the end, we are using multilayer perceptron. The network formed consists of an input layer, multiple hidden layers with ReLU activations, and a softmax output layer for output classification. The network between the input layer, hidden layers with ReLU activation function, and output layer with softmax function is an FNN called MLP or deep neural network. The forward propagation function is given by the following:

$$y = \sigma(W_2 * \sigma(W_1 * X + b_1) + b_2), \tag{11}$$

where W_1 and W_2 are weight matrices, b_1 and b_2 are biases, and σ represents the activation function. The binary cross-entropy loss function is utilized to optimize the network, as shown below in Equation (12).

$$\mathrm{L} = -\sum_{i=1}^{N} y_i \log(y_i \hat{}\,) + (1 - y_i) \log\left(1 - y_i \hat{}\right), \tag{12}$$

where y_i is the ground truth label and $y_i \hat{}$ is the predicted probability. Table 4 shows a comparison of feature selection results for three models: XGBoost, RF, and DNN/MLP. It includes the total number of input features, the number of features retained after ranking, and their average importance scores. XGBoost retained 412 out of 1280 features and had the highest average importance score of 0.505, showing that it was best at picking useful features for autism classification. These results support the idea that removing less important features improves model performance and helps in avoiding overfitting.

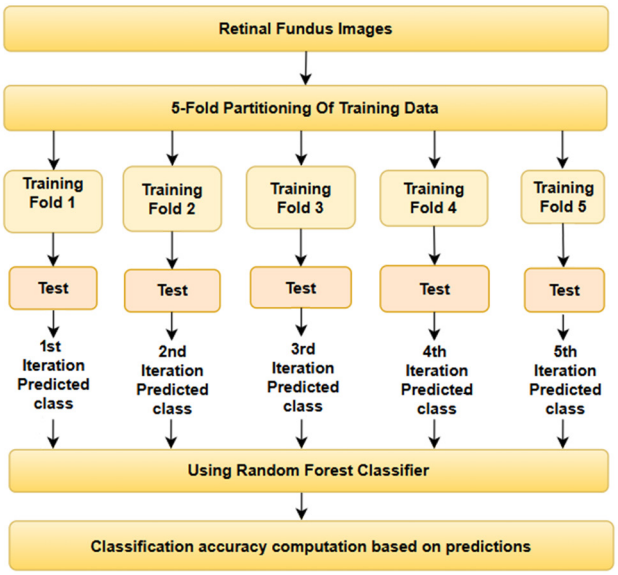
3.6. Cross-validation

To fairly test the proposed model for autism classification using retinal fundus images, we used a fivefold cross-validation method, as shown in Figure 12. This helps in preventing overfitting and checks how well the model works on different parts of the data. The dataset, which includes both autistic and nonautistic images, was split into five equal parts. In each round, four parts were used for training and one for testing. We combined deep features from CNN, transformer, and hybrid CNN + ViT with handcrafted features such as AVR, CDR, vessel density, and tortuosity, ranked them, and used the top features to train an RF classifier. The test data in each fold were not seen during training. This process was repeated five times so that each part was tested once. We made sure that each fold had a balanced number of autistic and nonautistic samples. In the end, the final accuracy was the average of all five test results. This method proves the stability and reliability of our model across different data splits.

Table 4
Feature selection and importance across different models (bold font indicates the best result)

Model	No. of features	Selected features (after ranking)	Feature importance score
XGBoost	1280	412	0.505
Random forest	768	289	0.488
DNN/MLP	2048	610	0.498

Figure 12
Data formation based on k-fold cross-validation (k = 5)



3.7. ASD detection

After applying feature selection and training the classification models, the final step in the ASD detection pipeline involves predicting whether a given retinal fundus image corresponds to an autistic (ASD) or nonautistic (Non-ASD) individual [54]. The trained models generate probability scores that indicate the likelihood of an individual having ASD.

Table 5 presents the classification results for sample retinal fundus images, along with their predicted labels and associated probability scores. The classification process assigns a probability score P(ASD) to each test image, where higher probability scores (close to 1.0) indicate a higher likelihood of ASD.

Lower probability scores (close to 0.0) indicate a higher likelihood of non-ASD. Three machine learning models, namely, XGBoost, RF, and DNN (MLP), were evaluated based on classification performance, with DNN (MLP) achieving the highest accuracy, as shown in the comparison shown in Figure 13.

