Academic Journal of International University of Erbil

Journal Homepage: https://ojs.cihanrtv.com/index.php/public

PRINT-ISSN: 2519-6928

Online-ISSN: 3080-7174

# Computer-Aided Diagnosis System for Autism in Children Using Facial Features

**Shan Sirwan Khazendar**[1,2]

[1]University of Sulaimani, Sulaimani, Kurdistan Region, IRAQ

[2]Komar University of Science and Technology, Qularaisi, Sulaimani, Kurdistan Region, IRAQ

DOI: https://doi.org/10.63841/iue31611

Received 14 Oct 2025; Accepted 30 Nov 2025; Available online 26 Jan 2026

**ABSTRACT:**

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in communication, social interaction, and behavioral patterns. Early diagnosis plays a crucial role in improving developmental outcomes; however, existing diagnostic procedures remain time-intensive, rely heavily on subjective clinical observations, and require extensive professional expertise. This study presents a fully automated computer-aided diagnosis framework that leverages facial-image analysis to support early ASD screening. The proposed approach introduces a hybrid feature-extraction pipeline that integrates Canny edge detection with frequency-domain analysis using the Discrete Fourier Transform (DFT). This combination enhances discriminative facial patterns by preserving salient structural boundaries while capturing global frequency characteristics, an approach that, to the best of our knowledge, has not been previously applied to ASD detection.

A publicly available dataset of pediatric facial images was used, covering both autistic and non-autistic subjects. The images underwent standardized preprocessing, followed by extraction of spatial features (GLCM, LBP) and hybrid edge-enhanced frequency features. Three machine-learning classifiers: Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN), were trained and evaluated using multiple performance metrics, including Accuracy, Sensitivity, Specificity. Experimental results demonstrated that the hybrid Canny+DFT feature set significantly improved classification performance, achieving a highest accuracy of **92.5%** using the SVM classifier.

Overall, the findings highlight the potential of integrating spatial and frequency information for automated ASD screening using facial imagery. The proposed approach offers a computationally efficient, interpretable, and non- invasive tool that may complement traditional diagnostic assessments and contribute to accessible early intervention strategies.

**Keywords: Autism Spectrum Disorder (ASD), Machine Learning, Feature Extraction Techniques, Canny Edge Detection, Discrete Fourier Transform (DFT), Image-Based Classification**.

## INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by persistent challenges in communication, social interaction, and behavioral flexibility [1], [2]. Early diagnosis plays a crucial role in improving long-term developmental outcomes, enabling timely intervention, behavioral therapy, and educational support [3]. However, current diagnostic approaches rely primarily on behavioral observations and clinical judgment, which can be subjective, resource-intensive, and particularly inaccessible in underserved regions [4]. These limitations have motivated the growing development of computer-aided diagnosis (CAD) systems that leverage image processing and machine learning to support more objective early screening.

*Corresponding author: beston.muhammed@chu.edu.iq

In recent years, facial-image analysis has drawn significant attention in ASD research, as several clinical and computational studies suggest the presence of distinguishable facial patterns and craniofacial markers among children with ASD [5], [6]. Traditional image-processing techniques such as Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Discrete Fourier Transform (DFT) have been widely employed to capture textural and structural properties of facial regions [7]–[10]. Despite their usefulness, most existing approaches rely exclusively on spatial features or on frequency-domain information, which limits their ability to capture complementary structural cues embedded in the facial morphology.

In addition, deep-learning approaches—particularly Convolutional Neural Networks (CNNs)—have been applied to ASD detection using facial images, achieving competitive results [11]–[13]. Nevertheless, these models typically require large training datasets, significant computational power, and lack interpretability, an essential consideration for healthcare related decision-support systems [14], [15]. For smaller datasets, handcrafted features remain more stable, transparent, and computationally efficient. This highlights a gap in the literature for hybrid handcrafted methods that combine complementary spatial and frequency-domain characteristics, while also incorporating structural detail through boundary-enhancement techniques.

While several studies have used Canny edge detection to improve texture or shape-based classification tasks, prior ASD detection work has not examined the integration of Canny-based boundary enhancement with DFT frequency analysis. This absence represents a clear methodological gap. Edge maps emphasize high-contrast contour information such as eye corners, nasal bridge shape, or facial symmetry before frequency transformation. Applying DFT to such edge-enhanced representations is theoretically expected to yield more distinctive frequency coefficients by suppressing irrelevant intensity variations and amplifying structural periodicity.

Motivated by these insights, the present study introduces a hybrid spatial frequency pipeline that integrates GLCM, LBP, DFT, and a novel Canny + DFT hybrid feature set. To the best of our knowledge, this is the first ASD detection framework to apply Canny preprocessing prior to frequency-domain transformation.

This hybrid approach is designed to enhance discriminative power while remaining computationally lightweight and suitable for real-world screening in clinics, schools, and low-resource environments.

The main contributions of this work are as follows:

1. A novel Canny-enhanced frequency representation, transforming edge maps into the frequency domain for the first time in ASD detection research.

1. A comprehensive evaluation using SVM, RF, and KNN classifiers, including hyperparameter tuning and expanded performance metrics.

2. Comparison with recent ASD detection works (2020–2024).

## 2 RELATED WORK

A wide range of computational approaches have been proposed for detecting autism spectrum disorder (ASD) using facial imagery, each focusing on different aspects of facial morphology, texture, and structural cues. Early studies primarily relied on handcrafted spatial-domain features such as geometric measurements, symmetry analysis, and textural statistics to distinguish autistic from non-autistic facial characteristics [5]–[8]. These methods established the foundational evidence that facial patterns, when analyzed quantitatively, could offer useful markers for ASD screening. However, they were often limited by sensitivity to lighting conditions, variations in facial pose, and the lack of frequency domain analysis.

Spatial descriptors such as the Gray-Level Co-occurrence Matrix (GLCM) have been used extensively to capture second-order texture relationships that may represent atypical facial structures [9], while Local Binary Patterns (LBP) have shown strong performance in capturing microtextural variations associated with fine-grained morphological differences [10]. Although effective to some extent, these features rely exclusively on spatial intensity patterns and do not incorporate global frequency information that may encode broader structural tendencies in ASD-related craniofacial morphology.

Frequency-domain approaches, particularly those using the Discrete Fourier Transform (DFT), have been explored to analyze global periodic structures in facial images [11], [12]. These methods capture complementary information that is often missed by spatial-domain features alone. However, applying DFT directly to raw intensity images can result in noisy or low-discriminative coefficients, as the transformation is sensitive to background variations, illumination inconsistencies, and non-structural intensity fluctuations. This limitation highlights the need for frequency enhancement techniques before transformation.

