 DOI: 10.2478/amma-2026-0005

REVIEW

Role of artificial intelligence in detecting and grading cataracts using color fundus photographs: A systematic review and meta-analysis

Pande Komang Wahyu Pradana[1*], Ida Ayu Prama Yanthi[1,2], Abdi Sastra Gunanegara[1]

1. Division of Ophthalmology, RS Dharma Kerti Tabanan, Indonesia
2. Department of Ophthalmology, Udayana University, Indonesia

Background: Cataracts are a leading cause of blindness and visual impairment worldwide, affecting millions of people. Early detection and accurate grading of cataracts are critical for timely intervention and improving patient outcomes. Artificial intelligence (AI), particularly deep learning, has emerged as a powerful tool for automating the detection and grading of cataracts using color fundus photographs.

Methods: A systematic review and meta-analysis was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines; a thorough literature search through databases such as PubMed, IEEE Xplore, and Google Scholar was conducted. The search parameters were restricted to studies published within the time frame of January 2020 to March 2025.

Results: A total of six studies were included in this systematic review and meta-analysis. Utilizing DTA meta-analysis, sensitivity ranged from 0.88 to 0.99, while specificity ranged from 0.89 to 0.99. Diagnostic Odds Ratio was estimated at 88.5, indicating that patients with cataracts are nearly 89 times more likely to be correctly identified by the AI model than non-cataract patients being misclassified.

Conclusion: AI particularly deep learning, has made significant strides in detecting and grading cataracts using color fundus photographs. The high accuracy, cost-effectiveness, and accessibility of AI models make them a valuable tool for improving cataract screening and management. As research continues to advance, AI has the potential to revolutionize cataract care, enabling early detection and timely intervention for millions of people worldwide.

Keywords: artificial intelligence, cataracts, color fundus photographs

Received 5 July 2025 / Accepted 3 December 2025

Introduction

Cataracts are a significant global health concern, recognized as a leading cause of vision impairment and blindness, affecting millions of individuals annually. According to the World Health Organization, cataracts account for approximately 51% of all cases of blindness worldwide.[1] Traditional cataract diagnosis typically requires in-person evaluation by an ophthalmologist using specialized equipment such as slit-lamp biomicroscopy. This requirement presents substantial barriers to early detection and timely treatment, particularly in regions with limited access to specialized care. The high prevalence of cataracts and limited accessibility of diagnostic resources highlight the need for more accessible screening methods.[2,3]

Color fundus photography (CFP) has become a potentially useful resource for cataract screening because it is more widely available than other eye imaging techniques. Fundus cameras are easier to access in primary care environments and demand less technical skill to use compared to slit-lamp biomicroscopes.[4] This unparalleled accessibility positions fundus photography as a powerful choice for creating automated screening tools that can revolutionize cataract detection far beyond the confines of specialized ophthalmology clinics.

The application of artificial intelligence (AI) in ophthalmology has grown rapidly in recent years, especially in the fast-paced internet-based electronic devices usage, with successful implementations for conditions including diabetic retinopathy, age-related macular degeneration, and glaucoma.[5] These applications illustrate the capacity of AI to enhance clinical decision-making processes and broaden screening functionalities. In the context of cataracts, the previous ten years have witnessed substantial progress in AI-driven techniques for the identification and grading of this condition through the analysis of fundus images.6 Hence, this literature systematically review the usage and development of AI-based technologies based on CFP, as a novel term in detecting cataract.

Methods

Literature search

A systematic literature review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as recommended by Cochrane. The review involved a comprehensive search of three medical-scientific databases: PubMed, ScienceDirect, and the Cochrane Library, covering the period from January 2020 to March 2025. The search strategy was developed using the Population, Intervention, Comparison, and Outcomes (PICO) framework, as outlined in Table 1. MeSH terms and keywords employed in

* Correspondence to: Pande Komang Wahyu Pradana
E-mail: pandekomang1@gmail.com

Table 1. Population, Intervention, Comparison, and Outcomes Search Strategy.

P (Population)	Individuals with Cataracts
I (Intervention)	Use of Artificial Intelligence (AI) and deep learning models for automated detection and grading of cataracts using color fundus photographs
C (Comparison)	Gold Standart “Slit Lamp” Examination
O (Outcome)	Improvement in the accuracy and efficiency of cataract detection and grading, leading to timely intervention and better patient outcomes.

the search related to (AI[Title/Abstract] OR artificial intelligence) AND (cataract[Title/Abstract]) AND ((detect*) OR (screen*) OR (grad*)). Studies were included for analysis if they met the following inclusion criteria: published in English, full-text availability, and publication date after January 2020. Literature search according to PRISMA was detailed in Figure 1.