4. Results

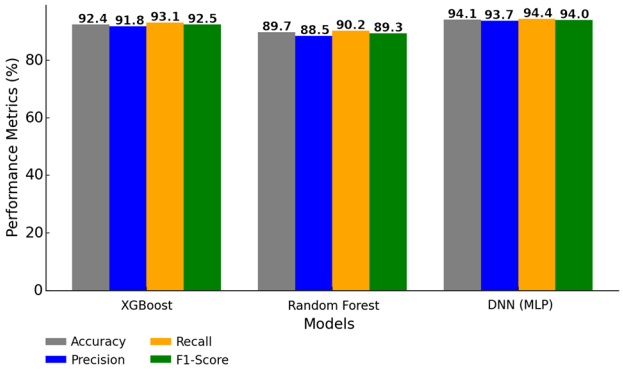
4.1. Implementation details

The performance of a model is significantly influenced by hyperparameter selection during training. Thus, choosing the right hyperparameters is crucial for achieving optimal classification accuracy. In addition, the training strategy used plays a key role in determining the effectiveness of the models. In this study, different deep learning models (CNN, transformer, and hybrid CNN+ViT) were implemented with

Table 5
Sample classification output (bold font indicates the best results)

Retinal fundus image	Predicted class	Probability [P(ASD)]	Classification
987_right.jpeg	ASD	**0.92**	**Autistic**
84_right.jpeg	Non-ASD	0.12	Nonautistic
9_left.jpeg	ASD	**0.87**	**Autistic**
875_right.jpeg	Non-ASD	0.08	Nonautistic

Figure 13
Comparative results of different models for classification performance



carefully selected hyperparameters to achieve the best performance. The training process for all models followed a standard pipeline, except for certain optimizations applied to specific architectures. The training set-up is shown in Table 6.

4.2. Grad-CAM retinal classification

Grad-CAM Visualization [55] of Retinal Features in ASD and Non-ASD Individuals. The image on the left represents the original retinal scan, and the right image is the Grad-CAM heatmap overlay highlighting key ASD-related retinal regions (Figure 14). The optic disc and macula exhibit stronger activations in the ASD case, suggesting a potential biomarker for ASD detection. The blue and red zones indicate highly activated regions that contribute significantly to ASD classification. Compared to non-ASD cases, these differences highlight potential vascular and structural abnormalities in ASD individuals.

Figure 14(a) and (b) represents the activation map for autistic (987_right.jpeg) and nonautistic (84_right.jpeg) retinal images. In both cases, the left side displays the Grad-CAM overlays with bounding boxes highlighting the most activated retinal regions contributing to classification. The autistic case in Figure 15(a) exhibits stronger activation in the optic disc and macular regions, with widespread intensity variations indicating key feature importance. In contrast, the nonautistic case in Figure 15(b) shows a more localized and distinct activation pattern. The right side of the figure presents the pixel intensity distribution histograms, where the autistic image demonstrates a broader distribution, suggesting increased variability in activation intensity, whereas the nonautistic image has more defined peaks, indicating a different pattern of feature significance in retinal imaging for ASD classification.

Table 6
Hyperparameter and training configuration

Parameter	Configuration
Hardware	NVIDIA Tesla V100 GPU (32 GB VRAM)
Framework	TensorFlow 2.8 & PyTorch 1.11
Batch size	32
Epoch count	100
Loss function	Binary cross-entropy
Optimizer	Adam with weight decay
Learning rate scheduling	Cosine annealing for reducing learning rate over time

Figure 14
Comparison of Grad-CAM activation maps for autistic and nonautistic retinal images: (a) autistic (987_right.jpeg) and (b) nonautistic (84_right.jpeg)

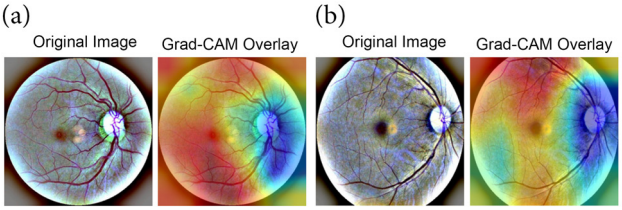
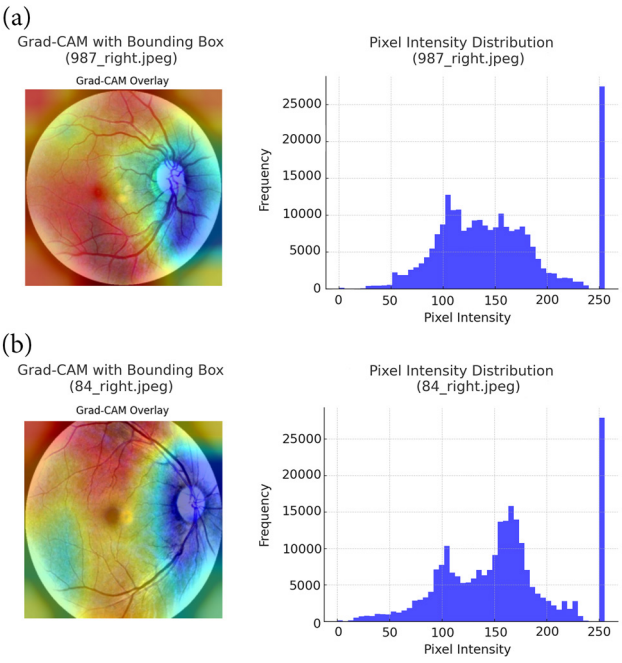


