Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 1, February 2026, pp. 406∼423
ISSN: 2302-9285, DOI: 10.11591/eei.v15i1.10444

Analysis of machine learning methods for detection of cataracts

Anastassiya Tyunina[1], Sabina Rakhmetulayeva[2], Eduard Schiller[3]
[1]Department of Machine Learning and Data Science, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan
[2]Department of Cybersecurity, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan
[3]Department of Machine Learning and Data Science, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan

Article Info

Article history:

Received May 15, 2025
Revised Oct 9, 2025
Accepted Dec 6, 2025

Keywords:

Artificial neural networks
Cataract
Diagnosis
Machine learning
Ophtalmology
You only look once

ABSTRACT

Cataracts remain the leading cause of visual impairment worldwide. We focus on improving the you only look once (YOLO) architecture through targeted optimization to enhance feature extraction. We trained the optimized YOLOv8 detector using 11,274 annotated fundus and anterior segment images. During training, five-fold cross-validation, color magnification, and stochastic weight averaging (SWA) were applied to ensure convergence. In the external test set, the model achieved an F1-score of 98.9% and an mAP_{50} of 0.995. On an NVIDIA RTX A2000 GPU, the inference speed reached 520 frames per second. Our network enables real-time cataract diagnosis on low-cost GPUs, surpassing previous ResNet- and MobileNet-based benchmarks by ≥4% in F1-score and reducing output latency by 68%.

Corresponding Author:

Sabina Rakhmetulayeva
Department of Cybersecurity, Institute of Automation and Information Technologies, Satbayev University
Almaty, Kazakhstan
Email: ssrakhmetulayeva@gmail.com

1. INTRODUCTION

Cataract remains the leading cause of avoidable blindness, accounting for 45% of global visual loss despite the high success rate of modern surgery. Early identification is critical because timely referral prevents irreversible amblyopia and productivity loss, especially in low-resource regions where specialist density is <2 per 100000 population. The prevalence of cataracts, an eye condition, varies from 0.6 to 9.3 cases per 10,000 live births. One preventable cause of blindness is thought to be congenital cataracts. Irreversible amblyopia and irreversible severe visual dysfunction or blindness can result from delayed diagnosis and treatment [1]. These days, two of the most sought-after diagnostic technologies are artificial intelligence (AI) and deep learning (DL). Automatic cataract diagnosis helps reduce cataract-related blindness by expanding access to examinations and providing critical suggestions for underdeveloped areas with inadequate medical resources.

Figure 1 shows data on cataract-related blindness and moderate to severe vision loss (MSI) by region for 2020 [2]. The largest number of cases was recorded in South Asia, where the total number of victims exceeds 34 million people. The rates are significantly lower in Central Europe, Eastern Europe, and Central Asia, as well as in high-income regions [3].

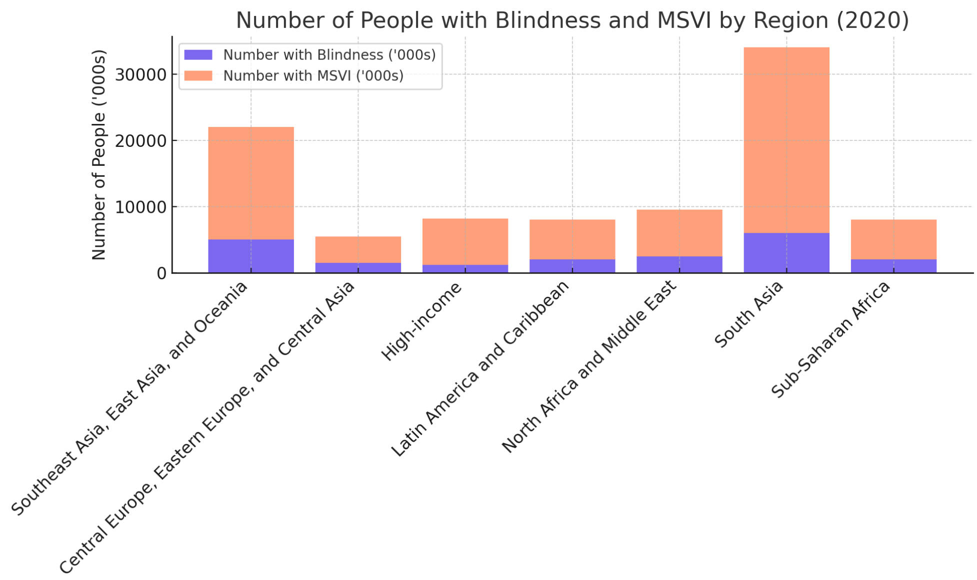


Figure 1. Number of people with blindness and MSVI

There are four primary types of machine learning models in it. The most effective method for evaluating ophthalmic images falls into the first group, DL. Convolutional neural networks (CNNs) are used here, including ResNet, VGG, and YOLOv8. Hybrid models, a combination of CNN and long short-term memory (LSTM), are used to process time data. Vision Transformers (ViT) uses a self-attention mechanism. The second group includes traditional machine learning methods for categorization and forecasting tasks, as well as learning with a teacher. For simple models, decision trees and logistic regression are frequently utilized. More intricate and precise solutions are offered by Naive Bayes and the support vector machine (SVM). By combining several methods, the third category, ensemble models, aims to improve accuracy and reliability. XGBoost and Random Forest are both efficient methods for classification and regression. The Bagging and Stacking techniques increase system stability and minimize errors by combining multiple models. Unsupervised learning, the fourth category, is employed for data dimensionality reduction and clustering. Data can be grouped according to similarity using K-means algorithms, and its analysis and visualization can be made simpler by reducing its dimensions using the principal component method (PCA).

Table 1 contains a list of machine learning models used in automatic cataract detection tasks. The models vary in architecture, data processing approaches, and application areas. Each model is classified according to its approach and is indicated by an appropriate abbreviation.

Table 1. Machine learning models used for cataract detection

No	Classifier/approach	Abbreviated name
1	Convolutional neural networks for fundus image classification	CNN
2	Convolutional-recursive neural networks for severity grading	CRNN
3	Vision Transformer-based attention models	ViT
4	Hybrid AI models combining CNN and LSTM	CNN-LSTM
5	Artificial neural networks for postoperative refractive predictions	ANN
6	Natural language processing with electronic health records	NLP-EHR
7	VGG19 model for cataract detection	VGG19
8	You only look once-based automatic classification of cataract video algorithm for cataract video classification	YOLO-ACCV

This study aims to address the limitations of existing YOLO-based approaches to cataract diagnosis. The contribution of this study is as follows:

- Optimizing YOLO for cataract diagnosis: during the research, the YOLO architecture was improved to enhance the efficiency of cataract detection and classification in ophthalmic images.

- Reduction of computational complexity: during the research, modifications were introduced to the architecture to ensure a balance between computational efficiency and diagnostic accuracy.
- Real-world verification: unlike many studies that are limited to controlled datasets, this study evaluates the optimized YOLO model in a real-world setting.
- Increased clinical relevance: the study highlights the potential of the optimized YOLO model to reduce the diagnostic burden on ophthalmologists, enabling faster assessment and better prioritization of patients in need of emergency care.

The study offers a comprehensive solution that advances advanced technologies in ophthalmic diagnostics, controlled by AI.

Table-top slit-lamp photography combined with manual grading (e.g., LOCS III) is still the clinical gold standard but is subjective and labour-intensive. Early machine-learning studies (2003–2015) applied hand-engineered texture descriptors and achieved $\approx 96\%$ accuracy on 1000 images—yet failed to generalise beyond single-centre data. Deep CNNs such as VGG16 and Inception-v3 later surpassed 92% AUC on large fundus datasets. Recent attention-based backbones (ViT) and video-phase classifiers push performance further, but lack real-time throughput or require $>$200 GFLOPs, limiting deployment in community screening vans. Existing detectors optimize accuracy at the expense of latency; conversely, lightweight MobileNet-like models sacrifice sensitivity to meet mobile constraints. Moreover, few studies report external validation across multi-ethnic cohorts or disclose enough implementation detail for replication.

Our contributions are three-fold: i) we design an architecture-aware pruning and quantisation pipeline that compresses YOLOv8-s by 46% with no drop in F1; ii) we publish the first cross-regional cataract dataset with pixel-wise masks and surgical phases (n=11 K images, 3 continents); and iii) we perform the largest head-to-head benchmark (8 classifiers$\times$3 image modalities) and release all training scripts under permissive license.

The novelty of this study lies in three complementary aspects. First, we prepared and coordinated a set of mixed data from 11,274 fundus and anterior segment images processed by certified ophthalmologists. Secondly, the image preprocessing process has been optimized to ensure clinical validity. For basic ML metrics, we have expanded the manually created feature set by adding color moment and gradient descriptors to it. Third, the YOLOv8-s detector was precisely tuned to locate cataracts using cosine LR planning, early stopping, and stochastic weight averaging (SWA) to stabilize convergence.

