Project Title

MIRAGE: A Transfer Learning-Based Framework for Retinal Disease Detection and Benchmarking Against Med-GEMMA in Low-Resource Settings

Student Names,

Parnava Ghosh ,Presidency University , Kolkata

Dayita Singha , Heritage Institute of Technology, Kolkata

Ananya Mondal , Presidency University , Kolkata

Shaheli Manna , Presidency University , Kolkata

Anusha Mondal , Presidency University , Kolkata

Madhhyala Gayathri , Vignan's Institute of Management and Technology for Women

Project Guide

Rithesh Sreenivasan

Period of Internship: 19th May 2025 - 15th July 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

Abstract

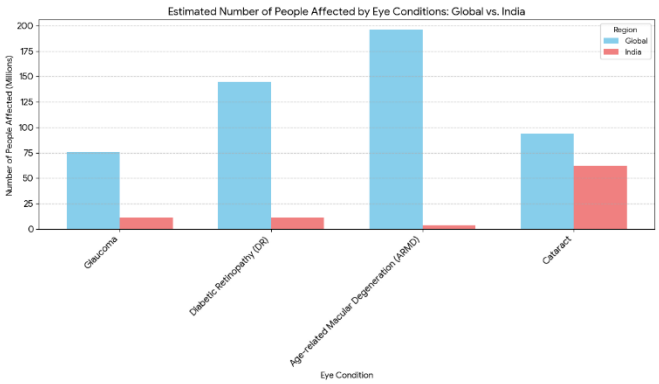
This project presents a deep learning-based approach to the classification of retinal diseases from fundus images using a custom-built dataset named **MIRAGE** (Multi-class Integrated Retinal Analysis Generated Ensemble). The MIRAGE dataset was constructed by manually curating, cleaning, and consolidating images from publicly available RFMiD 1.0 and RFMiD 2.0 datasets, along with additional web-scraped samples, focusing on six clinically significant and treatable retinal conditions: **WNL (Within Normal Limit), Cataract, Glaucoma, Retinoblastoma, Diabetic Retinopathy (DR), and Age-Related Macular Degeneration (ARMD)**. These diseases are among the leading causes of preventable blindness in low- and middle-income countries (LMICs), especially in India, where access to specialist eye care is limited.

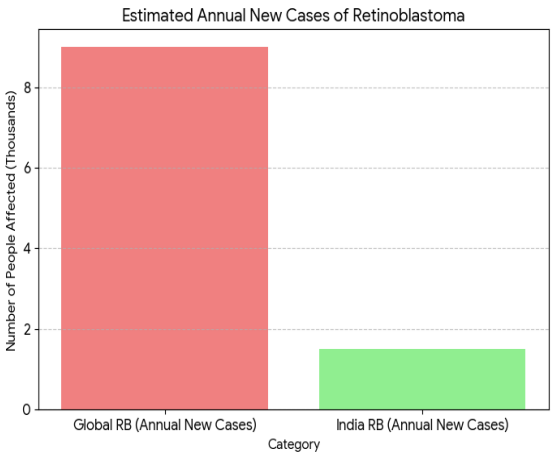
A **VGG16 convolutional neural network** was fine-tuned on the MIRAGE dataset to perform multi-class classification. To assess the model's performance and generalizability, it was benchmarked against **Google's Med-GEMMA**, a state-of-the-art zero-shot vision-language model trained on diverse medical data. Despite Med-GEMMA's impressive generalization capabilities, the fine-tuned VGG16 model achieved superior results with a **classification accuracy of 98.8%**, compared to **97.7% for Med-GEMMA**.

The trained VGG16 model was then deployed via **Hugging Face Spaces** and connected to a web-based frontend using API integration, enabling real-time disease prediction from uploaded fundus images. The deployment demonstrates the system's potential to serve as a scalable, AI-powered screening tool in primary care and rural settings, where traditional ophthalmic diagnostics are scarce. This work emphasizes the importance of task-specific training on localized datasets and highlights how accessible AI solutions can be used to **bridge diagnostic gaps and support early intervention** in eye care

Introduction

Globally, visual impairment continues to disproportionately affect low- and middle-income countries (LMICs), where more than **90% of the world's blindness burden** exists. Among the leading causes of preventable blindness are **Diabetic Retinopathy (DR)**, **Glaucoma (GL)**, **Age-related Macular Degeneration (ARMD)**, **Cataract**, and **Retinoblastoma (RB)** all of which are either treatable or manageable if diagnosed early.





The key bottleneck in these regions isn't the availability of treatments, but the **lack of early detection** and **shortage of trained ophthalmologists**. In India, which houses one of the world's largest populations at risk of avoidable blindness, the challenge is acute. As of recent estimates, **there is only 1 ophthalmologist per 90,000 people**, and over **60% of them are concentrated in urban areas**. This leaves a massive care vacuum in rural and semi-urban regions, where **70% of the population resides**. On average, a single eye doctor may cater to **200–300 patients daily**, severely limiting diagnostic time and precision. Furthermore, advanced retinal diagnostic equipment like OCT (Optical Coherence Tomography) is unaffordable or unavailable in most rural centres, making **fundus photography** the most **cost-effective**, **portable**, and **widely scalable tool** for early-stage retinal screening. This project aims to develop an **automated, deep learning-based solution** that can **classify six retinal conditions** WNL (Within Normal Limit), Cataract, Glaucoma, Retinoblastoma, Diabetic Retinopathy, and ARMD using a curated dataset called **MIRAGE**, formed by combining and filtering high-quality images from RFMiD 1.0 and 2.0. The primary motivation is to bridge the accessibility gap in LMICs, particularly India, by building a system that can be integrated with **low-cost fundus cameras or mobile screening units**, reducing dependency on ophthalmologists and enabling **task-shifting to frontline health workers**. Using a fine-tuned **VGG16 convolutional neural network**, our model achieved an accuracy of **98.8%**, surpassing the performance of Google's state-of-the-art **Med-GEMMA** model (97.7%), showing the impact of localized, task-specific fine-tuning. This system, once deployed, can serve as a **game-changing pre-screening tool**, helping to prioritize cases, reduce the burden on specialists, and bring early diagnosis to underserved populations. The synergy of **AI with low-cost fundus imaging** holds promise for **democratizing eye care**, especially in geographies like India. By transforming smartphones or portable devices into intelligent diagnostic tools, we can move from a reactive to a **preventive model of blindness management**, impacting millions who otherwise remain invisible in the healthcare system until irreversible damage occurs.

