Available online at sjuoz.uoz.edu.krd

Science Journal of University of Zakho

Vol. 13, No.4 pp 479–488October-December,2025

p-ISSN: 2663-628X
e-ISSN: 2663-6298

MULTI-CLASSIFICATION OF EYE DISEASES USING A CNN-HARALICK HYBRID FRAMEWORK

Oluwaseyi Ezekiel Olorunshola[1], Nanji Emmanuella Lakan[1,*] , Fatimah Adamu-Fika[2], Joshua Caleb Ishaya[3],

[1]Department of Computer Science,Faculty of Computing, Air Force Institute of Technology, Kaduna, Nigeria.
[2]Department of Cyber Security, Faculty of Computing, Air Force Institute of Technology, Kaduna, Nigeria.
[3]Department of International Relations, Faculty of Social and Management Science, Air Force Institute of Technology, Kaduna, Nigeria.

*Corresponding author email: lakannanji@gmail.com

Received: 24 May 2025 Accepted:20 Jul 2025 Published:2 Oct 2025 https://doi.org/10.25271/sjuoz.2025.13.4.1593

ABSTRACT:

Detecting and Classifying Ocular conditions like Diabetic Retinopathy and Cataract is critical for the early diagnosis of these eye diseases and their treatments. This study proposes a hybrid CNN-Haralick model, that leverages the lightweight MobileNetV2 CNN architecture for spatial feature extraction and Haralick statistical texture features extraction for texture analysis to improve the accuracy eye disease classification. A dual-branch architecture is modeled, which combines features extracted from both the Convolutional Neural Network and the Haralick-based texture analysis at an early stage and is then passed through the MobileNetV2 Neural Network. The model is then evaluated and the results show that the hybrid CNN-Haralick model achieves an overall accuracy of 98% on the validation set, outperforming traditional CNN models. The model demonstrates exceptional performance, with a macro average F1-score of 98% for the three classes, and AUC-ROC scores of 100% for each category. Additionally, the model's effectiveness is discussed in comparison with existing works, highlighting its superior performance in terms of both accuracy and multi-classification efficiency.

KEYWORDS: Eye Diseases, CNN, Haralick, Hybrid Model, Spatial Analysis, Texture Analysis, Dual-Branch Architecture, Multi-Classification.

1. INTRODUCTION

Beyond what is commonly known about Diabetes, diabetes mellitus is a chronic metabolic condition that also serves as a leading cause of ocular complications, distinctly Diabetic Retinopathy (DR) and Cataract, which are two of the most common and preventable causes of visual impairment and blindness (Cleveland Clinic, 2022; Shukla *et al*., 2023). These complications significantly affect the retina and lens of the eye, and if left untreated, can result in irreversible vision loss. Timely and precise detection of such retinal ailments is essential for avoiding permanent eye damage (Pratap *et al*., 2024). However, traditional diagnostic methods for retinal diseases still rely heavily on the availability and expertise of trained ophthalmologists and access to high-end imaging equipment, which may be unavailable or restricted in low-resource areas. This highlights the pressing need for autonomous, intelligent, and cost-effective screening tools that can assist healthcare professionals and expand access to early eye care.

In the 21st century, Machine Learning and Pattern Recognition have shown promise in medical imaging, with Convolutional Neural Networks (CNNs) excelling at extracting spatial features for disease classification (Mienye *et al*., 2025). Despite this, CNNs often miss subtle texture cues critical for identifying early-stage retinal abnormalities (Atcı *et al*., 2024). Such texture information, which can be essential for detecting micro-level changes in retinal images, may not be well-captured by standard CNN pipelines. To overcome this, we propose a hybrid model that integrates CNN-derived spatial features with Haralick texture descriptors—handcrafted features calculated from Gray-Level Co-occurrence Matrices (GLCMs) that capture local pixel intensity variations. By combining these two feature types, the model benefits from both global spatial structure and local textural nuance, enabling improved detection of complex changes, especially when dealing with fundus images.

This study introduces a dual-branch CNN-Haralick architecture designed for the multi-classification of retinal images. They include the DR, Cataract, and Normal classes. The model is trained on a balanced dataset of high-resolution fundus images, with over 1,000 images per class to ensure representativeness and generalizability. In contrast to earlier hybrid approaches that either apply handcrafted features independently or fuse them without optimization (Li *et al*., 2023), this model adopts a structured and tuned integration approach aimed at enhancing both classification performance and model interpretability. This careful fusion strategy ensures that the model is more sensitive to subtle visual anomalies, such as variations around the optic disc, which are often early indicators of disease.

The aim of this research is to validate a high-performing hybrid detection and classification methodology that leverages both deep and statistical features to improve diagnostic accuracy.

* Corresponding author

This approach aligns with ongoing advancements in AI-driven medical diagnostics and seeks to bridge the diagnostic gap in regions with limited clinical resources. By enhancing screening capabilities, it supports broader access to eye care and earlier intervention for individuals at risk of diabetic visual complications.