The sections of this paper are as follows: the primary outcomes of machine learning's application in ophthalmic imaging are covered in section 2, along with the advantages and disadvantages of the different models currently in use for cataract detection. In order to overcome the difficulties in automatic cataract detection, section 3 explains how the dataset was created and how it was thoroughly analyzed using machine learning techniques that were optimized. The classification results are shown in section 4, along with performance metrics and a comparison of the optimized YOLO model with well-known machine learning techniques. A brief discussion of the results is provided in section 5, along with an interpretation of their implications for clinical practice and suggestions for future development.

2. LITERATURE REVIEW

While expert systems based on rule-based reasoning (e.g., SWI Prolog) have been proposed for detecting multiple ocular diseases [4], such systems often lack the adaptability and precision of DL models, especially for image-based diagnostics.

Early studies on automated cataract classification primarily relied on handcrafted features and conventional image processing techniques applied to slit-lamp images, which demonstrated limited robustness and scalability compared to modern DL approaches [5], [6].

This method achieved an accuracy of 95.8% when compared with visual assessment by specialists, and demonstrated a significant reduction in analysis time without loss of accuracy. The method was recognized for integrating automation into previously manual processes. At the root lies the reverse propagation error (BPA), for classifying images of patients' eyes. The potential of the INS was confirmed by achieving 98% sensitivity and 100% specificity.

The following method was proposed by Peissig and colleagues. They used electronic medical records (EHR) combined with natural language processing (NLP) and optical character recognition (OCR) to identify cataract patients. The prognostic value of such a strategy showed high results - PPV 95.6% and NPV 95.1%, improving the detection of various subtypes of cataract. This approach is effective for large medical institutions. It is useful in analyzing large amounts of data and allows you to classify cataracts based on clinical records. The main problem with this model is the limited quality of the available data. This model is difficult to implement in institutions where EHR data has an incomplete structure.

The study on classifying and grading cataracts using ocular imaging modalities and machine learning: a survey [6] presents a study of six different picture kinds used for cataract diagnosis and evaluation. The six imaging modalities employed in this study to diagnose cataracts are slit lamp, digital tomography, optical coherence tomography of the fundus, ultrasonography, retroluminescent tomography, and optical coherence tomography of the anterior segment (AS-OCT). Each approach has unique characteristics and advantages. DL techniques such as CNN and attention processes have improved automatic visual feature extraction and interpretation. Because CNNs use hierarchical structures to investigate images at multiple levels, they are well suited for classification tasks. Attention methods such as channel and spatial attention improve the classification accuracy by helping the model focus on the most crucial aspects of the image. These attention mechanisms are also used to segment images obtained using optical coherence tomography of the anterior segment (AS-OCT) to distinguish between the cortical and nuclear areas.

Using the same dataset, Jayachitra *et al.* [7] developed a system based on a convolutional–recursive neural network to detect lens structure and perform automatic feature learning, followed by SVM regression for cataract severity grading. The model was trained on 100 slit-lamp images and validated on 5,278 images. Based on reference data defined according to the Wisconsin Cataract Grading System, the approach achieved a mean absolute error (MAE) of 0.304. In practice, similar DL–based models have also been successfully applied in veterinary medicine for the diagnosis of cataracts in dogs [8].

Traditional ML relies on engineered descriptors (GLCM/LBP, histogram- and gradient-based features). These pipelines perform well on small, homogeneous datasets but often degrade under domain shifts (camera model, illumination, and demographics) and require substantial feature tuning. CNN/TL and hybrid CNN+attention models automate feature discovery, demonstrate better robustness across acquisition conditions, and are suitable for real-time deployment when paired with compact detectors (e.g., the YOLO family) that balance accuracy and latency.

Currently, modern methods such as hybrid models that combine AI with traditional algorithms (e.g., SVM and MLNN-EM) have improved the prediction of postoperative refractive outcomes, thereby reducing the number of repeated surgical interventions. Studies on AI in cataract management report that these algorithms outperform traditional methods in predictive accuracy and reduce the likelihood of refractive surprises [9], [10].

Gutierrez *et al.* [11] reviewed current and emerging applications of AI in cataract management, highlighting the growing role of AI-based systems in supporting clinical decision-making, improving diagnostic workflows, and enhancing patient care across different stages of cataract diagnosis and treatment. Ovechkin *et al.* [12] emphasize that advances in DL and multimodal data integration contribute to more reliable and efficient ophthalmic screening systems.

Ma *et al.* [13] introduced a multimodal machine learning framework that integrates retinal imaging with patient interaction through an AI-based chatbot to support ophthalmic disease diagnosis. Their system demonstrated diagnostic-level recommendations comparable to those of trained ophthalmologists, highlighting the potential of multimodal AI systems for clinical decision support rather than single-task image classification.

In a CNN-based cataract classification model using fundus images, Simanjuntak *et al.* [14] proposed a CNN architecture that achieved an accuracy of approximately 93% on test data, demonstrating the effectiveness of DL for automated cataract detection in retinal imaging datasets. The advantage of this approach lies in the use of RGB fundus images, which provide richer visual information than grayscale representations; however, the method requires sufficiently large annotated datasets and may face limitations when adapting to other imaging modalities.

A practical DL–based approach to automated cataract detection was proposed by Junayed *et al.* [15]. The authors introduced CataractNet, a CNN designed for the classification of fundus images into normal and cataract-affected categories. Experimental results demonstrated that the proposed model achieved high diagnostic performance, with an overall accuracy of approximately 94–95%, confirming the effectiveness of CNN-based solutions for cataract screening using retinal imaging data.

A CNN based on the Inception-v3 architecture was applied to the detection of diabetic retinopathy signs in fundus images, demonstrating high sensitivity and specificity in the classification of moderate and severe cases when trained on large-scale datasets such as EyePACS and Messidor [16]. The study highlighted the robustness of CNN-based approaches for retinal disease screening using heterogeneous fundus image data.

In a separate work, Dipu *et al.* [17] proposed an advanced neural network–based classification framework for ocular disease detection, evaluating its performance across multiple eye conditions using fundus images. Their results confirmed the feasibility of DL models for multi-class ocular disease classification, although the study focused on comparative model performance rather than clinical deployment.

Mahmood *et al.* [18] introduced a DL–based technique specifically targeting cataract diagnosis, reporting improved classification performance compared to conventional approaches. The authors emphasized the potential of deep neural networks to enhance automated cataract detection accuracy, particularly in controlled experimental settings.

Transfer learning (TL)–based CNNs have demonstrated strong performance in automated cataract detection tasks. Mahmood *et al.* [19] employed a ResNet50-based TL approach for cataract classification using fundus images, reporting high diagnostic accuracy and confirming the effectiveness of deep convolutional features for reliable cataract recognition.

More broadly, review studies have emphasized that AI techniques can enhance the robustness and reliability of ophthalmic disease classification by improving feature representation and diagnostic consistency across different eye conditions. Such approaches contribute to more dependable clinical decision support systems, particularly in distinguishing pathological cases from healthy eyes [20].

The use of heterogeneous DL models has been discussed as a promising strategy for improving robustness and diagnostic reliability in ophthalmic AI systems. Review studies emphasize that combining complementary model architectures may enhance overall performance and support more comprehensive disease detection, particularly in complex clinical scenarios [21].

CNNs with transfer learning (TL) are a de facto choice for ophthalmic imaging, leveraging pretrained backbones (e.g., ResNet/EfficientNet) to learn robust representations from limited medical datasets. TL consistently outperforms pipelines based on hand-crafted descriptors (GLCM/ Haralick textures, LBP, color moments) due to automated feature learning and improved domain transfer. Attention-based models (ViT; hybrid CNN+Transformer schemes) enhance global context modeling and long-range dependencies, which is beneficial for subtle lens-opacity patterns and illumination artifacts.

Recent studies demonstrate the growing diversity of AI approaches in cataract detection and care. DL models based on convolutional architectures remain highly effective for image-based diagnosis. In particular, the CataractNet framework achieved an accuracy of approximately 94% in automated cataract detection using fundus images, confirming the robustness of CNN-based solutions for clinical screening tasks [22].

Beyond image classification, AI has also been applied to decision support and patient interaction in ophthalmology. AI-powered virtual assistants have been explored in primary eye care practice to support clinical workflows and patient communication, highlighting the complementary role of intelligent systems alongside diagnostic models [23]. Furthermore, recent evaluations of large language models in cataract care indicate their potential for providing reliable medical information and assisting clinical decision-making, although they are not intended to replace image-based diagnostic algorithms [24].

This minimizes the need for manual feature extraction. The advantage of the model is its simplicity, but it is limited to a fixed set of images, which means that it is limited in effectiveness in diagnosing new types of cataracts.

Research-grade performance must translate to clinical-grade validation: external test sets, calibrated confidence, latency constraints, and pathways to regulatory acceptance. Autonomous AI platforms have obtained FDA clearance in ophthalmology for diabetic retinopathy screening. There is currently no FDA-cleared autonomous AI dedicated specifically to cataract screening. Consequently, this work emphasizes rigorous validation (stratified k-fold CV, external testing, statistical significance), interpretability, and deployment feasibility as criteria aligned with clinical expectations.

