International Journal of Health Sciences and Research
[Indexed, Peer-Reviewed Journal]
Volume 15; Issue: 7; July 2025
Website: www.ijhsr.org
ISSN: 2249-9571

Original Research Article

Deep Learning-Based Classification of Retinal Diseases from OCT Images with LLM-Powered Patient Query Support

Sheik Imran[1], Annapurna L U[2], Divya M S[3], Tanishka Sharma[4], Tayyaba Khanum[5]

[1]Asst.Prof. Department of IS&E, BIET, Davangere, Karnataka, India.
[2,3,4,5]Students, Department of IS&E, BIET, Davangere, Karnataka, India.

Corresponding Author: Sheik Imran

DOI: https://doi.org/10.52403/ijhsr.20250738

ABSTRACT

Retinal diseases such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen often progress silently and may lead to permanent vision loss if not diagnosed early. Manual analysis of Optical Coherence Tomography (OCT) images is time-consuming and prone to human error, highlighting the need for automated diagnostic tools. This study proposes a deep learning-based platform using MobileNetV3, a lightweight CNN architecture, for the multi-class classification of retinal diseases. The model is trained on a publicly available Kaggle OCT dataset comprising 84,495 labeled images across four categories: CNV, DME, Drusen, and Normal. Extensive hyperparameter tuning was conducted by varying batch sizes (8, 16, 32) and optimizers (Adam, RMSProp, SGD). The optimal configuration—batch size 16 with Adam—achieved a test accuracy of 98.7%. Model performance was evaluated using accuracy, sensitivity, specificity, F1-score, and ROC-AUC. The results validate the model's robustness and its potential for scalable, accurate, and real-time retinal disease detection in clinical practice. To enhance user interaction and accessibility, a Large Language Model (LLM) was integrated into the system to handle patient and practitioner queries related to disease understanding, symptoms, risk factors, and treatment options.

Keywords: Optical Coherence Tomography (OCT), MobileNetV3, Retinal Disease Classification, Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, Deep Learning in Ophthalmology, Large Language Model (LLM).

INTRODUCTION

Retinal diseases such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen are among the leading causes of irreversible vision loss globally. According to the World Health Organization, over 2.2 billion people suffer from vision impairment, with a significant portion attributed to untreated or undiagnosed retinal disorders [1]. In India, diabetic retinopathy alone affects over 18% of known diabetic patients, posing a growing public health burden as the diabetic population continues to rise [2]. The asymptomatic nature of early-stage retinal diseases makes timely diagnosis challenging, especially in rural and underserved areas lacking access to trained ophthalmologists [3]. While several CNN-based studies have demonstrated promise,

many are limited to binary classification or focus on fundus images rather than OCT scans [4]. Moreover, crucial training hyperparameters, batch size and optimizer type remain underexplored in many works, despite their significant influence on model performance [5].
This research proposes a lightweight CNN framework based on MobileNetV3, trained on the Kaggle OCT dataset containing over 84,000 labeled images in four classes: CNV, DME, Drusen, and Normal. We systematically evaluate the effect of varying batch sizes (8, 16, 32) and optimizers (Adam, RMSProp, SGD). The best configuration (batch size 16 with Adam) achieves 98.7% accuracy. We further integrate this model into a real-time Streamlit-based web platform, offering multilingual support, AI-generated medical insights via Gemini, and voice-assisted feedback, designed to be deployable in both global and Indian clinical settings.
In this study, the deep learning classification of OCT images into Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal is performed using a lightweight and efficient Convolutional Neural Network (CNN) architecture—MobileNetV3. This model is selected due to its optimal trade-off between computational efficiency and classification accuracy, particularly suitable for real-time medical applications and edge deployment. MobileNetV3 combines elements from MobileNetV2, EfficientNet, and NAS (Neural Architecture Search). It integrates depthwise separable convolutions and squeeze-and-excitation (SE) modules along with the Hard-Swish activation function, which improves non-linearity while maintaining computational speed. Additionally, MobileNetV3 uses inverted residual blocks similar to those found in EfficientNet-B0, allowing for efficient channel-wise feature reuse and reducing the number of parameters without compromising model expressiveness.

