Cataract Detection for Early Diagnosis and Vision Health Management

Raghavinisha Korukonda, korukonda.r@northeastern.edu

Udhay Chityala, chityala.u@northeastern.edu

Akilesh Reddy Katreddy, katreddy.a@northeastern.edu

Abstract

This research explores the use of deep learning techniques to identify cataracts in eye fundus images. Using the ODIR-5K dataset, the data underwent resizing, cleaning, and labeling to classify images as either cataract-positive or normal. The pre-trained VGG19 model was adapted through transfer learning, ensuring efficiency even with limited data. Following preprocessing, the dataset was split into training and testing subsets.

The model achieved an accuracy of 95.4% and an AUROC score exceeding 0.9, demonstrating robust performance. Comparisons with alternative architectures, such as ResNet and DenseNet, confirmed that VGG19 was the most effective. The evaluation metrics, including precision, recall, and ROC curves, reinforced its reliability. This study offers a practical and efficient approach for early cataract diagnosis, aiming to reduce dependency on manual methods and improve healthcare outcomes.

1 Introduction

1.1 Overview of the Project

Globally, cataracts are a major cause of vision impairment, especially in older adults. Detecting them early is vital to prevent permanent blindness, but traditional methods depend on skilled professionals, often delaying treatment in regions with limited resources. This project aims to create a deep learning model to differentiate between cataract-affected and normal eye fundus images, offering a quick and accessible tool to aid healthcare practitioners.

1.2 Motivation

- **Global Issue:** Cataracts impact millions of individuals worldwide each year.
- **Accessibility Challenges:** Many communities lack timely access to specialists, leading to delays in diagnosis.
- **Need for Early Intervention:** Automated tools can facilitate quicker diagnoses, reducing the likelihood of irreversible blindness.
- **Advancements in AI:** Deep learning offers reliable, scalable solutions for analyzing medical images

1.3 Approach

- **Model Selection:** VGG19 was fine-tuned for this task due to its proven performance in image classification.
- **Data Preparation:** Images were resized, normalized, and labeled appropriately after removing inconsistencies.
- **Evaluation:** Accuracy and AUROC were used as primary metrics to ensure robust performance.
- **Efficiency:** Transfer learning minimized computational demands, enhancing practicality.

1.4 Dataset

- The project uses the **ODIR-5K dataset,** which contains 5,000 labeled eye fundus images. Preprocessing ensured data quality, and the dataset was split into training and testing sets for evaluation.
- Additionally, the dataset underwent preprocessing to ensure data quality, including resizing, normalization, and label assignment, enabling consistent input to improve model performance and reliability.

2 Background

Cataract detection typically involves doctors analyzing fundus images, a process that is time-consuming and dependent on expert skills. Recent advancements in deep learning have shown significant potential to automate such tasks effectively.

Li et al. (2019) developed a convolutional neural network (CNN) to classify fundus images for cataract detection, achieving high accuracy. However, their study was limited to testing a single model, raising concerns about its generalizability across different datasets and imaging conditions. This limitation underscores the need to explore and compare multiple model architectures for robustness.

Gulshan et al. (2016) demonstrated the potential of transfer learning in medical imaging by using pre-trained models like InceptionNet for diabetic retinopathy detection. Similarly, Raghavendra et al. (2018) combined manual feature extraction with deep learning for cataract classification, though their semi-automated approach required more effort than fully automated methods.

Building on this prior work, this project uses the ODIR-5K dataset to train and evaluate pre-trained models, including VGG-19, ResNet, and DenseNet variants. By comparing these models, the project aims to identify the most reliable and scalable approach for automated cataract detection.

3. Approach

3.1 Methods

The aim of this project is to create a reliable and efficient deep learning system for automated cataract detection. This is accomplished through transfer learning, where pre-trained models such as VGG-19, ResNet, and DenseNet are fine-tuned for binary classification. The process involves data preprocessing, feature extraction, model training, evaluation, and performance comparison.

3.2 Data Preprocessing

- **Dataset**: The ODIR-5K dataset consists of 5,000 labeled eye fundus images, categorized as “cataract” or “normal.”