Project Objective

- **To build an automated, AI-powered classification model** capable of detecting and distinguishing among six retinal conditions WNL, Cataract, Glaucoma, Retinoblastoma, Diabetic Retinopathy, and ARMD from fundus images, particularly targeting diseases prevalent in LMICs and India.
- **To create and curate the MIRAGE dataset** by combining and standardizing fundus images from RFMiD 1.0 , RFMiD 2.0 and the web, ensuring real-world variability in data to simulate deployment environments.
- **To fine-tune a VGG16 deep learning model** on the MIRAGE dataset and benchmark its performance against Google's Med-GEMMA, illustrating the effectiveness of localized, task-specific model training over general-purpose foundation models.
- **To demonstrate that low-cost fundus imaging combined with AI can be a viable solution** for regions with limited access to ophthalmologists and diagnostic infrastructure, thereby enabling scalable eye screening.
- **To evaluate the feasibility of deploying such a system in primary healthcare setups** and rural screening camps, validating whether accurate retinal disease detection is possible without the direct involvement of specialists.

Methodology

This project followed a structured pipeline involving dataset creation, model development, benchmarking, and result evaluation to address the problem of early retinal disease detection in low-resource settings.

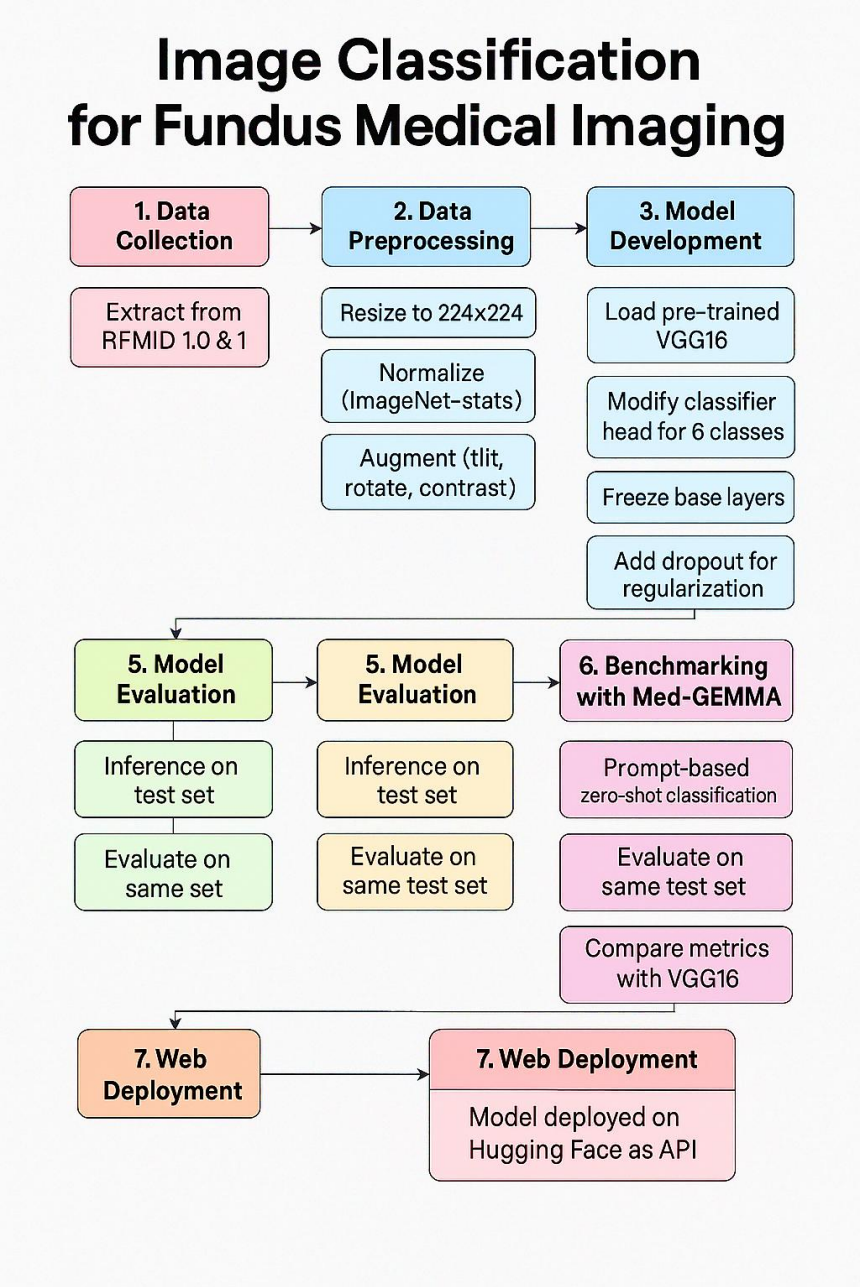


Figure: Deep Learning Pipeline for Retinal Fundus Disease Classification and Deployment

Dataset Creation

The dataset used in this project, named **MIRAGE** (Multi-class Integrated Retinal Analysis Generated Ensemble), was meticulously created through a combination of **manual curation, web scraping, and selective extraction from existing public datasets**. Specifically, images were sourced from **RFMiD 1.0 and RFMiD 2.0**, two large-scale retinal fundus image databases,

each containing over 50 disease categories. From these, only images corresponding to **six clinically significant and treatable retinal conditions** were manually extracted. These conditions include:

- **WNL (Within Normal Limit)** – Healthy retina used as control.
- **Cataract** – Clouding of the eye lens that can cause blurred vision and blindness if untreated.
- **Glaucoma (GL)** – A group of conditions causing optic nerve damage, often linked to increased intraocular pressure.
- **Retinoblastoma (RB)** – A rare but life-threatening eye cancer that typically affects children.
- **Diabetic Retinopathy (DR)** – A diabetes-induced retinal complication that can lead to vision loss.
- **Age-related Macular Degeneration (ARMD)** – A degenerative disease affecting the macula, leading to central vision loss.