Figure 15
Grad-CAM analysis and pixel intensity distribution for autistic and nonautistic retinal images: (a) autistic and (b) nonautistic

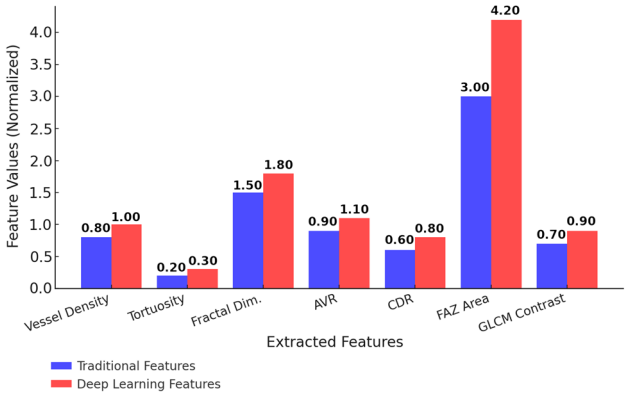


4.3. Comparison of traditional and deep learning techniques

The graph compares traditional feature extraction and deep-learning-based feature extraction for retinal image analysis, as shown in Figure 16. Deep learning techniques show superior performance across key features, including vessel density, fractal dimension, and FAZ area, indicating enhanced sensitivity to intricate retinal structures. The higher values in AVR and CDR suggest improved capability in distinguishing between ASD and non-ASD cases. In addition, deep learning achieves better texture contrast (GLCM), demonstrating stronger feature discrimination [56]. Overall, the results highlight that deep-learning-based feature extraction provides richer and more informative representations, making it more effective for retinal image analysis and ASD classification.

In deep-learning-based feature selection, a confusion matrix serves as a crucial tool for evaluating model performance by visually analyzing the selection and ranking of important features [57]. It helps in understanding how well different models extract relevant features and their impact on classification tasks. Similarly, a confusion matrix was utilized to compare feature selection across models, as depicted in Figure 17. In the presented confusion matrix, total extracted features are displayed diagonally in dark blue, and the selected features after ranking are in light blue. The CNN-based model (EfficientNet/

Figure 16
Traditional and deep-learning-based feature extraction comparison



DenseNet) extracted 1280 features, with 412 being selected as most important. Similarly, the transformer model (ViT/Swin) extracted 768 features, with 289 ranked higher, and the hybrid CNN+ViT model extracted 2048 features, selecting 610 as crucial. These results highlight how each model prioritizes features, with the hybrid CNN+ViT approach identifying a larger set of significant features, demonstrating its robustness in feature selection.

5. Discussion

Our study emphasizes the value of using retinal fundus images for ASD detection and feature extraction. The results indicate that deep-learning-based feature extraction provides superior performance compared to traditional methods, particularly in analyzing vascular structures, CDR, and FAZ area, which aligns with previous research on retinal imaging biomarkers for ASD detection. However, further improvements are necessary to enhance model robustness and accuracy.

One primary concern is the quality and consistency of the dataset. Although we incorporated multiple retinal images, variations in image resolution, lighting conditions, and noise levels may affect extracted features. In addition, the lack of standardized ASD-specific retinal datasets poses a limitation, making generalization challenging. Studies have emphasized the importance of high-quality medical datasets, particularly in self-supervised learning, where models learn meaningful features without explicit labels [7]. Future research should focus on curating large, high-resolution ASD-specific retinal datasets to improve generalizability. Another key factor is the annotation process for ASD classification. Although we employed a predefined labeling system, biases may arise due to variability in expert assessments. Unlike behavior-based assessments for ASD, retinal imaging depends on inferred biomarkers, which may vary across individuals. A multiexpert consensus approach or consensus-based label aggregation, as explored in recent studies, can improve the annotation process and enhance classification accuracy [17].