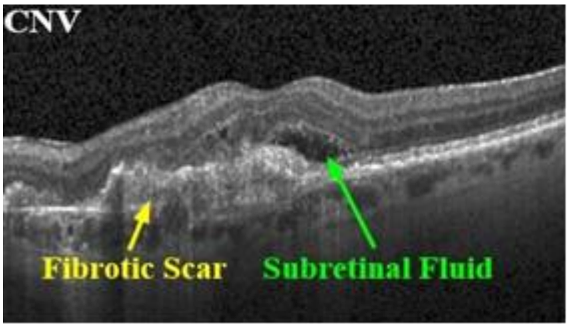

Figure 1. Choroidal Neovascularization

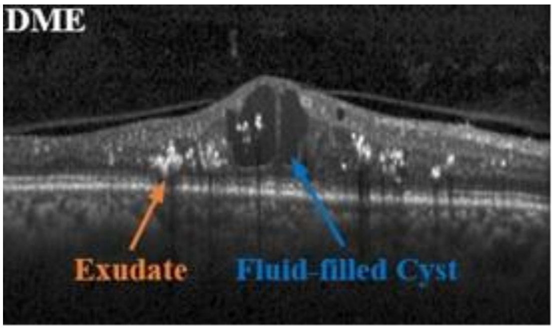

Figure 2. Diabetic Macular Edema

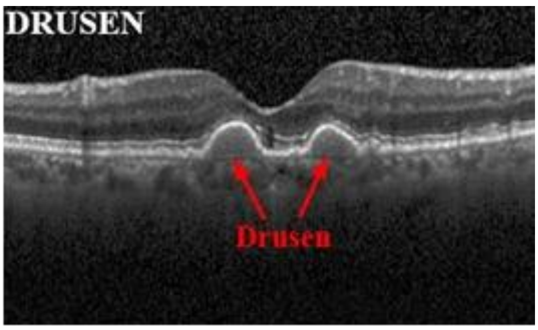

Figure 3. Drusen

Choroidal Neovascularization (CNV): It occurs when abnormal blood vessels grow beneath the retina and the macula. These vessels can leak blood and fluid, leading to vision distortion or loss. CNV is commonly associated with wet Age-related Macular

Degeneration (AMD). If untreated, it can cause severe and permanent central vision loss.

Diabetic Macular Edema (DME): DME is a complication of diabetic retinopathy, where fluid accumulates in the macula due to leaking blood vessels caused by high blood sugar levels. This swelling leads to blurred or distorted vision and can progress to blindness if not managed properly. DME is one of the leading causes of vision loss in diabetic patients.

Drusen: Drusen are yellow deposits that form under the retina, often associated with early stages of Age-related Macular Degeneration (AMD). While small drusen may not cause vision problems, larger or numerous drusen can be a sign of retinal damage and may increase the risk of progressing to advanced AMD, leading to vision loss.

Large Language Models (LLMs): such as GPT-based architectures, have revolutionized the way medical information is accessed and delivered by enabling intelligent, conversational interactions. In the context of ophthalmology, LLMs can serve as interactive tools to assist patients, caregivers, and clinicians by providing instant, reliable responses to queries related to eye diseases such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and other retinal conditions. By understanding natural language inputs, these models can explain complex medical terms, describe symptoms, suggest when to seek medical care, and clarify treatment options—bridging the gap between technical diagnostic outputs and user-friendly health communication. When integrated with automated diagnostic systems, LLMs enhance usability, patient education, and clinical decision support, contributing to a more informed and engaged healthcare experience.

BACKGROUND STUDY

The increasing prevalence of retinal diseases such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen has prompted significant research into AI-assisted diagnostic methods. Numerous deep learning models have been proposed for classifying Optical Coherence Tomography (OCT) images; however, several limitations still persist in terms of classification scope, deployment feasibility, and optimization strategies. A CNN-based approach presented by researchers in Frontiers in Computer Science (2023) [6] used a balanced dataset of 35,468 OCT images across CNV, DME, Drusen, and Normal classes, achieving high classification performance. However, their model lacks exploration of real-time deployment capabilities. A more recent study in Discover Applied Sciences (2025) [7] employed a modified VGG16 architecture with Grad-CAM for improved interpretability. Although effective, the model's complexity may hinder lightweight application in low-resource settings.

Transformer-based models such as T2T-ViT and Mobile-ViT were explored in *Lasers in Medical Science* (2024) [8], with Mobile-ViT achieving 99.17% accuracy. Despite this, the resource-intensive nature of transformers can limit practical deployment. A PubMed-indexed study (2023) [9] utilized smaller dataset of 1,528 OCT images for four-class classification and reported 98.7% accuracy, but was constrained by dataset scale and lacked a clinical interface. Additional research from *IRO Journals* (2024) [10] highlighted the role of preprocessing, using CNNs with batch normalization, yet lacked a deployment platform. Similarly, a hybrid CNN system proposed by *IIETA* (2024) [11] achieved high sensitivity for CNV, DME, and Drusen but did not include the Normal class.