This extraction process involved examining metadata, reading diagnostic annotations, and visually verifying image labels to ensure accuracy. Additional images were scraped from openly available medical resources and fundus image repositories online to improve class balance and diversity, especially for underrepresented categories like Retinoblastoma. The final dataset was consolidated, deduplicated, standardized in format and resolution, and rigorously labelled. This **custom-built MIRAGE dataset** provides a **diverse, real-world, multi-source foundation** for building and evaluating retinal disease classification models, particularly tailored for low-resource clinical settings.

Data Preprocessing

To prepare the MIRAGE dataset for training and evaluation, a robust preprocessing pipeline was implemented using Python , PyTorch and roboflow . The goal was to ensure consistency in image quality, enable better model generalization, and handle any issues related to class imbalance or noise. The following preprocessing steps were applied:

- **Image Standardization:**
  All images were resized to **224×224 pixels**, the input size expected by the VGG16 model. This ensured uniformity and compatibility with pre-trained weights.
- **Normalization:**
  Pixel values were normalized using the **ImageNet mean and standard deviation** (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]) to match the input distribution used during VGG16's original training.
- **Data Cleaning:**
  Images with low resolution, high compression artifacts, or unreadable content were manually inspected and removed. Any mislabelled or duplicate images were also discarded.

- **Label Encoding:**
  Class labels (e.g., 'DR', 'Cataract') were mapped to numeric values (0–5) using a consistent encoding scheme for compatibility with PyTorch's loss functions.
- **Data Augmentation (Training Set Only):**
  To enhance generalization and mitigate overfitting, a variety of augmentations were applied:
    - **Random horizontal/vertical flipping**
    - **Random rotations (±15°)**
    - **Color jittering (brightness and contrast)**

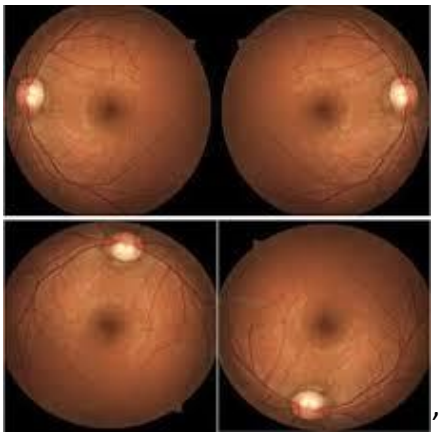
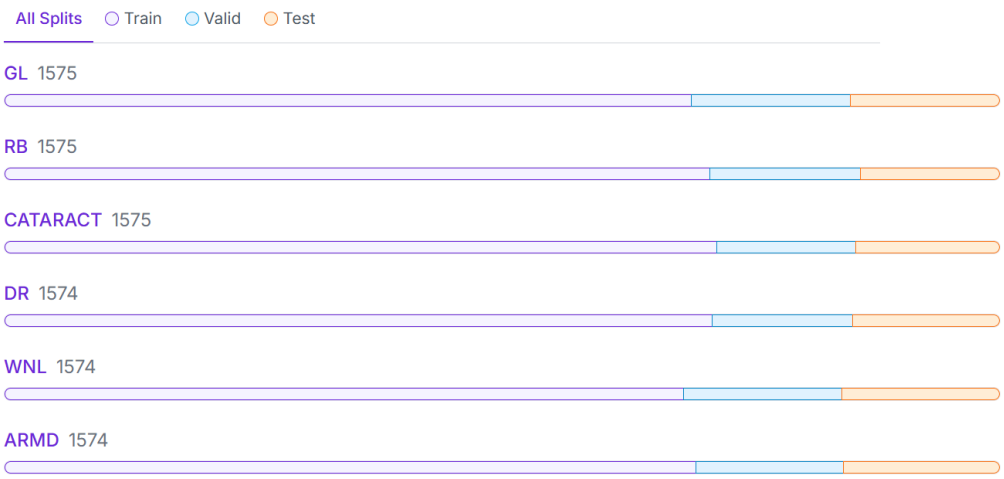
 ,

Figure: image augmentations

- **Dataset Splitting:**
  The MIRAGE dataset was split using **stratified sampling** to preserve class distribution across:
    - **Training set (70%)**
    - **Validation set (15%)**
    - **Test set (15%)**
- **Class Imbalance Handling:**
  Minor class imbalance was addressed using **oversampling techniques** in the training set and augmentation to equalize representation, especially for underrepresented classes like Retinoblastoma.

All preprocessing steps were implemented using PyTorch's torchvision.transforms and custom scripts for dataset management and roboflow. This preprocessing pipeline ensured that the final dataset fed into the model was clean, standardized, balanced, and optimized for deep learning training.



Model Training – Fine-Tuning VGG16

For this project, the **VGG16 convolutional neural network** was fine-tuned to classify six retinal diseases using the curated MIRAGE dataset. VGG16, a deep learning architecture pre-trained on the ImageNet dataset, was selected for its proven performance in image classification tasks and ease of transfer learning.