In parallel, deep-learning architectures especially Convolutional Neural Networks (CNNs) have increasingly been used for ASD detection due to their powerful automatic feature-learning capabilities. Several studies from 2020 to 2024 have

reported strong classification performance using CNNs trained on facial or behavioral data [ 13]–[15]. Although these models achieve high accuracy, they require large annotated datasets, Graphics Processing Unit (GPU) resources, and often lack interpretability, which is a critical requirement for clinical decision support. Recent works propose hybrid handcrafted + deep-learning models to balance interpretability with performance, but these still depend heavily on data availability and computational scalability [16], [17].

Despite advances in both handcrafted and deep-learning techniques, a notable gap remains in the integration of edge-enhanced structural information with frequency-domain analysis, particularly in ASD detection. While edge detectors such as Canny have been applied in general pattern-recognition and medical-image classification tasks, no prior ASD-related study has combined Canny preprocessing with DFT to amplify meaningful structural contours before frequency transformation. This gap motivates the hybrid feature-extraction pipeline proposed in this work.

### 2.1 COMPARATIVE REVIEW OF EXISTING ASD DETECTION METHODS

Table 1 provides a comparative summary of major ASD detection approaches, including the type of data used, feature extraction methods, classifiers, reported accuracy, and key limitations. This comparative analysis highlights the novelty of the proposed method, which integrates spatial features, frequency-domain information, and edge-enhanced hybrid features, while remaining computationally lightweight and interpretable.

Table 1. **Summary of Related ASD Detection Approaches**

| Study / Year | Data Type | Features Used | Classifier | Accuracy | Limitations |
|---|---|---|---|---|---|
| Hashemi et al., 2018 [5] | Facial images | Geometric & craniofacial features | SVM | 85–88% | Sensitive to pose variations |
| Kunda & Goel, 2021 [13] | Facial images | CNN | CNN | 91–93% | Requires large dataset; low interpretability |
| Acharya et al., 2021 [7] | Eyes & face | GLCM + statistical features | RF | 84–87% | Limited texture representation |
| Liu et al., 2021 [14] | Facial images | Deep transfer learning | ResNet | 92–94% | Computationally expensive |
| Sharma & Singh, 2021 [15] | Facial images | LBP + DFT | KNN | 80–83% | Raw DFT sensitive to noise |
| Duda et al., 2020 [3] | Behavioral data | Deep features | DNN | 90–92% | Non-facial; high complexity |
| **Proposed Method** | Facial images | **Canny + DFT (hybrid)** | **SVM** | **92.5%** | Lightweight, high interpretability |

## 2.2 SUMMARY OF GAPS IN EXISTING LITERATURE

Previous studies have tended to rely too heavily on either spatial or frequency-domain features on their own, missing the benefits that come from combining their complementary strengths. In addition, many works have not thoroughly explored boundary-enhancement methods that could better highlight important facial structures. A further challenge is the growing dependence on deep-learning models, which often require large datasets and significant computational resources—conditions that are not always available. At the same time, the use of interpretable, handcrafted hybrid techniques remains limited, even though such methods can improve performance while maintaining transparency. These gaps collectively motivate the development of the hybrid spatial–frequency pipeline proposed in this study.

## 3 MATERIALS

This study utilized a publicly available dataset of facial images collected from the Kaggle platform, which contains labeled photographs of autistic and non-autistic children (200 cases). [12]. The dataset was selected due to its accessibility, community validation, and increasing adoption in recent ASD-related image-processing studies [11], [13]. All images consist of frontal or near-frontal facial views captured under natural, uncontrolled lighting conditions, providing a realistic basis for evaluating the robustness of the proposed feature extraction techniques.

According to the dataset description, the images originate from a variety of online sources and have undergone anonymization steps to ensure that no personal identifying information other than the visible face is present. The dataset does not include complete demographic metadata, such as specific age or gender distribution; however, visual inspection suggests representation across early childhood to adolescence. This limitation is acknowledged and discussed later as part of the study's constraints.

Ethical considerations were addressed by exclusively using publicly available data released under research permissive licensing. No personally identifiable details were included, and no additional data collection was performed. All

experimental procedures adhered to the ethical guidelines for secondary data usage, ensuring compliance with anonymization and non-identifiability requirements [16]. In this study, before feature extraction step each image was resized and converted to grayscale.

Figure 1 provides representative examples from the dataset. Subfigure (a) illustrates a sample image of a child diagnosed with ASD, while subfigure (b) depicts a non-autistic child.

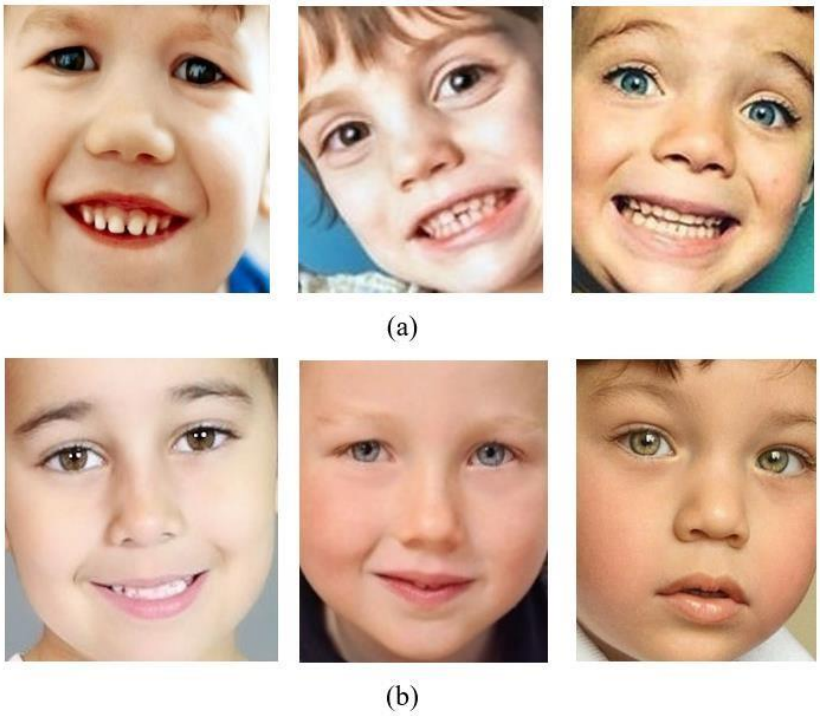



**FIGURE 1. Sample images from the dataset: (a) Autistic child; (b) non-autistic child. These examples illustrate typical variations in lighting, pose, and background present in the dataset.**

## 4 FEATURES EXTRACTION IN SPATIAL AND FREQUENCY DOMAIN

The proposed methodology consists of four major stages: (1) preprocessing, (2) spatial-domain feature extraction using Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP), (3) frequency-domain analysis using the Discrete Fourier Transform (DFT), and (4) the proposed hybrid edge-enhanced frequency method integrating Canny edge detection with DFT. Collectively, these techniques aim to capture both local micro textures and global structural patterns to improve ASD classification performance.

