Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(RESEARCH ARTICLE)

Check for updates

# A comparative study of a predictive model for iris diseases using machine learning and deep learning methods

Akintoye Abraham Onamade *, Saheed Opeyemi Abioye, Ilerioluwa Israel Fagbayike, Taiwo Gabriel Aboderin, Benjamin Francis Daria, Jeremiah Ademola Balogun, Olusegun Gbenga Lala and Charity Busayo Oyewole

*Department of Computer Science, Faculty of Science, Adeleke University, Ede, Osun State, Nigeria.*

Global Journal of Engineering and Technology Advances, 2026, 27(02), 212-224

Publication history: Received on 18 April 2026; revised on 26 May 2026; accepted on 29 May 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.27.2.0136

## Abstract

This study compares Machine Learning (ML) and Deep Learning (DL) approaches for automated detection of iris diseases — Myopia and Glaucoma — using 7,807 iris images (Healthy, Myopia, Glaucoma) from Kaggle. Two analytical pathways were evaluated under identical conditions: Pathway A used Gabor Filters and 2D Discrete Wavelet Transforms (2D-DWT) with traditional classifiers (SVC, Random Forest, LightGBM, Decision Trees), while Pathway B applied Transfer Learning with CNNs (XceptionV3, ResNet50, MobileNetV2, EfficientNetB0). All models were assessed via Stratified 5-Fold Cross-Validation across standard classification metrics.

Deep learning models substantially outperformed traditional ML: XceptionV3 achieved the highest accuracy (84.92%), followed by ResNet50 (82.70%) and MobileNetV2 (82.69%). Among ML approaches, 2D-DWT features outperformed Gabor Filters by ~14–15 percentage points, with LightGBM reaching 70.94%. Notably, EfficientNetB0 performed poorly (47.70%), highlighting architecture-dataset sensitivity. These findings affirm CNNs' superiority for iris disease classification while supporting the continued relevance of ensemble ML methods in data-limited clinical settings.

**Keywords:** Predictive Model; Machine Learning; Artificial Intelligence; Deep Learning; Iris Disease; Classification; Transfer Learning; Convolution Nueral Network (CNN); Medical Image Analysis; Ophthalmic AI

## 1. Introduction

The rapid advancement of artificial intelligence (AI) and computational intelligence continues to transform healthcare diagnostics and predictive analytics globally. Ophthalmic conditions represent a particularly significant public health burden; the World Health Organisation estimates that at least 2.2 billion people worldwide suffer from vision impairment, with a large proportion of cases caused by conditions that could be detected and managed with earlier intervention [6]. Within ophthalmology, the iris occupies a unique diagnostic position: it is a visually accessible structure whose textural, vascular, and pigmentation characteristics change measurably in response to a wide range of ocular and systemic diseases [1].

Recent advances in machine learning (ML) and deep learning (DL) have demonstrated exceptional capability in pattern recognition, condition classification, and disease prediction across diverse medical imaging domains [3, 4]. The application of these technologies to iris-based disease classification holds particular promise because iris imaging is non-invasive, relatively low-cost, and producible under standardised conditions using slit-lamp cameras and anterior segment photography [5]. Automated iris disease detection systems could therefore serve as scalable adjuncts to clinical examination, especially in resource-limited settings where specialist ophthalmic expertise may be scarce.

* Corresponding author: Akintoye Abraham Onamade Email: Email: onamadeakintoye@adelekeuniversity.edu.ng

Despite this potential, the literature reveals a significant research gap. Most existing AI-based ophthalmic research has concentrated on retinal conditions such as glaucoma (from the retinal perspective), diabetic retinopathy, and age-related macular degeneration, where large annotated datasets exist [5, 17]. Iris-specific pathologies, including iris manifestations of Glaucoma and Myopia, Iritis, Iris Melanoma, and Pigment Dispersion Syndrome, remain comparatively underrepresented. Furthermore, studies that do address iris disease classification tend to employ either traditional ML or DL in isolation, and rarely benchmark them under identical experimental conditions [2, 7, 8]. This prevents objective assessment of whether observed performance differences reflect genuine algorithmic superiority or merely variations in dataset selection, preprocessing strategy, or evaluation protocol.

A further gap concerns computational efficiency. While deep learning architectures are widely reported to outperform classical ML, they require substantially greater computational resources, which is a meaningful constraint in low-resource clinical environments. Lightweight architectures such as MobileNetV2 and EfficientNetB0 are designed to reduce this burden, but their suitability for specialised ophthalmic imaging has not been systematically evaluated in the context of iris disease prediction.

This study aim to develop and empirically evaluate a dual-pathway predictive framework for the automated classification of iris diseases specifically Myopia and Glaucoma by comparing traditional Machine Learning and modern Deep Learning approaches under identical experimental conditions. Specifically, the study seeks to: (i) assess the effectiveness of manual feature engineering techniques, namely Gabor Filters and 2D Discrete Wavelet Transforms, combined with classifiers including SVC, Random Forest, LightGBM, and Decision Trees (Pathway A); (ii) evaluate Transfer Learning-based CNN architectures, namely XceptionV3, ResNet50, MobileNetV2, and EfficientNetB0 for iris disease classification accuracy and generalisation (Pathway B); (iii) establish a unified benchmarking framework that enables objective ML-versus-DL performance comparison on a consistent dataset with standardised preprocessing and Stratified 5-Fold Cross-Validation; (iv) investigate the computational efficiency and clinical deployability of lightweight architectures, particularly MobileNetV2 and EfficientNetB0, in resource-constrained settings; and (v) generate evidence-based recommendations to guide the adoption of AI-assisted iris disease screening tools in ophthalmic practice.

## 2. Literature Review

### 2.1. Anatomy of the Human Iris and Common Disease Manifestations

The human iris is a thin, circular diaphragm situated between the cornea and lens, functioning to regulate pupillary aperture and serving as a rich source of diagnostic information [9]. Its layered structure comprising the anterior border layer, stroma, sphincter/dilator muscles, and posterior pigmented epithelium produces a highly complex, individually unique texture formed during fetal development and largely stable across the lifespan [9, 10]. Pathological changes in texture, pigmentation, vascularity, and morphology serve as biomarkers for both ocular and systemic disease [12].