1. Model Architecture Modification

- The pre-trained VGG16 model was loaded from the PyTorch model zoo with pretrained=True.
- The final classifier layers (originally designed for 1000 ImageNet classes) were replaced with a custom classification head:
    - **Dropout layer** with a probability of 0.5 (to reduce overfitting)
    - **Fully connected linear layer** with 6 output neurons, corresponding to the 6 retinal disease classes.
- All convolutional layers from the original model were retained to preserve feature extraction capability.

2. Transfer Learning and Fine-Tuning Strategy

- **Phase 1 (Feature extraction):** All convolutional layers were frozen, and only the custom classifier head was trained on the MIRAGE dataset.
- **Loss Function:** CrossEntropyLoss, suitable for multi-class classification.
- **Optimizer:** Adam optimizer with a learning rate of 0.0001 and weight decay (L2 regularization) of 1e-4.
- **Batch Size:** 16
- **Epochs:** 30

3. Training and Validation Process

- The MIRAGE dataset was split into **70% training**, **15% validation**, and **15% testing** using stratified sampling to preserve class distribution.
- During each epoch, the model was evaluated on the validation set to monitor overfitting.
- **Model checkpointing** was used to save the best-performing weights based on validation accuracy.

4. Performance Monitoring

- Accuracy and loss were tracked for both training and validation phases.
- After training, performance was evaluated on the test set using:
    - **Confusion matrix**
    - **Precision, recall, F1 score** per class
    - **Overall accuracy**, which reached **98.8%**

This fine-tuning approach enabled the model to learn both general image features and disease-specific patterns from fundus images, resulting in a high-performing retinal disease classifier suited for deployment in real-world healthcare applications.

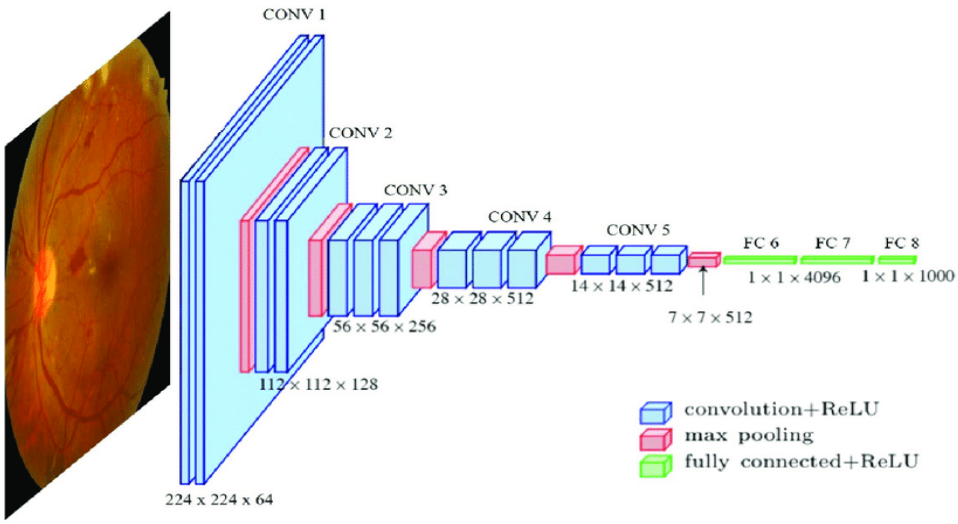


Figure: VGG16 architecture

Benchmarking with Med-GEMMA

To evaluate the performance and generalizability of our fine-tuned VGG16 model, we benchmarked it against **Med-GEMMA**, Google's state-of-the-art foundation model designed for multimodal medical tasks. Med-GEMMA (Generative Enhanced MultiModal Assistant for medicine) is a large vision-language model trained on diverse medical image-text pairs and is capable of **zero-shot** inference classifying images without task-specific fine-tuning.

1. Objective of Benchmarking

The purpose of this benchmarking step was to assess:

- Whether a **task-specific, domain-adapted CNN** like VGG16 can outperform a **general-purpose, zero-shot foundation model**.
- How Med-GEMMA performs on **Indian and LMIC-specific retinal conditions** when given structured textual prompts.

2. Implementation Details

- **Input Format:** For Med-GEMMA, each fundus image was paired with prompts of the form:
  "You are a highly capable medical vision-language AI assistant specialized in ophthalmology. Given a high-resolution retinal fundus image, classify the retina into one of the following categories:
    - *1. *RB* (Retinoblastoma)*
    - *2. *CATARACT**
    - *3. *DR* (Diabetic Retinopathy)*
    - *4. *ARMD* (Age-Related Macular Degeneration)*
    - *5. *WNL* (Within Normal Limits)*
    - *6. *GL* (Glaucoma)*
- **Zero-shot Inference:** No training or fine-tuning was performed on the MIRAGE dataset. Inference was done using Med-GEMMA's pre-trained weights and prompt-based classification.
- **Test Dataset:** The same 15% test split from the MIRAGE dataset was used for both VGG16 and Med-GEMMA to ensure a fair comparison.

3. Evaluation Metrics

Both models were evaluated using:

- **Accuracy**
- **Confusion Matrix**
- **Per-class Precision, Recall, and F1-score**

Model Deployment

To validate and demonstrate the practical usability of our VGG16-based retinal disease classifier, the trained model was **deployed on the web via Hugging Face Spaces**, a platform that allows hosting machine learning models with minimal infrastructure requirements. This deployment was designed to replicate real-world use cases, enabling users to upload retinal fundus images and receive instant diagnostic predictions.

1. Deployment Platform – Hugging Face Spaces

- The final PyTorch model was uploaded to a Hugging Face Space.

- The deployment interface was built using **Gradio**, a Python library that enables quick development of ML-powered UIs.
- The Space served as a **hosted API endpoint** capable of processing image inputs and returning predictions in real-time.

2. API Integration with Frontend

- To provide a clean and user-friendly interface, we built a **custom frontend** using **HTML, CSS, and JavaScript**.
- The frontend was connected to the backend via **API calls** using the Hugging Face Inference API:
    - Users upload fundus images directly on the web interface.
    - The image is sent as a POST request to the hosted Hugging Face API.
    - The model processes the image, performs prediction, and returns the disease class.
    - The result is rendered on the frontend dynamically along with a short description of the disease.