The paper is structured as follows: Section 1 presents the background of the study and an introduction of the proposed method; Section 2 is the literature review and existing research gaps; Section 3 details the proposed hybrid CNN-Haralick model; Section 4 covers the results, section 5 is the discussion and section 6 concludes with key findings, limitations, and recommendations for future work

2. .RELATED WORKS

Sushith *et al*. (2025) presented a hybrid model that puts together two known deep learning algorithms which are CNN and the Recurrent Neural Networks (RNNs) detecting DR using retinal fundus images. The Proposed model evaluation in terms of both sensitivity and specificity, achieved 97.5% accuracy on the DRIVE dataset, 94.04% on the Kaggle dataset, 96.9% on the Eyepacs Dataset.

Bidwai *et al*. (2024) presented the Dunnock-Scheduler optimization-based Light GBM (DkSO-Light GBM) for multimodal image fusion for DR detection. The experimental outcomes were evalauated and the accuracy, sensitivity, specificity, precision and F1_score, of the DkSO-Light GBM were 94.32 %, 94.94 %, 94.78 %, 94.78 % and 94.25 %, respectively. And for the k-fold 6 metrics we have 95.53 %, 94.72 %, 95.41 %, 94.16 %, 93.83 %, 95.07 %, and 92.00 %, respectively.

Babaqi *et al*. (2023) focused on classifying eye diseases including DR, cataract, and glaucoma using deep learning models. Using transfer learning, the study achieved a classification accuracy of 94%. The approach highlighted the need for accurate multi-class classification in ophthalmology.
Introducing a multi-class transfer learning approach for eye-disease classification, Bitto and Mahmud (2022) used a ResNet-50, Inception-v3, Visual Geometry Group (VGG-16) CNN models to classify between conjunctivitis eyes, normal eyes, and cataract eyes. With a detection speed of 485 seconds, the Inception-v3 model was recorded as the most accurate of all the models used for their eye disease detection approach, with an accuracy of 97.08%, ResNet-50 performed as the second-highest with 95.68% accuracy for 1090 seconds and finally, The VGG-16 performed with an accuracy of 95.48% with the longest time spent of 2510 seconds to detect eye diseases.

Ouda *et al*. (2022) introduced a multi-label classification approach for detecting different ocular diseases using retinal fundus images. The proposed model demonstrated high accuracy in classifying various eye conditions, showcasing the potential of multi-label classification in ophthalmology. The model was evaluated using Dice similarity coefficient (DSC), accuracy, precision, recall, and area under the curve (AUC). The results are 99%, 94.3%, 91.5%, 80% and 96.7%, respectively.

Sarki *et al*. (2021) developed a model using CNN for multi-classifiication of diabetic eye diseases. The study aimed to automate the diagnosis process, reducing the manual workload on ophthalmologists. The proposed model achieved a maximum accuracy of 81.33%, with both sensitivity and specificity reaching 100%, demonstrating its effectiveness in clinical settings.

Londhe (2021) proposed a hybrid CNN-RNN model for classifying eye diseases such as cataracts, glaucoma, and other retinal diseases. Utilizing transfer learning with architectures like InceptionV3, InceptionResNetV2, and DenseNet169, features were extracted and then classified using LSTM networks. Addressing the challenge of class imbalance in the dataset, data augmentation techniques were employed. The DenseNet169-LSTM model had the best performance with accuracy of 69.5%, precision of 87.4% and sensitivity of 69.5%.

Sarki *et al*. (2020) addressed the challenge of detecting diabetic eye diseases amidst the presence of mild feature difference using deep learning. Using a pretrained CNN model like VGG16, the study incorporated techniques such as fine-tuning, optimization, and contrast enhancement. The model had an accuracy of 88.3% for multi-class classification and 85.95% for mild multi-class classification, highlighting its robustness in handling varying disease severities.

Malik *et al*. (2019), introduced the development of a standardized system for collecting and processing diagnostic data, aiming to improve the accuracy and reliability of machine learning models in predicting diseases. multiple machine learning algorithms including Naive Bayes, Random Forest, Decision Tree, Random Forest, and Neural Network algorithms. The Random Forest and Decision Tree algorithms' accuracy was more than 90% as compared to just using Neural Networks and the Naïve Bayes algorithm.

Al-Bander *et al*. (2017) explored developing an automatic feature learning model for the detection of ocular conditions in colored retinal fundus images using CNN, a deep learning method. It was developed to distinguish between normal and glaucomatous patterns for diagnostic decisions. Unlike traditional methods where the optic disc features are handcrafted, feature extraction was carried out using raw images by CNN and fed to an SVM classifier to classify the images into normal or glaucomatous. The model was evaluated and had accuracy, specificity and sensitivity of 88.2%, 90.8%, and 85%.

While CNNs, transfer learning, and hybrid deep learning architectures have proven effective for classifying eye diseases, they still face limitations. Most models focus on spatial or deep semantic features extracted from retinal images, often neglecting subtle texture patterns essential for distinguishing early-stage diseases such as Cataract from Normal retina (Babaqi *et al*., 2023; Bidwai *et al*., 2024). Additionally, multi-class and multi-label classification techniques, despite improving general diagnostic accuracy, often struggle with class-specific performance, particularly in underrepresented or visually similar categories (Ouda *et al*., 2022; Sarki *et al*., 2021).

