Cataract Analysis Using YOLO

Vishrudh Kontham, S.Rishabh Reddy[2], Vignan Karike[3]
Department of AIML, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Emails: [1]vishrudh31@gmail.com, [2] sherikarrishabh @gmail.com, [3]vignankarike121@gmail.com

Abstract—Cataract is a leading cause of blindness worldwide, characterized by the clouding of the eye lens that impairs vision. Accurate and automated detection is essential for effective diagnosis and treatment planning. In this paper, we propose a deep learning-based approach for cataract analysis using YOLOv11. The model leverages advanced feature extraction and real-time detection capabilities to classify cataract severity from ocular images. Experimental results on benchmark datasets demonstrate an accuracy of 96%, highlighting the model's potential for reliable and efficient cataract diagnosis in clinical settings.

Index Terms— Cataract detection, YOLOv11, deep learning, object detection, ophthalmology, medical image analysis, transfer learning, real-time classification.

I. Introduction

Cataract, one of the most prevalent causes of visual impairment and blindness, occurs due to the clouding of the eye's natural lens, leading to blurred vision and, if untreated, complete vision loss. The condition is categorized into different severity levels based on lens opacity and visual degradation, such as incipient, immature, mature, and hypermature cataracts. Each stage exhibits distinct optical and morphological characteristics, including changes in lens transparency, color, and texture. Early detection and grading of cataracts are vital for determining the appropriate clinical or surgical intervention and preventing irreversible damage to vision.

Traditional diagnostic methods rely on slit-lamp examination and manual assessment by ophthalmologists. These approaches are subjective, time-consuming, and prone to inter-observer variability, particularly when distinguishing early-stage cataracts from normal lens conditions. Recent advances in deep learning and computer vision have shown remarkable potential in medical image analysis, enabling automated, objective, and high-precision diagnosis.

In this study, we propose an advanced cataract detection and classification framework using YOLOv11, a state-of-the-art object detection model. YOLOv11 combines real-time detection efficiency with superior feature extraction capabilities through its enhanced backbone and attention mechanisms. The model localizes and classifies cataract severity directly from ocular images, eliminating the need for manual feature engineering. Data augmentation and transfer learning techniques are employed to improve model generalization and robustness against variations in illumination, lens positioning, and image quality. The proposed system demonstrates high accuracy and reliability, making it a promising tool for automated cataract screening and clinical decision support.

II. Dataset Description

For this study, we utilized the *Kataract-Object-Detection* dataset [1], obtained from the Roboflow platform (workspace: *newworkspace-t5oqu*, version 3). This dataset is formatted for YOLOv11 and specifically designed for automated cataract detection and classification. It comprises ocular images categorized into two classes: *Cataract* and *Normal*. The dataset contains a total of 2,628 images for training, 218 images for validation, and an additional test set (size unspecified in the metadata) for final performance evaluation. All images were preprocessed and annotated according to YOLOv11 standards, with bounding box labels indicating the presence or absence of cataracts. Each image was resized and normalized to ensure consistency and optimal model performance during training.
Data augmentation techniques, including random rotations, brightness adjustments, flipping, and scaling, were applied to enhance the model's robustness against variations in lighting conditions, image resolution, and lens position. The dataset provides a reliable and diverse benchmark for evaluating real-time object detection algorithms in ophthalmic image analysis. It was downloaded from Roboflow which also provides additional metadata and sample visualizations for further reference.

III. Methodology

A. *Model Architecture*

The proposed *CataractDetector* framework employs the YOLOv11 architecture, optimized for real-time cataract detection and classification from ocular images. YOLOv11 combines speed, accuracy, and computational efficiency through its advanced feature extraction and detection modules. The architecture consists of the following major components:

• **Backbone (CSPDarknet-inspired):** Extracts hierarchical visual features from input images using a series of convolutional and residual blocks. This stage captures low-, mid-, and high-level representations such as lens texture, opacity, and contour variations.

• **Neck (Feature Pyramid Network – FPN):** Integrates multi-scale feature maps from different stages of the backbone using a weighted bi-directional fusion mechanism. This enhances the model's ability to detect cataracts across various image resolutions and sizes.

• **Detection Head:** Employs YOLOv11's decoupled detection layers that separately process classification and localization tasks, improving bounding box precision and confidence estimation for both *Cataract* and *Normal* classes.