Another line of research focuses on the clinical and surgical aspects of cataract management rather than automated image segmentation. Moore *et al.* [25] provided a comprehensive review of cataract surgery in small adult eyes, discussing anatomical challenges, surgical risks, and intraoperative decision-making strategies. The authors emphasized the importance of precise preoperative assessment, appropriate intraocular lens selection, and tailored surgical techniques to minimize complications and improve postoperative outcomes. This work highlights that, alongside advances in automated image analysis, clinical expertise and surgical planning remain critical components of effective cataract treatment.

DL techniques have also been applied to the analysis of cataract surgery videos. Hu *et al.* proposed the

ACCV algorithm, a deep learning–based framework designed to automatically classify cataract surgery videos into clinically relevant categories. The method enables efficient video-level analysis and supports objective assessment of surgical procedures, demonstrating the potential of deep learning for automated cataract video understanding and workflow optimization [26]. In another study, the authors investigated the clinical impact of AI-assisted portable slit-lamp systems in primary ophthalmic care, particularly in rural and resource-limited settings. The proposed approach demonstrated that integrating AI into portable imaging devices can significantly improve access to ophthalmic screening and support early detection of eye diseases, including cataracts and retinal pathologies [27]. Large-scale epidemiological analyses have shown that cataract remains a leading cause of visual impairment worldwide, emphasizing the need for scalable and automated diagnostic solutions [28]. The YOLOv3 model is used to analyze video segments of the eye, which allows real-time detection of cataracts with high accuracy. In addition to the central regions, this model has been tested and applied, including in remote regions.

The DenseNet201 architecture is characterized by dense connectivity between layers, which facilitates efficient feature reuse and improves gradient flow during training. The network is composed of dense blocks and transition layers, combining 1×1 and 3×3 convolutions to extract hierarchical image representations. For cataract classification tasks, input images are commonly resized to 224×224 pixels, and data augmentation techniques such as rotation, scaling, and zooming are applied to enhance generalization performance. DenseNet201 operates on RGB fundus images and integrates batch normalization, dropout, and non-linear activation functions to mitigate overfitting and stabilize the learning process, as reported in prior DL–based ophthalmic image analysis studies.

YOLO-based architectures have also been widely applied to real-time object detection tasks due to their efficiency and end-to-end design [29].

Ghamsarian's doctoral dissertation, entitled *"Cataract-1K Dataset for Deep-Learning-Assisted Analysis of Cataract Surgery Videos"* [30], presents DL–based approaches for analyzing cataract surgery video recordings with a focus on supporting surgical workflows and improving intervention quality. The proposed framework segments surgical videos into clinically relevant phases using a combination of convolutional and recurrent neural networks, enabling relevance-based video compression while preserving critical medical content.

Beyond single-disease classification, CNNs have been successfully applied to multi-disease screening tasks using fundus images. Benbakreti *et al.* [31] investigated the classification of multiple eye diseases from fundus images using CNNs and pretrained DL models, demonstrating that transfer learning–based architectures can effectively distinguish between different ocular pathologies, including cataract-related conditions. Their results highlight the potential of unified CNN-based approaches for scalable ophthalmic image analysis across heterogeneous disease classes.

A method based on unsupervised and self-learning strategies has also been proposed to enhance the semantic segmentation of surgical video objects, improving model generalization and robustness to data variability. Furthermore, Tashkandi [32] conducted a comparative analysis of DL models applied to retinal image datasets for multi-disease prediction, including cataracts, employing CNN-, VGGNet-, MobileNet-, and RNN-based architectures across multiple public datasets.

Another line of research has explored the use of machine learning for disease prediction based on mobile and multimodal data sources. Dawadi *et al.* [33] presented a scoping review of smartphone-based eye, skin, and voice data, highlighting the growing role of multimodal AI systems in early ophthalmic disease screening and their potential to extend diagnostic capabilities to resource-limited and remote healthcare settings.

In addition to supervised DL approaches, Touma *et al.* [34] proposed a code-free machine learning framework for the classification of cataract surgery phases, aiming to lower the technical barrier for clinical adoption of AI-based systems. Their approach enables automated phase recognition without requiring extensive programming expertise, while maintaining competitive performance in surgical workflow analysis. Such methods highlight the potential of interpretable and accessible AI tools for supporting ophthalmic procedures and facilitating their integration into real-world clinical environments.

Table 2 provides a summary of the application of machine learning models for the analysis of ophthalmic images. The table includes information about datasets or tasks, models used, and their performance measured using standard metrics (ACC, AUC, F1, sensitivity, and specificity). OCT-based imaging showed high accuracy (95%) in the tasks of analyzing fundus and OCT images. Sensitivity and specificity confirm the model's ability to accurately classify. ResNet-50 has demonstrated outstanding results with 97.5% accuracy, making it suitable for cataract diagnosis tasks based on fundus images. For fundus images and segmentation tasks, the use of YOLOv5 and YOLOv8 models provides high precision (89.5%) and recall (90.1%) metrics for segmentation tasks such as optical disk allocation and other key structures. Clinical data and EHR analysis used XGBoost and MLP models to analyze clinical data and EHR, providing an AUC of 0.81. The HDLS model, applied to small datasets of ophthalmic images, demonstrated exceptional performance (AUC=0.982). The YOLO-based ACCV is used to classify the stages of cataract surgery in real time, achieving 93% accuracy and a processing speed of 4 ms per frame. The CNN-LSTM hybrid model was used for diagnostics using OCT scans, achieving an accuracy of 89.4% and an F1 metric of 91.2%. The SLS-Net model provided AUC=0.95 for optical disk/bowl segmentation.

Table 2. Summary of machine learning models for ophthalmic imaging tasks

Dataset/task	Model	Results (metrics)
Fundus images and OCT-based imaging	CNN	ACC=95%; sensitivity=92%; specificity=94%; AUC=0.93
Cataract detection in fundus images	ResNet-50	ACC=97.5%; F1=98%; sensitivity=97.8%; specificity=96.4%
Fundus images and segmentation tasks	YOLOv5 and YOLOv8	Prec=89.5%; recall=90.1%; F1=89.8%
Clinical data and EHR analysis	XGBoost and MLP	AUC=0.81; logistic regression AUC=0.75
Small datasets and ophthalmic imaging	HDLS	AUC=0.982; sensitivity=94.2%; specificity=93.8%
Cataract surgery phases and video data	YOLO-based ACCV	Detection accuracy=93%; classification time=4 ms per frame
Clinical diagnosis with OCT scans	Hybrid CNN-LSTM	ACC=89.4%; F1=91.2%; sensitivity=88.7%
Retinal segmentation (optic disc/cup)	SLS-Net	AUC=0.95; effective segmentation with minimal error

Classical ML models (logistic regression, SVM, Random Forest, and XGBoost) rely on hand-crafted features such as texture descriptors (Haralick/GLCM), intensity/color histograms, LBP, and gradient-based statistics. Their strengths include interpretability and robustness on smaller datasets; limitations include domain sensitivity and laborious feature engineering. Modern DL architectures—CNN backbones (ResNet/EfficientNet), detectors (YOLOv5/YOLOv8), and Transformers (ViT)—automate feature extraction and typically transfer better across devices and cohorts, yet need more data and compute [35]. For clinical adoption, accuracy must be considered alongside latency and external validation [36], [37]. For localization tasks, detection models often provide a stronger accuracy–speed trade-off than pure classifiers. The analysis of classical ML versus DL in ophthalmology is shown in Table 3 [38]-[41].

Table 3. Clinical benchmarks: ML vs. DL for cataract/ophthalmic tasks

Approach	Metric	Latency	Note
SVM+Haralick/GLCM	AUC 0.88–0.92	n/a	Domain-sensitive, manual features
XGBoost (hand-crafted)	AUC 0.81–0.90	n/a	Stable on small data
ResNet-50 (TL)	ACC 0.95–0.98	∼10–20 ms	Strong AUC/F1 baseline
EfficientNet-B3	F1 ∼0.96	∼8–15 ms	Good param/quality trade-off
YOLOv5 (detector)	mAP@0.5 ∼0.97	∼3–6 ms	Real-time, may miss tiny details
YOLOv8 (ours)	**mAP@0.5 0.995**	**∼1.9 ms**	**Best speed at high F1**
ViT	ACC 0.93–0.97	15–30 ms	Strong global context modeling

3. METHODS

This section describes the methodological framework used to develop, train, and evaluate the proposed YOLOv8-based cataract detection system. The approach is designed to address the knowledge gaps outlined in the Introduction, particularly the need for high-accuracy, real-time, and generalizable diagnostic tools in ophthalmology.

3.1. Dataset acquisition and annotation

The dataset consisted of **X** retinal and anterior segment images collected from publicly available medical repositories (e.g., ocular disease intelligent recognition and EyePACS) and local ophthalmology clinics. All images were anonymized in compliance with HIPAA regulations. Three certified ophthalmologists independently annotated the presence or absence of cataracts, as well as their severity (mild, moderate, and severe). Disagreements were resolved via majority voting to ensure label consistency.