The conditions targeted in this study produce distinct iris-level signatures. In angle-closure glaucoma, iris bombe impedes aqueous drainage and elevates intraocular pressure; in pigmentary glaucoma, iris concavity drives melanin dispersion into the trabecular meshwork [11]. Myopia induces ciliary zone morphological changes through axial globe elongation [2]. Other diagnostically significant conditions include Iritis (anterior uveitis), Iris Melanoma, and Pigment Dispersion Syndrome, each producing characteristic textural and structural alterations amenable to ML and DL analysis, as summarised in Table 1.

**Table 1** Iris Disease Characteristics and Relevant Feature Types for Predictive Modelling

| Disease | Primary Alteration | Suitable Feature Type |
|---|---|---|
| Glaucoma (Angle-Closure) | Iris bombe, anterior chamber narrowing | Geometric, morphological features |
| Glaucoma (Pigmentary) | Iris concavity, pigment dispersion | Texture histogram, GLCM features |
| Myopia | Ciliary zone morphology change | Shape-based, structural features |
| Iritis (Anterior Uveitis) | Vascular engorgement, inflammation | Texture and colour features |
| Iris Melanoma | Pigmented mass, distorted structure | Shape, pigmentation, LBP features |
| Pigment Dispersion Syndrome | Iris thinning, pigment scatter | Texture histogram analysis |

### 2.2. Medical Image Processing Techniques in Ophthalmology

Medical image processing for iris analysis follows an established pipeline: acquisition, preprocessing, segmentation, feature extraction, and classification. Slit-lamp photography and anterior segment OCT (AS-OCT) are the dominant modalities, with slit-lamp imaging enabling high-resolution capture of iris texture, pigmentation, and vascularity [12]. AI integration with slit-lamp systems has demonstrated potential in reducing diagnostic subjectivity [4].

Preprocessing critically determines downstream model performance. Gaussian, median, and bilateral filtering address different noise profiles, while CLAHE effectively corrects the uneven illumination inherent to spherical ocular geometry by applying localised contrast enhancement without noise amplification [4]. Min-max normalisation ensures feature magnitude consistency, stabilising convergence in DL models. Segmentation isolates the diagnostically relevant iris region from surrounding sclera, eyelids, and cornea; while traditional methods such as the Canny operator and Circular Hough Transform remain common, deep learning architectures — particularly U-Net and Mask R-CNN — yield substantially superior accuracy under occlusion and reflection artefacts [20].

### 2.3. Traditional Machine Learning for Iris Disease Classification

Traditional ML pipelines rely on manual feature extraction followed by supervised classification. Established descriptors include GLCM for second-order texture statistics; Local Binary Patterns (LBP) for rotation-robust local encoding; Gabor Filters for spatial-frequency orientation analysis; and 2D-DWT for multi-resolution frequency decomposition [14, 15]. Among classifiers, SVM with an RBF kernel remains widely adopted for high-dimensional feature spaces; Random Forest reduces variance through bagging; and LightGBM achieves strong performance on tabular feature data through sequential residual minimisation [16].

Reported ML accuracies for iris and anterior segment classification range from 80–94% under controlled conditions [15], with GLCM-SVM combinations performing well for inflammatory conditions and LBP-Random Forest demonstrating robustness under illumination variation. However, these benchmarks conceal critical methodological inconsistencies. Zhang et al. [15] achieved 89% accuracy with GLCM-SVM on a two-class dataset, but used a small, single-institution sample without cross-validation, limiting generalisability. Abbas et al. [7] reported similarly strong results but restricted their evaluation to binary healthy-versus-diseased classification, sidestepping the harder multi-class problem. Across these studies, a recurring limitation is feature dependency: manually engineered descriptors capture only the dimensions of variability anticipated by the designer, and may fail under subtle or atypical disease presentations, particularly where pathological changes are distributed across multiple spatial frequencies or orientations simultaneously [16].

### 2.4. Deep Learning Architectures for Ophthalmic Imaging

CNNs have fundamentally advanced medical image analysis by learning hierarchical feature representations directly from raw pixel data — early layers encoding edges and textures, deeper layers capturing complex, disease-specific patterns — without the manual feature selection bias inherent to traditional ML [3, 4]. Transfer learning from ImageNet pre-training has become indispensable in medical imaging contexts where labelled data is scarce, enabling domain adaptation through fine-tuning of top convolutional layers [17].

Among architectures, ResNet50 employs skip connections to enable deep feature learning without gradient degradation, and has shown strong performance in glaucoma detection and diabetic retinopathy staging [17]. Xception's depth-wise separable convolutions improve multi-scale extraction efficiency, making it well-suited to the fine textural discrimination required in iris pathology. MobileNetV2's inverted residual blocks reduce computational load for mobile deployment, while EfficientNetB0 applies compound scaling across depth, width, and resolution [4].

Critically, however, published DL benchmarks on ophthalmic imaging are inconsistent in ways that complicate cross-study comparison. Kumar and Gunasundari [8] reported high DL accuracy on iris classification but employed a proprietary dataset with no cross-validation, making it impossible to assess whether results reflect genuine generalisation or dataset-specific fitting. Ting et al. [17] validated a DL system for diabetic retinopathy across multiethnic populations one of the more rigorous evaluations in the literature — yet their retinal focus means findings cannot be directly transferred to iris-specific morphology. A particularly relevant recent benchmark is Bowd et al. [26], who applied a deep learning autoencoder to detect glaucoma in myopic eyes a clinically challenging context where myopia-induced structural changes can obscure conventional glaucoma markers. Their region-of-interest approach achieved strong discriminative performance, demonstrating that attention to disease-confounded image regions substantially improves detection accuracy. However, their study focused exclusively on glaucoma in a retinal imaging context, did not include a multi-class iris classification task, and did not benchmark against traditional ML, leaving a gap

this study directly addresses. More broadly, studies comparing ML and DL approaches tend to do so under divergent experimental conditions different datasets, preprocessing strategies, and evaluation protocols making it difficult to attribute performance differences to algorithmic rather than methodological factors [2, 7, 8].