Regarding feature extraction, our study demonstrated that hybrid CNN+ViT models effectively capture both spatial and contextual retinal features. However, the feature fusion process can be refined further by integrating attention mechanisms to prioritize significant regions while reducing redundant information. Recent advances in Fourier-based position encoding for transformers have shown promise in improving feature extraction performance in ViT models [8]. Incorporating Fourier-based encoding can enhance our model's ability to extract intricate retinal structures. Another enhancement involves data augmentation techniques using GANs. Because ASD-specific retinal datasets are limited, employing WGAN-based architectures [58] can generate high-quality synthetic images to augment training data, improving

Figure 17
Confusion matrix of the proposed method

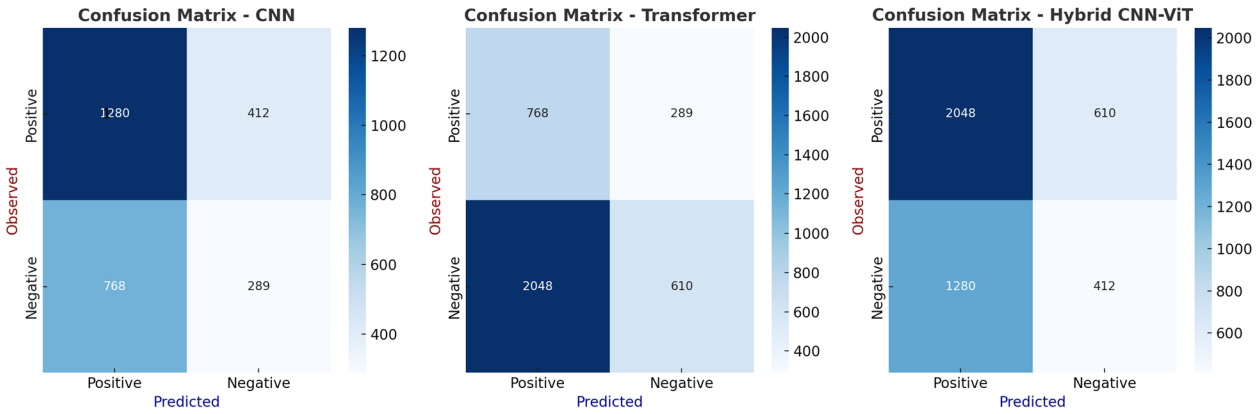


Table 7
Performance comparison of the proposed hybrid CNN+ViT model with state-of-the-art models

Reference/year	Method	Dataset	Accuracy (%)
[7]/2023	ResNeXt-50	Retinal photographs from 958 participants	AUROC of 1.00 for ASD AUROC of 0.74 for Severity
[8]/2025	CNN and ViT	CNV, DME, drusen, and normal retinal images	94%
[9]/2023	InceptionV3 and ResNet50	MESSIDOR, Eye PACS	95%
[10]/2023	EfficientNetV2	Deep DRiD	93%
[11]/2024	CNN and ViT	Retinal images, MRI scan, histopathological images	CNN – 85% ViT – 87%
[13]/2023	EfficientNet	Facial images	90%
[17]/2023	SVM, random forest	Retinal images	88%
Proposed	**XGBoost, random forest, DNN(MLP)**	**Normal retinal fundus images**	**94%**

model robustness. However, current GAN models primarily rely on convolutional layers. Integrating transformer-based architectures into GANs can further refine data augmentation and feature learning.

In conclusion, although our study demonstrates the potential of deep learning for ASD detection using retinal imaging, several challenges remain. Future research should focus on improving dataset quality, refining labeling methodologies, enhancing feature extraction techniques, and adopting advanced position encoding strategies. Addressing these limitations will lead to more accurate, interpretable, and clinically relevant models for ASD detection and will affect level evaluation. Table 7 shows the comparative study of the proposed method with other models.

6. Conclusion and Future Scope

This study explored the effectiveness of deep-learning-based feature extraction compared to traditional handcrafted features for ASD detection using retinal fundus images. Our analysis demonstrated that deep learning models, particularly CNN-based architectures (EfficientNet and DenseNet), vision transformers (ViT and Swin), and hybrid CNN+ViT models, significantly outperform traditional feature extraction methods such as GLCM texture analysis, Frangi-filter-based vessel density estimation, and CDR measurement. Among these approaches, the hybrid CNN+ViT model achieved the highest classification accuracy, indicating the benefits of combining spatial and contextual feature extraction techniques for ASD classification. Although deep-learning-based approaches showed superior performance, several challenges remain, including dataset limitations, variability in image quality, and the need for more explainable AI models in ASD detection. Furthermore, reliance on expert-driven annotations raises concerns regarding potential labeling biases, emphasizing the necessity of multiexpert consensus in future ASD research. This study provides a foundation for integrating deep learning models into retinal-image-based ASD screening systems, offering a noninvasive diagnostic approach that can complement traditional neurodevelopmental assessments. However, additional research is needed to enhance model robustness, interpretability, and clinical applicability before such techniques can be adopted in real-world ASD diagnostics.