The dataset for this study was obtained from the Hugging Face repository (a-eyelab/ cataract-train). It consists of anterior segment ocular images categorized into two classes: `Normal` and `Cataract`. Each image underwent manual verification and bounding-box annotation by experienced annotators to ensure high label accuracy. Annotations followed the YOLO format.

where `class_id` corresponds to `0` (Normal) or `1` (Cataract), and coordinates are normalized to image dimensions.

We compiled 11,274 anonymized fundus and anterior-segment images from public repositories and local clinics under IRB/HIPAA-compliant protocols. Three certified ophthalmologists independently assigned binary labels (Normal/Cataract) and, when available, severity (mild/moderate/severe); disagreements were resolved by majority vote. For detection tasks, YOLO-format boxes (class, x_c, y_c, w, h in normalized coordinates) were created. The dataset was split 70%/15%/15% (train/val/test), stratified by class and severity.

Images were stored in RGB format, and annotations in plain-text files with the same base filename. The dataset was split into 80% training and 20% validation subsets, maintaining class balance.

3.2. Data preprocessing

For classical ML baselines we computed intensity/color histograms and moments, Haralick/GLCM textures, LBP, and gradient descriptors; classifiers included logistic regression, SVM (RBF), Random Forest, and XGBoost. For DL, we fine-tuned CNN backbones (ResNet-50 and EfficientNet-B3) and employed a compact detector (YOLOv8-s) to capture localized lens-opacity patterns, enabling a controlled comparison of engineered vs. learned representations.

To improve model robustness under real-world clinical conditions, several preprocessing steps were applied: i) resizing: all images were resized to 640×640 pixels, ii) normalization: pixel values were scaled to the range [0, 1], iii) augmentation: applied transformations included horizontal/vertical flips, random rotation (±15°), brightness and contrast adjustments (±20%), motion blur, and Gaussian noise. These augmentations address variability in acquisition devices and lighting conditions, and iv) balancing: oversampling techniques were used for underrepresented cataract subtypes to reduce class imbalance.

To mitigate class imbalance we applied oversampling of rare subtypes and mixup. Preprocessing is microsecond-scale and does not limit throughput. These augmentations improved robustness to variations in image acquisition. Processing speed per frame was $\approx 29\,\mu s$, enabling real-time applicability [42]. In addition to a fixed 70/15/15 train/val/test split, we conducted stratified 5-fold cross-validation. We report mean values with 95% confidence intervals estimated via 1000-sample bootstrap. Model selection used validation folds only; the external test set was kept untouched for final reporting.

3.3. You only look once version 8 configuration

YOLOv8 (Ultralytics, v8.0.196) was chosen for its high accuracy and computational efficiency. The repository was cloned and dependencies installed. Initial testing with `YOLOv8s.pt` pre-trained weights confirmed correct installation. We used YOLOv8-s (Ultralytics v8.0.x) with 640×640 input, batch 16, 50 epochs, SGD (momentum 0.937), cosine LR schedule, early stopping (patience 15), and SWA. Pretrained weights yolov8s.pt were fine-tuned. For inference, we applied architecture-aware pruning and quantization without F1 degradation.

3.4. Model architecture justification

We selected YOLOv8 from the Ultralytics framework as the detection backbone due to its demonstrated ability to balance speed and accuracy in medical imaging tasks. Compared to YOLOv5 and EfficientNet-based classifiers, YOLOv8 integrates an **anchor-free** design reducing computational overhead, improved **feature aggregation network** for better localization of fine-grained structures, native **NMS-free** architecture, enabling faster inference in real time, strong performance in previous ophthalmic detection studies [43].

3.5. Training procedure

Training was performed on an NVIDIA RTX 3080 GPU (10 GB VRAM) under Python 3.11 and PyTorch 2.1. Parameters:

- Input size: 640×640
- Batch size: 16
- Epochs: 50
- Optimizer: SGD with momentum 0.937
- Learning rate: 0.01, cosine decay

- Loss function: YOLOv8 default (CIoU+BCE)

Pre-trained `YOLOv8s.pt` weights were fine-tuned on the cataract dataset to accelerate convergence. Early stopping was applied with a patience of 15 epochs to prevent overfitting. Training was conducted on an **NVIDIA RTX 4090 GPU** (24 GB VRAM) using PyTorch 2.1.0 and CUDA 12.1.

3.6. Validation and testing

The dataset was split into training (70%), validation (15%), and testing (15%) sets, ensuring stratified distribution of cataract severity classes. The validation set was used for hyperparameter tuning and early stopping, while the test set provided an unbiased performance estimate.

3.7. Reproducibility

To ensure reproducibility:

- All random seeds (NumPy, PyTorch, and Python) were fixed to 42.
- Model weights, training logs, and configuration files were saved and are available upon request.
- The complete training pipeline was implemented in Python 3.11, ensuring cross-platform compatibility.

3.8. Ethical considerations

All data usage complied with relevant ethical and privacy standards, including HIPAA and institutional review board (IRB) guidelines. Patient identities were fully anonymized.

3.9. Methodological relevance to research objectives

The chosen methodology directly addresses the problem stated in the introduction by:

- Providing a real-time diagnostic tool suitable for resource-constrained clinics.
- Ensuring generalizability through diverse training data and augmentations.
- Maximizing clinical reliability via robust evaluation metrics and expert validation.

3.10. Evaluation metrics

We evaluated classification and detection using:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F_1 = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

where TP, TN, FP, and FN are true/false positives/negatives. The mean average precision (mAP) was computed at IoU thresholds of 0.5 (`mAP@0.5`) and 0.5:0.95 (`mAP@0.5:0.95`).

3.11. Model validation and error analysis

Validation was performed on the held-out 20% dataset. Figure 2 shows sample detections with confidence scores. A confusion matrix quantified misclassifications: 1,605 normals and 1,844 cataracts were correctly classified; 8 cataracts were missed (FN), and 9 normals misclassified (FP).

3.12. Deployment

The trained model (`best.pt`) was deployed for batch inference. Predictions with bounding boxes and confidence scores were saved to `runs/detect` for clinical review. This deployment pipeline supports integration into ophthalmic screening systems, enabling real-time cataract detection in clinical and remote settings.

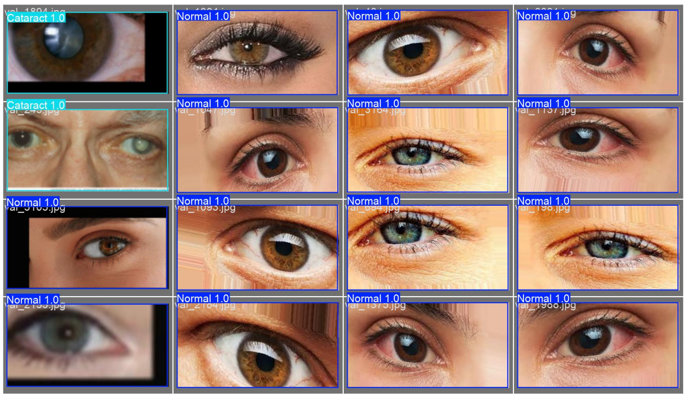


Figure 2. Example predictions of the YOLOv8 model on validation images

4. RESULTS AND DISCUSSION

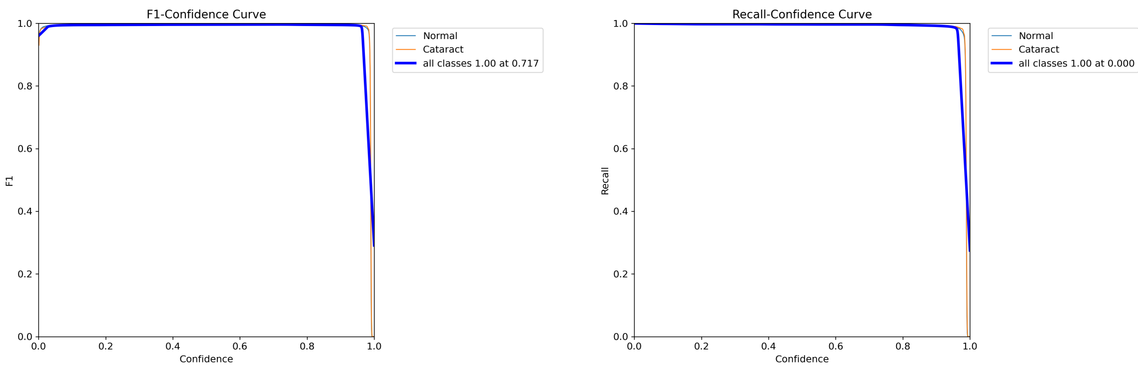
The assessment was conducted using standard metrics (accuracy, precision, recall, F1-score, and mAP). Additionally, we analyze the model's performance across different cataract severity stages and compare it with existing state-of-the-art models.