Unlike previous work, the present study uses MobileNetV3—a lightweight, real-time deployable model—trained on a large OCT dataset (84,495 images) with comprehensive hyperparameter optimization. It further integrates a Streamlit-based diagnostic platform featuring Gemini AI-generated insights, multilingual support, and voice assistance, thus addressing both accuracy and accessibility in retinal disease screening.

METHODOLOGY

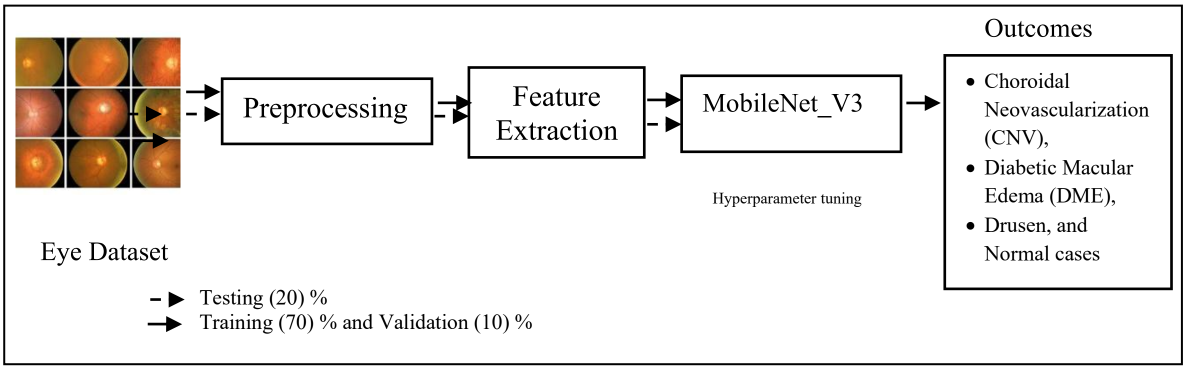

Fig 4. The proposed methodology for the classification of retinal diseases using MobileNetV3

In this study, an optimized deep learning model based on MobileNetV3 is proposed for the multiclass classification of retinal diseases, specifically Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal cases, using OCT (Optical Coherence Tomography) images. The overall workflow is depicted in Figure 4, illustrating the model training pipeline, optimization strategy, and deployment framework. We employ transfer learning, wherein the MobileNetV3 model—pretrained on the ImageNet dataset—is fine-tuned for the classification of retinal diseases. Transfer learning enables efficient training by leveraging prior knowledge from large-scale datasets, thus minimizing computational costs and improving generalization, especially when working with specialized medical images. The pretrained MobileNetV3 is adapted by replacing its final fully connected layer with a custom classification head suited for four output classes.

Data Set Acquisition:

In this research, we utilize the Labeled Optical Coherence Tomography (OCT) dataset publicly available on the Kaggle repository (Kaggle, 2021) [12] . This dataset comprises 84,495 grayscale OCT images, categorized into four distinct retinal conditions: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal. The dataset was originally compiled by Kermany et al. (2018) [13]and has become a widely recognized benchmark for evaluating deep learning models in retinal disease classification.

Authors	Year	Method	Accuracy
Jaimes *et. al.*	2025	VGG16 + Grad-CAM	Improved interpretability with an accuracy of 95.19%
Akça S *et. al.*	2024	T2T-ViT, Mobile-ViT	Mobile-ViT: 99.17%
Babu *et. al.*	2024	CNN with BN	Better performance with preprocessing
Tuncer *et. al.*	2024	Hybrid CNN	High sensitivity/specificity
Riazi Esfahani P *et. al.*	2023	CNN	Accuracy: 98.7%
Elkholy M *et. al.*	2023	CNN	High accuracy 98%

Table 1: Comparative Review of Recent Studies on OCT-Based Retinal Disease Classification

Data Preprocessing and Augmentation:

The images were captured using clinical OCT devices and vary in resolution; for this work, all images were uniformly resized to 224×224 pixels to align with the input requirements of the MobileNetV3 model.