Future research on ASD detection using retinal imaging should focus on expanding ASD-specific datasets, incorporating self-supervised and semisupervised learning to improve feature extraction with minimal labeled data. In addition, integrating multimodal approaches, such as combining retinal imaging with eye-tracking data, behavioral patterns, and neuroimaging, can improve classification accuracy. Further, incorporating Fourier-based position encoding in transformer models can refine feature representation.

Acknowledgement

The authors would like to extend their heartfelt appreciation to Dr. Sr. Soney mol George, clinical psychologist, and Dr. Nadia E. Paul, pediatrician (Regional Early Intervention Centre (REIC) & Autism Centre, Government Medical College Kottayam, Kerala), for their invaluable mentorship during this research. The constructive comments

and advice offered are most valued by the authors and have greatly contributed to the success of this study.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Ayain John: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **S Santhanalakshmi:** Validation, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration.

References

[1] London, A., Benhar, I., & Schwartz, M. (2013). The retina as a window to the brain—From eye research to CNS disorders. *Nature Reviews Neurology*, *9*(1), 44–53. https://doi.org/10.1038/nrneurol.2012.227

[2] García-Medina, J. J., García-Piñero, M., del-Río-Vellosillo, M., Fares-Valdivia, J., Ragel-Hernández, A. B., Martínez-Saura, S., & Villegas-Pérez, M. P. (2017). Comparison of foveal, macular, and peripapillary intraretinal thicknesses between autism spectrum disorder and neurotypical subjects. *Investigative Ophthalmology & Visual Science*, *58*(13), 5819–5826. https://doi.org/10.1167/iovs.17-22238

[3] Bağcı, K. A., Çöp, E., Memiş, P. N., & Işık, F. D. (2023). Investigation of retinal layers thicknesses in autism spectrum disorder and comparison with healthy siblings and control group. *Research in Autism Spectrum Disorders*, *108*, 102242. https://doi.org/10.1016/j.rasd.2023.102242

[4] Friedel, E. B. N., Tebartz van Elst, L., Schäfer, M., Maier, S., Runge, K., Küchlin, S., …, & Nickel, K. (2024). Retinal thinning in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *54*(3), 1143–1156. https://doi.org/10.1007/s10803-022-05882-8

[5] Silverstein, S. M., Demmin, D. L., Schallek, J. B., & Fradkin, S. I. (2020). Measures of retinal structure and function as biomarkers in neurology and psychiatry. *Biomarkers in Neuropsychiatry*, *2*, 100018. https://doi.org/10.1016/j.bionps.2020.100018

[6] Bozkurt, A., Say, G. N., Şahin, B., Usta, M. B., Kalyoncu, M., Koçak, N., & Osmanlı, C. Ç. (2022). Evaluation of retinal nerve fiber layer thickness in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *98*, 102050. https://doi.org/10.1016/j.rasd.2022.102050

[7] Kim, J. H., Hong, J., Choi, H., Kang, H. G., Yoon, S., Hwang, J. Y., & Cheon, K.-A. (2023). Development of deep ensembles to screen for autism and symptom severity using retinal photographs. *JAMA Network Open*, *6*(12), e2347692. https://doi.org/10.1001/jamanetworkopen.2023.47692

[8] Zhang, T., Xu, W., Luo, B., & Wang, G. (2025). Depth-wise convolutions in vision transformers for efficient training on small datasets. *Neurocomputing*, *617*, 128998. https://doi.org/10.1016/j.neucom.2024.128998

[9] Pan, Y., Liu, J., Cai, Y., Yang, X., Zhang, Z., Long, H., & Tan, Z. (2023). Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases. *Frontiers in Physiology*, *14*, 1126780. https://doi.org/10.3389/fphys.2023.1126780

[10] Tummala, S., Thadikemalla, V. S. G., Kadry, S., Sharaf, M., & Rauf, H. T. (2023). EfficientNetV2 based ensemble model for quality estimation of diabetic retinopathy images from DeepDRiD. *Diagnostics*, *13*(4), 622. https://doi.org/10.3390/diagnostics13040622