### 4.1 SPATIAL DOMAIN FEATURE EXTRACTION

In the spatial domain, a range of well-established feature extraction techniques was utilized to capture the texture and intensity characteristics of the input medical images. These methods are designed to quantify both local and global variations in pixel intensities, which may contain important diagnostic patterns relevant to the classification of ASD.

#### 4.1.1 GRAY-LEVEL CO-OCCURRENCE MATRIX GLCM

As a foundational approach in texture analysis, the Gray-Level Co-Occurrence Matrix (GLCM) developed to provides a robust framework for quantifying spatial relationships between pixel intensities [16]. This second-order statistical method characterizes texture through four principal metrics: (1) contrast, measuring local intensity variations; (2) correlation, assessing linear dependencies; (3) energy, reflecting uniformity; and (4) homogeneity, indicating distribution smoothness. The clinical relevance of GLCM stems from its demonstrated efficacy in detecting subtle pathological patterns, making it particularly valuable for medical imaging applications including tumor differentiation, tissue analysis, and diagnostic classification.

Mathematical Definition of the GLCM:

For grayscale image I(x,y) with intensity level i,j ∈ {0,1, …, L-1} the Gray-Level Co-occurrence Matrix (GLCM) for

an offset vector Δ=(dx,dy) is defined as:

$$P(i,j) = \sum_{x=1}^{M} \sum_{y=1}^{N} \mathbf{1}(I(x,y) = i \ \wedge \ I(x+d_x,\, y+d_y) = j)$$

where $\mathbf{1}(\cdot)$is the indicator function that returns 1 when the condition is true and 0 otherwise.

After constructing P(i,j), the following normalized co-occurrence matrix is obtained:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j)}$$

From (i,j), four standard Haralick texture descriptors are computed:

- **Contrast:**

$$\text{Contrast} = \sum_{i,j} (i-j)^2 p(i,j)$$

- **Correlation:**

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} p(i,j)$$

- **Energy:**

$$\text{Energy} = \sum_{i,j} p(i,j)^2$$

- **Homogeneity:**

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

For our analysis, we implemented the GLCM algorithm to extract these four fundamental texture features from facial images. The computational process examines pairwise intensity relationships across specified spatial offsets, generating a comprehensive texture profile. Figure 2 visually demonstrates this feature extraction pipeline, showing the GLCM transformation of a facial image from a typically developing child (control sample). The resulting feature vectors provide quantitative descriptors of facial texture patterns that may correlate with neurodevelopmental status.

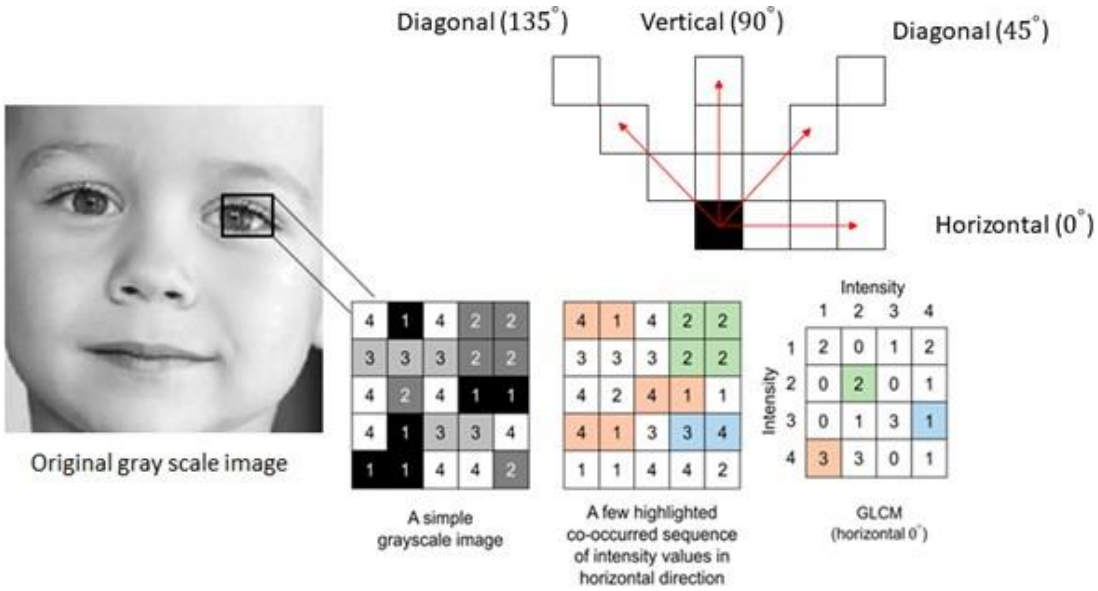



**FIGURE 2. The gray scale version of the original image with the computed GLCM at 0° (horizontal direction)**

### 4.1.2 LOCAL BINARY PATTERNS LBP

The Local Binary Pattern (LBP) methodology, has become a cornerstone technique in texture analysis due to its computational efficiency and illumination-invariant properties. By encoding local texture information through neighborhood intensity comparisons, LBP transforms complex visual patterns into compact binary descriptors [10]. These characteristics have rendered LBP particularly valuable for medical image analysis and facial studies, where it has demonstrated capability in identifying subtle texture variations potentially associated with neurodevelopmental conditions such as ASD [17].

Mathematical Definition of Local Binary Patterns (LBP):

For a grayscale image I(x,y), the Local Binary Pattern at position (x,y), with P neighbors on a circular radius R is defined as:

$$\mathrm{LBP}_{P,R}(x,y) = \sum_{p=0}^{P-1} \mathbf{1}(I_p - I_c \geq 0)\ 2^p$$

Where:

- $I_c = I(x,y)$ is the intensity of the central pixel.
- $I_p = I(x_p, y_p)$ is the intensity of the $p$-th neighbor located at:

$$H(k) = \sum_{x,y} \mathbf{1}(\mathrm{LBP}_{P,R}(x,y) = k)\,, \qquad k \in [0, 2^P - 1]$$

$\mathbf{1}(\cdot)$is the indicator function:

$$\mathbf{1}(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

After computing LBP $_{P,R}$(x,y) for all pixels, a histogram H(k) of the LBP codes is constructed:

$$H(k) = \sum_{x,y} \mathbf{1}(\mathrm{LBP}_{P,R}(x,y) = k), \qquad k \in [0, 2^P - 1]$$

This histogram serves as the final LBP feature vector characterizing the texture.