The dataset is well-balanced across the four categories, ensuring minimal class bias during model training and evaluation. Table 1 provides the class-wise distribution of the image samples.

Data augmentation synthetically increases the diversity of the training dataset and simulates real-world variability in patient data, thereby improving the model's generalizability. All preprocessing steps were implemented using the TensorFlow/Keras data pipeline, allowing real-time augmentation during training. These transformations ensured the input to the network was not only standardized but also representative of the wide variations found in real clinical scenarios. By incorporating these preprocessing techniques, the model was better equipped to extract robust and meaningful features from OCT scans, ultimately contributing to improved classification accuracy and clinical reliability.

Model Training and Testing

For training, the dataset—comprising 84,495 labeled OCT images—was split into 80% training and 20% testing sets. Additionally, 20% of the training data was set aside as a validation set to monitor model performance during training. This results in approximately 67,596 images for training and 16,899 for testing, with further division into 54,076 training images and 13,520 validation images.

To mitigate the risks of data leakage and overfitting, we employed two cross-validation strategies:

- Stratified K-Fold Cross-Validation (k=5) was used during model training to preserve class distribution in each fold.
- Leave-One-Out Cross-Validation (LOOCV) was applied during hyperparameter tuning to assess model stability and generalization performance on varied subsets.

All images underwent preprocessing steps, including normalization, resizing, and data augmentation techniques (random rotation, horizontal flipping) to enhance model robustness and prevent overfitting.

Due to its scale, diversity, and clinical relevance, this dataset is ideally suited for training deep learning models for retinal disease detection. The effectiveness of our trained model on this dataset underscores its practical value for real-time ophthalmic screening systems.

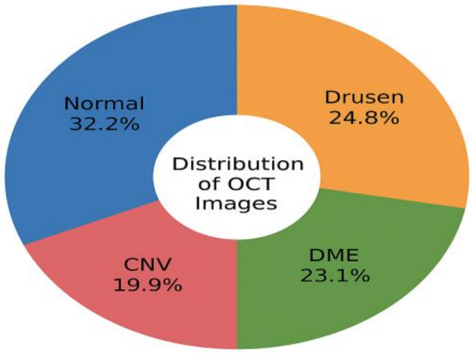

Figure 5. Distribution of OCT images

Preprocessing is a crucial step in medical image analysis as it standardizes input data, reduces noise, and enhances the model's ability to learn discriminative features. In this study, all Optical Coherence Tomography (OCT) images were subjected to a series of preprocessing operations to improve model performance and generalization. Initially, the original grayscale OCT images—acquired at varying resolutions—were resized to 224×224 pixels to match the input requirement of the MobileNetV3 architecture. Resizing ensures consistency in spatial dimensions across the dataset and reduces computational overhead without significant loss of clinical features.

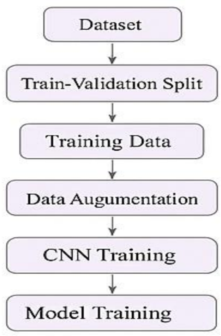

Figure 6. Training Process

Next, image normalization was applied to scale pixel intensity values to the range [0,1]. This standardization stabilizes and accelerates the training process by ensuring uniform feature scaling across images, as recommended in CNN-based medical image classification tasks. To enhance the robustness of the model and reduce the risk of overfitting, data augmentation techniques were employed.

Performance Evaluation:
To evaluate the effectiveness of the proposed deep learning model for multiclass classification of retinal diseases, we conducted a comprehensive performance analysis using key metrics: accuracy, precision, recall, F1-score, ROC-AUC, and a confusion matrix. The classification task involved distinguishing among four retinal classes: Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal using the MobileNetV3 architecture.

Accuracy and Loss Analysis:
The model was trained for 15 epochs with a batch size of 50 using the Adam optimizer. As shown in Fig. 5, the training and validation loss consistently decreased over epochs, indicating effective learning and minimal overfitting. The model achieved an overall test accuracy of 97% on the unseen dataset comprising 10,933 OCT images.

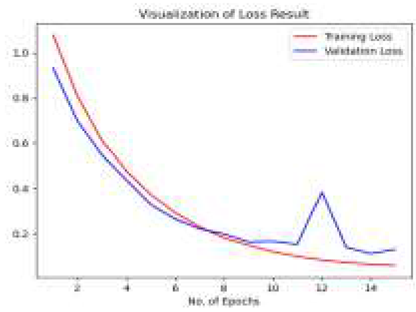

Figure 7. Loss Curve showing Training and Validation Loss over 15 epochs

Confusion Matrix:
A confusion matrix was generated to examine the model's classification behaviour (Fig. 6). The matrix reveals strong diagonal dominance, demonstrating high true positive rates across all four classes.