[11] Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., & Hamamoto, R. (2024). Comparison of vision transformers and convolutional neural networks in medical image analysis: A systematic review. *Journal of Medical Systems*, *48*(1), 84. https://doi.org/10.1007/s10916-024-02105-8

[12] Razzak, R., Taki, S. T. O., Mim, M. R., Patwary, M. S. H., & Pavel, M. I. (2023). Vulgar comments classification: Comparison between CNN, XGBoost and SVM. *Research Square*. https://doi.org/10.21203/rs.3.rs-3185443/v1

[13] Alam, M. S., Rashid, M. M., Faizabadi, A. R., Mohd Zaki, H. F., Alam, T. E., Ali, M. S., & Ahsan, M. M. (2023). Efficient deep learning-based data-centric approach for autism spectrum disorder diagnosis from facial images using explainable AI. *Technologies*, *11*(5), 115. https://doi.org/10.3390/technologies11050115

[14] Lai, M., Lee, J., Chiu, S., Charm, J., So, W. Y., Yuen, F. P., & Zee, B. (2020). A machine learning approach for retinal images analysis as an objective screening method for children with autism spectrum disorder. *EClinicalMedicine*, *28*, 100588. https://doi.org/10.1016/j.eclinm.2020.100588

[15] Rasul, R. A., Saha, P., Bala, D., Karim, S. M. R. U., Abdullah, M. I., & Saha, B. (2024). An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder. *Healthcare Analytics*, *5*, 100293. https://doi.org/10.1016/j.health.2023.100293

[16] Simeoli, R., Rega, A., Cerasuolo, M., Nappo, R., & Marocco, D. (2024). Using machine learning for motion analysis to early detect autism spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders*. Advance online publication. https://doi.org/10.1007/s40489-024-00435-4

[17] Wei, Q., Cao, H., Shi, Y., Xu, X., & Li, T. (2023). Machine learning based on eye-tracking data to identify autism spectrum disorder: A systematic review and meta-analysis. *Journal of Biomedical Informatics*, *137*, 104254. https://doi.org/10.1016/j.jbi.2022.104254

[18] Bahathiq, R. A., Banjar, H., Bamaga, A. K., & Jarraya, S. K. (2022). Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging. *Frontiers in Neuroinformatics*, *16*, 949926. https://doi.org/10.3389/fninf.2022.949926

[19] John, A., & Singh, T. (2023). MOD-DHGN for autism segmentation. *Procedia Computer Science*, *218*, 621–630. https://doi.org/10.1016/j.procs.2023.01.044

[20] Li, M., Wang, Y., Gao, H., Xia, Z., Zeng, C., Huang, K., & Zhang, W. (2024). Exploring autism via the retina: Comparative insights in children with autism spectrum disorder and typical development. *Autism Research*, *17*(8), 1520–1533. https://doi.org/10.1002/aur.3204

[21] Subah, F. Z., Deb, K., Dhar, P. K., & Koshiba, T. (2021). A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI. *Applied Sciences*, *11*(8), 3636. https://doi.org/10.3390/app11083636

[22] Nawaz, A., Ali, T., Mustafa, G., Babar, M., & Qureshi, B. (2023), Multi-class retinal diseases detection using deep CNN with minimal memory consumption. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3281859

[23] Badar, M., Haris, M., & Fatima, A. (2020). Application of deep learning for retinal image analysis: A review. *Computer Science Review*, *35*, 100203. https://doi.org/10.1016/j.cosrev.2019.100203

[24] Lee, L., & Ingram, K. (2024). Retinal image analysis for simultaneous classification and severity grading of attention-deficit hyperactivity disorder and autism spectrum disorder using deep learning. *Journal of Student Research*, *13*(2), 1–10. https://doi.org/10.47611/jsrhs.v13i2.6482

[25] Amrutha, C. V., Jyotsna, C., & Amudha, J. (2020). Deep learning approach for suspicious activity detection from surveillance video. In *2020 2nd International Conference on Innovative Mechanisms for Industry Applications*, 335–339. https://doi.org/10.1109/ICIMIA48430.2020.9074920

[26] Rajatha & Ashoka, D. V. (2025). EffiViT: Hybrid CNN–transformer for retinal imaging. *Computers in Biology and Medicine*, *191*, 110164. https://doi.org/10.1016/j.compbiomed.2025.110164

[27] Dutta, P., Sathi, K. A., Hossain, M. A., & Dewan, M. A. A. (2023). Conv-ViT: A convolution and vision transformer-based hybrid feature extraction method for retinal disease detection. *Journal of Imaging*, *9*(7), 140. https://doi.org/10.3390/jimaging9070140