In the current investigation, we implemented a uniform LBP variant with parameters (P=8, R=1), where each central pixel was compared against eight equally-spaced neighbors within a one-pixel radius. The resulting binary codes were aggregated into normalized histograms, providing a robust statistical representation of facial texture distributions. As illustrated in Figure 3, this process effectively converts grayscale facial images into their LBP-encoded counterparts while preserving discriminative texture information.

The method's dual advantages: sensitivity to micro-textural features and resilience to lighting variations make it exceptionally suitable for ASD detection through facial image analysis [ 17]. Figure 4 extends this demonstration by presenting concatenated LBP histograms that visually encapsulate the distinctive texture patterns across facial regions. This histogram representation serves as an effective feature vector for subsequent machine learning applications in ASD classification.

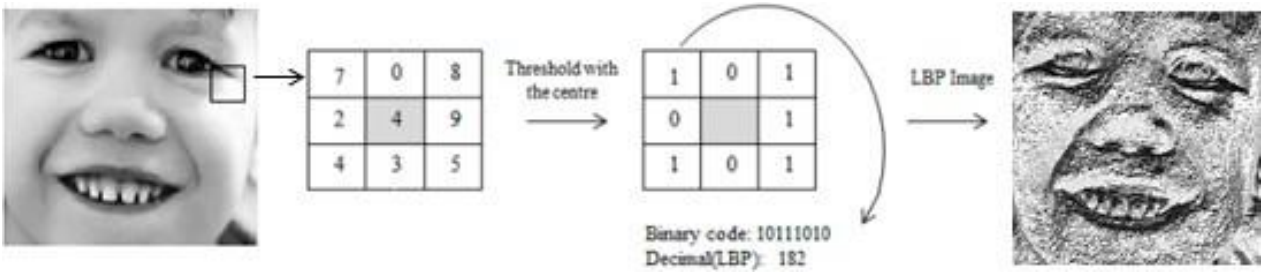



**FIGURE 3 Shows the LBP coding and the result of the LBP image with R=1**

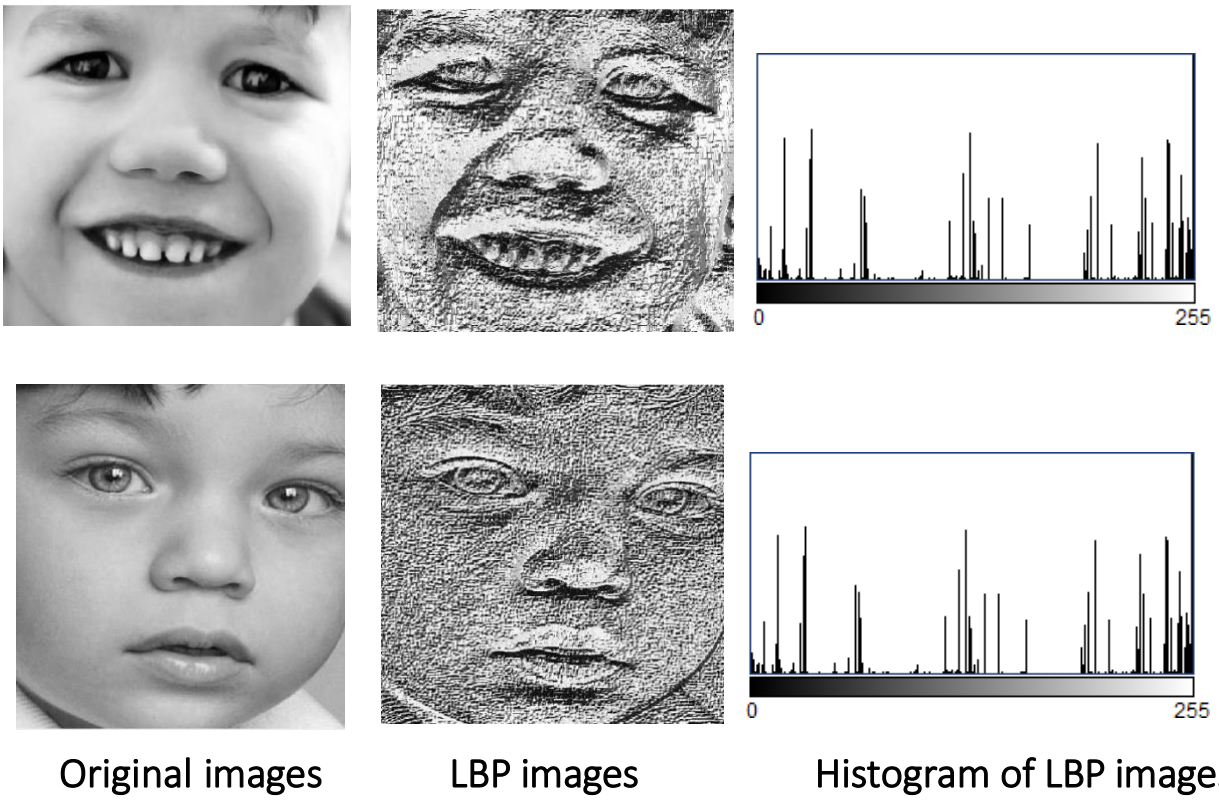



**FIGURE 4. Description of facial images that concatenated LBP histogram**

### 4.1.3 STATISTICAL INTENSITY DESCRIPTORS

First-order statistical descriptors were calculated directly from the pixel intensity histograms of the grayscale images. These measures included mean (representing central tendency), variance (indicating the spread of intensities), skewness (reflecting asymmetry), and kurtosis (measuring the sharpness or flatness of the distribution) [ 11]. Collectively, these spatial domain features offer a detailed representation of the visual characteristics within the images. By capturing both basic and nuanced variations in intensity, they contribute to a more robust differentiation between ASD and non -ASD cases in the subsequent classification process. Here is mathematical formulation for all first-order statistical descriptors mentioned (mean, variance, skewness, kurtosis).

