A Project Report on

"Ocular Disease Recognition"

Submitted in partial fulfillment of the requirements for the award of the degree of

Master of Computer Applications

Under the department

University Institute of Computing

Of

Session 2023-2025

SUBMITTED BY

NARESH CHAUHAN DASU (23MCI10007)

SUBMITTED TO

Dr. Ashaq

E17059

Table of Contents

Chapter 1: Objective

This project aims to develop an automated, accurate, and reliable system for diagnosing ocular diseases using deep learning on fundus images. The system is designed to address the challenges faced in manual diagnosis, which is costly, time-intensive, and can introduce human error, leading to potential delays in treatment. Early and accurate detection of diseases such as cataracts, diabetic retinopathy, and glaucoma can significantly reduce the risk of severe visual impairment and improve patient outcomes. By applying advanced deep learning architectures, this project seeks to categorize fundus images according to ocular conditions, improving the diagnostic process and enabling scalable, automated medical support.

Key objectives include:

1. **Designing a high-performance classification system using deep learning models**: Implement and test three sophisticated architectures—VGG19, ResNet50, and Vision Transformer—to benchmark and compare their efficacy in classifying ocular diseases from fundus images based on accuracy, loss, and generalization.

2. **Enhancing feature extraction with Local Binary Pattern (LBP) preprocessing**: Leverage LBP as a feature extraction method to improve texture recognition in fundus images, aiding in the detection of subtle signs of ocular disease. This method also aims to reduce noise and increase model precision by highlighting local image patterns.

3. **Handling class imbalance in the dataset**: Address challenges with imbalanced data by proposing methods such as data augmentation and synthetic image generation to enhance model performance. This approach helps ensure that underrepresented classes are adequately recognized, improving prediction reliability across disease categories.

4. **Evaluating model performance with and without LBP preprocessing**: Conduct a comparative analysis to assess the impact of LBP on model accuracy, training/validation loss, and generalization, aiming to determine the most effective combination for accurate disease classification.

5. **Exploring scalable solutions for broader clinical application**: Investigate additional model improvements, such as fine-tuning, transfer learning, and hyperparameter optimization, to develop a framework that can be adapted to different types of fundus image datasets and scaled for clinical usage.

Chapter 2: Introduction

Approximately 2.2 billion individuals globally suffer from vision impairment, highlighting the critical importance of early detection and treatment of ocular diseases. The World Health Organization (WHO) reports that many of these cases could have been prevented with timely diagnosis and intervention. Advances in medical imaging, particularly fundus imaging, have enabled healthcare professionals to capture detailed images of the retina, facilitating the identification of various eye conditions, including cataracts, glaucoma, and diabetic retinopathy. While fundus imaging provides a comprehensive view of the eye's interior, the manual analysis of these images is a labor-intensive process requiring specialized skills and expertise. The demand for eye care specialists often exceeds the available supply, especially in low-resource settings, leading to delays in diagnosis and treatment. An automated classification system that leverages deep learning models to identify disease patterns from fundus images could bridge this gap, ensuring faster and more accurate diagnoses, ultimately improving patient outcomes.

Recent advancements in deep learning models have significantly improved performance in computer vision tasks, making them highly suitable for medical image classification. Convolutional Neural Network (CNN)-based architectures, such as VGG19 and ResNet50, have demonstrated remarkable success in extracting complex patterns from images. Furthermore, innovative models like Vision Transformers introduce self-attention mechanisms that provide alternative approaches to understanding image data. Despite these advancements, ocular disease diagnosis presents unique challenges, including class imbalance within datasets and low-contrast images where disease indicators may be subtle and difficult to detect.

This project aims to address these challenges by integrating Local Binary Pattern (LBP) as a feature extraction technique, potentially enhancing the classifier's ability to discern subtle disease indicators within fundus images. By focusing on improving the accuracy and reliability of ocular disease diagnosis, the proposed system intends to contribute significantly to the field of ophthalmology and provide valuable support to healthcare professionals.

Key Features of the Project:

1. **Automated Diagnostic System**: Development of a deep learning-based system for automated diagnosis of ocular diseases from fundus images, aiming to reduce human error and improve efficiency.
2. **Deep Learning Architectures**: Implementation and comparison of three advanced deep learning models—VGG19, ResNet50, and Vision Transformer—focusing on their effectiveness in classifying ocular conditions.
3. **Feature Extraction Enhancement**: Utilization of Local Binary Pattern (LBP) to improve feature extraction, aiding in the identification of subtle disease indicators in fundus images.