Inclusion and Exclusion Criteria

Studies were included if they: (1) used AI to detect or grade cataracts from fundus photographs; (2) reported diagnostic performance metrics (sensitivity, specificity); (3) involved human subjects; and (4) were cohort, cross-sectional, or case-control studies. We excluded conference abstracts without full texts, animal studies, review articles, cataract congenital, and single case reports.

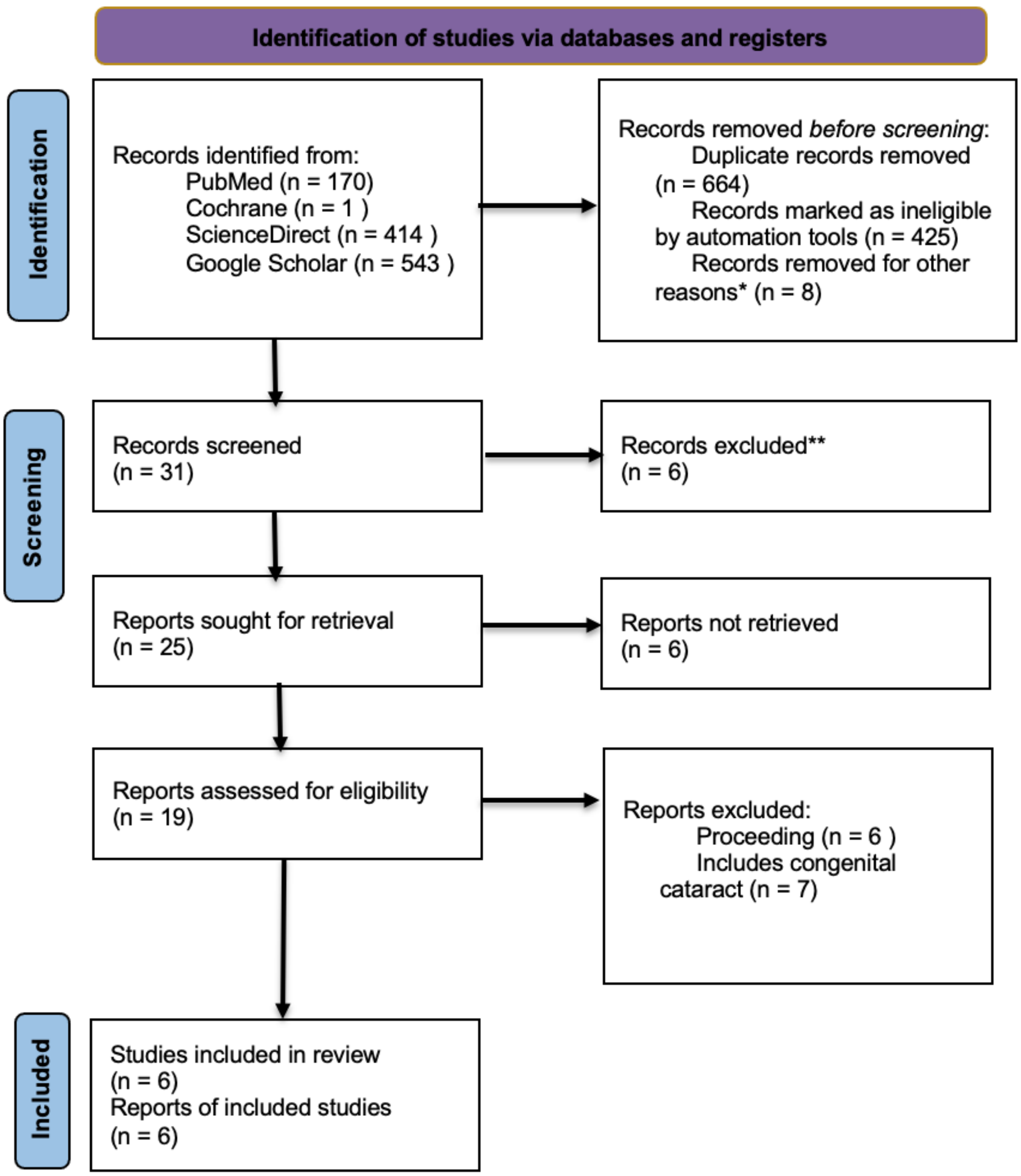

Fig. 1. PRISMA flowchart

Quality Assesment

The articles that passed the primary screening were then reviewed by the two reviewers (PKP, IAP) individually. Authors independently assessed the quality of the studies according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was applied for the risk of bias assessment of the included studies. The QUADAS-2 scale consists of 4 aspects for risk of bias including patient selection, index test, reference standard, and flow & timing as well as 3 domains for applicability concerns including patient selection, index test, and reference standard. The risk of bias was classified into 3 categories (i.e. low, high, and unclear risk bias). Studies with low quality or with evident defects in design and procedure were excluded from this review. All studies were in low to moderate risk of bias. Any disagreements between the authors were resolved by consensus.