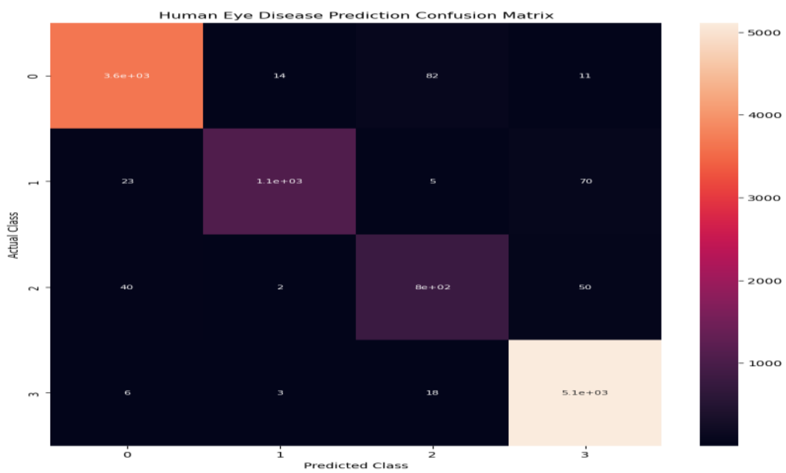

Figure 8. Confusion Matrix for multi-class classification

- Macro Average F1-score: 0.95
- Weighted Average F1-score: 0.97

These results confirm the model's excellent capability in identifying all classes, with slightly lower precision for the Drusen class likely due to overlapping features with other conditions.

ROC-AUC Analysis:
The model's ROC curves (not shown here for brevity) indicated AUC scores above

0.98 for all four classes, reflecting excellent discriminative performance. A one-vs-rest approach was used to compute the multi-class AUC.

RESULTS

The classification report indicates strong overall model performance across four classes: CNV, DME, Drusen, and Normal. The precision, recall, and F1-scores are all quite high, reflecting the model's ability to make accurate and reliable predictions. Specifically, the CNV (0) and Normal (3) classes demonstrate excellent performance, with precision and recall values close to or above 0.98, suggesting the model can both correctly identify and consistently detect these classes with very few false positives or negatives. The DME (1) class also shows high precision (0.98), indicating the model is very confident when it predicts DME. However, its lower recall (0.92) suggests that it occasionally fails to detect some true DME cases, leading to a slightly reduced F1-score of 0.95.

Class	Precision	Recall	F1-Score
CNV (0)	0.98	0.97	0.98
DME (1)	0.98	0.92	0.95
Drusen (2)	0.88	0.90	0.89
Normal (3)	0.98	0.99	0.98
Overall	0.97	0.97	0.97

Table 2: Result analysis

The Drusen (2) class has comparatively lower precision (0.88) and recall (0.90), which may be due to class imbalance, feature similarity with other retinal diseases, or insufficient training samples for this category. Despite this, the overall performance metrics—precision, recall, and F1-score all at 0.97—indicate that the model is highly effective for multi-class classification, though minor improvements could be made to enhance detection of the Drusen class.

```
import google.generativeai as genai

# Setup
genai.configure(api_key="AIzaSyCJweqKO1Yv8mMqbfIMlb0Ykqnub7sYT88")

# Load model (no 'models/' prefix)
model = genai.GenerativeModel('gemini-2.0-flash')

def get_disease_insights(predicted_class, language='en'):
    prompt = f"""Provide a detailed, medically accurate explanation and recommended
    medications (as of 2025) for the retinal disease: {predicted_class}. Include:
    - A short introduction
    - Common symptoms
    - Causes and risk factors
    - Latest treatment options
    - Real-time medication names and their purposes (generic + branded if available)
    - Follow-up care suggestions
    Use clear, patient-friendly language.
    Language: {language}
"""
    response = model.generate_content(prompt)
    return response.text
```

Fig 9. Code snippet for generating retinal disease insights using the Gemini 2.0 Flash model

The above figure indicates the python implementation of the get disease insights function that interacts with the Gemini 2.0 Flash model. This function dynamically generates a medically informative prompt based on the predicted retinal disease class and user language preference. It retrieves AI-generated insights, including symptoms, treatment options, and recommended medications (as of 2025), to assist patients and clinicians with detailed yet accessible information.