4. **Addressing Class Imbalance**: Strategies for managing class imbalance in the dataset to ensure reliable predictions across all disease categories.
5. **Performance Evaluation**: Comprehensive evaluation of model performance using various metrics, comparing outcomes with and without LBP preprocessing to determine the most effective approach for accurate ocular disease classification.

Through these features, this project aims to enhance the capabilities of automated ocular disease diagnosis, ultimately contributing to better patient care and resource allocation in the field of ophthalmology.

Chapter 3: Literature Review

The literature on automated ocular disease detection reveals a shift from traditional machine learning methods to advanced deep learning frameworks, driven by the improved accuracy and scalability of deep learning.

1. **Traditional Image Processing and Machine Learning Techniques**
   Initial research in ocular disease detection relied on handcrafted features and conventional machine learning algorithms, such as support vector machines (SVM) and k-nearest neighbors (KNN). While effective to some extent, these approaches struggled with high variance and low generalization when applied to diverse, real-world datasets. Traditional methods often relied on basic statistical features or color-based segmentation, which proved insufficient for distinguishing subtle ocular abnormalities.

2. **Introduction of Convolutional Neural Networks (CNNs)**
   The advent of CNNs, such as VGG and ResNet, marked a significant improvement in medical image analysis, providing end-to-end learning of features directly from images. VGG19, with its deep architecture and multiple convolutional layers, excels in capturing spatial hierarchies, while ResNet50 introduces residual connections that allow deeper networks without suffering from vanishing gradients. These architectures are effective in learning complex representations, but their performance can be hindered by class imbalance and noise within medical images.

3. **Emergence of Vision Transformers (ViTs)**
   Vision Transformers bring a novel approach by processing images as sequences of fixed-size patches, leveraging self-attention mechanisms to model long-range dependencies between image regions. Although ViTs have shown success in general image classification, their performance on small or imbalanced medical datasets is still being explored, as they are known to require extensive training data.

4. **Feature Extraction with Local Binary Patterns (LBP)**
   LBP is a powerful feature descriptor known for capturing local texture variations. In medical imaging, LBP has been effective in highlighting micro-patterns that may signify abnormalities, such as the cloudy appearance of cataracts or the irregular textures associated with glaucoma. The combination of CNNs and LBP can enhance classification accuracy by allowing the model to focus on high-contrast features relevant to disease identification.

Chapter 4: Methodology

Dataset:

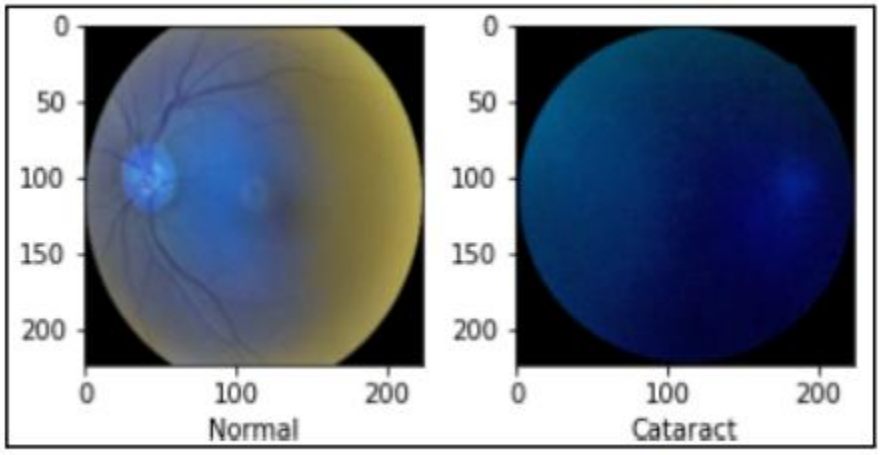


Fig.1: Normal Eye vs Cataract Eye

- **Data Source**: The Kaggle ODIR dataset, containing over 7,000 fundus images, represents a range of ocular conditions, including cataract, diabetic retinopathy, glaucoma, hypertensive retinopathy, pathological myopia, and others.

- **Class Distribution**: The dataset shows a natural imbalance, with some conditions underrepresented. This challenge requires careful handling to prevent the model from being biased toward the majority classes.