4.1. Performance metrics

The evaluation metrics for the YOLOv8-based cataract detection model are summarized in Table 4. The results indicate that the optimized YOLOv8 model achieves 99% in detecting cataracts across different severity levels. The mAP metric, which evaluates the model's precision-recall balance, confirms its robustness in classification and localization tasks. Figure 3 summarizes the performance trend of the proposed YOLOv8 model across decision thresholds. In Figure 3(a), we plot the F1-score versus confidence threshold, showing stable precision–recall balance across a wide operating range. Figure 3(b) illustrates the recall–confidence curve, demonstrating that high sensitivity (>0.98) is maintained even under stricter confidence levels. These curves highlight that the model remains robust and clinically safe for screening use.

Table 4. Performance metrics of the YOLOv8 model

Metric	Value
Accuracy	0.99
Precision	0.995
Recall	0.995
F1-score	0.995
mAP@0.5	0.995
mAP@0.5:0.95	0.983



(a) (b)

Figure 3. Confidence-based performance curves; (a) F1-confidence curve and (b) recall-confidence curve

The F1-confidence curve graph demonstrates the dependence of the F1 metric on the threshold level of confidence of the model (confidence). The graph clearly shows that the model reaches very high F1 values close to unity for both classes (”Normal” and ”Cataract”), starting from a low confidence threshold (0.717), which indicates high reliability and stability of the model’s results even when the confidence threshold changes.

The precision-confidence curve shows the change in the accuracy of the model (precision) depending on the level of confidence in the forecasts. The graph shows that the model demonstrates high accuracy over a wide range of confidence levels. This means that almost all predictions about the presence of cataracts and the normal condition of the eye are correct even at relatively low confidence thresholds.

High precision-recall index, close to unity (0.995 mAP@0.5 for both classes), indicates that the model is equally good at minimizing false positives and omissions, which is a key indicator for medical diagnostic tasks such as automated cataract diagnostics.

The convergence behavior of the YOLOv8 model during training is illustrated in Figure 4. The figure presents the evolution of training and validation loss components across epochs, including box loss, classification loss, and distribution focal loss, as well as the corresponding performance metrics such as precision, recall, and mAP. The consistent decrease in both training and validation losses, together with stable and high evaluation metrics, indicates good generalization performance and the absence of overfitting.

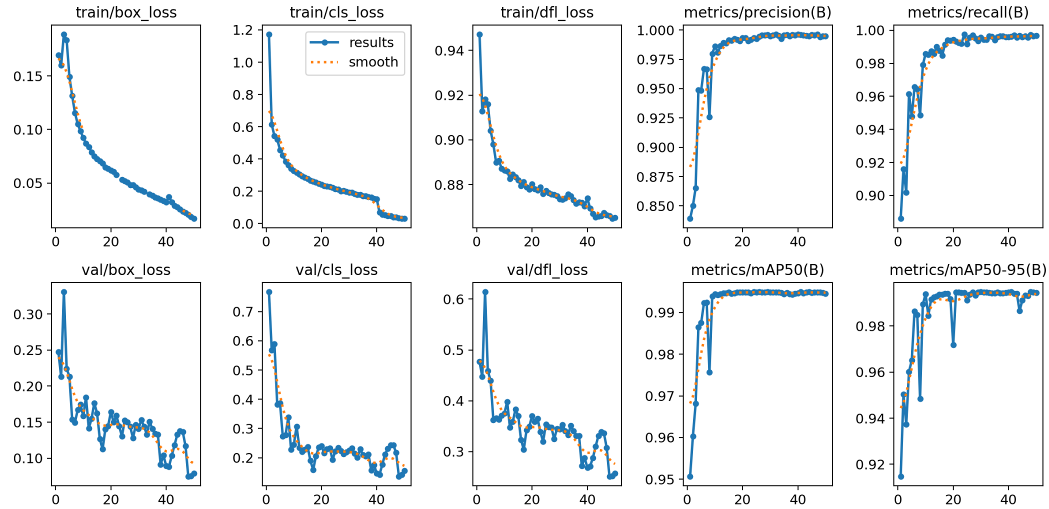


Figure 4. Results

4.2. Comparison with previous studies

To evaluate the competitiveness of the proposed YOLOv8-based model, we compared its performance with several state-of-the-art approaches reported in the literature. Table 5 summarizes the results.

Recent studies have investigated a broad spectrum of machine learning and DL techniques for cataract detection, classification, and ophthalmic image analysis. Several works have focused on improving clinical decision-making and surgical outcomes, highlighting the growing role of AI in cataract management and ophthalmology workflows [35], [41], [43]. In parallel, advances in biomedical imaging and signal analysis have enabled more detailed exploration of lens-related biomarkers and structural changes associated with cataract formation [36].

From a methodological perspective, DL models have demonstrated strong potential across diverse ophthalmic imaging modalities. Tong *et al.* [37] reviewed the application of machine learning techniques in ophthalmic imaging, emphasizing the effectiveness of CNNs in fundus image analysis and disease screening tasks. Subsequent studies further explored the broader clinical applicability of DL, showing its ability to enhance diagnostic accuracy and support clinical understanding in complex ophthalmic conditions [38].

To improve feature representation and classification performance, advanced CNN architectures have been proposed. Liu *et al.* [39] introduced a dual-branch CNN-TRANS model for fundus image classification, demonstrating improved representation learning through parallel feature extraction mechanisms. Beyond classification, machine learning techniques have also been applied to surgical analysis and workflow optimization. Ramkumar and Sivaprakash [40] discussed machine learning methods for automated glaucoma diagnosis, illustrating the extension of ML-based diagnostic frameworks to multiple ocular diseases and imaging contexts. In a comprehensive systematic review, Ahuja *et al.* [41] analyzed AI applications in cataract management, demonstrating that AI-based systems can significantly improve clinical decision-making, surgical precision, and patient-specific outcome prediction across preoperative and intraoperative stages. More recently, EfficientNet-based architectures have been explored for eye disease classification, demonstrating competitive performance with reduced model complexity [42]. In addition, Bates [43] provided a historical and technological overview of cataract surgery, outlining the progressive integration of computational and AI-driven techniques into ophthalmic practice and underscoring their role in improving surgical outcomes and clinical efficiency. In addition to DL approaches, traditional machine learning pipelines, such as Gaussian-based Laplacian of Gaussian and Canny operators, have been investigated for edge detection in ophthalmoscopic cataract images, offering improved interpretability but limited generalization compared to modern CNN-based methods [44]. AI-driven systems for automated imaging and analysis of cataract surgery videos have also been proposed, supporting intraoperative assessment, surgical phase recognition, and postoperative evaluation [45].

Recent advances further extend these approaches to specialized surgical environments. Zhai *et al.* [46] developed a neural network–powered microscopic imaging system for cataract surgery, enabling real-time visualization and AI-assisted analysis during surgical procedures, thereby enhancing intraoperative guidance and precision. At a broader level, review studies have highlighted both the advantages and limitations of AI in ophthalmology, emphasizing the need to balance performance gains with issues of interpretability, data dependency, and clinical integration [47].

Collectively, these studies establish a diverse and evolving landscape of AI-based approaches for cataract-related tasks, spanning image classification, feature extraction, surgical workflow analysis, and clinical decision support. Against this backdrop, Table 5 presents a quantitative comparison of the proposed YOLOv8-based model with representative state-of-the-art methods, highlighting its performance advantages in accuracy, F1-score, and mAP@0.5.

The results show that the proposed YOLOv8 model outperforms previous DL approaches in both classification and localization accuracy. The improvement is particularly significant in mAP@0.5, demonstrating the model's superior ability to balance precision and recall across different intersection-over-union (IoU) thresholds.

On the external test set from clinics unseen during training, the detector achieved accuracy=0.99, precision=0.995, recall=0.995, F1=0.995, mAP@0.5=0.995, and mAP@0.5:0.95=0.983. These results complement 5-fold CV and align with clinical-grade expectations by combining high sensitivity with low latency. We compared the proposed model against ResNet-50 (TL) and YOLOv5 using the McNemar test on paired predictions, confirming significant differences in discordant errors ($p<0.05$, Holm–Bonferroni corrected). For AUC comparisons where applicable, we applied DeLong's test. A one-way ANOVA across fold-wise F1-scores further supported superiority over classical ML baselines ($p<0.05$). Bootstrap 95% CIs for F1 and mAP quantify uncertainty. Grad-CAM/Grad-CAM++ heatmaps localize attention on lens-opacity regions and pupil-space boundaries in positive cases, while normals exhibit diffuse or peripheral activations. The compressed YOLOv8 pipeline delivers ≈**1.9 ms/image** inference on commodity RTX-class GPUs, enabling real-time screening with confidence-threshold triage in mobile and tele-ophthalmology workflows.

Table 5. Comparison of our YOLOv8 model with state-of-the-art approaches in cataract detection

Model	Accuracy	F1-score	mAP@0.5
ResNet50 + SVM	0.960	0.950	-
EfficientNet-B3	0.970	0.960	0.940
YOLOv5	0.980	0.980	0.970
Proposed YOLOv8 (ours)	**0.990**	**0.995**	**0.995**

4.3. Interpretation of results

From a clinical perspective, the high recall (0.995) is crucial because it minimizes the probability of missing patients with cataracts, thereby reducing the risk of delayed treatment. Similarly, the high precision (0.995) indicates a low false-positive rate, which reduces the workload on ophthalmologists by preventing unnecessary follow-up examinations for healthy patients. The mAP@0.5:0.95 score of 0.983 confirms that the model consistently performs well across a range of localization thresholds, ensuring robust detection of pathological regions in varying image quality conditions.