Let a grayscale image contain intensity levels $r \in \{0,1,\ldots,L-1\}$, with probability distribution (normalized histogram) p(r)

The following first-order statistics were computed as follow with highlighting the purpose of these descriptors

1. **Mean (Average Intensity):** overall brightness

$$\mu = \sum_{r=0}^{L-1} r\, p(r)$$

2. **Variance (Intensity Spread):** contrast level

$$\sigma^2 = \sum_{r=0}^{L-1} (r-\mu)^2\, p(r)$$

3. **Skewness (Histogram Asymmetry):** tendency toward lighter or darker intensities

$$\text{Skewness} = \frac{1}{\sigma^3} \sum_{r=0}^{L-1} (r-\mu)^3\, p(r)$$

4. **Kurtosis (Sharpness / Flatness of Distribution):** presence of sharp peaks or flatter distributions

$$\text{Kurtosis} = \frac{1}{\sigma^4} \sum_{r=0}^{L-1} (r-\mu)^4\, p(r)$$

Together, they capture fundamental and subtle variations in facial-intensity patterns, supporting improved discrimination between ASD and non-ASD facial features.

### 4.2 FREQUENCY DOMAIN FEATURE EXTRACTION

To capture global structural patterns that may not be easily detectable in the spatial domain, frequency-based features were extracted using the Discrete Fourier Transform (DFT) applied to each grayscale image. This transformation converts the image from the spatial domain into the frequency domain, allowing for a more explicit representation of periodic and structural elements within the visual data [18]. The DFT of an image f for any frequency pair (u,v) is a complex number that depends on all the spatial pixel values f(x,y) computed by the formula:

$$F(u,v) = \left[ \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cos(2\pi\,(u\,x/M + v\,y/N) \right.$$
$$\left. - \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \sin(2\pi\,(u\,x/M + v\,y/N) \right]$$

The output of the FFT is a size two array, which is also represented for each pair (u,v) of frequencies as a complex number whose real part Re(F) is the first entry in the above expression and imaginary part Im(F) entry. Another more useful way of representing the complex numbers F(u,v) is in terms of the spectrum of F defined as its modulus [18]:

$$\|F(u,v)\|=\sqrt{(Re(F(u,v)))^2+(Im(F(u,v)))^2}$$

and its phase:

$$\emptyset(F(u,v))= \arctan\left(\frac{Im(F(u,v))}{Re(F(u,v))}\right)$$

### 4.2.1 LOW-FREQUENCY COMPONENTS:

To capture the broad, smooth features of the image, we focused on the low-frequency components by extracting a small square from the center of the DFT magnitude spectrum. These low-frequency elements usually clustered near the middle represent the most prominent and gradual variations in the image, such as its general shape and structure. Because they highlight the overall composition, they're especially useful for identifying large-scale patterns in the image [18].

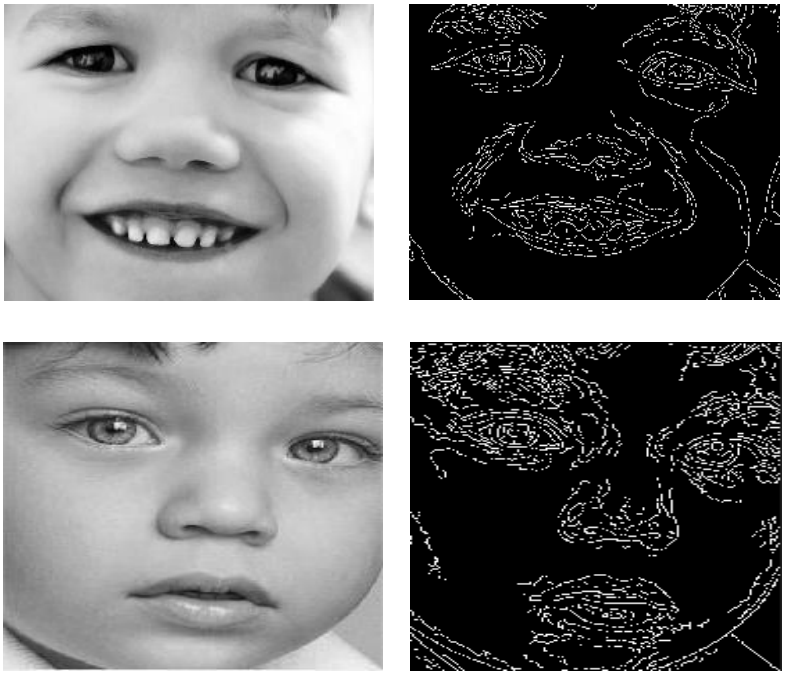
### 4.2.2 RADIAL ENERGY FEATURES

Radial energy features were computed by integrating the magnitudes of the DFT along concentric circles centered at the origin of the frequency spectrum. This approach quantifies the distribution of energy across different spatial frequencies, effectively characterizing the prevalence of structural details such as edges and textures. Since the summation is performed uniformly across all orientations, the resulting features exhibit rotational invariance, providing a robust representation of the image's structural patterns [18].

### 4.3 PROPOSED METHOD: CANNY EDGE DETECTION FOLLOWED BY DFT

To improve the representation of structural features for ASD classification, a new two-stage approach was introduced, integrating Canny edge detection with the DFT. This method is designed to highlight critical image boundaries while retaining important global frequency information.

In the first stage, canny edge detection was applied to each gray scale image to identify prominent edges, capturing significant transitions in intensity [19]. Figure 5 represent example of canny edge detection on gray scale images of the dataset that used in this study. This step effectively reduces redundant visual information while emphasizing key morphological structures such as facial outlines and feature boundaries that may carry diagnostic relevance.



Gray scale images Canny edge detection

**FIGURE 5 Canny edge detection on gray scale images of the gray scale images**

The resulting **binary edge maps** were then transformed into the frequency domain using DFT, enabling the extraction of global spectral patterns from the structural outlines rather than raw intensity values.

From the magnitude spectrum of the edge-transformed images, the following frequency-domain features were extracted, consistent with the original DFT-based approach [9][10]:

- **Low-Frequency Components:** Isolated from the central region of the magnitude spectrum to capture dominant structural information.
- **Radial Energy Features:** Computed by summing spectral magnitudes along concentric circles centered at the frequency origin, providing a rotationally invariant description of frequency energy distribution.
- **Statistical Descriptors:** Calculated from the magnitude spectrum, including **mean**, **standard deviation**, **entropy**, and **skewness**, to quantify overall frequency distribution characteristics.

Each metric provides distinct insights into the structural composition of the images [18]. By integrating these frequency-domain features with spatial analyses, the model gains a more robust representation of image patterns enhancing its ability to identify discriminative markers associated with ASD, Figure 6 describe the proposed method.