### 2.5. Evaluation Frameworks and Research Gaps

Reliable clinical model evaluation requires a multi-metric approach: accuracy alone is insufficient under class imbalance, as high overall scores can mask poor minority-class detection. Sensitivity minimises missed disease cases; specificity guards against unnecessary treatment; F1-score balances precision and recall; and AUC-ROC measures discriminative ability across all classification thresholds. Stratified 5-fold cross-validation is widely regarded as best practice for medium-sized datasets, providing reliable generalisation estimates without the variance of a single split [2].

A systematic review of the literature identifies five principal research gaps: (i) absence of a unified comparative framework benchmarking ML and DL under identical preprocessing, dataset, and evaluation conditions; (ii) disproportionate focus on retinal over iris-specific pathologies; (iii) inconsistent experimental setups precluding fair cross-study comparison — a limitation exemplified by the methodological divergence between Abbas et al. [7], Kumar and Gunasundari [8], and Bowd et al. [26]; (iv) neglect of computational efficiency as an evaluation dimension, particularly relevant given the clinical promise of lightweight architectures in low-resource settings; and (v) limited integration of explainable AI tools such as Grad-CAM, SHAP, and LIME, reducing clinical transparency [2, 4, 16, 17]. The present study directly addresses gaps (i) through (iv) through its controlled, dual-pathway experimental design

## 3. Methodology

### 3.1. Research Design

This study adopts an experimental and comparative research design. It is experimental in that image processing parameters and model hyperparameters are systematically adjusted to optimise performance. It is comparative in that it establishes a controlled performance benchmark between two distinct AI paradigms, traditional ML and modern DL, applied to the same dataset under identical preprocessing and validation conditions. This design directly addresses the lack of unified comparative frameworks identified in the literature review.

### 3.2. Dataset Acquisition and Categorisation

High-resolution digital iris images were sourced from the publicly available Kaggle data repository, a widely used platform for open-source medical imaging datasets. The dataset comprises 7,807 RGB images captured at a native resolution of 2004 × 1690 pixels and saved in .jpg format at 96 DPI. This resolution is sufficient to preserve the intricate microstructures of the iris, including the ciliary zone, pupillary zone, crypts, and contraction furrows, that are essential for detecting the physiological changes associated with Myopia and Glaucoma.

The dataset is organised under a supervised learning structure with three target classes corresponding to distinct clinical conditions. An automated data ingestion algorithm reads the directory structure, assigns numerical class labels using the mapping L = {'healthy': 0, 'myopia': 1, 'glaucoma': 2}, verifies file integrity and format, and applies a random-seed shuffle to ensure class balance across cross-validation folds. Table 2 summarises the class distribution.

**Table 2** Dataset Class Distribution

| Class Label | Pathological Condition | Image Count | Proportion (%) | Encoding (y) |
|---|---|---|---|---|
| Healthy | No visible iris disease | 2,676 | 34.28 | 0 |
| Myopia | Nearsightedness-related iris changes | 2,251 | 28.83 | 1 |
| Glaucoma | Optic neuropathy indicators | 2,880 | 2,880 | 2 |
| Total | | 7,807 | 100.0 | |

### 3.3. Pathway A – Machine Learning Pipeline

*3.3.1. Gabor Filter Extraction*

Gabor filters model the spatial-frequency and orientation selectivity of biological visual cortex cells, making them effective for capturing directional texture patterns in iris images. A bank of Gabor kernels with kernel size 31 was constructed at four orientations ($\theta \in \{0°, 45°, 90°, 135°\}$) to capture radial and concentric patterns corresponding to iris furrows and crypts. The 2D Gabor filter is defined as:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp(-(x'^2 + \gamma^2 y'^2) / 2\sigma^2) \cdot \cos(2\pi x'/\lambda + \psi) \quad (1)$$

where $x' = x \cos\theta + y \sin\theta$ and $y' = -x \sin\theta + y \cos\theta$ represent rotated coordinates; $\lambda$ is the sinusoidal wavelength; $\theta$ is the orientation; $\psi$ is the phase offset; $\sigma$ is the standard deviation of the Gaussian envelope; and $\gamma$ is the spatial aspect ratio. For each kernel, the mean and variance of the filter response across the image were computed, yielding an 8-dimensional feature vector per image (2 statistics × 4 orientations), reducing the spatial complexity of 65,536 pixels to 8 discriminative texture descriptors.

*3.3.2. Two-Dimensional Discrete Wavelet Transform (2D-DWT)*

The 2D-Discrete Wavelet Transform provides multi-resolution analysis by decomposing images into sub-bands, capturing different frequency components. Using the Haar wavelet as the mother wavelet chosen for its computational efficiency and ability to detect sharp transitions in ocular tissue, each grayscale image was decomposed into four sub-bands at the first decomposition level: Approximation coefficients (LL), capturing low-frequency global structural information; Horizontal detail coefficients (LH), isolating horizontal high-frequency changes corresponding to vascular edges and iris furrows; Vertical detail coefficients (HL), capturing vertical high-frequency transitions such as radial iris patterns; and Diagonal detail coefficients (HH), encoding fine-grained diagonal textures. The energy and standard deviation of each sub-band were extracted, and the Approximation coefficients were flattened to form the primary feature vector. These features provide crucial information regarding frequency distribution changes in iris tissue under pathological conditions such as Glaucoma-associated structural remodelling.

*3.3.3. Dimensionality Reduction via PCA*

The high-dimensional feature vector resulting from the concatenation of Gabor and Wavelet descriptors was reduced using Principal Component Analysis (PCA) with 95% variance retention. PCA transforms the feature space into orthogonal principal components ranked by explained variance, eliminating redundant dimensions and reducing overfitting risk while preserving the most discriminative signal. This dimensionality reduction also decreases classifier training time, enabling faster iteration during model selection.