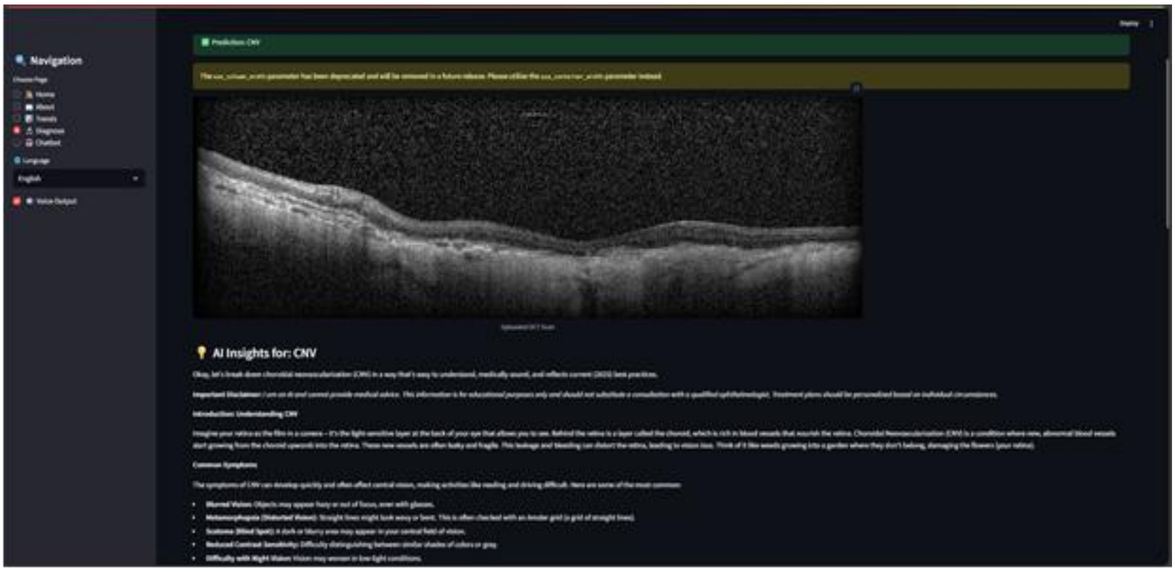

Fig 10. AI-generated diagnostic insights interface for retinal OCT scans

The above figure indicates the AI-generated diagnostic insights interface for retinal OCT scans. The platform classifies uploaded retinal images into one of four major retinal conditions—Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), Drusen, and Normal. After classification, the system utilizes the Gemini 2.0 Flash model to provide comprehensive, medically aligned information including symptoms, causes, treatments, and medication recommendations (as of 2025). The insights are designed to be understandable for patients while retaining clinical relevance.

Comparison With State-of-the-Art Techniques

Method	Dataset	Accuracy
VGG16 + Grad-CAM	Retinal OCT Images (Kermany et al., 2018)	95.19%
Mobile-ViT		96.17%
CNN		96.7%
MobileNet_V3		97.5%

Table 3: Comparison With State-of-the-Art Techniques

Various deep learning methods have been applied to the Retinal OCT Images dataset by Kermany et al. (2018) to classify retinal diseases such as CNV, DME, Drusen, and Normal. Among these, MobileNet_V3 achieved the highest accuracy of 97.5%, demonstrating its efficiency in handling medical image classification with fewer parameters and fast inference. The standard CNN model also performed strongly with 96.7% accuracy, reflecting its capability in capturing spatial features in OCT images. The Mobile-ViT model, which combines the strengths of CNNs and Transformers, achieved a competitive accuracy of 96.17%, indicating its effectiveness in learning both local and global representations. Meanwhile, VGG16 integrated with Grad-CAM achieved 95.19% accuracy, providing not only strong classification performance but also model interpretability through visual explanations. These results highlight that lightweight and hybrid models like MobileNet_V3 and Mobile-ViT can outperform traditional CNNs and provide accurate, real-time diagnostic support for retinal diseases.