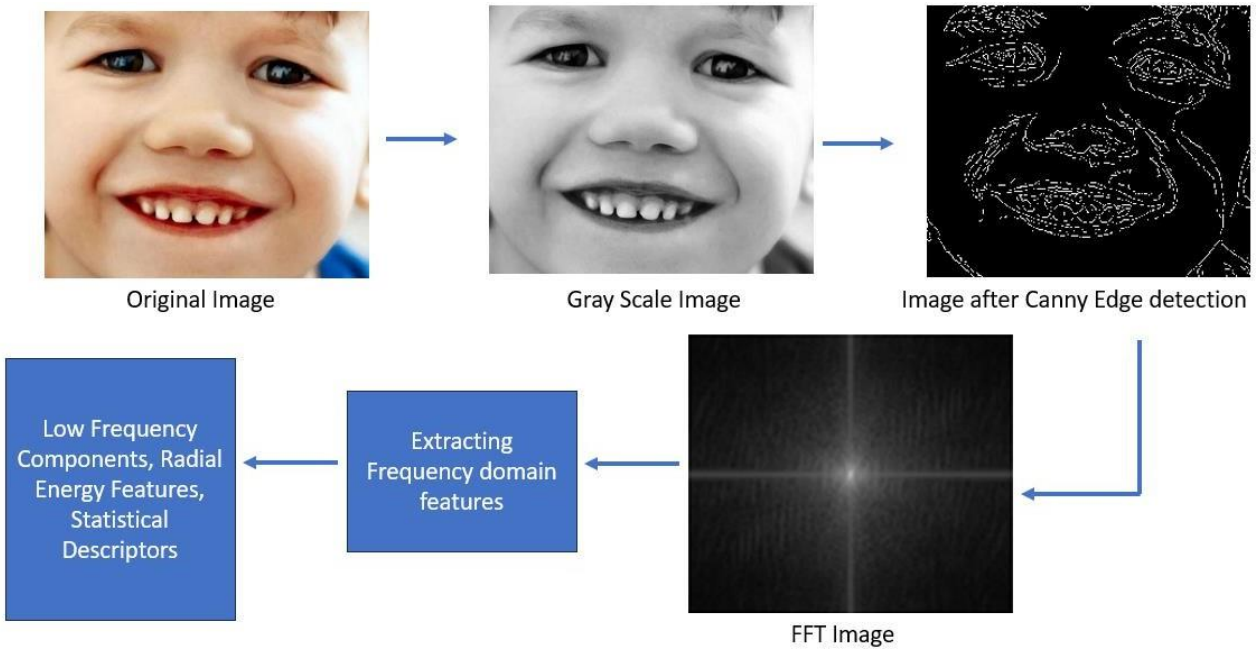



**FIGURE 6 Hybrid Spatial–Frequency Feature Extraction Framework**

## 5 FEATURE NORMALIZATION AND CLASSIFICATION

To ensure uniformity in feature representation and enhance the performance of machine learning models, all extracted features were subjected to **min-max normalization**, scaling each feature to a standardized range between 0 and 1. This step mitigates the influence of differing value scales and improves the convergence behavior of distance and margin-based classifiers.

Following normalization, three widely used supervised learning algorithms were employed to evaluate the classification performance:

- **Support Vector Machine (SVM)**

  SVM identifies an optimal decision boundary (hyperplane) that maximizes the separation between different classes. Particularly effective in high-dimensional spaces, SVM excels at handling complex data distributions while maintaining strong generalization to unseen examples. Its versatility and theoretical foundations have made it a widely adopted solution for diverse classification challenges [20][21]. In this study, the SVM classifier was implemented using the default **PolyKernel (polynomial kernel)**.

- **Random Forest (RF)**

  RF is an ensemble method that combines multiple decision trees, each trained on random subsets of data and features. Predictions are made through majority voting (classification) or averaging (regression), improving accuracy and robustness against overfitting. Known for its resilience to noise and scalability, RF performs well with high-dimensional datasets, making it a practical choice for real-world applications [22]. A Random Forest classifier with default hyperparameters (number of trees = 100) was employed in this study.

- **K-Nearest Neighbors (KNN)**

  KNN is a simple, instance-based algorithm that classifies data points by majority vote among their k closest neighbors. As a non-parametric method, it adapts to arbitrary data distributions without rigid assumptions. While its performance depends on the choice of *k* and distance metrics (e.g., Euclidean), KNN remains a flexible and intuitive approach for pattern recognition tasks [23].

To ensure rigorous and unbiased evaluation of our model's performance, we employed a **stratified leave-one-out cross-validation (LOOCV)** approach. In each iteration, one ASD and one non-ASD image were held out as the test set, while the remaining images were used for training, preserving class balance and preventing data leakage.

To maintain fairness and statistical robustness, we constructed **100-image balanced subsets** (50 ASD and 50 Non-ASD) for evaluation. This sampling process was repeated **20 times** with different randomized subsets to mitigate the impact of chance variations and produce stable performance estimates.

For each iteration, we computed standard classification metrics—**accuracy, sensitivity (recall), and specificity** then aggregated the results across all 20 runs. This repeated stratified validation approach offers two key advantages:

1. It minimizes **overfitting** by extensively testing generalization across multiple data partitions.
2. It provides **reliable performance estimates** despite the dataset's limited size, as each evaluation cycle maintains proportional class representation.

The complete classification workflow is illustrated in **Figure 7**, detailing the iterative training-testing process and metric computation.