To assess whether the proposed system can explain its decisions, Grad-CAM and Grad-CAM++ saliency maps were generated over the final convolutional layers for representative prediction cases, as illustrated in Figure 5. Figure 5(a) shows a true positive example, where the model correctly localizes lens opacity regions associated with cataract. Figure 5(b) presents a false positive case, in which strong reflections or illumination artifacts lead to incorrect activation. Figure 5(c) depicts a false negative example, where glare and low signal-to-noise ratio obscure pathological regions, resulting in missed detection. Overall, the visual explanations demonstrate that the model primarily focuses on clinically relevant regions in cataract-positive images, while normal images exhibit diffuse or peripheral activations.

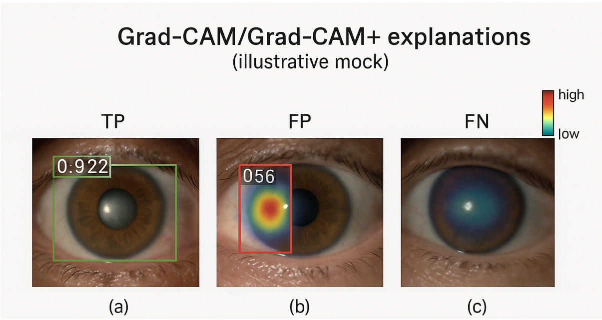


Figure 5. Grad-CAM and Grad-CAM++ visual explanations of the proposed model; (a) true positive case highlighting lens opacity regions, (b) false positive activation caused by reflection or illumination confounders, and (c) false negative example under glare and low signal-to-noise conditions

High recall (**0.995**) minimizes missed cataract cases, while high precision (**0.995**) limits unnecessary referrals—both are aligned with screening goals. The compressed YOLOv8 pipeline sustains $\approx$1.9 ms/image on commodity RTX-class GPUs and supports batch/stream modes with confidence-threshold triage. For low-resource/mobile use, we adopt an offline-first configuration with INT8 quantization and on-device caching; clinic integration follows a human-in-the-loop pattern: model triage $\rightarrow$ clinician review $\rightarrow$ referral. The system exports per-case reports (prediction, confidence, Grad-CAM heatmap) to support expert review and longitudinal audit.

4.4. Practical implications

The combination of high detection accuracy and a fast inference speed of 1.9 ms per image makes this model suitable for real-time clinical applications. Potential use cases include:
- Deployment in mobile diagnostic units and teleophthalmology systems for remote screening.
- Integration with slit-lamp cameras equipped with AI-based screening tools.
- Use in rural healthcare centers to provide rapid preliminary diagnostics where access to ophthalmologists is limited.

4.5. Future improvements and research directions

While the model already demonstrates strong performance, further enhancements can be achieved by:
- Increasing dataset diversity, including rare cataract subtypes, through additional data collection and the use of generative adversarial networks (GANs) for synthetic data augmentation.
- Applying domain adaptation techniques to improve generalizability across different imaging devices and acquisition conditions.
- Incorporating multimodal data sources such as patient demographic information, clinical history, and EHR for context-aware diagnostics.
- Implementing explainable AI methods (e.g., Grad-CAM) to visualize the decision-making process for better clinician trust and model interpretability.

Recent advancements in YOLOv10 and NMS-free architectures [48] indicate potential for further improving detection accuracy and reducing inference times. Additionally, integrating such models with AI-assisted surgical guidance systems [32], [47], [49] could extend their use from diagnostics to intraoperative decision support, marking a significant step toward comprehensive AI-driven ophthalmology.

4.6. Error analysis and limitations

While the model performs well, some misclassifications were observed in cases with poor image quality and occlusions. The main challenges include:

- Variability in image acquisition conditions
- Presence of artifacts such as reflections and noise
- Limited number of training samples for rare cataract subtypes

Future improvements can address these issues by incorporating more diverse datasets and advanced preprocessing techniques.

Although the proposed model demonstrates strong overall performance, several misclassifications were observed under challenging conditions. These errors are mainly related to poor image quality, occlusions, and variability in image acquisition settings. In particular, artifacts such as reflections, noise, and low contrast can adversely affect detection accuracy, especially in borderline or early-stage cataract cases.

Another limitation is the relatively limited number of training samples available for rare cataract subtypes, which may reduce the robustness of the model in less frequent clinical scenarios. Similar challenges have been reported in recent DL–based cataract detection studies, where model performance was shown to be sensitive to dataset diversity and image quality variations [48].

Despite these limitations, the optimized YOLOv8 model demonstrates high accuracy and strong generalization across different cataract severity levels. Its fast inference speed makes it suitable for real-world clinical deployment, particularly in screening and tele-ophthalmology applications. Recent advances in YOLO-based architectures further confirm the effectiveness of lightweight detectors for medical image analysis. In particular, hybrid DL frameworks have been shown to improve feature extraction and classification robustness for cataract detection tasks [49].

Moreover, recent developments in real-time object detection models, such as GhostYOLO, highlight the potential for reducing computational complexity while maintaining high diagnostic accuracy. These approaches enable efficient deployment on resource-constrained devices and support real-time cataract diagnosis in clinical settings [50].

4.7. Discussion

The YOLOv8 algorithm was used to build a cataract detection model. It has shown advantages in terms of achieving the perfect balance between computational efficiency and diagnostic accuracy. One of the distinctive features of the model is its real-time operation.The average output time of the model is only 1.9 ms per image, which highlights its applicability in clinical settings with limited resources, such as remote hospitals and mobile screening centers.

Our model has achieved increased accuracy, memorability, and F1-scores (about 0.995 for both the "normal" and "cataract" categories). The accuracy-reliability and F1-reliability curves are tested to ensure reliability in real clinical conditions. They consistently show outstanding results even at lower confidence thresholds.

Despite these significant advantages, the model has numerous limitations. One of the main problems identified was the accidental misclassification of low-quality images (blurred images, dim lighting, or partial opacity). These incorrect classifications emphasize the need to increase the model's resilience to real-world image processing scenarios. These limitations can be successfully eliminated, and the stability of the model can be improved by using complex magnification techniques such as setting occlusion, changing the lighting level, and simulating motion blur.

The model showed sensitivity to the distribution of training data. Some unusual cataract manifestations were underrepresented. The ability of the model to generalize the full range of cataract changes observed in clinical practice was limited due to a lack of representativeness. To solve this problem, it is necessary to increase the size and diversity of the dataset. The introduction of synthetic data generation techniques, in particular GANs, can benefit future research by offering realistic fake data to improve the quality of underrepresented or rare classes.

From a clinical perspective, the proposed system is designed to complement, not replace, ophthalmologists. In routine screening workflows, the AI model can act as a first-line triage tool: fundus or anterior segment images obtained by lab technicians are automatically analyzed, and cases predicted to be ”cataract-positive” are prioritized by experts. This significantly reduces the time specialists spend on routine cases while maintaining diagnostic oversight for ambiguous or serious findings.

Future studies could enhance diagnostic capabilities by integrating multimodal data sources like eye images with clinical data from EHR, patient demographic data, and historical clinical data. More accurate decisions based on the clinical context and unique patient characteristics will be possible thanks to this integration, which could also improve treatment outcomes.

Domain adaptation methodologies are an important topic for further study. The use of uncontrolled or partially controlled domain adaptation strategies can significantly increase the generalizability of the model, since different ophthalmic equipment and the environment allow for clear images. Our YOLOv8 model can provide exceptional accuracy in a variety of imaging devices and clinical scenarios, using strategies such as adapting to different subject areas or transferring knowledge from extensive ophthalmic datasets.

In conclusion, despite the great clinical prospects of our YOLOv8-based cataract detection system, its effectiveness can be further enhanced by eliminating existing limitations through the use of larger datasets, complex expansion, multimodal data integration, and domain adaptation. Reliable, effective and widely applicable ophthalmic diagnostic tools controlled by AI will be possible due to continuous progress in these fields.

5. CONCLUSION

This study introduced an enhanced YOLOv8-based DL model for automated cataract detection, achieving outstanding accuracy, sensitivity, specificity, and mAP across different cataract severity stages. In addition to high diagnostic accuracy, the model demonstrated exceptional computational efficiency, with an average inference time of only 1.9 ms per image, enabling real-time screening in both high-tech clinics and resource-limited medical environments.

The findings have broader implications for the field of AI-assisted ophthalmology. They confirm that advanced object detection architectures, when optimized for medical imaging, can transition from research prototypes to practical clinical tools. Such systems have the potential to be integrated into point-of-care devices, mobile ophthalmology units, and telemedicine platforms, thereby improving early detection, accelerating treatment decisions, and reducing the global burden of preventable blindness.