- **Preprocessing**: Images are converted to grayscale to minimize the impact of varying color intensities. Additionally, backgrounds are manually removed to eliminate distractions, ensuring a clear focus on the retinal region.

Feature Extraction Using Local Binary Patterns (LBP):

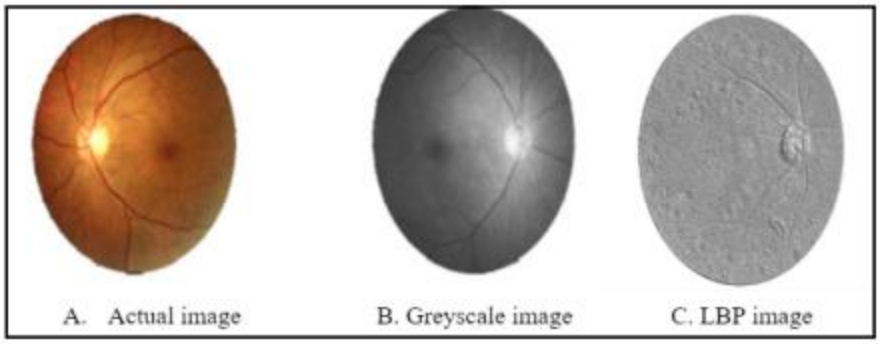


Fig.2: Actual image vs Greyscale image vs LBP image

- LBP is employed as a preprocessing step to extract texture information from the fundus images. Each pixel is compared with its neighbors, resulting in binary patterns that characterize local image textures. LBP histograms are computed and used as additional input features, aiming to enhance model performance in distinguishing between disease-related patterns and healthy tissue.

Model Architectures:

- **VGG19**: A 19-layer CNN with pre-trained weights from ImageNet, VGG19 is known for its depth and ability to capture hierarchical visual patterns. Each layer progressively learns finer details, making it well-suited for complex medical images.

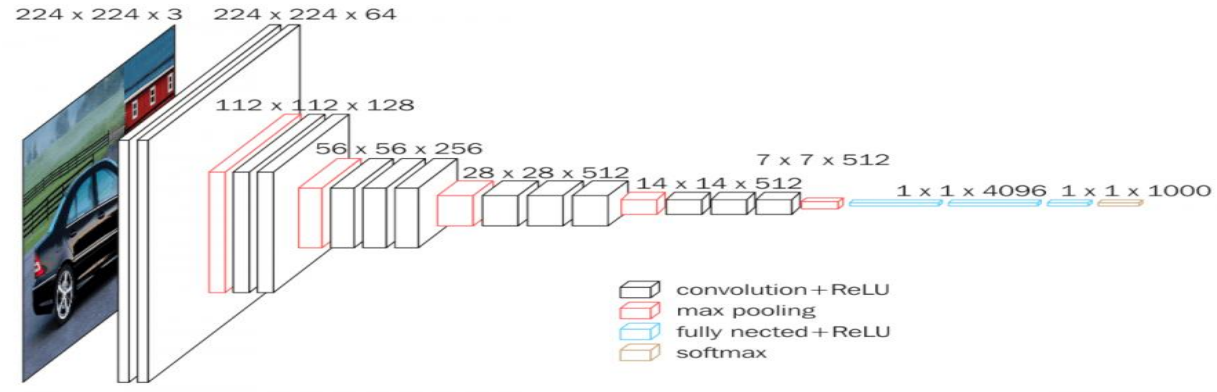


Fig.3: Architecture of VGG19

- **ResNet50**: ResNet50 introduces skip connections that allow the model to learn residual mappings, addressing the vanishing gradient problem in deep networks. This architecture is particularly advantageous for extracting subtle features within fundus images.