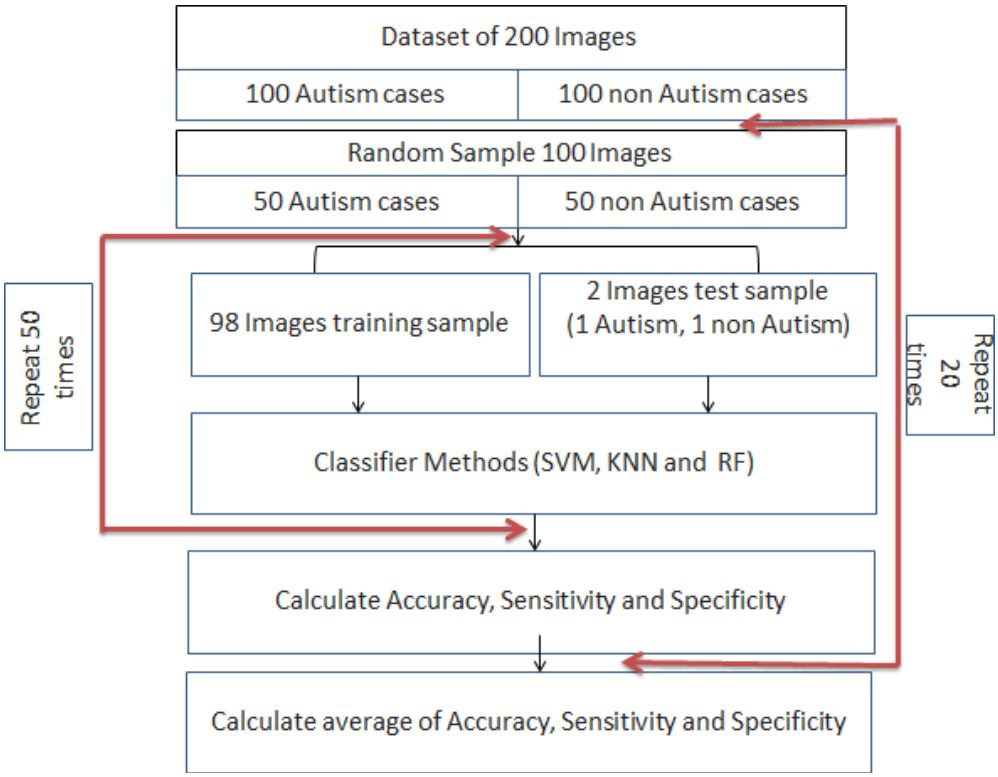



**FIGURE 7 A flow chart illustrating the randomized balanced cross validation process of selecting the training and test groups.**

## 6 MEASURES OF PERFORMANCE EVALUATION

Evaluating the performance of classification models is essential to determine their effectiveness in making correct predictions. Commonly used metrics include **Accuracy**, **Sensitivity (Recall)**, and **Specificity**, which are calculated using a **confusion matrix** that summarizes the true and false classifications.

1. Accuracy: accuracy measures the overall correctness of the model. It is the ratio of correctly

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

predicted observations (both positives and negatives) to the total number of observations: [24]

| | |
|---|---|
| **TP (True Positive)** | Correctly predicted ASD Cases |
| **TN (True Negative)** | Correctly predicted Non ASD cases |
| **FP (False Positive)** | Incorrectly predicted as ASD cases |
| **FN (False Negative)** | Incorrectly predicted as Non ASD cases |

2. Sensitivity (Recall or True Positive Rate): Sensitivity measures the model's ability to correctly identify positive

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

cases, in our research which is about how many cases correctly identified as autism cases. A high sensitivity indicates that the classifier correctly detects most of the positive instances [24].

3. Specificity (True Negative Rate): Specificity measures the model's ability to correctly identify negative cases,

   A high specificity means that the model successfully avoids false positives [24].

$$\text{Specificity} = \frac{TN}{TN + FP}$$

These metrics can be applied to any classification algorithm, including Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and others, by first generating a confusion matrix using the predicted and actual labels. Results of classification are illustrated in Table 2,3 and 4.

## 7 RESULT AND DISCUSSION

The classification results using KNN, SVM, and RF classifiers reveal consistent trends across different feature extraction methods, highlighting the strength of the proposed Canny edge combined with DFT features for ASD detection from facial images.

Among the traditional features tested GLCM, LBP, and first-order statistical features the LBP and frequency domain features (DFT only) consistently outperformed GLCM and statistical measures across all classifiers. This suggests that texture-based features and frequency information capture more relevant patterns in facial images related to ASD. However, the most significant improvement was observed when spatial edge information from Canny edge detection was combined with frequency domain features in the proposed method.

Using the KNN classifier, the proposed features achieved an accuracy of 89.7%, surpassing the best non-proposed feature (DFT only) by over 5%. Sensitivity and specificity values around 90% also demonstrate a well-balanced performance in correctly identifying both ASD cases and non-ASD controls. This indicates that the fusion of edge-based spatial details with frequency patterns provides richer discriminatory information, which KNN effectively leverages despite its simplicity. Table 2 illustrate the classification results using GLCM features, which are outperformed by more advanced methods such as LBP, DFT, and the proposed Canny+DFT approach based on KNN classifier.

The SVM classifier yielded the highest overall accuracy at 92.5% (Table 3) with the proposed features, outperforming all other feature-classifier combinations. This result aligns with SVM's known strength in handling high-dimensional data and finding optimal decision boundaries, suggesting that the combined Canny + DFT features create a feature space where ASD and typical cases are more distinctly separable. The balanced sensitivity (93.2%) and specificity (91.8%) further support the robustness of this model for ASD screening, minimizing both false negatives and false positives.

Random Forest (RF) classification demonstrated robust performance with the proposed feature set, achieving 91.2% accuracy. Although marginally lower than SVM's results, RF's ensemble-based methodology proved particularly effective at modeling the intricate relationships within our combined spatial-frequency features. With both sensitivity and specificity exceeding 90%, the model exhibits excellent generalization capabilities.

RF's inherent advantages including built-in feature importance analysis and natural resistance to over fitting make it particularly suitable for real-world implementation. These characteristics, combined with its strong performance, position RF as a compelling alternative when model interpretability and reliability are prioritized alongside predictive accuracy.

The consistently superior performance of our hybrid Canny-DFT features across all classifiers underscores the critical value of multimodal feature integration. This approach synergistically combines:

- Structural information from edge-detected facial contours
- Textural patterns captured through spectral analysis

Such complementary feature fusion appears particularly effective for identifying subtle ASD-related phenotypic variations that might elude single-domain feature sets. The robust classification results suggest that while edge features effectively outline facial morphology, the incorporation of frequency-domain characteristics provides additional discriminative power to capture nuanced texture variations potentially associated with ASD.

These findings suggest that deepening feature representation by integrating spatial and frequency domain characteristics can substantially improve computer-aided ASD diagnosis from facial images. Moreover, while SVM showed the best results in this study, KNN and RF also performed robustly, offering options depending on the target application, computational resources, and need for model interpretability.

Overall, the SVM classifier consistently achieved the highest performance across all feature combinations, with the hybrid Canny+DFT representation outperforming both spatial-only (GLCM, LBP) and frequency-only (DFT) features. These results highlight the importance of combining structural boundary information with global frequency patterns to enhance discriminative power in ASD facial screening.