[28] Jesu Mariyan Beno Ranjana, J., & Muthukkumar, R. (2025). ADET MODEL: Real time autism detection via eye tracking model using retinal scan images. *Technology and Health Care*, *33*(4), 1661–1678. https://doi.org/10.1177/09287329241301678

[29] Minissi, M. E., Altozano, A., Marín-Morales, J., Chicchi Giglioli, I. A., Mantovani, F., & Alcañiz, M. (2024). Biosignal comparison for autism assessment using machine learning models and virtual reality. *Computers in Biology and Medicine*, *171*, 108194. https://doi.org/10.1016/j.compbiomed.2024.108194

[30] John, A., & Santhanalakshmi, S. (2025). Explainable AI solutions for emotion understanding in autism spectrum disorder. In *ICT Systems and Sustainability: Proceedings of ICT4SD 2024*, *6*, 255–268. https://doi.org/10.1007/978-981-97-9523-9_22

[31] Kulkarni, N., & Amudha, J. (2018). Eye gaze-based optic disc detection system. *Journal of Intelligent & Fuzzy Systems*, *34*(3), 1713–1722. https://doi.org/10.3233/JIFS-169464

[32] Fernandez-Lanvin, D., Gonzalez-Rodriguez, M., de-Andres, J., & Camero, R. (2023). Towards an automatic early screening system for autism spectrum disorder in toddlers based on eye-tracking. *Multimedia Tools and Applications*, *83*(18), 55319–55350. https://doi.org/10.1007/s11042-023-17694-8

[33] Nag, A., Haber, N., Voss, C., Tamura, S., Daniels, J., Ma, J., & Wall, D. P. (2020). Toward continuous social phenotyping: Analyzing gaze patterns in an emotion recognition task for children with autism through wearable smart glasses. *Journal of Medical Internet Research*, *22*(4), e13810. https://doi.org/10.2196/13810

[34] Nguyen, T. D., Le, D.-T., Bum, J., Kim, S., Song, S. J., & Choo, H. (2024). Retinal disease diagnosis using deep learning on ultra-wide-field fundus images. *Diagnostics*, *14*(1), 105. https://doi.org/10.3390/diagnostics14010105

[35] Leung, F. Y. N., Stojanovik, V., Micai, M., Jiang, C., & Liu, F. (2023). Emotion recognition in autism spectrum disorder across age groups: A cross-sectional investigation of various visual and auditory communicative domains. *Autism Research*, *16*(4), 783–801. https://doi.org/10.1002/aur.2896

[36] Atlam, E.-S., Masud, M., Rokaya, M., Meshref, H., Gad, I., & Almars, A. M. (2024). EASDM: Explainable autism spectrum disorder model based on deep learning. *Journal of Disability Research*, *3*(1), e20240003. https://doi.org/10.57197/JDR-2024-0003

[37] Tamuly, S., Jyotsna, C., & Amudha, J. (2020). Deep learning model for image classification. In *Computational Vision and Bio-Inspired Computing*, 312–320. https://doi.org/10.1007/978-3-030-37218-7_36

[38] Huynh, N., & Deshpande, G. (2024). A review of the applications of generative adversarial networks to structural and functional MRI based diagnostic classification of brain disorders. *Frontiers in Neuroscience*, *18*, 1333712. https://doi.org/10.3389/fnins.2024.1333712

[39] Yang, Y., Zhang, B., Guo, D., Du, H., Xiong, Z., Niyato, D., & Han, Z. (2024). Generative AI for secure and privacy-preserving mobile crowdsensing. *IEEE Wireless Communications*, *31*(6), 29–38. https://doi.org/10.1109/MWC.004.2400017

[40] Yang, Y., Wang, W., Yin, Z., Xu, R., Zhou, X., Kumar, N., & Gadekallu, T. R. (2022). Mixed game-based AoI optimization for combating COVID-19 with AI bots. *IEEE Journal on Selected Areas in Communications*, *40*(11), 3122–3138. https://doi.org/10.1109/JSAC.2022.3215508

[41] Linchundan. (2019). *1000 fundus images with 39 categories* [Data set]. Kaggle. https://www.kaggle.com/datasets/linchundan/fundusimage1000

[42] Alnowami, M., Taha, E., Alsebaeai, S., Muhammad Anwar, S., & Alhawsawi, A. (2022). MR image normalization dilemma and the accuracy of brain tumor classification model. *Journal of Radiation Research and Applied Sciences*, *15*(3), 33–39. https://doi.org/10.1016/j.jrras.2022.05.014