Although some hybrid models, such as CNN-RNN and CNN-SVM, have been proposed, few effectively combine handcrafted statistical texture features, like Haralick descriptors, with deep features in a unified architecture. Most existing approaches either use these features independently. Therefore, there is a pressing need for an interpretable, multi-class classification system that seamlessly integrates spatial deep features with robust statistical texture representations. Combining Haralick features with CNN-derived features could improve diagnostic accuracy and enhance predictability, particularly for texture-dependent classes.

3. METHODOLOGY

Research Design:

The model being proposed had its development and training conducted on a Jupyter Notebook environment powered by an Intel Core i7 processor with 16 GB of RAM. This research is structured into four sequential phases: preprocessing of retinal images to standardize inputs, extraction of Haralick texture features to capture statistical regularities, deep feature learning via fine-tuned MobileNetV2 for semantic representation, and fusion of both feature types through a dual-branch architecture, followed by joint training and classification.This layered design combines low-level statistical regularities and high-level deep features to enhance diagnostic precision for retinal diseases. This architecture leverages both domain-agnostic statistical descriptors and hierarchical deep features for enhanced diagnostic performance.

Dataset Preparation and Preprocessing:

The model is trained and evaluated on a balanced dataset comprising high-resolution fundus images grouped into three diagnostic categories: Diabetic Retinopathy, Cataract, and Normal. These images were sourced from publicly available repositories, including the Indian Diabetic Retinopathy Image Dataset (IDRiD), the Ocular Disease Recognition dataset, and the High-Resolution Fundus (HRF) database. The IDRiD dataset, curated specifically for diabetic eye disease detection, includes both pathological and normal fundus images from Indian patients, capturing real-world variability. The Ocular Disease Recognition dataset provides a broad range of labeled eye conditions, while the HRF database offers high-resolution clinical images aimed at retinal vessel segmentation and anomaly detection. Each diagnostic class comprises over 1,000 images, ensuring class balance and promoting robust generalization across varied imaging conditions.

All images were uniformly resized to 128×128 pixels and standardized using the preprocess_input function from MobileNetV2 to align with the model's expected input distribution. This preprocessing step also normalized color and contrast values across all samples, thereby reducing domain-specific biases. Stratified sampling was applied to split the dataset into training (70%), validation (15%), and test (15%) subsets, ensuring proportional class representation throughout.
MobileNetV2, used as the CNN backbone, employs depthwise separable convolutions with 3×3 kernels and strides of 1 or 2, designed for computational efficiency. Its convolutional blocks progressively downsample the input while increasing filter depth from 32 to 1,280. The resulting feature maps are condensed using a Global Average Pooling layer, followed by a dense layer with 128 units activated by the Swish function. The final classification head is a fully connected softmax layer with three output neurons, corresponding to the Diabetic Retinopathy, Cataract, and Normal classes.

Model Architecture:

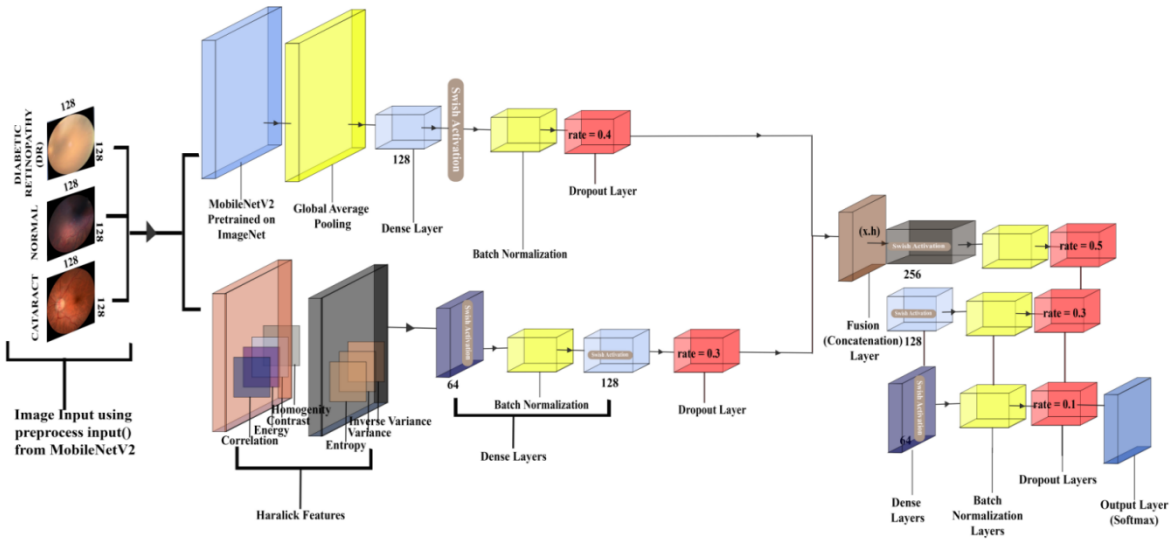

Figure 1: CNN-Haralick Architecture for the Multi-Classification of Eye Diseases

The model's architecture represented in Figure 1 above is divided into two parallel branches for features extraction, The CNN Feature extraction Branch and the Haralick Feature Branch which are later fused to make a final prediction and classification.