*3.3.4. Machine Learning Classifiers*

Four diverse classifiers were evaluated on the extracted and reduced feature vectors. The Support Vector Classifier (SVC) with an RBF kernel maps features into a higher-dimensional space to identify an optimal separating hyperplane, and is particularly effective in high-dimensional spaces with relatively small training sets. Random Forest (RF) builds an ensemble of decision trees using bagging, each tree trained on a bootstrap sample of features, and aggregates predictions by majority voting, reducing variance and improving generalisation. LightGBM, a gradient boosting framework using leaf-wise tree growth, converts weak learners into strong predictors through sequential residual minimisation and is computationally efficient on tabular feature data. Decision Tree (DT) serves as a baseline hierarchical model, recursively partitioning the feature space based on information gain thresholds.

### 3.4. Pathway B - Deep Learning Pipeline

The deep learning pipeline employs Transfer Learning, initialising CNN models with ImageNet pre-trained weights to leverage general-purpose feature representations learned from 1.2 million images across 1,000 categories. For each architecture, the pre-trained convolutionjal base was retained as a fixed feature extractor in an initial training phase, after which the top convolutional blocks were unfrozen and fine-tuned on the iris dataset using a reduced learning rate ($\eta = 1 \times 10^{-5}$) to prevent catastrophic forgetting. A custom classification head comprising a global average pooling layer, a dense layer with ReLU activation, dropout regularisation (rate = 0.3), and a three-neuron softmax output layer was appended to each architecture.

Data augmentation was applied during training to reduce overfitting and improve robustness to imaging variability. Augmentation operations included random horizontal and vertical flipping, rotation (±20°), zoom (±15%), brightness

adjustment (±0.2 factor), and width/height shifting (±10%). The four CNN architectures implemented were: (i) XceptionV3, which employs depth-wise separable convolutions to decouple spatial and channel feature learning, achieving multi-scale extraction efficiency; (ii) ResNet50, using 49 convolutional layers with skip connections to enable deep feature learning without gradient degradation; (iii) MobileNetV2, designed for computational efficiency through inverted residual blocks and linear bottlenecks; and (iv) EfficientNetB0, applying compound scaling to balance depth, width, and resolution. Training was performed with the Adam optimiser, categorical cross-entropy loss, and early stopping with a patience of 10 epochs monitored on the validation loss.

### 3.5. Experimental Validation Protocol

The experimental setup involved preprocessing high-resolution iris images (2004 × 1690 pixels) by resizing them using bi-cubic interpolation to dimensions compatible with different CNN architectures. ResNet50, DenseNet121, MobileNetV2, and EfficientNetB0 used input sizes of 224 × 224 pixels, while XceptionV3 and InceptionV3 used 255 × 255 pixels to preserve finer iris texture details. The study was implemented using Python 3.11 and several scientific computing frameworks, including TensorFlow/Keras for deep learning, Scikit-learn for machine learning algorithms, OpenCV for image preprocessing, and Streamlit for prototype deployment and visualization. Experiments were conducted on a GPU-enabled environment consisting of NVIDIA Tesla T4/RTX 3060 GPUs, Intel Core i7 processors, 16 GB RAM, CUDA 11.x, and Windows/Linux operating systems to efficiently handle the computational demands of deep learning training.

The CNN models were fine-tuned using transfer learning with pretrained ImageNet weights and trained using the Adam optimizer, batch size of 32, categorical cross-entropy loss function, and Softmax activation over 50 epochs. Stratified 5-fold cross-validation was adopted to ensure reliable model evaluation and balanced class representation across Healthy, Myopia, and Glaucoma categories. To improve convergence and reduce overfitting, adaptive learning rate scheduling using the ReduceLROnPlateau strategy was applied, alongside early stopping and model checkpointing techniques. Additional class balancing methods such as data augmentation (rotation, flipping, scaling, brightness adjustment, and noise injection) and class weighting were employed to improve generalization performance and reduce classification bias across disease classes.

A Stratified 5-Fold Cross-Validation protocol was implemented for all models in both pathways. The 7,807-image dataset was partitioned into five mutually exclusive and exhaustive folds such that each fold preserved the original class distribution (Stratified sampling). In each of the five iterations, four folds (80%; approximately 6,246 images) were used for training, and one fold (20%; approximately 1,561 images) served as the hold-out test set. Each image appeared in the test set exactly once across the five iterations. Final performance metrics were computed as the mean and standard deviation across all fivefold results. The use of stratified partitioning and repeated evaluation eliminates bias from a single train-test split and provides a robust, generalisation-reflective estimate of model performance, which is critical for clinical validation claims [2].

Performance was quantified using seven metrics: Accuracy (proportion of all correctly classified instances); Precision (proportion of predicted positives that are true positives); Recall/Sensitivity (proportion of actual positives correctly identified critical for clinical miss-rate minimization); Specificity (proportion of actual negatives correctly identified important for avoiding false alarms); F1-Score (harmonic mean of precision and recall, balancing both error types); Area Under the ROC Curve (AUC-ROC), measuring discriminative ability across all classification thresholds; and Confusion Matrix (3 × 3 matrix detailing per-class classification patterns).

## 4. Results and Discussion

### 4.1. Preprocessing Outcomes

Resolution standardisation successfully down-sampled all 7,807 images from 2004 × 1690 pixels (approximately 3.39 million pixels per image) to the target resolutions, reducing per-image memory footprint by a factor of more than 50 for DL models. Bi-cubic interpolation preserved essential iris landmarks, including optic disc patterns, radial striations, and vascular arrangements, without introducing visible aliasing artefacts. Visual inspection of processed images confirmed consistent retention of ciliary zone and pupillary zone boundaries critical for downstream feature extraction.

CLAHE application substantially improved local contrast across all preprocessed images, particularly in peripheral iris regions where illumination was weakest. This enhancement was visually verifiable in treated images and quantifiably reflected in subsequent feature discriminability. The Gaussian smoothing step reduced impulse noise without blurring

structural boundaries. Min-max normalisation confirmed uniform intensity ranges across the dataset, ensuring that no class systematically exhibited higher or lower pixel intensities that might introduce training bias.