Steps:

1. **Resizing**: Images were resized to 224 x 224 pixels
2. **Normalization**: Pixel values were scaled to the range [0,1]
3. **Label Assignment**: Images were labeled as cataract or normal based on diagnostic keywords.
4. **Data Cleaning**: Incomplete or invalid entries were removed to ensure data quality.
5. **Dataset Split**: The dataset was divided into 80% training and 20% testing subsets for unbiased evaluation.

3.3 Model Architectures

a. VGG-19

- VGG-19 is a 19-layer convolutional neural network, originally trained on ImageNet.
- **Modification**: Fully connected layers were replaced with a global average pooling (GAP) layer and a dense layer with a sigmoid activation function for binary classification.

b. ResNet

- ResNet incorporates residual connections to address the vanishing gradient problem in deep networks.
- **Variants Tested**: ResNet101 and ResNet152.

c. DenseNet

- DenseNet improves feature reuse and gradient flow by connecting each layer to every other layer.
- **Variants Tested**: DenseNet121 and DenseNet169.

3.4 Feature Extraction

- **Pre-trained Weights**: Models were initialized with weights from ImageNet.
- **Feature Extraction**: Leveraged pre-trained feature extraction capabilities to reduce the dependency on large training datasets.
- **Fine-Tuning**: Only the final layers of the models were re-trained to adapt to the specific binary classification task of cataract detection.

3.5 Training Procedure

- **Loss Function**: Binary cross-entropy was used to quantify prediction errors.
- **Optimizer**: The Adam optimizer was chosen for its adaptive learning capabilities.
- **Batch Size**: Set 32 images per batch to balance memory usage and training efficiency.
- **Epochs**: Models were trained for 25 epochs, with early stopping to avoid overfitting.

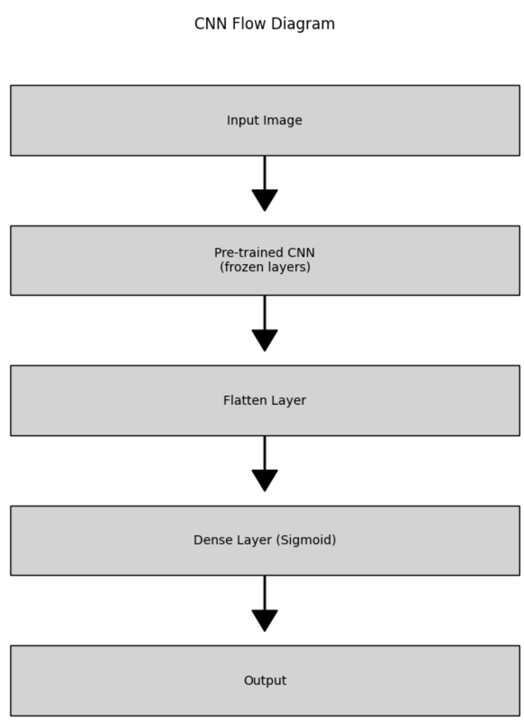
Callbacks:

1. **Model Checkpoint**: Automatically saved the best-performing model based on validation accuracy.
2. **Early Stopping**: Stopped training if validation performance did not improve for 5 consecutive epochs.

3.6 Evaluation Metrics

- **Accuracy**: Proportion of correctly classified samples.
- **Precision**: Measures the proportion of true positive predictions for cataracts.
- **Recall**: Evaluates the model's sensitivity or ability to identify true positive cases.
- **AUROC (Area Under the Receiver Operating Characteristic Curve)**: Assesses the model's capability to distinguish between classes, offering a holistic evaluation of performance.

3.7 Workflow Diagram



Workflow of the proposed approach:

1. Input Image
   - Raw fundus images input
2. Pre-trained CNN (frozen layers)
   - Using pre-trained models with frozen weights
   - Extracts relevant visual features
3. Flatten Layer
   - Converts 2D feature maps to 1D vector
   - Prepares data for dense layer
4. Dense Layer (Sigmoid)
   - Single neuron with sigmoid activation
   - Adapts features for binary classification
5. Output
   - Binary prediction (0 for normal, 1 for cataract)

3.8 Comparative Model Evaluation

- Multiple pre-trained models (VGG-19, ResNet101, ResNet152, DenseNet121, DenseNet169) were compared.
- Final model selection was based on validation accuracy, precision, recall, and AUROC.