CNN Feature Extraction Branch:

The CNN branch receives images with input shape (128, 128, 3) and utilizes a pretrained MobileNetV2 base. The initial layers (up to layer 100) are frozen to retain general-purpose low-level features, while the final 20 layers are unfrozen for domain-specific fine-tuning. After convolutional processing, a Global Average Pooling (GAP) layer compresses the feature maps into a 1D representation, followed by a dense layer of 128 units with the Swish activation function. Batch normalization is applied, and a dropout layer with a rate of 0.4 prevents overfitting. This branch effectively extracts vascular patterns, optic disc anomalies, and macular irregularities commonly found in retinal disease.

Haralick Feature Branch:

The second branch operates on grayscale-converted fundus images, which are rescaled to 256 intensity levels. Gray-Level Co-occurrence Matrices (GLCMs) are computed across four orientations (0°, 45°, 90°, and 135°) at multiple pixel distances (1, 2, 4, and 8). From these matrices, thirteen Haralick features are derived, of which seven are retained based on their clinical

and statistical relevance: contrast, correlation, homogeneity, energy, variance, inverse variance, and entropy. These features were selected for their ability to capture critical retinal texture patterns. For example, entropy quantifies randomness in pixel distribution and can indicate localized retinal degeneration or hemorrhages, while correlation measures the linear dependency of gray levels, helping identify structured patterns such as aligned blood vessels in healthy retinas. After extraction, the features are standardized and passed into a dedicated subnetwork comprising two fully connected layers with 64 and 128 units respectively, both using the Swish activation function. To improve generalization, batch normalization is applied between layers, and a dropout layer with a rate of 0.3 is added for regularization.

Haralick Texture Features:

Contrast:

An interaction between a pixel and its neighbors which often causes the intensity:

$$Contrast = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2 \cdot P(i,j) \tag{1}$$

Entropy:

measures the level of randomness among pixels within the image:

$$Entropy = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P(i,j) \cdot \log_2(P(i,j)+\epsilon\ ) \tag{2}$$

Energy:

The sum of the squared elements within the Matrix, with its ranging up to 1.

$$Energy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P(i,j)^2 \tag{3}$$

Correlation:

Indicates the extent to which a pixel's neighboring values are correlated with the overall structure of the image.

$$Correlation = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j} \tag{4}$$

Homogeneity:

reflects the uniformity of the texture. As the GLCM elements shift further from the diagonal, the homogeneity value increases in a geometric manner.

$$\text{Homogeneity} = \sum_{i,j}\frac{P(i,j)}{1+|i-j|} \tag{5}$$

Variance: Represents the degree to which intensity values vary or deviate from the mean.

$$\text{Variance} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left[(i-\mu_i)^2+(j-\mu_j)^2\right]P(i,j) \tag{6}$$

Inverse Variance:

Highlights elements close to the diagonal in the GLCM, providing a precise assessment of texture similarity.

$$\text{Inverse Variance} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{P(i,j)}{1+(i-j)^2} \tag{7}$$

Where:

P(i,j) = Probability value at position (i,j) in the GLCM.

N = Number of gray levels in the image.
= row i and column j.
σ_i, σ_j = Standard deviations of row i and column j.
ϵ is a small value added to prevent log(0).

Feature Fusion and Classification Layer:

The outputs from both branches—each of shape (128,)—are concatenated to form a 256-dimensional feature vector. This combined vector is passed through two dense layers of 256 and 128 units, each followed by batch normalization and dropout layers (rates = 0.5 and 0.3 respectively). The final classification layer consists of 3 output units corresponding to Diabetic Retinopathy, Cataract, and Normal classes, using a Softmax activation function to generate class probabilities. This configuration is appropriate for mutually exclusive multi-class classification tasks. The model is compiled using categorical cross-entropy as the loss function and the AdamW optimizer with an initial learning rate of 0.01. Learning rate scheduling is used to reduce the rate when the validation loss plateaus, improving convergence.

Training Protocol:

The training protocol includes 30 epochs with early stopping (patience = 5) and real-time data augmentation to prevent overfitting. A batch size of 32 is used, and input images are consistently formatted to dimensions (128, 128, 3). The dropout rates across different layers were not selected arbitrarily but were the result of empirical tuning through iterative experimentation using validation accuracy and loss as performance indicators. A higher dropout 0.5 was assigned to the fusion layer to mitigate overfitting from the high-dimensional joint feature vector. The Haralick and CNN branches used 0.3 and 0.4, respectively, as these provided the best trade-off between stability and regularization during trials. The 0.1 dropout in the final dense layer was retained after testing values between 0.0 and 0.3, as it preserved gradient flow without compromising convergence. Its minor regularization was sufficient due to upstream regularization layers and batch normalization.

Integration of Techniques and Novelty:

The key novelty of this approach lies in the synergistic fusion of deep and statistical features within a unified learning framework:

The CNN branch abstracts complex spatial relationships, capturing semantic cues like exudates and hemorrhages.

The Haralick branch quantifies micro-textural variations, which are often early indicators of diseases like cataract.

Instead of ensemble voting or post-hoc feature concatenation, an early fusion strategy is employed in a multi-input neural network, allowing the model to learn joint feature representations in an end-to-end fashion.

The inclusion of interpretable Haralick descriptors contributes to transparency and aligns with clinical reasoning, thus improving the model's trustworthiness.

The synergy between CNN and Haralick features boosts performance in distinguishing visually similar classes (e.g., Cataract vs. Normal), a known limitation in conventional CNNs.