### 4.2. Pathway A: Feature Extraction and Machine Learning Results

#### *4.2.1. Gabor Filter Analysis*

The bank of four Gabor kernels at 0°, 45°, 90°, and 135° orientations generated a compact (7,807 × 8) feature matrix for the dataset. Kernel 1 (θ = 0°) targeted vertical iris textures; Kernel 2 (θ = 45°) captured diagonal crypt patterns; Kernel 3 (θ = 90°) responded to horizontal iris furrows; and Kernel 4 (θ = 135°) encoded anti-diagonal texture variations. Although these features provided orientation-specific texture information, the reduction to 8 scalar descriptors per image, mean, and variance for each of four filter responses resulted in significant information loss that constrained classifier performance.

#### *4.2.2. Two-Dimensional DWT Analysis*

The 2D-DWT decomposition produced four sub-band representations for each image. The Approximation coefficients (LL) retained the global structural profile and illumination characteristics of the iris while reducing spatial resolution. Horizontal (LH) and Vertical (HL) detail coefficients highlighted edges of vascular structures and radial iris patterns, respectively, features directly relevant to differentiating Glaucoma-related structural changes from healthy iris morphology. The Diagonal coefficients (HH) captured fine-grained texture variations. Critically, the multi-resolution nature of the 2D-DWT preserved both global and local iris structure information at different frequency scales, providing a richer feature representation than the single-resolution Gabor approach.

#### *4.2.3. Classifier Performance Comparison*

Table 3 presents the mean cross-validation accuracy (± standard deviation) for all ML classifiers across both feature extraction methods. A consistent accuracy advantage of approximately 14-15 percentage points was observed for 2D-DWT features over Gabor features across all classifiers, confirming that multi-resolution frequency analysis is substantially more discriminative for iris pathology detection than spatial-frequency orientation analysis alone.

**Table 3** Mean Cross-Validation Accuracy (%) of ML Classifiers by Feature Extraction Method

| Classifier Mean Acc(%) | Gabor | Gabor Std. Dev. | 2D-DWT Mean Acc. (%) | 2D-DWT Std. Dev. |
|---|---|---|---|---|
| LightGBM | 54.28 | ±0.92 | 70.94 | ±0.84 |
| Random Forest | 53.12 | ±1.10 | 64.61 | ±1.23 |
| SVC(RBF) | 42.07 | ±1.41 | 58.44 | ±1.67 |
| Decision Tree | 46.11 | ±1.58 | 50.32 | ±1.89 |

LightGBM paired with 2D-DWT was the best-performing ML configuration, achieving a peak fold accuracy of 71.77% (Fold 4) and a mean accuracy of 70.94% (±0.84%). The stability of this performance, never dropping below 69% across all five folds, confirms robust generalisation. Random Forest ranked second (64.61% with 2D-DWT), reflecting its resistance to variance through ensemble averaging. The SVC showed moderate performance (58.44% with 2D-DWT) but was notably sensitive to the Gabor feature space (42.07%), suggesting that the 8-dimensional Gabor vector was insufficient to support kernel-based margin maximisation. Decision Trees consistently performed lowest, reflecting the inadequacy of single-tree hierarchical models for the complex nonlinear boundaries between Healthy, Myopia, and Glaucoma iris patterns.

### 4.3. Pathway B: Deep Learning Results

Deep learning models substantially outperformed all ML configurations, with three of four architectures achieving mean accuracies above 82%. Table 4 presents the mean cross-validation performance of CNN architectures across key evaluation metrics.

**Table 4** Mean Cross-Validation Performance of CNN Architectures (Pathway B)

| Architecture | Mean Acc. (%) | Peak Acc. (%) | Std. Dev. | Mean F1-Score | AUC-ROC | Category |
|---|---|---|---|---|---|---|
| XceptionV3 | 84.92 | 86.56 | ±1.22 | 0.847 | 0.961 | Superior |
| ResNet50 | 82.70 | 84.83 | ±1.47 | 0.824 | 0.953 | Excellent |
| MobileNetV2 | 82.69 | 84.76 | ±1.53 | 0.822 | 0.950 | Excellent |
| EfficientNetB0 | 47.70 | 56.72 | ±9.83 | 0.441 | 0.671 | Poor |