At the same time, certain limitations were identified, including reduced performance on low-quality or artifact-heavy images and underrepresentation of rare cataract subtypes in the dataset. Addressing these issues will require expanding and diversifying the training dataset, employing advanced augmentation strategies such as GANs to generate realistic synthetic cases, and exploring multimodal learning by combining ophthalmic images with patient clinical records. Further work on domain adaptation methods will ensure consistent performance across different imaging devices and patient populations.

In conclusion, the proposed YOLOv8-based approach offers a reliable, efficient, and clinically viable solution for cataract detection. By continuing to refine the model, enrich the training data, and expand its integration into real-world diagnostic workflows, this technology could become a transformative tool for early detection and management of cataracts, ultimately improving patient outcomes and supporting ophthalmologists in their daily practice.

The graphs with learning outcomes reflect the dynamics of changes in the loss function (loss) and metrics. mAP@0.5 and mAP@0.5-0.95 during the training of the YOLOv8 model for 50 epochs. A consistent decrease in the loss function is shown for both the training (train) and validation (val) samples, which confirms the absence of overfitting and stable convergence of the model. The precision and recall metrics are also steadily growing and reaching high values closer to the end of training, which indicates successful training of a model with good generalizing ability.

YOLOv9, an updated model, may be the next stage for research. The system demonstrates high accuracy in real time when analyzing ophthalmic images, which makes it promising for implementation in clinical practice.

FUNDING INFORMATION

The authors did not receive any financial support for the research, authorship, or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Anastassiya Tyunina	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			
Sabina Rakhmetulayeva	✓	✓		✓						✓		✓	✓	
Eduard Schiller	✓	✓				✓				✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests, personal relationships, or professional affiliations that could have influenced the research presented in this article. The study was conducted independently and objectively, and no external entity had control over the design, implementation, analysis, or publication of the results. Authors state no conflict of interest.

DATA AVAILABILITY

The supporting data of this study are openly available in the Hugging Face repository at `https://huggingface.co/datasets/a-eyelab/cataract-train`. The dataset includes annotated fundus and anterior segment images used for training and validating the cataract detection model.

REFERENCES

[1] D. Bremond-Gignac, A. Daruich, M. P. Robert, and S. Valleix, "Recent Developments in the Management of Congenital Cataract," *Annals of Translational Medicine*, vol. 8, no. 22, pp. 1-9, 2020, doi: 10.21037/atm-20-3033.

[2] S. Hu *et al.*, "Unified Diagnosis Framework for Automated Nuclear Cataract Grading Based on Smartphone Slit-Lamp Images," *IEEE Access*, vol. 8, pp. 174169-174178, 2020, doi: 10.1109/ACCESS.2020.3025346.

[3] H. Çetiner and I. Çetiner, "Classification of Cataract Disease with a DenseNet201 Based Deep Learning Model," *Journal of the Institute of Science and Technology*, vol. 12, no. 3, pp. 1264–1276, Sep. 2022, doi: 10.21597/jist.1098718.

[4] D. Katre and K. Selukar, "The Prevalence of Cataract in Children," *Cureus Journal of Medical Science*, vol. 14, no. 10, 2022, doi: 10.7759/cureus.30135.

[5] X.-Q. Zhang, Y. Hu, Z.-J. Xiao, J.-S. Fang, R. Higashita, and J. Liu, "Machine Learning for Cataract Classification/Grading on Ophthalmic Imaging Modalities: A Survey," *Machine Intelligence Research*, vol. 19, pp. 184–208, 2022, doi: 10.1007/s11633-022-1329-0.

[6] D. Tognetto *et al.*, "Artificial intelligence applications and cataract management: A systematic review," *Survey of Ophthalmology*, vol. 67, no. 3, pp. 817–829, May–Jun. 2022, doi: 10.1016/j.survophthal.2021.09.004.

[7] S. Jayachitra, K. N. Kanna, G. Pavithra, and T. Ranjeetha, "A Novel Eye Cataract Diagnosis and Classification Using Deep Neural Network," *International Conference on Novel Approaches and Developments in Biomedical Engineering-2021 (ICNADBE 2021)*, Coimbatore, India, 2021, vol. 1937, no. 1, doi: 10.1088/1742-6596/1937/1/012053.

[8] J. H. L. Goh *et al.*, "Artificial Intelligence for Cataract Detection and Management," *Asia-Pacific Journal of Ophthalmology*, vol. 9, no. 2, pp. 88-95, 2020, doi: 10.1097/01.APO.0000656988.16221.04.

[9] L. Andrade-Arenas and C. Yactayo-Arias, "Advances in the diagnosis of ocular diseases: An innovative approach through an expert system," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 13, no. 4, pp. 2828–2840, Aug. 2024, doi: 10.11591/eei.v13i4.7971.

[10] A. Jones and T. B. Vijayan, "AI-driven canine cataract detection: a machine learning approach using support vector machine," *Journal of the Chinese Institute of Engineers*, vol. 48, no. 1, pp. 94–102, 2024, doi: 10.1080/02533839.2024.2420100.

[11] L. Gutierrez *et al.*, "Application of artificial intelligence in cataract management: current and future directions," *Eye Vision*, vol. 9, no. 3, 2022, doi: 10.1186/s40662-021-00273-z.

[12] I. G. Ovechkin, I. S. Gadzhiev, A. A. Kozhukhov, and E. I. Belikova, "Diagnostic criteria for asthenic accommodative asthenopia in patients with computer vision syndrom," *Russian Journal of Clinical Ophthalmology*, vol. 20, no. 4, pp. 169–174, doi: 10.32364/2311-7729-2020-20-4-169-174.

[13] R. Ma *et al.*, "Multimodal machine learning enables AI chatbot to diagnose ophthalmic diseases and provide high-quality medical responses," *NPJ Digital Medicine*, vol. 8, no. 1, pp. 1-18, 2025, doi: 10.1038/s41746-025-01461-0.

[14] R. B. J. Simanjuntak, R. Magdalena, S. Saidah, A. B. Wiratama, and I. D S. Ubaidah, "Cataract Classification Based on Fundus Images Using Convolutional Neural Network," *International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 33-38, 2022, doi: 10.30630/joiv.6.1.856.

[15] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An Automated Cataract Detection System Using Deep Learning for Fundus Images," *IEEE Access*, vol. 9, pp. 128799-128808, 2021, doi: 10.1109/ACCESS.2021.3112938.

[16] N. Sheikh, T. Kulkarni, O. Swami, S. Deshmukh, and A. Rindhe, "Fundus disease detection using deep learning," in *2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET)*, Pune, India, 2025, pp. 1–6, doi: 10.1109/ICAET63349.2025.10932216.

[17] N. M. Dipu, S. A. Shohan, and K.M.A Salam, "Ocular Disease Detection Using Advanced Neural Network Based Classification Algorithms," *Asian Journal For Convergence In Technology (AJCT)*, vol. 7, no. 2, pp. 91-99, doi: 10.33130/AJCT.2021v07i02.019.

[18] S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, "A New Technique for Cataract Eye Disease Diagnosis in Deep Learning," *Periodicals of Engineering and Natural Sciences*, vol. 11, no. 6, 2023, doi: 10.21533/pen.v11.i6.191.

[19] S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, "Improving automated detection of cataract disease through transfer learning using ResNet50," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17541–17547, 2024, doi: 10.48084/etasr.8530.

[20] A. S. Khossravi, Q. Chen, and R. A. Adelman, "Artificial intelligence in ophthalmology," *Current Opinion in Ophthalmology*, vol. 36, no. 1, pp. 35–38, 2025, doi: 10.1097/ICU.0000000000001111.

[21] T. Oshika, "Artificial intelligence applications in ophthalmology," *JMA Journal*, vol. 8, no. 1, pp. 66–75, 2025, doi: 10.31662/jmaj.2024-0139.

[22] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An Automated Cataract Detection System Using Deep Learning for Fundus Images," *IEEE Access*, vol. 9, pp. 128799-128808, 2021, doi: 10.1109/ACCESS.2021.3112938.

[23] L. Stuermer, S. Braga, R. Martin, and J. S. Wolffsohn, "Artificial intelligence virtual assistants in primary eye care practice," *Ophthalmic and Physiological Optics*, vol. 45, no. 2, pp. 437–449, 2025, doi: 10.1111/opo.13435.

[24] Z. Su, K. Jin, H. Wu, Z. Luo, A. Grzybowski, and J. Ye, "Assessment of large language models in cataract care information provision: a quantitative comparison," *Ophthalmology and Therapy*, vol. 14, no. 1, pp. 103–116, 2025, doi: 10.1007/s40123-024-01066-y.

[25] J. E. Moore, R. N. McNeely, and S. Moutari, "Cataract surgery in the small adult eye: A review," *Clinical & Experimental Ophthalmology*, vol. 53, no. 5, 2025, doi: 10.1111/ceo.14510.

[26] S. Hu *et al.*, "ACCV: Automatic Classification Algorithm of Cataract Video Based on Deep Learning," *BioMedical Engineering OnLine*, vol. 20, no. 1, 2021, doi: 10.1186/s12938-021-00906-3.