Evaluation of the Model:

The proposed model was evaluated on a separate validation set using metrics such as accuracy, precision, recall and F1-score. To validate the effectiveness of the hybrid approach. The performance metrics are:

Accuracy:

This refers to the proportion of correctly predicted instances, TP+TN, to the total number of samples, TP+TN+FP+FN, in the dataset:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Recall (Sensitivity):

The proportion of true positive predictions relative to the total number of instances predicted as positive:

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Precision:

The proportion of correctly predicted instances within all instances classified under that specific class:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

F1-Score:

The harmonic average of precision and recall, for reducing the impact of large disparities between them.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{11}$$

Where:

TP is True Positive
TN is True Negative
FP is False Positive
FN is False Negative.

4. RESULTS AND DISCUSSIONS

In this section, we present the results obtained from the proposed hybrid CNN-Haralick model for eye disease classification. The performance of the model is evaluated using various metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. We also analyze the contribution of Haralick features to the model's performance and compare the results with existing methods to demonstrate the effectiveness of our approach.

Confusion Matrix Analysis:

The confusion matrix (Table 1 and Figure 2) provides insight into the model's classification performance across the three classes—Cataract, DR, and Normal. The model achieved nearly perfect classification for DR and high performance for Cataract and Normal classes.

Table 1: Confusion Matrix for the predicted classes

Class	Cataract	DR	Normal
Cataract	154	0	2
DR	0	165	0
Normal	7	1	153

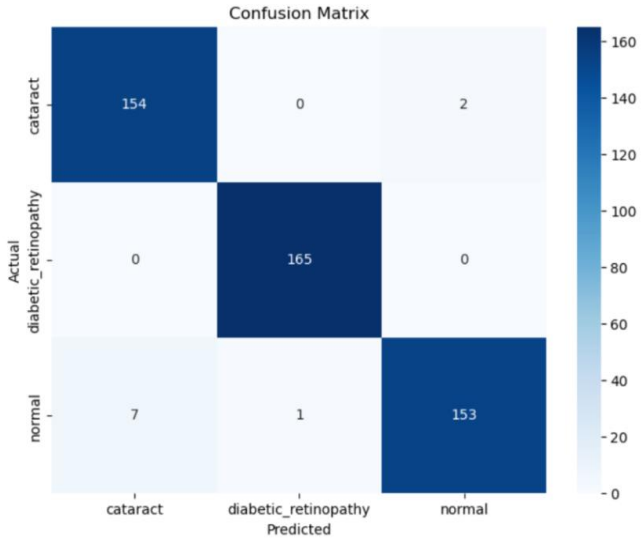

Figure 2: Confusion matrix illustrating classification results on the validation set.

The model accurately classified 165 images of DR with no false positives or negatives, reflecting strong discriminative capability in detecting this class. For Cataract, only 2 Normal images were misclassified, while 7 Normal samples were wrongly predicted as Cataract, revealing minor overlap in feature representation between these classes.

Classification Metrics:

The classification report in Table 2 presents the F1-score, recall, precision, and AUC-ROC for each class. Figure 3a is a screenshot of the classification report which includes the weighted and macro average of each of the metrics for all classses, highlighting the model's strong accuracy across the eye disease categories.

Figure 3b displays the AUC-ROC curve, demonstrating the model's ability to distinguish between the classes. The high AUC values indicate reliable performance in multi-class classification, effectively identifying subtle differences between similar eye conditions.

Table 2: Classification Report of the CNN-Haralick Model Performance on the Validation Set.

Class	F1-Score	Recall	Precision	AUC-ROC
Cataract	97%	99%	96%	100%
DR	100%	100%	99%	100%
Normal	97%	97%	99%	100%

```
✅ Test Accuracy: 0.9793
16/16 ━━━━━━━━━━━━━━━━━━━━ 5s 263ms/step

Classification Report:
                    precision    recall  f1-score   support

          cataract       0.96      0.99      0.97       156
diabetic_retinopathy       0.99      1.00      1.00       165
            normal       0.99      0.95      0.97       161

          accuracy                           0.98       482
         macro avg       0.98      0.98      0.98       482
      weighted avg       0.98      0.98      0.98       482
```

Figure 3a: Classification Report Visualizing Precision, Recall, and F1-Score.

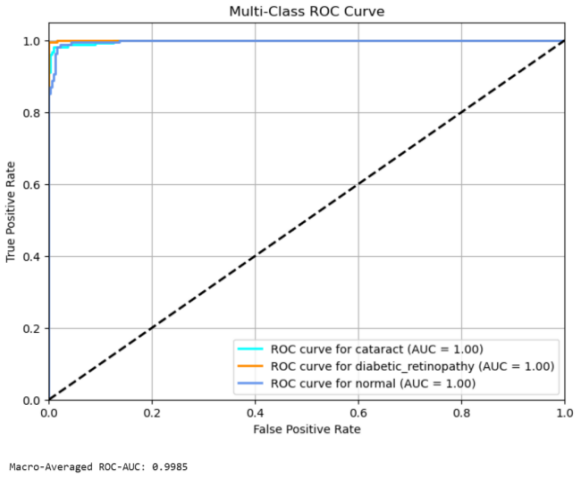

Figure 3b: ROC-AUC curve for the 3 classes.