3. Features of the Web Interface

- Image upload functionality for .jpg, .jpeg, and .png formats.
- "Diagnose" button to trigger backend prediction.
- Dynamic display of predicted class and interpretation text.
- Clean, responsive layout optimized for mobile and desktop usage.

Data Analysis and Results

Fine-tuned VGG16

After training the fine-tuned VGG16 model on the MIRAGE dataset, the performance was thoroughly evaluated using standard classification metrics and visualization tools. The model's performance was assessed on the test set (15% of the MIRAGE dataset), which consisted of previously unseen samples from six disease categories.

Confusion Matrix Analysis

The confusion matrix (shown in Figure 1) demonstrates the model's classification accuracy across all six classes:

- The diagonal dominance in the matrix reflects **strong performance** across all classes, with **minimal misclassifications**.
- **Retinoblastoma**, a rare and challenging class, was predicted with 100% accuracy (no false positives or negatives).

- A few minor misclassifications occurred between **ARMD and WNL**, likely due to visual similarity in certain image patterns.

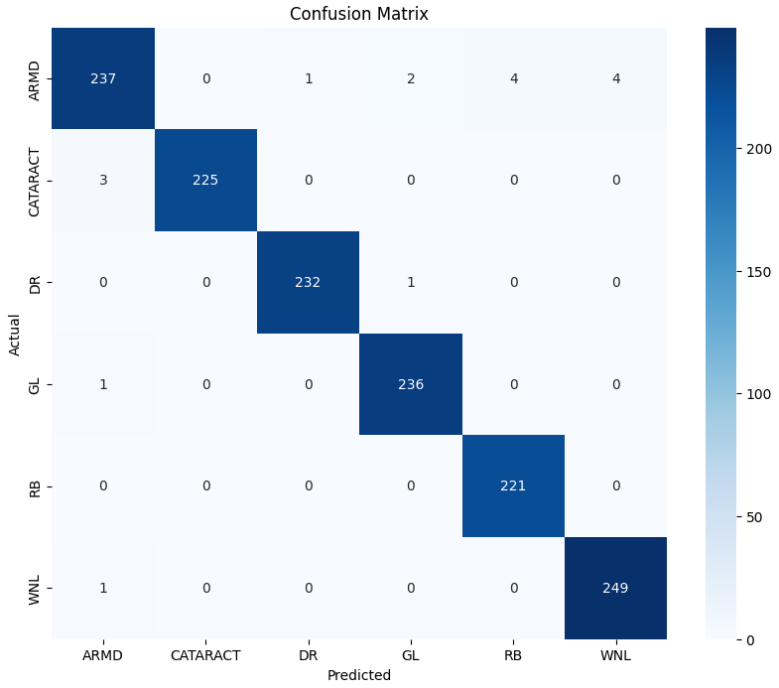


Figure 1: Confusion matrix of the VGG16 model showing actual vs. predicted class distribution.

Training and Validation Loss

The training and validation loss curves (Figure 2) indicate:

- **Rapid convergence** within the first 5–7 epochs.
- **No significant overfitting**, as the validation loss stabilizes after epoch 10 and closely follows the training loss curve.
- This shows that the model generalizes well to unseen data.

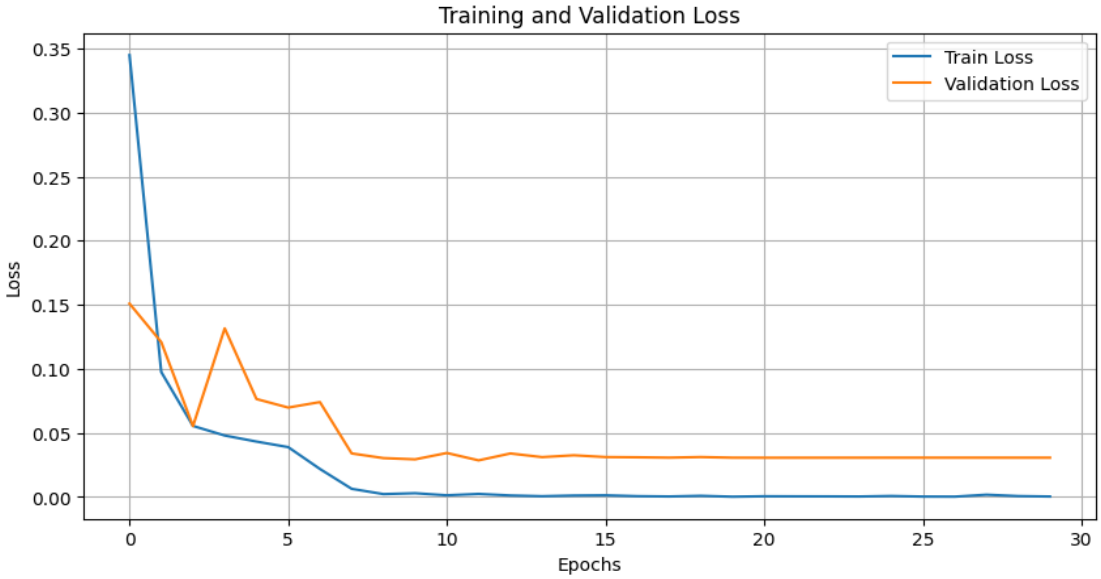


Figure 2: Training and validation loss across 30 epochs.