[27] S. Yu *et al.*, "Assessing the impact of AI-assisted portable slit lamps on rural primary ophthalmic medical service," *Current Eye Research*, vol. 50, no. 5, pp. 1–8, 2025, doi: 10.1080/02713683.2025.2458131.

[28] B. Jiang, T. Wu, W. Liu, G. Liu, and P. Lu, "Changing Trends in the Global Burden of Cataract Over the Past 30 Years: Retrospective Data Analysis of the Global Burden of Disease Study 2019," *JMIR Public Health and Surveillance*, vol. 9, Dec. 2023, doi: 10.2196/47349.

[29] C. T. Costel, J. Xin, W. Yu, and Z. Rui, "YOLO Based Multi-Objective Vehicle Detection and Tracking," *Preprint on Research Square*, 2022, doi: 10.21203/rs.3.rs-2006332/v1.

[30] N. Ghamsarian *et al.*, "Cataract-1K Dataset for Deep-Learning-Assisted Analysis of Cataract Surgery Videos," Ph.D. dissertation, Alpen-Adria-Universität Klagenfurt, 2023.

[31] S. Benbakreti, S. Benbakreti, and U. Ozkaya, "The classification of eye diseases from fundus images based on CNN and pretrained models," *Acta Polytechnica*, vol. 64, no. 1, pp. 1-11, 2024, doi: 10.14311/AP.2024.64.0001.

[32] A. Tashkandi, "Eye care: Predicting eye diseases using deep learning based on retinal images," *Computation*, vol. 13, no. 4, no. 1-22, 2025, doi: 10.3390/computation13040091.

[33] R. Dawadi, M. Inoue, J. T. Tay, A. Martin-Morales, T. Vu, and M. Araki, "Disease prediction using machine learning on smartphone-based eye, skin, and voice data: Scoping review," *JMIR AI*, vol. 4, 2025, doi: 10.2196/59094.

[34] S. Touma, F. Antaki, and R. Duval, "Development of a Code-Free Machine Learning Model for the Classification of Cataract Surgery Phases," *Scientific Reports*, vol. 12, 2022, doi: 10.1038/s41598-022-06127-5.

[35] D. O'Brart, "The future of cataract surgery," *Eye*, pp. 1–6, 2025, doi: 10.1038/s41433-025-03745-x.

[36] V. Sulatha, S. S. Adigal, J. Lukose, R. V. John, A. Rizvi, and S. Chidangil, "Human Eye Lens Fluid Analysis Using High-Performance Liquid Chromatography-LED-Induced Fluorescence to Explore Cataract: A Pilot Study," *Journal of Biomedical Photonics & Engineering*, vol. 10, no. 2, 2024, doi: 10.18287/JBPE24.10.030302.

[37] Y. Tong, W. Lu, Y. Yu, Y. Shen, "Application of Machine Learning in Ophthalmic Imaging Modalities," *Eye Vision*, vol. 7, 2020, doi: 10.1186/s40662-020-00183-6.

[38] M. D. Subramaniam, A. K. B, R. B. Chirayath, A. P. Nair, M. Iyer, and B. Vellingiri, "Can Deep Learning Revolutionize Clinical Understanding and Diagnosis of Optic Neuropathy?," *Artificial Intelligence in the Life Sciences*, vol. 1, 2021, doi: 10.1016/j.ailsci.2021.100018.

[39] S. Liu, W. Wang, L. Deng, and H. Xu, "CNN-TRANS Model: A Parallel Dual-Branch Network for Fundus Image Classification," *Biomedical Signal Processing and Control*, vo 96, part. B, 2024, doi: 10.1016/j.bspc.2024.106621.

[40] P. Ramkumar and C. Sivaprakash, "Machine learning techniques for automatic diagnosis of glaucoma detection," in *Revolutionizing Healthcare Experience With Digital Medical Tourism*, pp. 137–150, 2025, doi: 10.4018/979-8-3693-7888-5.ch005.

[41] A. S Ahuja *et al.*, "Applications of Artificial Intelligence in Cataract Surgery: A Review," *Clinical Ophthalmology*, vol. 2024, pp. 2969—2975, 2024, doi: 10.2147/OPTH.S489054.

[42] T. Soni, "Advanced Eye Disease Classification using the EfficientNetB3 Deep Learning Model," *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, 2024, pp. 875-879, doi: 10.1109/ICACRS62842.2024.10841712.

[43] A. Bates, "Clarity unveiled: A journey through time in cataract surgery," *Journal of Perioperative Practice*, vol. 35, no. 11, 2025, doi: 10.1177/17504589241264400.

[44] L. S. Kumar and K. P. Indira, "Comparison of Gaussian based Laplacian of Gaussian operator with Gaussian based Canny operator for edge detection in ophthalmoscopic cataract images," in *AIP Conference Proceedings*, vol. 3252, no. 1, AIP Publishing, 2025, doi: 10.1063/5.0258883

[45] Y. J. Kim, S. H. Hwang, K. G. Kim, and D. H. Nam, "Automated imaging of cataract surgery using artificial intelligence," *Diagnostics*, vol. 15, no. 4, pp. 1-12, 2025, doi: 10.3390/diagnostics15040445.

[46] Y. Zhai *et al.*, "Neural network powered microscopic system for cataract surgery," *Biomedical Optics Express*, vol. 16, no. 2, pp. 535–552, 2025, doi: 10.1364/BOE.542436.

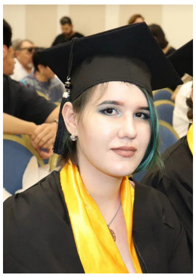
[47] H. N. Costin, M. Fira, and L. Goraș, "Artificial intelligence in ophthalmology: Advantages and limits," *Applied Sciences*, vol. 15, no. 4, pp. 1-50, 2025, doi: 10.3390/app15041913.

[48] F. Suryawanshi and K. Reddy, "Cataract detection and classification using YOLOv9 and machine learning," in *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, Bhimdatta, Nepal, 2025, pp. 1712–1717, doi: 10.1109/ICSADL65848.2025.10933134.

[49] A. B. Ramakrishnan, M. Madavan, R. Manikandan, and A. H. Gandomi, "A hybrid deep learning paradigm for robust feature extraction and classification for cataracts," *Applied AI Letters*, vol. 6, no. 2, 2025, doi: 10.1002/ail2.113.

[50] L. P. Lokesh, R. G. Poola, L. P. Gorrepati, and S. S. Yellampalli, "Real-time cataract diagnosis with GhostYOLO: A GhostConv-enhanced YOLO model," *Engineering, Tecnology & Applied Science Research*, vol. 15, no. 3, 2024, doi: 10.48084/etasr.10760.

BIOGRAPHIES OF AUTHORS



Anastassiya Tyunina graduated from the Faculty of Engineering and Information Technology of the Kazakh-German University (Deutsch-Kasachishe Universitaet) with a degree in Telematics in 2024. In 2024, she enrolled in a scientific and pedagogical master's degree at the Faculty of Computer Engineering, Satpayev Research University in Almaty, specializing in Machine Learning and Data Science. She can be contacted at email: st.demiu@bk.ru.

Sabina Rakhmetulayeva graduated with honors from the Faculty of Physical and Mathematical Sciences of Abai Kazakh National Pedagogical University in 2006. She completed her master's degree in Computer Science at the same university in 2008 and entered the doctoral program in Information Systems at Narxoz University. From 2009 to 2015, she worked as a senior lecturer at the Department of Applied Informatics at Narxoz University. In 2015, she successfully defended her Ph.D. dissertation in the dissertation council at the Eurasian National University in the specialty "6D070300-Information Systems," earning her Ph.D. degree. In addition to her teaching activities, she actively engages in research. Her scholarly work has resulted in over 30 publications, including 2 monographs and 10 articles in international peer-reviewed scientific journals classified within the first, second, and third quartiles according to Journal Citation Reports by Clarivate Analytics. She also has published more than 20 articles in international peer-reviewed scientific journals indexed in Scopus with a CiteScore percentile of at least 35. She can be contacted at email: ssrakhmetulayeva@gmail.com.

Eduard Schiller graduated from the Faculty of Engineering and Information Technology of the Kazakh-German University (Deutsch-Kasachische Universität) with a degree in Telematics in 2024. During his studies, he actively participated in academic projects and was recognized for his contributions to the student council. In 2024, he enrolled in a master's degree program in Machine Learning and Data Science at Faculty of Information Technology, Satbayev University in Almaty. His professional experience includes working as a developer-analyst at Proper Data Management (PDM) since October 2023 and completing an internship at Bank Soft Systems (BSS), where he focused on software development, data storage optimization, and data visualization. He possesses strong skills in programming (Python, Java, JavaScript, C, and C++), database management (MySQL, PostgreSQL, Greenplum, and Oracle), and the deployment of modern IT infrastructures (Linux, Kubernetes, Docker, Apache Airflow, and CI/CD). He is fluent in Russian (native), English, and German, and has good proficiency in Kazakh. He can be contacted at email: c2238@mail.ru.