The model achieved an impressive overall accuracy of 98%. Notably, the AUC-ROC of 100% for each of the classes demonstrates the model's outstanding ability to distinguish nuanced patterns. This perfect score though seemingly questionable is supported by the balanced precision and recall that each class got showing that the model doesn't struggle with predicting accurately for each of the classes. This also indicates that the model consistently ranks positive cases higher than negative ones, regardless of the disease type. The model's ability reflects how well it leverages both spatial and texture features, captures subtle patterns that helps it differentiate between the different eye conditions with high precision and Sensitivity.

Training and Validation Dynamics:

The Training and Validation Accuracy/Loss curve over 30 epochs are shown below in Figures 4a and 4b. The graphs show a gradual and then instant convergence with no signs of overfitting, further validating the model's robustness and generalization capability.

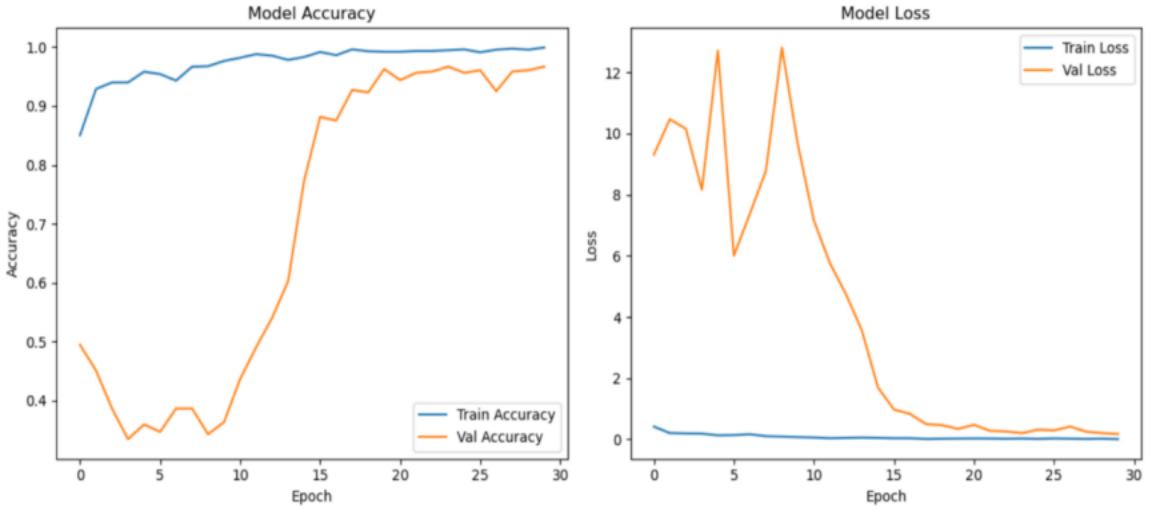

Figure 4: Performance (a) training vs validation accuracy and (b) training vs validation loss for the proposed model

The minimal gap between training and validation accuracy (less than 2%) and the consistent decline in validation loss confirm that the model effectively learns meaningful features without overfitting, even with the hybrid fusion of CNN and Haralick features.

5. DISCUSSION

The proposed CNN-Haralick hybrid model leverages two distinct types of features: spatial features from the CNN branch and textural features from the Haralick descriptors. Combining These features to be used for the classification of multiple eye diseases has achieved an outstanding classification ability for eye diseases, with results that surpass that of benchmark models that rely on either CNN or texture-based features alone.

MobileNetV2, a lightweight CNN architecture pretrained on ImageNet, was employed to extract deep spatial features. The model's ability to recognize high-level subtleties—such as vascular anomalies, lesions, and other structural abnormalities—was important and a key feature in distinguishing between Cataract, DR, and Normal conditions. The inclusion of a pretrained backbone enables the model to retain generalizable low-level features, which are then fine-tuned for the specific task of retinal disease classification. The use of global average pooling (GAP) further ensures that the model captures global image features while reducing the risk of overfitting.

Incorporating Haralick texture features into the model significantly enhances its performance. Haralick features capture microscopic textural variations in the images that may not be easily discernible through spatial CNN filters alone. For example, Contrast and Homogeneity offer insights into the uniformity of pixel intensities and patterns of texture that may be early indicators of diseases like Cataract. Energy and Entropy provide additional measures of textural regularity and information randomness, which are critical for identifying conditions that alter the fine structure of the retina.

The fusion of these texture features with the deep features from the CNN model enables the network to capture both high-level semantic cues (e.g., lesions, vascular patterns) and fine-grained details (e.g., textural irregularities) simultaneously. This dual-branch architecture results in a model that can better handle the inter-class variability and intra-class subtle differences that often challenge conventional CNN-only models.

Comparison with Existing Models:

When compared to existing state-of-the-art approaches for retinal disease classification, the CNN-Haralick hybrid model stands out in several key aspects:

Many existing models rely solely on CNNs for retinal disease detection. While CNNs excel at capturing global patterns, they often struggle with fine-grained texture details that are crucial for differentiating between diseases like Cataract and Normal. The proposed model addresses this by incorporating Haralick features, which offer an additional layer of textural sensitivity. Previous studies that used pure CNN models typically report accuracy in the range of 85%-93%, with misclassification between visually similar conditions such as Cataract and Normal being common. Table 3 shows the comparison of the proposed hybrid model with other state of the art models used in classification of animal diseases.