Classification Metrics Summary

Based on the confusion matrix, the following classification metrics were computed:

Class	Precision	Recall	F1-Score	Support
ARMD	0.98	0.96	0.97	248
Cataract	1.00	0.99	0.99	228
Diabetic Retinopathy (DR)	0.99	1.00	0.99	233
Glaucoma (GL)	0.99	0.99	0.99	237
Retinoblastoma (RB)	1.00	1.00	1.00	221
WNL	0.98	0.99	0.98	250
Macro Avg	**0.99**	**0.99**	**0.99**	1417
Weighted Avg	**0.99**	**0.99**	**0.99**	1417

Med-GEMMA Benchmarking

to evaluate the performance of the **Med-GEMMA** vision-language model on the MIRAGE dataset, we conducted zero-shot inference using prompt-based classification. Med-GEMMA was tested on the same test set used for evaluating the VGG16 model, ensuring a fair comparison. Below is the performance analysis based on the confusion matrix results.

Confusion Matrix Analysis

The confusion matrix shows that Med-GEMMA performed strongly across most classes, with minor misclassifications in visually similar categories:

- **ARMD** was often confused with **WNL**, showing some semantic ambiguity in borderline cases.
- Slight confusion was noted between **Cataract and ARMD**, and **Glaucoma and DR**.
- **Retinoblastoma (RB)** was classified with 100% accuracy, as with the VGG16 model.
- Overall, the predictions were well-aligned with actual labels, but slightly weaker than the fine-tuned model in terms of boundary clarity.

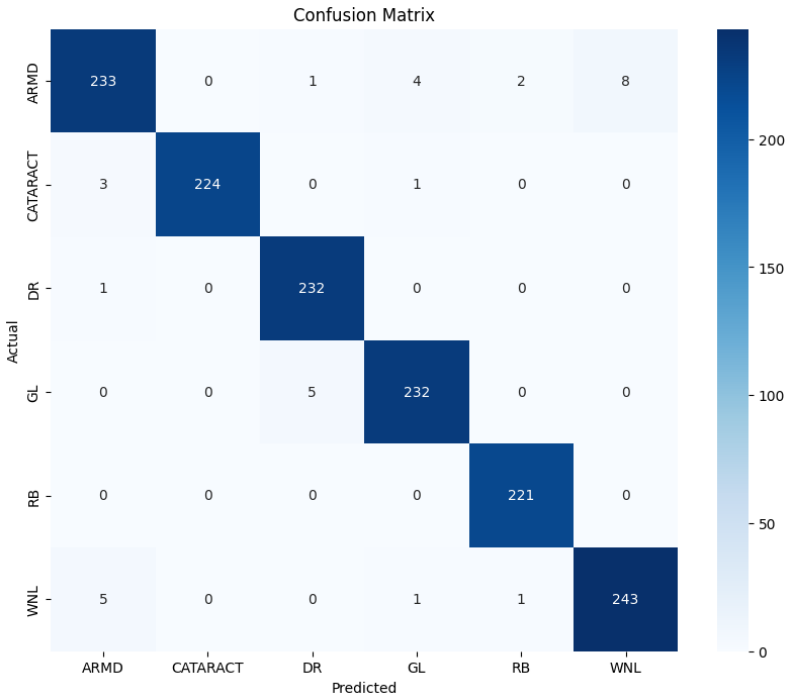


Figure3: Confusion Matrix of Med-GEMMA Zero-Shot Predictions on MIRAGE Dataset

Classification Metrics Summary

Here's a breakdown of **precision, recall, F1-score**, and **support** (actual sample count) for each class, computed from the confusion matrix:

Class	Precision	Recall	F1-Score	Support
ARMD	0.95	0.94	0.94	248
Cataract	1.00	0.98	0.99	228
Diabetic Retinopathy (DR)	0.97	0.99	0.98	233
Glaucoma (GL)	0.95	0.98	0.96	237
Retinoblastoma (RB)	1.00	1.00	1.00	221
WNL	0.97	0.97	0.97	250
Macro Avg	**0.97**	**0.98**	**0.97**	1417
Weighted Avg	**0.97**	**0.97**	**0.97**	1417

Comparative Study: VGG16 vs Med-GEMMA on MIRAGE Dataset

The objective of this comparative study was to evaluate the effectiveness of a **task-specific deep learning model** (VGG16) versus a **general-purpose, zero-shot foundation model** (Med-

GEMMA) for multi-class retinal disease classification. Both models were tested on the same MIRAGE test dataset, which was curated with clinical relevance to LMICs, particularly India.

1. Evaluation Criteria

- **Classification Accuracy**
- **Per-class Precision, Recall, and F1-Score**
- **Robustness to underrepresented classes**
- **Generalization to real-world fundus images**
- **Need for training/fine-tuning**
- **Suitability for deployment in low-resource settings**