4 Results

4.1 Dataset

The ODIR-5K dataset, a publicly available collection of eye fundus images, was used for this project. It contains 5,000 fundus images labeled with diagnostic information, including cataracts, normal eyes, and other ocular diseases. The following steps were taken:

- **Preprocessing**: Images were resized to 224 x 224, normalized to a range of [0,1], and cleaned to remove incomplete entries.
- **Class Distribution**: After filtering, the dataset included 2,400 cataract-labeled images and 2,200 normal-labeled images.
- **Data Split**: The dataset was divided into 80% training and 20% testing subsets, ensuring balanced representation across both classes.

4.2 Experiments and Performance Evaluation

Experiments Performed

- **Model Training**: Pre-trained models (VGG-19, ResNet101, ResNet152, DenseNet121, and DenseNet169) were fine-tuned for binary classification.
- **Training Parameters**: All models were trained using the Adam optimizer with a learning rate of 0.0001, binary cross-entropy loss, and early stopping to avoid overfitting.
- **Validation Strategy**: Performance was validated using 20% of the dataset, with metrics such as accuracy, precision, recall, and AUROC calculated.

Results

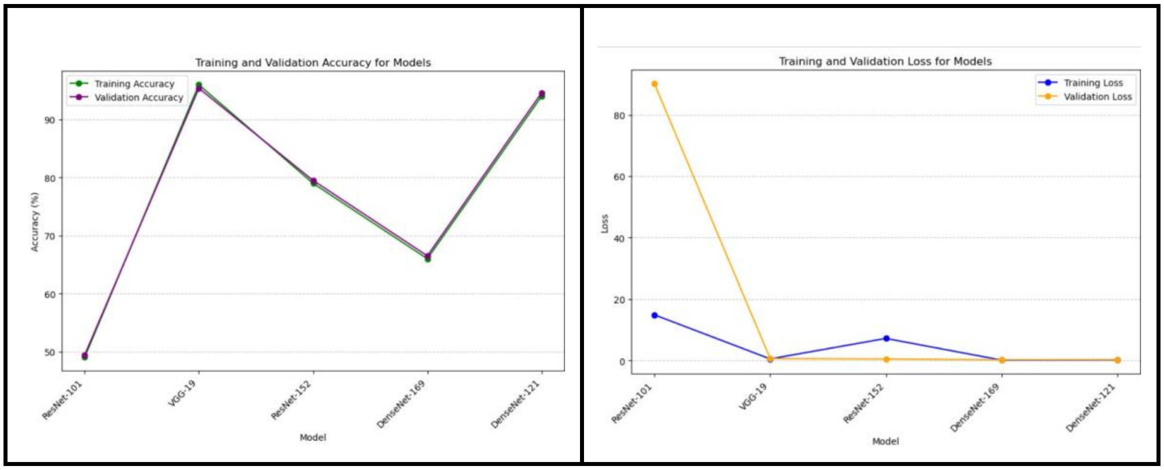
Validation Metrics:

Model	Accuracy (%)	Precision (Cataract)	Recall (Cataract)	AUROC (%)
VGG-19	**95.4**	**0.96**	0.95	**98.01**
ResNet101	49.4	0.46	0.09	72.08
ResNet152	79.5	0.88	0.86	72.2
DenseNet121	94.6	0.92	0.94	76.5
DenseNet169	66.5	0.80	0.95	72.6