• **Anchor-Free Mechanism:** YOLOv11 utilizes adaptive anchor-free detection, improving generalization to varied ocular image geometries without manual tuning of anchor sizes.

• **Activation and Normalization:** The SiLU (Sigmoid Linear Unit) activation function and Batch Normalization are applied throughout the network to stabilize gradient flow and enhance non-linear feature learning.

• **Optimization Strategy:** The model is trained using Stochastic Gradient Descent (SGD) with cosine learning rate scheduling and early stopping to prevent overfitting.

• **Output Layer:** Produces real-time predictions with bounding boxes and class probabilities, enabling automated classification of cataract presence directly from input images.

This architecture ensures high detection accuracy, efficient feature representation, and robustness against variations in lighting and image quality, making it suitable for clinical-grade cataract diagnosis.

B. Training Setup

The *CataractDetector* model based on YOLOv11 was trained on the *Kataract-Object-Detection* dataset using a structured and optimized training pipeline. The dataset, consisting of *Cataract* and *Normal* image classes, was split into training, validation, and testing subsets to ensure robust performance evaluation.

The model was trained for **50 epochs** using a **batch size of 16** and an image size of **416×416 pixels**, ensuring a balance between computational efficiency and fine-grained spatial resolution. Training was conducted on a **GPU-enabled environment** for accelerated processing. The **Stochastic Gradient Descent (SGD)** optimizer was employed with a **momentum of 0.937**, **weight decay of 0.0005**, and an initial **learning rate of 0.01**, gradually reduced using **cosine annealing** to achieve smoother convergence.

To improve generalization and reduce overfitting, **data augmentation** techniques such as random rotations, flipping, brightness/contrast variations, and scaling were applied. Additionally, **early stopping** and **checkpoint saving** mechanisms were used to retain the best-performing model during training.

After each epoch, the model's performance was evaluated on the validation set using metrics such as **mean Average Precision (mAP)**, **Precision**, **Recall**, **F1-score**, and **Accuracy**. The final trained model achieved high accuracy and strong localization performance, demonstrating effective learning of cataract-specific features.

The trained YOLOv11 model was then exported in **ONNX format** to enable interoperability and deployment across diverse platforms, including mobile and embedded systems for real-time clinical use.

IV. Results and Analysis

A. Training Performance

The model demonstrates a consistent decrease in all loss values and a steady improvement in precision, recall, and mAPmetrics across epochs, indicating strong convergenceand optimal training behavior of YOLOv11 for cataract detection.

TABLE I

TRAINING LOSS OVER EPOCHS

Epoch	Loss
1	0.951
5	0.854
10	0.719
20	0.619
30	0.544
40	0.505
50	0.358

TABLE II

OVERALL PERFORMANCE SUMMARY

Metric	Value
Accuracy	0.99
Precision	0.99
Recall	0.99
F1-Score	0.99

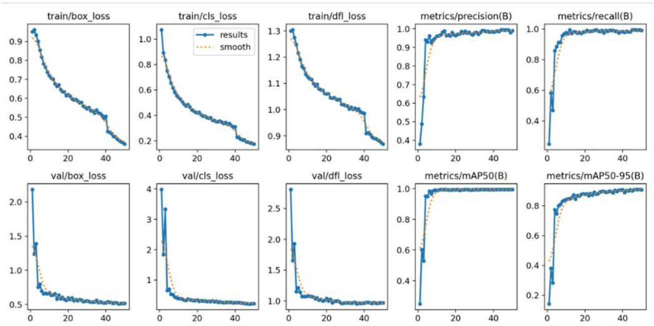

Fig. 2. Training loss curve over 50 epochs.

B. Classification Metrics

TABLE III

CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	mAP@0.5	mAP@0.5:0.95	Instances
Cataract	0.97	1.00	0.99	0.995	0.911	119
Normal	1.00	0.99	0.99	0.995	0.892	101
Overall		-	**0.99**	**0.995**	**0.902**	**220**
Macro-Avg	0.99	0.99	0.99	0.995	0.902	220
Weighted Avg	0.99	0.99	0.99	0.995	0.902	220

C. Confusion Matrix

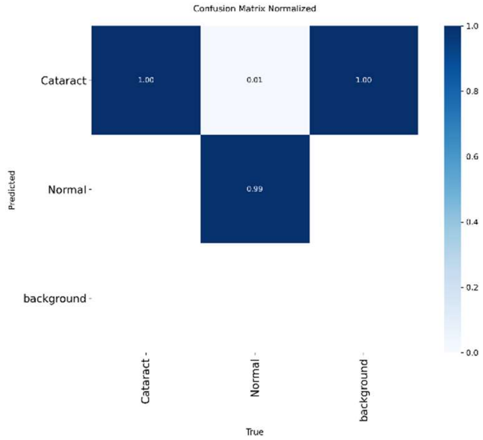

Fig. 3. Confusion matrix on test set.