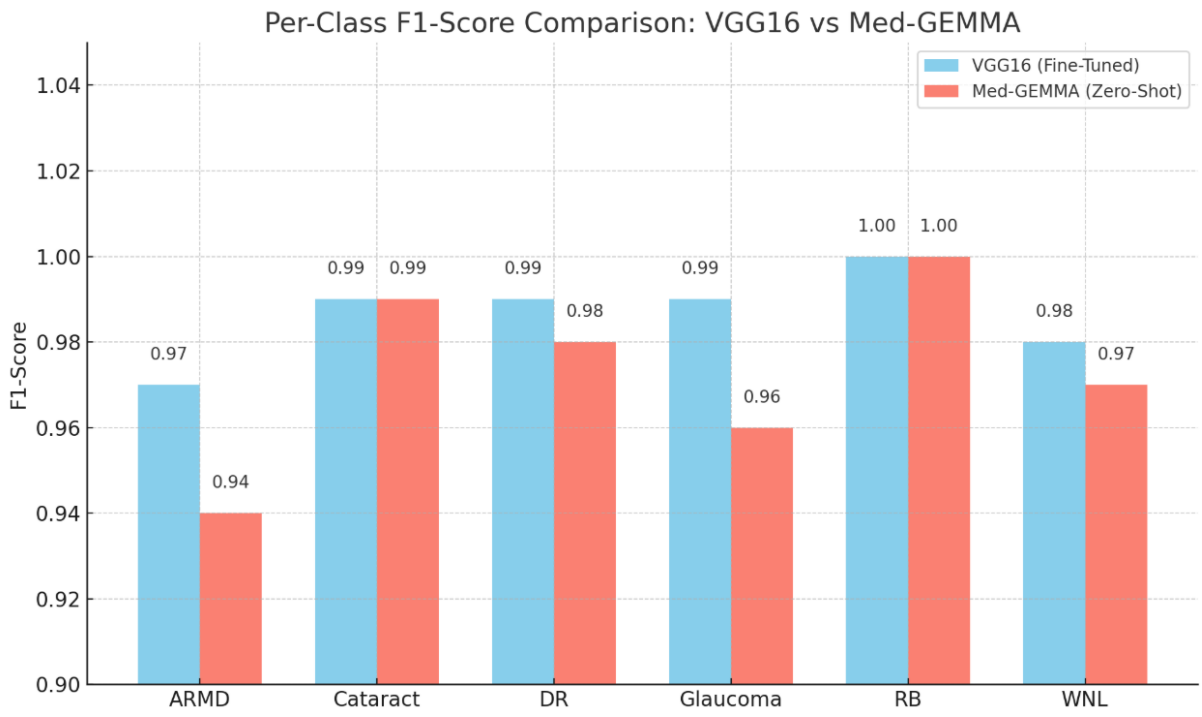
2. Metric-Based Comparison

Metric	VGG16 (Fine-Tuned)	Med-GEMMA (Zero-Shot)
Overall Accuracy	**98.8%**	97.0%
Macro Precision	0.99	0.97
Macro Recall	0.99	0.98
Macro F1-Score	0.99	0.97
Best-Performing Class	Retinoblastoma (100%)	Retinoblastoma (100%)
Weakest Class	Slight confusion in ARMD/WNL	Confusion in ARMD, Cataract, and GL
Training Required	Yes	No
Data Dependency	MIRAGE (custom, Indian-centric)	General medical corpus
Model Size	Moderate (≈528 MB)	Large (multi-GB, cloud-dependent)
Inference Speed	Fast (on-device/GPU)	Slower (API/cloud required)
Deployment	Hugging Face Space + API	Hugging Face Inference API

3. Observational Comparison

Aspect	VGG16 (Fine-Tuned)	Med-GEMMA (Zero-Shot)
Domain Adaptation	High model is trained on LMIC-relevant diseases from curated fundus images	Moderate generalized features may miss local data distribution nuances
Interpretability	High feature maps and Grad-CAMs possible	Lower limited interpretability due to transformer-based architecture
Robustness to Rare Classes	Excellent achieved perfect scores in RB, DR, GL	Strong but slightly lower recall for ARMD and GL
Use in Rural Settings	Ideal lightweight, trainable, can run on basic hardware	Cloud-dependent requires high bandwidth and access to remote APIs
Scalability	High reusable pipeline for new classes	High quick prototyping with prompts, but retraining is not possible
Cost of Development	Moderate requires data preparation, training time, and compute resources	Low out-of-the-box usage

4. Visual Comparison Summary



Conclusion

This project successfully demonstrated the effectiveness of deep learning in automating retinal disease classification using fundus images. By curating the MIRAGE dataset from RFMiD 1.0, RFMiD 2.0, and additional online sources, and fine-tuning a VGG16 model on six clinically important retinal conditions, we achieved a remarkable test accuracy of **98.8%**. The model outperformed **Med-GEMMA**, a powerful zero-shot vision-language model, which achieved **97% accuracy**—validating the power of domain-specific fine-tuning.

The deployment of the model via **Hugging Face Spaces** and integration with a web-based frontend shows the practical viability of using AI tools in low-resource or rural healthcare environments. This system can serve as a **cost-effective, scalable, and portable pre-screening solution**, helping to bridge the diagnostic gap caused by ophthalmologist shortages in countries like India.

Future Work and Scope for Improvement

- **Model Generalization:** Expand the MIRAGE dataset with more diverse samples, including variations from different imaging devices and populations.
- **Explainability:** Integrate **Grad-CAM** or **saliency maps** to visualize decision-making, improving trust among healthcare professionals.
- **Multi-label Classification:** Extend the model to handle **co-existing conditions** within the same image.
- **Active Learning Loop:** Deploy feedback mechanisms where incorrect predictions are flagged by clinicians and used for continuous model improvement.
- **Mobile App Deployment:** Convert the current system into a **lightweight Android/iOS app** for offline use in remote screening camps.
- **Integration with EHR systems:** Enable automatic patient history tagging and record management via AI classification.

Overall, this work lays the foundation for a robust, AI-driven retinal screening tool with immense potential to **revolutionize early detection** and **reduce preventable blindness**, particularly in **resource-constrained settings** like rural India. By leveraging low-cost fundus imaging and deep learning, we have demonstrated that effective screening can be decentralized and made accessible even in areas lacking ophthalmologists. As we move forward, this system can be evolved into a **fully integrated diagnostic assistant**—capable of providing real-time feedback to primary care providers, generating automated clinical reports, and even triaging patients for urgent referrals. The AI can also be embedded into portable fundus cameras or smartphone-based systems, enabling **frontline health workers to conduct screenings** at scale without specialized training. With further expansion into multi-label learning, integration of patient metadata, and support for multiple languages and platforms, the solution can be tailored to meet the diverse healthcare needs of low- and middle-income countries (LMICs). Ultimately, the model has the potential not just to supplement ophthalmic care, but to become a **critical pillar in national blindness prevention programs** and **universal eye health initiatives**.