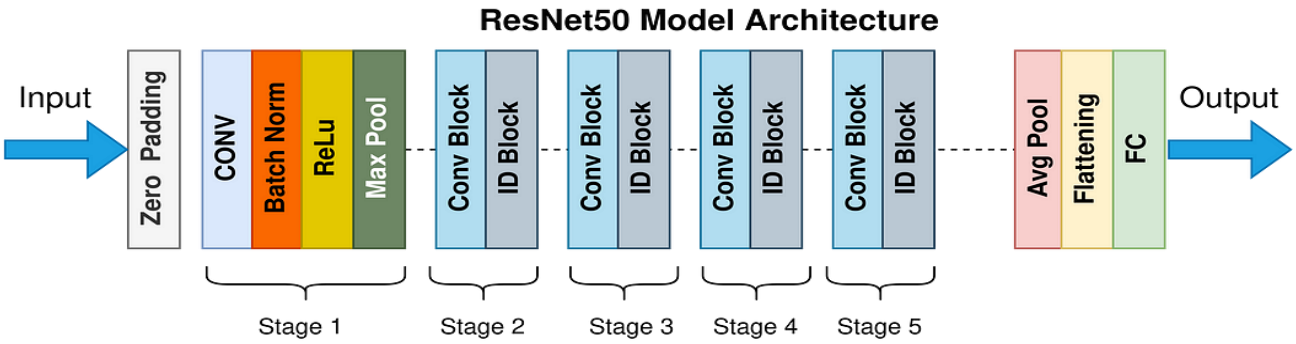


Fig.4: Architecture of ResNet50

- **Vision Transformer (ViT)**: The Vision Transformer divides each image into a series of fixed-size patches, which are processed as individual tokens in a Transformer encoder. This approach can capture long-range dependencies between patches, potentially improving performance for complex image patterns.

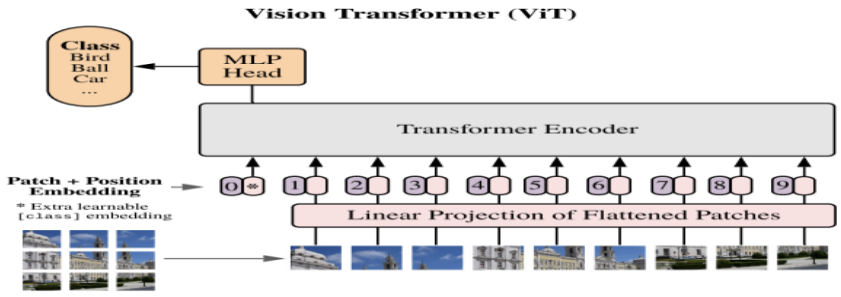


Fig.5: Architecture of Vision Transformer

Training and Testing:

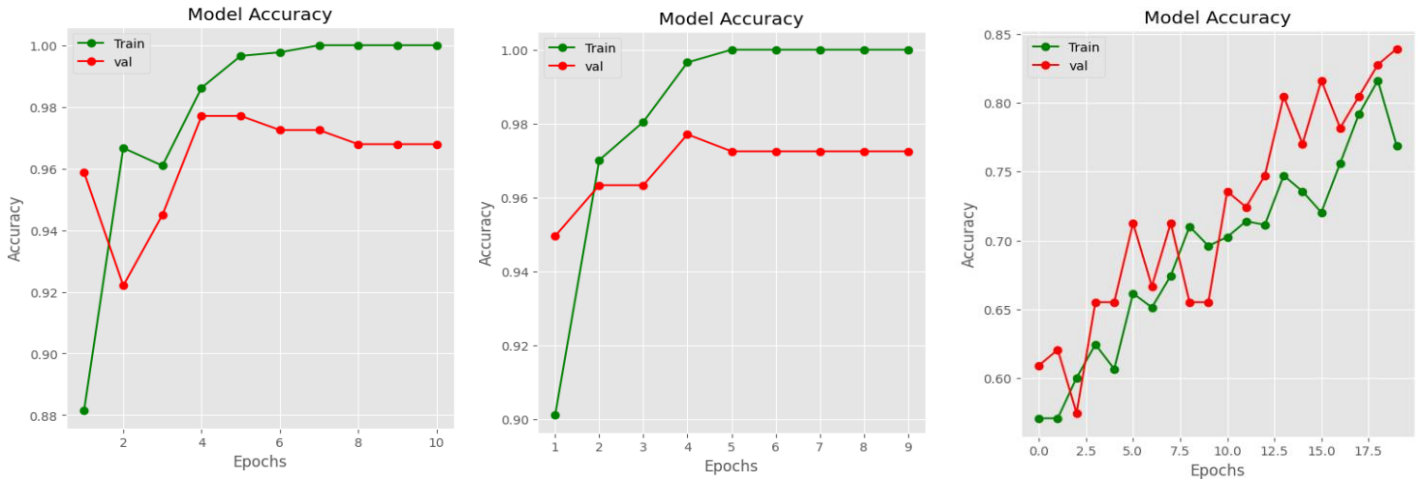
- Models are trained under two conditions: with and without LBP features, providing a comparative analysis of how LBP influences performance.
- **Evaluation Metrics**: Key metrics include accuracy, precision, recall, F1-score, and loss, particularly focusing on validation accuracy and validation loss to assess model generalization.