Table 3: Comparison of Proposed Work with existing works

Author(s) and Year	Proposed Methodology	Diseases Focused On	Result
Sushith *et al*. (2025)	A hybrid deep learning framework combining CNN and RNN	Diabetic Retinopathy	The model Achieved 97.5% accuracy on the DRIVE dataset, 94.04% on the Kaggle dataset, 96.9% on the Eyepacs Dataset.
Bidwai *et al*. (2024)	DkSO-Light GBM for multimodal image fusion (ResNet 101 + DkSO + Light GBM)	Diabetic Retinopathy	Accuracy: 94.32%, Sensitivity: 94.94%, Specificity: 94.78%, Precision: 94.78%, F1-Score: 94.25%, MCC: 91.77% (TP=90%)
Babaqi *et al*. (2023)	Transfer learning with deep learning models	Diabetic Retinopathy, Cataract, Glaucoma	Accuracy: 94%, CNN Models (Traditional: 84%)
Bitto and Mahmud (2022)	Multi-categorical transfer learning (VGG-16, ResNet-50, Inception-v3)	Normal, Conjunctivitis, Cataract	Accuracy: Inception-v3: 97.08%, ResNet-50: 95.68%, VGG-16: 95.48%
Ouda *et al*. (2022)	Multi-label deep learning classification (fundus images)	Multiple Ocular Diseases	Accuracy: 94.3%, Recall: 80%, Precision: 91.5%, Dice Similarity: 99%, AUC: 96.7%
Sarki *et al*. (2021)	Convolutional Neural Network (CNN) for multi-class classification	Diabetic Eye Diseases	Accuracy: 81.33%, Sensitivity: 100%, Specificity: 100%
Londhe (2021)	Hybrid CNN-RNN model with transfer learning (InceptionV3, InceptionResNetV2, DenseNet169 + LSTM)	Cataracts, Glaucoma, Retinal Diseases	Accuracy: 69.5%, Specificity: 87.4%, Sensitivity: 69.5%
Sarki *et al*. (2020)	Pretrained CNN models (VGG16), fine-tuning, and optimization	Diabetic Eye Diseases	Accuracy: 88.3%, Multi-Class Accuracy: 85.95%
Proposed Model	CNN-Haralick Hybrid Framework	Diabetic Retinopathy, Cataract	Accuracy: 98% Recall: 98% Precision: 98% AUC-ROC: 100%

Model Robustness and Generalization:

The model's high performance on the validation set, with an accuracy of 98% and F1-scores ranging from 97% to 100%, indicates its ability to generalize across different types of eye diseases. The AUC-ROC score of 100% for all three classes further demonstrates that the model maintains high sensitivity and specificity, crucial for clinical applications where false negatives (missed diagnoses) can be life-threatening. This robustness is attributed to several factors: The integration of deep spatial features from MobileNetV2 with handcrafted Haralick texture features, the use of balanced training data through class weighting to address potential class imbalance, the application of data augmentation to enhance model generalization, and the fine-tuning of the CNN backbone to adapt to subtle differences among the disease classes. Additionally, the combination of batch normalization, dropout, and the AdamW optimizer helped maintain model stability and prevented overfitting, leading to consistent performance across the validation set.

CONCLUSION

Conclusion:

This study introduced a novel hybrid CNN-Haralick model for multi-class eye disease classification, achieving superior performance over existing models by combining the strengths of spatial deep features from CNNs and textural features from Haralick descriptors. The model demonstrated 98% accuracy and outperformed traditional CNN-only approaches in detecting diseases like Cataract and DR. The results indicate the potential of the hybrid model to contribute meaningfully to early disease detection and clinical decision-making.

Limitations:

However, potential limitations include differences in imaging equipment, population-specific characteristics (such as ethnic variations in retinal appearance), and possible label inconsistencies across datasets. These factors may affect the generalizability of the model when applied to other fundus datasets or real-world screening environments. Although, the use of multiple data sources enhances the model's robustness across multiple scenarios. In addition to dataset-related limitations, the model faces challenges related to feature fusion and

hyperparameter tuning. Specifically, integrating handcrafted Haralick features with deep features required careful normalization and dimensionality alignment to avoid redundancy and overfitting. Training the dual-branch architecture also posed computational constraints, particularly during early experiments where the model showed unstable convergence across certain folds. Moreover, the limited interpretability of deep learning decisions, even with statistical feature fusion, remains a concern for real-world clinical trust and acceptance. These technical and methodological challenges highlight the complexity of developing reliable hybrid diagnostic tools.

Recommendations:

To further enhance the model's capabilities, the following recommendations are made:

Multimodal Data Integration: Future work should explore the integration of additional modalities, such as optical coherence tomography (OCT) images or patient demographics, to enrich feature extraction and improve diagnostic accuracy for complex cases.

Real-time Inference: Efforts should be directed toward optimizing the model for real-time inference in clinical settings, ensuring that it can be used in resource-limited environments without sacrificing accuracy.