[43] Guo, J., Ma, J., García-Fernández, Á. F., Zhang, Y., & Liang, H. (2023). A survey on image enhancement for low-light images. *Heliyon*, *9*(4), e14558. https://doi.org/10.1016/j.heliyon.2023.e14558

[44] Ashraf, A., Zhao, Q., Bangyal, W. H., Raza, M., & Iqbal, M. (2025). Female autism categorization using CNN based NeuroNet57 and ant colony optimization. *Computers in Biology and Medicine*, *189*, 109926. https://doi.org/10.1016/j.compbiomed.2025.109926

[45] Lartseva, A., Dijkstra, T., Kan, C. C., & Buitelaar, J. K. (2014). Processing of emotion words by patients with autism spectrum disorders: Evidence from reaction times and EEG. *Journal of Autism and Developmental Disorders*, *44*(11), 2882–2894. https://doi.org/10.1007/s10803-014-2149-z

[46] Zhang, G., Zhang, R., Zhou, G., & Jia, X. (2018). Hierarchical spatial features learning with deep CNNs for very high-resolution remote sensing image classification. *International Journal of Remote Sensing*, *39*(18), 5978–5996. https://doi.org/10.1080/01431161.2018.1506593

[47] Khan, K., & Katarya, R. (2025). MCBERT: A multi-modal framework for the diagnosis of autism spectrum disorder. *Biological Psychology*, *194*, 108976. https://doi.org/10.1016/j.biopsycho.2024.108976

[48] Wang, Y., Zhu, Z., Wang, Y., Li, M., Ma, X., Huang, K., …, & Ke, X. (2024). Relationship between autism spectrum disorder

and peripapillary intraretinal layer thickness: A pediatric retrospective cross-sectional study. *Quantitative Imaging in Medicine and Surgery*, *14*(12), 8347–8360. https://doi.org/10.21037/qims-24-753

[49] Vidivelli, S., Padmakumari, P., & Shanthi, P. (2025). Multimodal autism detection: Deep hybrid model with improved feature level fusion. *Computer Methods and Programs in Biomedicine*, *260*, 108492. https://doi.org/10.1016/j.cmpb.2024.108492

[50] Alam, M. Z., Rahman, M. S., & Rahman, M. S. (2019). A random forest based predictor for medical data classification using feature ranking. *Informatics in Medicine Unlocked*, *15*, 100180. https://doi.org/10.1016/j.imu.2019.100180

[51] Farooq, S., He, H., Guo, D., Feng, Y., Hang, J., Kong, D., & He, S. (2025). Advanced autism detection and visualization through XGBoost algorithm for fNIRS hemo-dynamic signals. *Expert Systems with Applications*, *275*, 127013. https://doi.org/10.1016/j.eswa.2025.127013

[52] Haarika, R., & John, A. (2024). Enhancing customer retention through feature selection and XGBoost classification for churn prediction. In *2024 International Conference on Smart Systems for Applications in Electrical Sciences*, 1–7. https://doi.org/10.1109/ICSSES62373.2024.10561442

[53] Khudhur, D. D., & Khudhur, S. D. (2023). The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups. *Measurement: Sensors*, *27*, 100774. https://doi.org/10.1016/j.measen.2023.100774

[54] Tokala, S., Hajarathaiah, K., Gunda, S. R. P., Botla, S., Nalluri, L., Nagamanohar, P., & Enduri, M. K. (2023). Liver disease prediction and classification using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, *14*(2), 871–878. https://doi.org/10.14569/IJACSA.2023.0140299

[55] Zhang, Y., Hong, D., McClement, D., Oladosu, O., Pridham, G., & Slaney, G. (2021). Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *Journal of Neuroscience Methods*, *353*, 109098. https://doi.org/10.1016/j.jneumeth.2021.109098

[56] Wiratsin, I.-O., & Narupiyakul, L. (2021). Feature selection technique for autism spectrum disorder. In *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence*, 53–56. https://doi.org/10.1145/3448218.3448241

[57] Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 52. https://doi.org/10.1186/s40537-020-00327-4

[58] Anaya-Sánchez, H., Altamirano-Robles, L., Díaz-Hernández, R., & Zapotecas-Martínez, S. (2024). WGAN-GP for synthetic retinal image generation: Enhancing sensor-based medical imaging for classification models. *Sensors*, *25*(1), 167. https://doi.org/10.3390/s25010167

How to Cite: John, A., & Santhanalakshmi, S. (2025). AI-Driven Diagnosis of Autism Spectrum Disorder Using Retinal Fundus Imaging: A Comparison of Traditional and Deep Learning Feature Extraction Methods. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE52026045