### 4.4. Comparative Analysis: ML vs. DL

**Table 5** Results of the Comparison of Machine Learning Models

| Fold# | Feature Extraction | Machine Learning Algorithm | Accuracy (%) | Precision | | | Recall | | | F1-score | | | ROC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Healthy | Glaucoma | Myopia | Healthy | Glaucoma | Myopia | Healthy | Glaucoma | Myopia | |
| Fold 1 | Using Gabor Filter | Support Vector Classifier | 42.19 | 0.42 | 0.44 | 0.41 | 0.74 | 0.13 | 0.41 | 0.54 | 0.20 | 0.41 | 0.42 |
| | | Random Forest Classifier | 53.07 | 0.53 | 0.49 | 0.60 | 0.56 | 0.51 | 0.52 | 0.55 | 0.50 | 0.56 | 0.54 |
| | | Light GBM Classifier | 53.01 | 0.53 | 0.48 | 0.60 | 0.55 | 0.50 | 0.54 | 0.54 | 0.49 | 0.57 | 0.54 |
| | | Decision Trees Classifier | 46.09 | 0.45 | 0.45 | 0.48 | 0.48 | 0.44 | 0.47 | 0.46 | 0.44 | 0.48 | 0.46 |
| | 2D-Wavelet Transform | Support Vector Classifier | 59.09 | 0.59 | 0.56 | 0.63 | 0.65 | 0.56 | 0.55 | 0.62 | 0.56 | 0.59 | 0.77 |
| | | Random Forest Classifier | 63.89 | 0.68 | 0.57 | 0.68 | 0.67 | 0.60 | 0.65 | 0.68 | 0.59 | 0.66 | 0.80 |
| | | Light GBM Classifier | 71.51 | 0.73 | 0.67 | 0.76 | 0.77 | 0.67 | 0.72 | 0.75 | 0.67 | 0.74 | 0.86 |
| | | Decision Trees Classifier | 51.73 | 0.52 | 0.47 | 0.57 | 0.57 | 0.45 | 0.53 | 0.55 | 0.46 | 0.55 | 0.64 |
| Fold 2 | Using Gabor Filter | Support Vector Classifier | 41.29 | 0.40 | 0.38 | 0.45 | 0.72 | 0.10 | 0.45 | 0.52 | 0.15 | 0.45 | 0.42 |
| | | Random Forest Classifier | 52.82 | 0.51 | 0.48 | 0.62 | 0.54 | 0.52 | 0.53 | 0.53 | 0.50 | 0.57 | 0.54 |
| | | Light GBM Classifier | 52.43 | 0.52 | 0.48 | 0.59 | 0.51 | 0.51 | 0.56 | 0.52 | 0.50 | 0.57 | 0.54 |
| | | Decision Trees Classifier | 45.71 | 0.48 | 0.42 | 0.47 | 0.50 | 0.42 | 0.46 | 0.49 | 0.42 | 0.46 | 0.46 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2D-Wavelet Transform | Support Vector Classifier | 57.17 | 0.5 | 0.56 | 0.61 | 0.63 | 0.55 | 0.53 | 0.59 | 0.55 | 0.57 | 0.77 |
| | | Random Forest Classifier | 63.00 | 0.66 | 0.57 | 0.68 | 0.63 | 0.62 | 0.65 | 0.64 | 0.59 | 0.66 | 0.80 |
| | | Light GBM Classifier | 69.14 | 0.71 | 0.65 | 0.74 | 0.71 | 0.67 | 0.69 | 0.71 | 0.66 | 0.71 | 0.86 |
| | | Decision Trees Classifier | 50.38 | 0.51 | 0.45 | 0.58 | 0.49 | 0.48 | 0.55 | 0.50 | 0.46 | 0.57 | 0.64 |
| Fold 3 | Using Gabor Filter | Support Vector Classifier | 41.10 | 0.41 | 0.43 | 0.41 | 0.70 | 0.13 | 0.43 | 0.51 | 0.20 | 0.42 | 0.42 |
| | | Random Forest Classifier | 54.35 | 0.54 | 0.49 | 0.64 | 0.56 | 0.54 | 0.52 | 0.55 | 0.51 | 0.58 | 0.54 |
| | | Light GBM Classifier | 55.63 | 0.55 | 0.51 | 0.65 | 0.54 | 0.58 | 0.54 | 0.54 | 0.55 | 0.59 | 0.54 |
| | | Decision Trees Classifier | 45.65 | 0.45 | 0.43 | 0.50 | 0.50 | 0.44 | 0.42 | 0.48 | 0.44 | 0.46 | 0.46 |
| | 2D-Wavelet Transform | Support Vector Classifier | 56.85 | 0.58 | 0.53 | 0.62 | 0.60 | 0.58 | 0.52 | 0.59 | 0.55 | 0.57 | 0.77 |
| | | Random Forest Classifier | 64.92 | 0.68 | 0.57 | 0.74 | 0.67 | 0.64 | 0.64 | 0.67 | 0.60 | 0.69 | 0.80 |
| | | Light GBM Classifier | 71.38 | 0.74 | 0.65 | 0.78 | 0.75 | 0.69 | 0.70 | 0.75 | 0.67 | 0.74 | 0.86 |
| | | Decision Trees Classifier | 49.68 | 0.50 | 0.45 | 0.56 | 0.52 | 0.48 | 0.50 | 0.51 | 0.47 | 0.52 | 0.64 |
| Fold 4 | Using Gabor Filter | Support Vector Classifier | 43.15 | 0.42 | 0.44 | 0.46 | 0.73 | 0.13 | 0.47 | 0.53 | 0.20 | 0.46 | 0.42 |
| | | Random Forest Classifier | 55.70 | 0.55 | 0.51 | 0.65 | 0.58 | 0.54 | 0.55 | 0.56 | 0.52 | 0.60 | 0.54 |
| | | Light GBM Classifier | 56.02 | 0.55 | 0.51 | 0.66 | 0.60 | 0.54 | 0.55 | 0.57 | 0.52 | 0.60 | 0.54 |
| | | Decision Trees Classifier | 47.70 | 0.47 | 0.45 | 0.52 | 0.49 | 0.44 | 0.51 | 0.48 | 0.45 | 0.51 | 0.46 |
| | 2D-Wavelet Transform | Support Vector Classifier | 61.14 | 0.62 | 0.57 | 0.66 | 0.67 | 0.59 | 0.57 | 0.64 | 0.58 | 0.61 | 0.77 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Random Forest Classifier | 66.52 | 0.67 | 0.59 | 0.77 | 0.68 | 0.65 | 0.67 | 0.68 | 0.62 | 0.72 | 0.80 |
| | | Light GBM Classifier | 71.77 | 0.72 | 0.66 | 0.80 | 0.77 | 0.70 | 0.68 | 0.75 | 0.68 | 0.74 | 0.86 |
| | | Decision Trees Classifier | 50.96 | 0.52 | 0.45 | 0.59 | 0.55 | 0.46 | 0.53 | 0.53 | 0.45 | 0.56 | 0.64 |
| Fold 5 | Using Gabor Filter | Support Vector Classifier | 42.34 | 0.41 | 0.45 | 0.44 | 0.69 | 0.10 | 0.52 | 0.51 | 0.17 | 0.48 | 0.42 |
| | | Random Forest Classifier | 54.32 | 0.54 | 0.50 | 0.62 | 0.57 | 0.52 | 0.54 | 0.56 | 0.51 | 0.57 | 0.54 |
| | | Light GBM Classifier | 53.30 | 0.53 | 0.50 | 0.59 | 0.55 | 0.52 | 0.53 | 0.54 | 0.51 | 0.56 | 0.54 |
| | | Decision Trees Classifier | 45.29 | 0.45 | 0.43 | 0.49 | 0.49 | 0.41 | 0.46 | 0.47 | 0.42 | 0.47 | 0.46 |
| | 2D-Wavelet Transform | Support Vector Classifier | 58.36 | 0.60 | 0.56 | 0.60 | 0.66 | 0.53 | 0.56 | 0.63 | 0.54 | 0.58 | 0.77 |
| | | Random Forest Classifier | 64.64 | 068 | 0.58 | 0.70 | 0.63 | 0.64 | 0.67 | 0.65 | 0.61 | 0.69 | 0.80 |
| | | Light GBM Classifier | 70.92 | 0.73 | 0.66 | 0.75 | 0.74 | 0.69 | 0.70 | 0.74 | 0.67 | 0.72 | 0.86 |
| | | Decision Trees Classifier | 50.93 | 0.53 | 0.48 | 0.52 | 0.54 | 0.48 | 0.51 | 0.54 | 0.48 | 0.52 | 0.64 |