Fine-Grained Diagnosis: The model could be expanded to include more granular classifications, such as subtypes of DR or the early stages of Glaucoma, which are often difficult to detect with standard methods.

Model Explainability: Incorporating explainability frameworks such as SHAP or Grad-CAM will help visualize and interpret predictions, thereby increasing clinician trust and regulatory acceptability.

Cross-dataset Evaluation: Future work should involve rigorous testing across geographically and demographically diverse datasets to further assess model robustness and improve generalizability.

DECLARATIONS

Ethical approval:

Does not apply.

Publication consent:

Does not apply.

Data availability:

The data used in this study are available from the corresponding author on request.

Conflicts of interest:

The authors have no conflicts of interest to declare.

REFERENCES

Al-Bander, B., Al-Nuaimy, W., Al-Taee, M., & Zheng, Y. (2017). Automated glaucoma diagnosis using deep learning approach. *2017 14th International Multi-Conference on Systems, Signals & Devices (SSD)*. https://doi.org/10.1109/SSD.2017.8166974

Bitto, A. K., & Mahmud, I. (2022). Multi categorical of common eye disease detect using convolutional neural network: a transfer learning approach. *Bulletin of Electrical Engineering and Informatics*, *11*(4), 2378–2387. https://doi.org/10.11591/eei.v11i4.3834

Cleveland Clinic. (2022, November 14). *Blindness (Vision Impairment): Types, Causes and Treatment*. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/24446-blindness

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A Guide to Deep Learning in Healthcare. *Nature Medicine*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Flaxman, S. R., Bourne, R. R. A., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T. Y., Taylor, H. R., & Bourne, R. (2017). Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*, *5*(12), e1221–e1234. https://doi.org/10.1016/s2214-109x(17)30393-5

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973a). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/tsmc.1973.4309314

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973b). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/tsmc.1973.4309314

Ibomoiye Domor Mienye, Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep Convolutional Neural Networks in Medical Image Analysis: A Review. *Information*, *16*(3), 195. https://doi.org/10.3390/info16030195

Kiziltoprak, H., Tekin, K., Inanc, M., & Goker, Y. S. (2019). Cataract in diabetes mellitus. *World Journal of Diabetes*, *10*(3), 140–153. https://doi.org/10.4239/wjd.v10.i3.140

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. Neural Information Processing Systems; Curran Associates, Inc.

Londhe, M. (n.d.). *Classification of Eye Diseases using Hybrid CNN-RNN Models MSc Research Project Data Analytics*. Retrieved October 21, 2022, from

Malik, S., Kanwal, N., Asghar, M. N., Sadiq, M. A. A., Karamat, I., & Fleury, M. (2019). Data Driven Approach for Eye Disease Classification with Machine Learning. *Applied Sciences*, *9*(14), 2789. https://doi.org/10.3390/app9142789

Mateen, M., Wen, J., Hassan, M., Nasrullah, N., Sun, S., & Hayat, S. (2020, March 11). *Automatic Detection of Diabetic Retinopathy: A Review on Datasets, Methods and Evaluation Metrics | IEEE Journals & Magazine | IEEE Xplore*. Ieeexplore.ieee.org.

Ouda, O., AbdelMaksoud, E., Abd El-Aziz, A. A., & Elmogy, M. (2022). Multiple Ocular Disease Diagnosis Using Fundus Images Based on Multi-Label Deep Learning Classification. *Electronics*, *11*(13), 1966. https://doi.org/10.3390/electronics11131966

Pooja Bidwai, Shilpa Gite, Pahuja, N., Kishore Pahuja, Kotecha, K., Jain, N., & Sheela Ramanna. (2024). Multimodal Image Fusion for the Detection of Diabetic Retinopathy Using Optimized Explainable AI-based Light GBM Classifier. *Information Fusion*, *111*, 102526–102526. https://doi.org/10.1016/j.inffus.2024.102526

Pratap, U., Surico, P. L., Singh, R. B., Romano, F., Salati, C., Spadea, L., Musa, M., Gagliano, C., Mori, T., & Zeppieri, M. (2024). Artificial Intelligence (AI) for Early Diagnosis of Retinal Diseases. *Medicina-Lithuania*, *60*(4), 527–527. https://doi.org/10.3390/medicina60040527

Sarki, R., Ahmed, K., Wang, H., & Zhang, Y. (2020). Automated detection of mild and multi-class diabetic eye diseases using deep learning. *Health Information Science and Systems*, *8*(1). https://doi.org/10.1007/s13755-020-00125-5

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2021). VU Research Repository. *VU Research Repository | Victoria University| Melbourne Australia*.

Shukla, U. V., & Tripathy, K. (2023a). *Diabetic Retinopathy*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK560805/

Shukla, U. V., & Tripathy, K. (2023b). *Diabetic Retinopathy*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK560805/

Şükran Yaman Atcı, Güneş, A., Metin Zontul, & Arslan, Z. (2024). Identifying Diabetic Retinopathy in the Human Eye: A Hybrid Approach Based on a Computer-Aided Diagnosis System Combined with Deep Learning. *Tomography*, *10*(2), 215–230. https://doi.org/10.3390/tomography10020017

Sushith, M., Sathiya, A., Kalaipoonguzhali, V., & Sathya, V. (2025). A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-99309-w