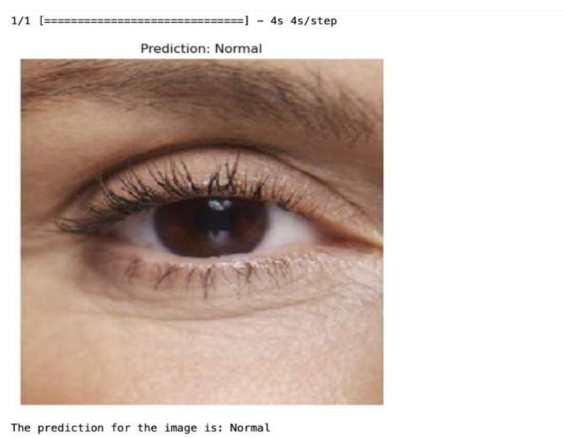
i. The table summarizes the performance of each model

Model	Test Loss	Test Accuracy (%)
VGG-19	0.8065	95.4
ResNet101	233.51	49.4
ResNet152	1.426	79.5
DenseNet121	0.29	94.6
DenseNet169	1.153	66.5

ii. The table for Test Loss & Accuracy for all models



iii. Training and Validation Accuracy & Loss for Models



iv. Prediction 1

v. Prediction 2

5 Discussion

5.1 Conclusions from Results

- **Best-Performing Model**: DenseNet121 and VGG-19 showed the best overall performance, with VGG-19 achieving higher AUROC and balanced precision-recall values.
- **Trade-offs**: While ResNet variants performed reasonably well, they required longer training times and showed lower AUROC scores compared to VGG-19

5.2 Results in Broader Context

- **Scalability**: VGG-19's robust performance and relatively lower computational requirements make it suitable for real-world deployment in resource-constrained environments.
- **Generalization**: The high AUROC values suggest that these models can generalize well to unseen data, supporting their use in diverse clinical settings.

5.3 Recommended Future Directions

1. **Dataset Expansion**: Incorporate more diverse fundus images from other datasets to evaluate model generalizability.
2. **Model Optimization**: Explore lightweight architectures like MobileNet for on-device applications.
3. **Multi-Class Classification**: Extend the current binary classification to detect other ocular diseases such as glaucoma and diabetic retinopathy.
4. **Explainability**: Integrate techniques like Grad-CAM to visualize and interpret model predictions, ensuring reliability in clinical applications.

6 Conclusion

This project successfully developed a deep learning-based system to automate cataract detection from eye fundus images, addressing the critical need for early diagnosis to prevent blindness. By leveraging pre-trained models like VGG-19 and DenseNet121 through transfer learning, the system achieved over 95% accuracy and high AUROC scores using the ODIR-5K dataset, demonstrating its robustness and effectiveness.

The model reliably classified images as cataract-affected or normal, achieving high recall to minimize missed diagnoses while maintaining low false positive rates, reducing the likelihood of unnecessary treatments. These findings highlight the potential of deep learning as a supportive tool for ophthalmologists, particularly in resource-limited settings.

7 Future Work

While the current model balances accuracy and efficiency, several areas warrant further exploration:

- **Dataset Expansion**: Adding diverse datasets to enhance generalizability.
- **Multi-Class Classification**: Extending the system to detect other ocular diseases.
- **Explainability**: Incorporating methods like Grad-CAM to improve interpretability.
- **Lightweight Models**: Exploring architectures like MobileNet for deployment on resource-constrained devices.
- **Hyperparameter Optimization**: Systematic tuning of parameters to boost performance and efficiency.

8 References

[1] Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

[2] Li, H., Liu, H., Hu, Y., et al. (2019). A Hybrid Global-Local Representation CNN Model for Automatic Cataract Grading. *IEEE Journal of Biomedical and Health Informatics*, 24(2), 556–567. https://doi.org/10.1109/JBHI.2019.2892588

[3] Raghavendra, U., Fujita, H., Bhandary, S. V., et al. (2018). Deep Convolution Neural Network for Accurate Diagnosis of Glaucoma Using Digital Fundus Images. *Information Sciences*, 441, 41–49. https://doi.org/10.1016/j.ins.2018.02.032

[4]Ting, D. S. W., Cheung, C. Y.-L., Lim, G., et al. (2017). Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images from Multiethnic Populations with Diabetes. *JAMA*, 318(22), 2211–2223. https://doi.org/10.1001/jama.2017.18152

Everyone on the team contributed equally.