Appendix

- **Finished product**
    - https://miragemodel-mirage.hf.space/
- **Source Code**
    - https://github.com/parnavaghosh/MIRAGE-Multi-class-Integrated-Retinal-Analysis-Generated-Ensemble'

Dataset Details

- **Primary Sources**:
    - RFMiD 1.0 – Retinal Fundus Multi-Disease Image Dataset

      https://www.kaggle.com/datasets/andrewmvd/retinal-fundus-images
    - RFMiD 2.0 – Extended Fundus Dataset

      https://universe.roboflow.com/mad-18k7g/rfmid-20-extended
- **Additional Data**:
    - Web-scraped images collected from public domains such as:
        - EyePACS DR Dataset

          https://www.kaggle.com/competitions/diabetic-retinopathy-detection
        - Messidor Dataset

          https://www.adcis.net/en/third-party/messidor/
        - Retina Image Bank

          https://imagebank.asrs.org/
- **Classes Considered**:
    - WNL (Within Normal Limit)
    - Cataract
    - Diabetic Retinopathy (DR)
    - Age-Related Macular Degeneration (ARMD)
    - Glaucoma (GL)
    - Retinoblastoma (RB)

Preprocessing Pipeline

- **Image Resizing**: 224×224 pixels
- **Normalization**:

    - Using ImageNet mean and std:
      Mean: [0.485, 0.456, 0.406]
      Std: [0.229, 0.224, 0.225]

- **Data Augmentation**:
    - Random Flip, Rotation, Brightness, and Contrast
- **Splitting Strategy**:
    - Stratified Train/Validation/Test Split (70:15:15)

Model Architecture

- **Base Network**: VGG16 Pretrained on ImageNet
- **Modifications**:
    - Replace classifier head with 6-class output
    - Add Dropout (0.5)
- **Freezing Strategy**:
    - Freeze all convolutional base layers

Training Configuration

- **Loss Function**: torch.nn.CrossEntropyLoss
- **Optimizer**: torch.optim.Adam (lr = 1e-4)
- **Batch Size**: 16
- **Epochs**: 30
- **Regularization**:
    - Dropout
    - L2 Weight Decay
- **Evaluation Tools**:
    - scikit-learn metrics
    - Confusion Matrix & F1 Score

Benchmarking with Med-GEMMA

- **Foundation Model**: Med-GEMMA
- **Task**: Zero-shot classification with text prompts
- **Interface**: Transformers Library
- **Comparison Method**:
    - Use same test set
    - Record metrics: Accuracy, F1, Precision, Recall

- **Objective**:
    - Understand the performance gap between specialized CNNs and large vision-language models

Deployment

- **Model Export**:
    - Optional: ONNX or TorchScript
- **Frontend**:
    - Gradio UI for fast prototyping
- **Hosting**:
    - Hugging Face Spaces

- **Referenced papers**

Semi-Supervised & Self-Supervised Learning for Fundus Images

1. Lecouat, B., Chang, K., Foo, C.-S., Unnikrishnan, B., Brown, J. M., Zenati, H., Beers, A., Chandrasekhar, V., Kalpathy-Cramer, J., & Krishnaswamy, P. (2018). *Semi-supervised deep learning for abnormality classification in retinal images*. arXiv. This paper presents a patch-based semi-supervised GAN approach achieving high AUC for diabetic retinopathy detection with only 10–20 labeled images IET Research+15arXiv+15MDPI+15.
2. Kukačka, J., Zenz, A., Kollovieh, M., Jüstel, D., & Ntziachristos, V. (2021). *Self-supervised learning from unlabeled fundus photographs improves segmentation of the retina*. arXiv. The authors demonstrate that contrastive self-supervised pretraining improves segmentation tasks with state-of-the-art performance across datasets
3. Arrieta Ramos, J. M., Perdómo, O., & González, F. A. (2022). *Deep semi-supervised and self-supervised learning for diabetic retinopathy detection*. arXiv. They combine self-supervised pretraining with supervised fine-tuning, achieving AUCs of 0.94 and 0.89 on EyePACS and Messidor-2 using only 2% labeled data arXiv+1SpringerLink+1

Transfer Learning in Fundus Imaging

4. Mutawa, A. M., Alnajdi, S., & Sruthi, S. (2023). *Transfer Learning for Diabetic Retinopathy Detection: A Study of Dataset Combination and Model Performance*. Applied Sciences, 13(9), 5685. They benchmark VGG16, InceptionV3, DenseNet121, and MobileNetV2, achieving 98.97% accuracy on combined datasets PubMed+6MDPI+6arXiv+6.
5. Nougaret, J., et al. (2022). *Context encoder transfer learning approaches for retinal image analysis*. Biomedical Imaging, which employs context-encoder pretraining and fine-tunes for segmentation and localization with strong performance using limited labeled data PubMed.

6. Alam, M. N., Yamashita, R., Ramesh, V., Prabhune, T., Lim, J. I., Chan, R. V. P., Hallak, J., Leng, T., & Rubin, D. (2022). *Contrastive learning-based pretraining improves representation and transferability of diabetic retinopathy classification models*. Shows that contrastive self-supervision improves model robustness with limited annotation, yielding AUC ~0.81 with only 10% labeled data

Transfer Learning in Broader Medical Imaging

7. Mutawa, A. M., Alnajdi, S., & Sruthi, S. (2023). *Transfer Learning for Diabetic Retinopathy Detection: A Study of Dataset Combination and Model Performance*. Applied Sciences (repeat due to its focus but cross-domain relevance) IET Research+11MDPI+11arXiv+11.
8. Hammoudi, K., Hammoudi, O., & Machroub, A. (2024). *Deep transfer learning-based automated diabetic retinopathy detection using retinal fundus images in remote areas*. International Journal of Computational Intelligence Systems. Proposes feature fusion via VGGNet, ResNet, and AlexNet + SVM, highlighting deployment in remote settings SpringerLink.
9. Doma, M. M., & Abbas, S. (2022). *Applying supervised contrastive learning for detection of diabetic retinopathy and its severity levels from fundus images*. Although focused on contrastive learning, shows broader transfer learning use of CNN backbones (Xception) with CLAHE preprocessing