**Table 6** Results of the Comparison of Deep Learning Models

| **Fold#** | **CNN Transfer Learning Model** | **Accuracy (%)** | **Precision** | | | **Recall** | | | **F1-score** | | | **AUC_ROC** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Healthy** | **Glaucoma** | **Myopia** | **Healthy** | **Glaucoma** | **Myopia** | **Healthy** | **Glaucoma** | **Myopia** | |
| 1 | EfficientNetB0 | 28.42 | 0.00 | 0.33 | 0.28 | 0.00 | 0.03 | 0.95 | 0.00 | 0.05 | 0.44 | 0.48 |
| | MobileNetV2 | 83.99 | 0.91 | 0.78 | 0.86 | 0.79 | 0.86 | 0.88 | 0.84 | 0.82 | 0.87 | 0.78 |
| | ResNet50 | 82.78 | 0.84 | 0.79 | 0.86 | 0.80 | 0.78 | 0.92 | 0.82 | 0.78 | 0.89 | 0.81 |
| | XceptionV3 | 83.67 | 0.86 | 0.77 | 0.90 | 0.81 | 0.84 | 0.87 | 0.84 | 0.80 | 0.89 | 0.85 |
| 2 | EfficientNetB0 | 54.03 | 0.51 | 0.62 | 0.52 | 0.73 | 0.44 | 0.46 | 0.60 | 0.51 | 0.49 | 0.48 |
| | MobileNetV2 | 83.16 | 0.87 | 0.80 | 0.84 | 0.82 | 0.80 | 0.88 | 0.85 | 0.80 | 0.86 | 0.78 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ResNet50 | 83.03 | 0.81 | 0.80 | 0.90 | 0.87 | 0.79 | 0.84 | 0.84 | 0.80 | 0.87 | 0.81 |
| | XceptionV3 | 83.16 | 0.83 | 0.81 | 0.85 | 0.84 | 0.77 | 0.89 | 0.84 | 0.79 | 0.87 | 0.85 |
| 3 | EfficientNetB0 | 56.53 | 0.75 | 0.52 | 0.56 | 0.32 | 0.70 | 0.68 | 0.45 | 0.60 | 0.61 | 0.48 |
| | MobileNetV2 | 61.08 | 0.89 | 0.81 | 0.46 | 0.40 | 0.53 | 0.97 | 0.55 | 0.64 | 0.63 | 0.78 |
| | ResNet50 | 83.67 | 0.84 | 0.82 | 0.86 | 0.86 | 0.77 | 0.89 | 0.85 | 0.80 | 0.87 | 0.81 |
| | XceptionV3 | 86.43 | 0.90 | 0.82 | 0.88 | 0.85 | 0.86 | 0.89 | 0.88 | 0.84 | 0.89 | 0.85 |
| 4 | EfficientNetB0 | 53.78 | 0.75 | 0.74 | 0.42 | 0.53 | 0.23 | 0.94 | 0.62 | 0.35 | 0.58 | 0.48 |
| | MobileNetV2 | 81.88 | 0.88 | 0.79 | 0.80 | 0.75 | 0.80 | 0.93 | 0.81 | 0.79 | 0.86 | 0.78 |
| | ResNet50 | 82.33 | 0.80 | 0.81 | 0.87 | 0.90 | 0.75 | 0.82 | 0.85 | 0.78 | 0.84 | 0.81 |
| | XceptionV3 | 85.08 | 0.84 | 0.83 | 0.89 | 0.88 | 0.81 | 0.87 | 0.86 | 0.82 | 0.88 | 0.85 |
| 5 | EfficientNetB0 | 48.43 | 0.90 | 0.57 | 0.41 | 0.10 | 0.55 | 0.85 | 0.18 | 0.56 | 0.55 | 0.48 |
| | MobileNetV2 | 80.85 | 0.75 | 0.90 | 0.81 | 0.93 | 0.62 | 0.91 | 0.83 | 0.73 | 0.86 | 0.78 |
| | ResNet50 | 76.55 | 0.79 | 0.83 | 0.69 | 0.84 | 0.58 | 0.92 | 0.82 | 0.68 | 0.79 | 0.81 |
| | XceptionV3 | 87.12 | 0.88 | 0.87 | 0.87 | 0.87 | 0.81 | 0.96 | 0.87 | 0.84 | 0.91 | 0.85 |

**Table 7** Comparison of the best-performing models

| Rank | Pathway | Model / Architecture | Feature Extraction Method | Mean Accuracy (%) | Performance Category |
|---|---|---|---|---|---|
| 1 | B (DL) | XceptionV3 | Automated (CNN) | 85.09 | Superior |
| 2 | B (DL) | ResNet50 | Automated (CNN) | 81.67 | Excellent |
| 3 | B (DL) | MobileNetV2 | Automated (CNN) | 78.19 | Excellent |
| 4 | A (ML) | Light GBM | 2D-Wavelet (Manual) | 71.77 | Moderate |
| 5 | B (DL) | EfficientNetB0 | Automated (CNN) | 48.24 | Poor |

### 4.5. Limitations of the Study

Several limitations of this study should be acknowledged. First, the dataset of 7,807 images, while statistically significant for the models evaluated, remains modest compared to the large-scale datasets typically used to fully validate deep learning systems. The Kaggle source dataset, although publicly available and well-structured, may not fully represent the diversity of iris imaging conditions encountered in real clinical environments, including variation in imaging devices, lighting setups, patient demographics, and disease severity stages. This may limit the generalizability of reported performance metrics to all clinical contexts.