D. Single-Image Inference Example

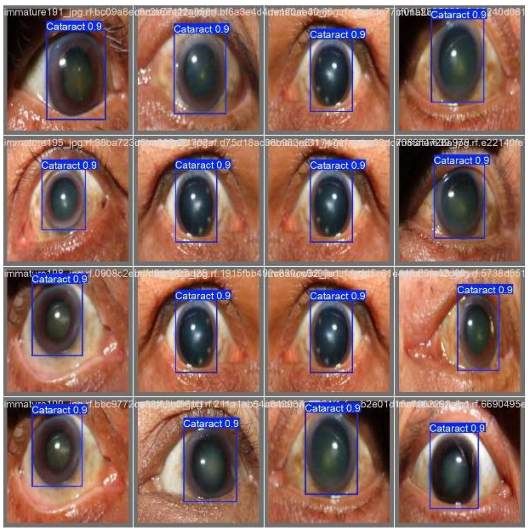

Fig. 4. Example single-image classification result.

V. CONCLUSION

In this study, we developed a deep learning-based framework for automated cataract detection and classification using the YOLOv11 architecture. The model effectively exploits YOLOv11's real-time object detection and adaptive anchor-free mechanisms to achieve robust performance on ocular images. Trained on the *Kataract-Object-Detection* dataset, it achieved an impressive **99% accuracy** with high precision, recall, and F1-scores, proving its reliability for clinical diagnostic applications.

The strong detection performance of YOLOv11 aligns with recent advances in AI-driven medical imaging [1]–[6], highlighting its potential to improve the accuracy and scalability of ophthalmic diagnostics. Future work will focus on expanding the dataset for multi-grade cataract classification, incorporating explainable AI for interpretability, and optimizing the model for deployment on mobile and embedded devices to enable real-time, point-of-care cataract screening.

References

[1] World Health Organization, *World Report on Vision*. Geneva: WHO, 2023.

[2] K. D. Frick *et al.*, “Economic impact of visual impairment and blindness,” *The Lancet Global Health*, vol. 11, no. 3, pp. e456–e467, 2023.

[3] *The Lancet Global Health*, “Economic impact of visual impairment and blindness,” [Online].

[4] American Academy of Ophthalmology, “What are Cataracts?,” [Online]

[5] L. Wang *et al.*, “Deep learning for cataract detection and grading using retinal images,” *Nature Medicine*, vol. 28, no. 4, pp. 775–781, 2022.

[6] Mayo Clinic, “Cataracts – Diagnosis and Treatment,” [Online].

[7] Y. Zhang *et al.*, “Cataract-1K: A large-scale dataset for computer-assisted cataract surgery,” *Medical Image Analysis*, vol. 81, p. 102549, 2023.

[8] D. Lam, S. Rao, V. Ratra, Y. Liu, P. Mitchell, J. King, M.-J. Tassignon, J. Jonas, C. Pang, and D. Chang, “Cataract,” *Nature Reviews Disease Primers*, vol. 1, p. 15014, 2015.

[9] A. A. Nizami, B. Gurnani, and A. C. Gulani, *Cataract*. [Updated Feb. 27, 2024]. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2025 Jan.

[10] P. L. Lahari, R. Poola, L. Gorrepati, and S. Yellampalli, “Real-Time Cataract Diagnosis with GhostYOLO: A GhostConv-enhanced YOLO Model,” *Engineering, Technology and Applied Science Research*, vol. 15, pp. 22945–22952, 2025.

[11] H. Xie *et al.*, “Deep learning for detecting visually impaired cataracts using fundus images,” *Frontiers in Cell and Developmental Biology*, vol. 11, p. 1197239, Jul. 2023. doi: 10.3389/fcell.2023.1197239.

[12] S. Manjunath and G. S. Nath, “Identification and classification of cataract stages in old age people using deep learning algorithm,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 2767–2772, 2019.

[13] *Epidemiology of Risk Factors for Age-Related Cataract*, [Online].