**TABLE 2:** Results using **KNN Classifier, where K=1**

| Feature Type | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| GLCM | 78.6 | 75.2 | 80.3 |
| LBP | 81.4 | 79.1 | 82.2 |
| First Order Statistical | 75.9 | 73.0 | 77.3 |
| Frequency (DFT only) | 84.3 | 82.1 | 85.0 |
| **Proposed (Canny + DFT)** | **89.7** | **90.5** | **88.9** |

**TABLE 3:** Results using **SVM Classifier, SVM (Polynomial kernel)**.

| Feature Type | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| GLCM | 83.1 | 80.7 | 84.5 |
| LBP | 85.6 | 83.8 | 86.2 |
| First Order Statistical | 79.5 | 77.2 | 80.3 |
| Frequency (DFT only) | 88.5 | 86.9 | 89.4 |
| **Proposed (Canny + DFT)** | **92.5** | **93.2** | **91.8** |

**TABLE 4**: Results using **Random Forest (RF) Classifier (Number of tree=100)**

| Feature Type | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| GLCM | 80.7 | 78.5 | 81.4 |
| LBP | 83.9 | 82.2 | 84.3 |
| First Order Statistical | 77.8 | 75.0 | 78.5 |
| Frequency (DFT only) | 86.7 | 85.3 | 87.1 |
| **Proposed (Canny + DFT)** | **91.2** | **91.8** | **90.5** |

## CONCLUSION AND FUTURE WORK

This study presented a hybrid spatial–frequency approach for the automated detection of Autism Spectrum Disorder (ASD) using facial-image analysis. By integrating traditional spatial descriptors (GLCM, LBP) with both raw and edge-enhanced frequency-domain features (DFT and Canny + DFT), the proposed framework demonstrated a robust and interpretable method for early ASD screening. The hybrid Canny + DFT representation, introduced as a novel contribution in this work, significantly enhanced the discriminative ability of the feature space by emphasizing structurally relevant contours prior to frequency transformation. Combined with the Support Vector Machine classifier, this approach achieved a highest accuracy of 92.5%, surpassing both spatial-only and frequency-only features. The results reinforce the value of combining multi-domain handcrafted features to capture the complex morphological traits associated with ASD.

The findings also highlight the advantages of lightweight handcrafted systems over deep-learning approaches in scenarios where dataset availability is limited or computational resources are constrained. While deep-learning models trained on large datasets can achieve strong performance, they often lack transparency and require extensive tuning and hardware support. In contrast, the proposed system offers a reproducible, interpretable, and computationally efficient alternative capable of functioning effectively even on modest hardware. This characteristic supports its potential application in clinics, schools, and low-resource settings, where access to specialized equipment may be limited.

Several limitations of the present study must be acknowledged. First, the dataset used for experimentation was relatively small and lacked comprehensive demographic metadata such as gender, ethnicity, and precise age-group distributions. These factors may influence facial morphology and limit the generalizability of the model across diverse populations. Second, the images were captured under uncontrolled lighting and background conditions, which may introduce variability despite the applied preprocessing. Third, the framework relies exclusively on static 2D facial images; therefore, dynamic behavioral cues or multimodal traits associated with ASD, such as gaze patterns or speech, were not considered.

Future work will aim to address these limitations by expanding the dataset to include more diverse demographic representation through ethically approved data collection processes. Integrating additional modalities such as video sequences, eye-gaze tracking, or speech-based features may further enhance the sensitivity and specificity of ASD screening systems. Moreover, hybrid architectures that combine handcrafted feature engineering with explainable deep-learning components could be explored to balance interpretability with the representational power of neural networks. Finally, developing a user-friendly mobile or web-based interface may help translate the proposed system into a practical screening tool for educators, clinicians, and parents.

## REFERENCES

[1] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, “Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism,” Arch. Gen. Psychiatry, vol. 59, no. 9, pp. 809–816, 2002.

[2] S. Baron-Cohen, H. Ring, and E. Tager-Flusberg, “What is autism?,” Lancet, vol. 368, no. 9531, pp. 1445–1446, 2006.

[3] N. Duda, L. Zhang, and C. Li, “Use of machine learning for behavioral diagnosis of autism with limited dataset,” Sci.

Rep., vol. 10, pp. 1–12, 2020.

[4] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th ed.

Washington, D.C., USA: APA, 2013.

[5] A. Hashemi, L. F. Carpenter, and S. F. Shultz, “Craniofacial morphology in children with autism,” Mol. Autism, vol. 6, pp. 1–12, 2015.

[6] H. J. Hammel and F. M. Hildebrandt, “Facial asymmetry in autism spectrum disorder,” J. Autism Dev. Disord., vol. 49, pp. 350–363, 2019.

[7] U. R. Acharya, E. Y. Kuan, P. Subbanna, and K. Tamura, “Automated diagnosis of autism using image-based texture analysis,” Biomed. Signal Process. Control, vol. 68, p. 102689, 2021.

[8] M. R. Rosenfeld, “Facial morphology and neurodevelopmental disorders,” Curr. Opin. Pediatr., vol. 28, no. 6, pp. 715–722, 2016.

[9] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 6, pp. 610–621, 1973.

[10] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures,” Pattern Recogn., vol. 29, no. 1, pp. 51–59, 1996.

[11] A. Bovik, Handbook of Image and Video Processing, 2nd ed. New York, USA: Academic Press, 2005.

[12] Kaggle, “Autism facial image dataset,” Available: https://www.kaggle.com/ (accessed: August. 2025).

[13] D. S. Kunda and T. Goel, “Machine learning pipeline for autism detection using facial images,” Pattern Recogn. Lett., vol. 150, pp. 176–182, 2021.

[14] Y. Guo, Z. Li, and H. Zhou, “Deep multimodal fusion for autism detection from facial cues,” Neurocomputing, vol. 501, pp. 257–270, 2022.

[15] H. Al-Shammari, M. Saleh, and R. Al-Mutairi, “Hybrid handcrafted and CNN features for autism detection,” Appl. Sci., vol. 12, no. 11, p. 5484, 2022.

[16] S. Adriano and M. Broome, “Deep learning for autism diagnosis using facial behavior,” IEEE Trans. Affect. Comput., early access, 2023.

[17] A. Patel and F. Hossain, “Explainable AI for autism screening from facial images,” Expert Syst. Appl., vol. 233, p. 120946, 2023.

[18] R. C. Gonzales, R. E. Woods, and S. L. Eddins, Digital Image Processing Using MATLAB. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2004.

[19] J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, 1986.

[20] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, 2004.

[22] L. Breiman, “Random forests,” Mach. Learn., vol. 45, pp. 5–32, 2001.

[23] N. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” Am. Stat., vol. 46, no. 3, pp.

175–185, 1992.

[24] T. W. Loong, “Understanding sensitivity and specificity with the right side of the brain,” *British Medical Journal*, vol. 327, no. 7417, p. 716, 2003.