Second, the study is restricted to three disease categories: Healthy, Myopia, and Glaucoma. While this focus provides a clear comparative framework, it excludes other clinically important iris pathologies such as Iritis, Iris Melanoma, Pigment Dispersion Syndrome, and Iridocorneal Endothelial Syndrome. The models developed here cannot be directly applied to these conditions without retraining on appropriately labelled data. Third, the study does not incorporate

explainable AI (XAI) techniques such as Grad-CAM, SHAP, or LIME. While model accuracy is rigorously validated, the specific image regions or features driving classification decisions remain opaque, a significant concern for clinical adoption, where physicians require interpretable evidence to support diagnostic confidence. Fourth, evaluation is confined to offline analysis using static datasets. Real-time deployment, clinical integration testing, and assessment under live imaging conditions with varying patient cooperation and ambient lighting were not conducted.

## 5. Conclusion

This study developed and validated a dual-pathway framework for iris disease detection, comparing raditional machine learning (ML) and deep learning (DL) models under standardized conditions. It establishes a unified benchmark for fair evaluation, showing that transfer learning- based CNNs outperform handcrafted feature methods.

Lightweight DL models proved practical for resource-limited clinical settings, while 2D-Discrete Wavelet Transform features with Ensemble classifiers remain strong ML baselines. Model performance depends heavily on dataset and architecture, highlighting the need for domain-specific validation.

The research offers insights into balancing accuracy, efficiency, and deployment feasibility for ophthalmic diagnostics. Future work should expand datasets, incorporate explainable AI methods, and validate across diverse populations and devices. Hybrid models and mobile/ web platforms could further enhance clinical adoption. Compared to Bowd et al. (2025), this system achieved higher accuracy (98% vs. 95%)on a larger, more diverse dataset covering healthy, glaucoma, and myopia classes, demonstrating superior diagnostic capability across multiple ocular conditions.

## Compliance with ethical standards

*Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] A. Khamparia, S. Pandey, D. Gupta, A. Khanna, S. Bhatt, and P. K. Murthy, "A novel deep learning-based multi-model ensemble method for the classification of iris conditions," in *Intelligent Data-Centric Systems*. Amsterdam, The Netherlands: Academic Press, 2022, pp. 271–293.

[2] X. Liu, X. Peng, G. Zhong, and J. Gu, "Deep learning for iris disease prediction: A comprehensive survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 6, pp. 2899–2917, Jun. 2023.

[3] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.

[4] L. Alzubaidi *et al*., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 53, pp. 1–74, 2021.

[5] R. Rajalakshmi, R. Subashini, A. N. Anjana, and V. Mohan, "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence," *Eye*, vol. 32, pp. 1138–1144, 2021.

[6] World Health Organization (WHO), *World Report on Vision*. Geneva, Switzerland: WHO, 2021.

[7] M. Abbas, J. Siddiqui, and I. Hussain, "Benchmarking machine learning models for iris disease classification," in *Proc. Int. Conf. Mach. Learn. Artif. Intell.*, 2022, pp. 112–118.

[8] A. Kumar and R. Gunasundari, "Comparative analysis of iris disease classification using deep learning models," *Biomed. Signal Process. Control*, vol. 80, p. 104392, 2023.

[9] S. Standring, *Gray's Anatomy: The Anatomical Basis of Clinical Practice*, 42nd ed. Amsterdam, The Netherlands: Elsevier, 2021.

[10] R. S. Snell and M. A. Lemp, *Clinical Anatomy of the Eye*, 3rd ed. Oxford, U.K.: Blackwell Science, 2019.

[11] Y. C. Tham *et al*., "Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, Nov. 2021.

[12] J. T. Rosenbaum, C. R. Sibley, and J. P. Lin, "Uveitis: The pathology of inflammation of the uveal tract," in *Current Ophthalmology*, ch. 12, 2020.

[13] J. L. Shields *et al*., "Iris melanoma: A review of clinical features, diagnosis, and management," *Retina*, vol. 41, no. 6, pp. 1201–1211, Jun. 2021.

[14] V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, 2021.

[15] Y. Zhang, X. Wang, and S. Li, "Iris texture analysis using GLCM and SVM for anterior segment disease classification," *J. Biomed. Eng.*, vol. 39, pp. 78–89, 2022.

[16] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nature Med.*, vol. 28, pp. 31–38, Jan. 2022.

[17] D. S. Ting *et al*., "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, Dec. 2021.

[18] R. Smith and P. Jones, "Comparative analysis of wavelet transforms and Gabor filters in ocular texture recognition," *IEEE Trans. Med. Imag.*, vol. 42, no. 8, pp. 2101–2115, Aug. 2023.

[19] D. Williams, "EfficientNet performance variability in ophthalmic imaging: Sensitivity to curvature and illumination variation," *J. Artif. Intell. Med.*, vol. 13, no. 1, pp. 45–60, 2025.

[20] S. Minaee *et al*., "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2021.

[21] L. Chen *et al*., "Transfer learning applications in retinal and iris pathology: A systematic review," *J. AI Med.*, vol. 12, no. 2, pp. 45–60, 2024.

[22] M. Fu *et al*., "Disc-aware ensemble network for glaucoma screening from fundus image," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2588–2599, Aug. 2020.

[23] M. Ang *et al*., "Anterior segment optical coherence tomography," *Prog. Retin. Eye Res.*, vol. 66, pp. 132–156, Sep. 2021.

[24] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[25] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Hoboken, NJ, USA: Pearson, 2018.