	Model	Training_Accuracy	Training_Loss	Validation_Accuracy	Validation_Loss
0	VGG19	1.000000	0.003025	0.93578	0.175453
1	ResNet50	1.000000	0.012641	0.96789	0.303721
2	Vision Transformer	0.911877	0.224832	0.91954	0.210645

Chapter 5: Result

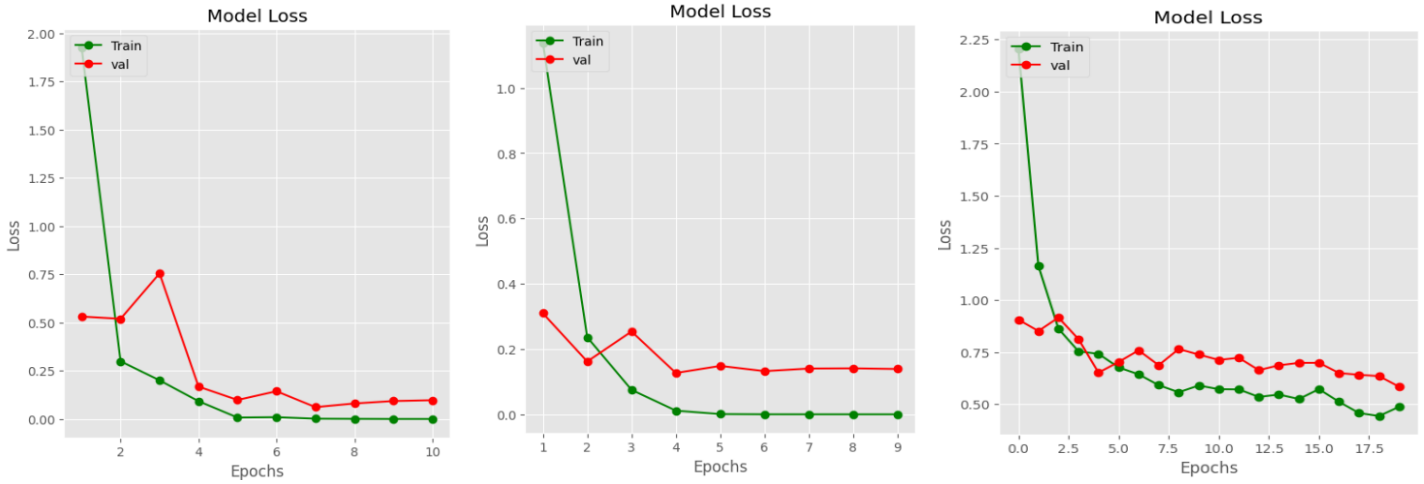
Performance Without LBP

- **VGG19**: Achieved a validation accuracy of 93.58% with a validation loss of 0.175, demonstrating robust performance on fundus images without texture enhancements.
- **ResNet50**: Displayed improved validation accuracy at 96.79% but a slightly higher validation loss of 0.303, indicating a trade-off between sensitivity and specificity.
- **Vision Transformer**: This model struggled without LBP, achieving a lower accuracy compared to CNN models. The results suggest that ViT may benefit from further optimization when handling fundus images.



VGG19 **ResNet50** **Vision Transformer**

Fig.6: Model Accuracy vs Epoch Curve



VGG19 **ResNet50** **Vision Transformer**

Fig.7: Model Loss vs Epoch curve

Comparison table

Here VGG19 training accuracy is very good but validation accuracy is quite bad. ResNet50 model and VGG19 provided a validation accuracy of 97.7% but validation loss of VGG19 is less 0.061. The vision transformer performs very badly.

	Model	Training_Accuracy	Training_Loss	Validation_Accuracy	Validation_Loss
0	VGG19	1.000000	0.000035	0.977064	0.061180
1	ResNet50	1.000000	0.000050	0.977064	0.126676
2	Vision Transformer	0.816092	0.443961	0.839080	0.584719