CONCLUSION

The comparative analysis of deep learning models on the Retinal OCT Images dataset (Kermany et al., 2018) demonstrates that modern lightweight and hybrid architectures outperform traditional models in both accuracy and efficiency. MobileNet_V3 achieved the highest classification accuracy of 97.5%, showcasing its superior performance for real-time, resource-constrained medical applications. Mobile-ViT and conventional CNNs also showed competitive results, with accuracies of 96.17% and 96.7%, respectively, confirming their effectiveness in feature extraction from OCT images. Although VGG16 + Grad-CAM offered slightly lower accuracy (95.19%), it remains valuable due to its interpretability and visual explanation capabilities. Overall, these findings suggest that optimized and explainable deep learning models are highly capable of supporting accurate and interpretable diagnosis of retinal diseases from OCT images.In addition to high diagnostic accuracy, this study integrates a Large Language Model (LLM) to provide interactive query support related to retinal diseases. The LLM enhances the system's usability by allowing patients and clinicians to ask natural language questions and receive understandable, medically relevant answers about conditions like CNV, DME, and Drusen. This hybrid approach combines robust image-based diagnosis with conversational AI, offering a comprehensive and user-friendly platform for retinal disease detection and patient education.

Declaration by Authors

Ethical Approval: Not applicable
Acknowledgement: None
Source of Funding: None
Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

1. World Health Organization. (2019). World report on vision. https://www.who.int/publications-detail/world-report-on-vision
2. All India Ophthalmological Society. (n.d.). Prevalence of diabetic retinopathy in India. https://www.aios.org
3. Robinson, R. (2001). The impact of rural access on retinal disease diagnosis. Social Determinants of Health and Impact on Screening, Prevalence, 17(3), 145–150.All India Ophthalmological Society. (n.d.). Prevalence of diabetic retinopathy in India. https://www.aios.org
4. Deepak, S., & Bhat, S. (2023a). Deep learning-based CNN for multiclassification of ocular diseases. Journal of Medical Imaging and Health Informatics, 13(1), 26–34. https://doi.org/10.1166/jmihi.2023.3983
5. Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Maskeliūnas, R. (2020). Detection of diabetic retinopathy using convolutional neural networks. Healthcare, 8(4), 348. https://doi.org/10.3390/healthcare8040348
6. Elkholy M and Marzouk MA (2024) Deep learning-based classification of eye diseases using Convolutional Neural Network for OCT images. Front. Comput. Sci. 5:1252295. https://doi.org/10.3389/fcomp.2023.1252295
7. Jaimes, W.J., Arenas, W.J., Navarro, H.J. et al. Detection of retinal diseases from OCT images using a VGG16 and transfer learning. Discov Appl Sci 7, 160 (2025). https://doi.org/10.1007/s42452-025-06565-6
8. Akça S, Garip Z, Ekinci E, Atban F. Automated classification of choroidal neovascularization, diabetic macular edema, and drusen from retinal OCT images using vision transformers: a comparative study. Lasers Med Sci. 2024 May 27;39(1):140. https://doi.org/10.1007/s10103-024-04089-w. PMID: 38797751; PMCID: PMC11128386.
9. Riazi Esfahani P, Reddy AJ, Nawathey N, Ghauri MS, Min M, Wagh H, Tak N, Patel R. Deep Learning Classification of Drusen, Choroidal Neovascularization, and Diabetic Macular Edema in Optical Coherence Tomography (OCT) Images. Cureus. 2023 Jul 9;15(7):e41615. doi: 10.7759/cureus.41615. PMID: 37565126; PMCID: PMC10411652.
10. Babu, T. R. Ganesh, R. Praveena, Gurram Puneeth, T. Satish, and Chemala Vinod Kumar. "Classifications of CNV, DME, Drusen and Normal in Retinal Optical Coherence using CNN." Journal of Artificial Intelligence and Capsule Networks 6, no. 2 (2024): 149-157
11. Tuncer, S.A., Çınar, A., Fırat, M. (2021). Hybrid CNN based computer-aided diagnosis system for choroidal neovascularization, diabetic macular edema,

drusen disease detection from OCT images. Traitement du Signal, Vol. 38, No. 3, pp. 673-679. https://doi.org/10.18280/ts.380314
12. https://www.kaggle.com/datasets/anirudhcv/labeled-optical-coherence-tomography-oct/data
13. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), 1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

How to cite this article: Sheik Imran, Annapurna L U, Divya M S, Tanishka Sharma, Tayyaba Khanum. Deep learning-based classification of retinal diseases from OCT images with LLM-powered patient query support. *Int J Health Sci Res.* 2025; 15(7):314-323. DOI: *https://doi.org/10.52403/ijhsr.20250738*

******