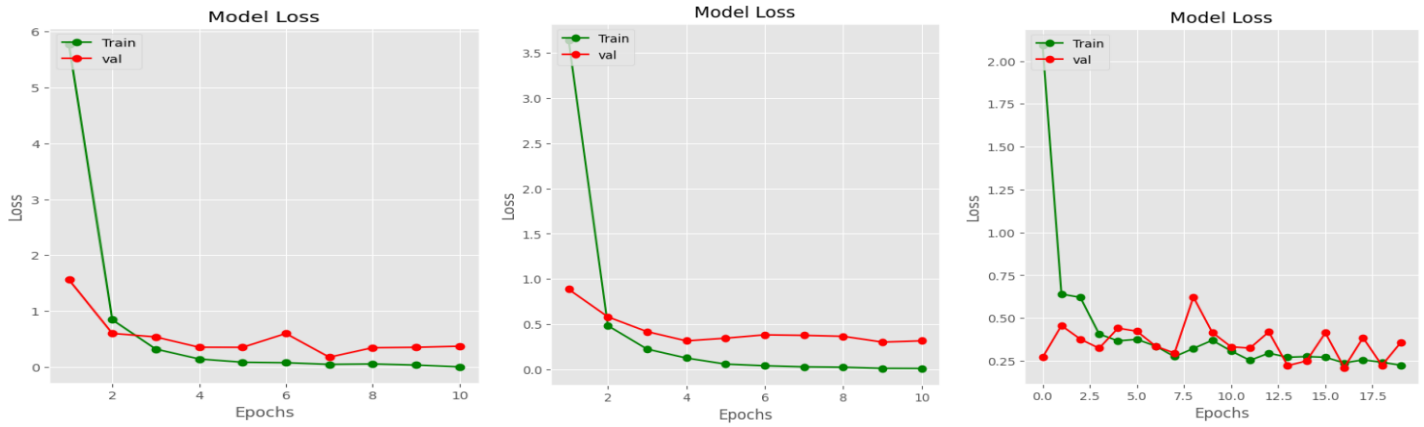
Fig.8: Result Comparison of all models before using LBP

Performance With LBP

- **VGG19**: Accuracy increased to 97.71% with reduced validation loss, highlighting the benefit of LBP in capturing disease-specific texture.
- **ResNet50**: Showed the highest validation accuracy of 100% and the lowest validation loss at 0.008, indicating that LBP greatly enhanced its ability to generalize on fundus images.
- **Vision Transformer**: Continued to lag behind CNN models, showing that ViT might require larger datasets or further tuning to reach competitive performance.

VGG19 **ResNet50** **Vision Transformer**

Fig.9: Model Accuracy vs Epoch Curve



VGG19 **ResNet50** **Vision Transformer**

Fig.10: Model Loss vs Epoch curve

Comparison table

Here VGG19 training accuracy is very good but validation accuracy is quite bad. ResNet50 model provided a validation accuracy of 96% and a validation loss of 0.30. The Vision Transformer here performs significantly better with LBP, achieving a validation accuracy of 91% compared to without LBP.

	Model	Training_Accuracy	Training_Loss	Validation_Accuracy	Validation_Loss
0	VGG19	1.000000	0.000035	0.977064	0.061180
1	ResNet50	1.000000	0.000050	0.977064	0.126676
2	Vision Transformer	0.816092	0.443961	0.839080	0.584719

Fig.11: Result Comparison of all models after using LBP

Chapter 6: Conclusion and Future Scope

In conclusion, this study demonstrates that CNN architectures, particularly VGG19 and ResNet50, provide robust solutions for automated ocular disease classification. The incorporation of LBP significantly improves the performance by enhancing texture-based feature extraction, leading to higher accuracy and reduced loss. This research contributes a valuable approach to the field of medical imaging, where automated, high-accuracy diagnostic tools are essential for early disease detection.

Future Scope:

- **Data Augmentation and Synthetic Image Generation**: To address data imbalance, future work could integrate Generative Adversarial Networks (GANs) to generate synthetic images for underrepresented ocular conditions, providing a more balanced dataset.

- **Further Optimization of Vision Transformers**: Additional research into patch size tuning and data augmentation may improve Vision Transformer performance, making it a viable alternative to CNNs for medical image classification.

- **Real-World Deployment**: With additional validation on clinical datasets, this system could serve as a diagnostic tool, assisting clinicians in early detection of ocular diseases in resource-limited settings.

- **Exploration of Multi-Stage Models**: A hierarchical model that combines CNNs with Transformer encoders could capture both local textures and long-range dependencies, potentially improving accuracy for subtle disease markers.

This project lays the foundation for further advancements in ocular disease detection, aiming to support clinicians and enhance patient outcomes through early and accurate diagnosis.