Data Extraction

The following data were extracted: (1) the basic characteristics of the included studies and participants, including the methods, algorithms, databases, sample sizes, outcomes, and procedures; (2) grading system, and; (3) the evaluation indices of the algorithms, including the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FP) outcomes as well as the accuracy, and area under the curve (AUC).

When raw confusion matrix values (TP, FP, TN, FN) were not reported, we estimated them based on the reported sensitivity, specificity, total sample size, and disease prevalence using standard diagnostic test formulas. To estimate TP, FP, TN, and FN values, we used the following approach:

1. Estimate the number of actual positive and negative cases based on reported or assumed disease prevalence:
   Positives = N × Prevalence,   Negatives = N − Positives
   where N is the total sample size.
2. Apply the definitions of sensitivity and specificity:
   Sensitivity = TP / (TP + FN)   ⇒ TP = Sensitivity × Positives
   Specificity = TN / (TN + FP)   ⇒ TN = Specificity × Negatives
   From these, the remaining values are derived as:
   FN = Positives − TP,   FP = Negatives − TN
3. If prevalence was not explicitly reported, we assumed a baseline prevalence of 50% for balanced datasets, unless otherwise inferred from the study text or confusion matrix visualizations.

Weighted Metrics

To allow comparison across studies with varying prevalences, we also calculated weighted sensitivity and specificity as:

Weighted Sensitivity = (Sensitivity × (TP + FN)) / N

Weighted Specificity = (Specificity × (TN + FP)) / N

Statistical analysis

For the DTA meta-analysis, we utilize Meta Disc 2.0 (Cochrane Symposium, Barcelona, Spain). We performed a bivariate random-effects meta-analysis to compute pooled sensitivity, specificity, and generate a summary receiver operating characteristic (SROC) curve. Forest plots were created to visualize sensitivity and specificity across studies. Heterogeneity was evaluated using between-study variance estimates and correlation of logit-transformed sensitivity and specificity.

Results

Study Characteristics

A total of six studies were included in this systematic review and meta-analysis, spanning publication years from 2022 to 2024. These studies evaluated the diagnostic accuracy of artificial intelligence (AI) algorithms in detecting and grading cataracts using color fundus photographs. Sample sizes ranged from 1,000 to 8,681 participants. All studies used reference standards validated by ophthalmologists and implemented machine learning or deep learning models.

Pooled Sensitivity and Specificity

Forest plots are presented in Figure 2 (sensitivity) and Figure 3 (specificity). Sensitivity ranged from 0.88 to 0.99, while specificity ranged from 0.89 to 0.99. The highest performance was observed in Lu et al (2022) and Xie et al, reflecting the effectiveness of their AI models.[7,8]

Beyond pooled sensitivity and specificity, several secondary diagnostic indicators further support the accuracy and clinical utility of AI for cataract detection:

The pooled sensitivity and specificity of AI-based detection of cataracts were derived from a bivariate random-effects meta-analysis model:

- Pooled sensitivity (logit scale): 3.27 (corresponding to a sensitivity of ~96%)
- Pooled specificity (logit scale): 2.81 (corresponding to a specificity of ~94%)
- Between-study variance: 0.776 (sensitivity) and 0.647 (specificity)
- Correlation between logit sensitivity and specificity: 0.765

These values suggest a high level of diagnostic accuracy, with strong discriminatory power across the AI algorithms.

- Diagnostic Odds Ratio (DOR): The DOR was estimated at 88.5, indicating that patients with cataracts are nearly 89 times more likely to be correctly identified by the AI model than non-cataract patients being misclassified. A DOR greater than 1 implies good discriminatory test performance, and values above 20 are considered strong.
- Positive Likelihood Ratio (PLR): The pooled PLR was 16.85, meaning that a positive AI prediction is ~17 times more likely in a patient with cataracts than without. This value exceeds the threshold of 10, typically considered strong evidence to rule in the diagnosis.

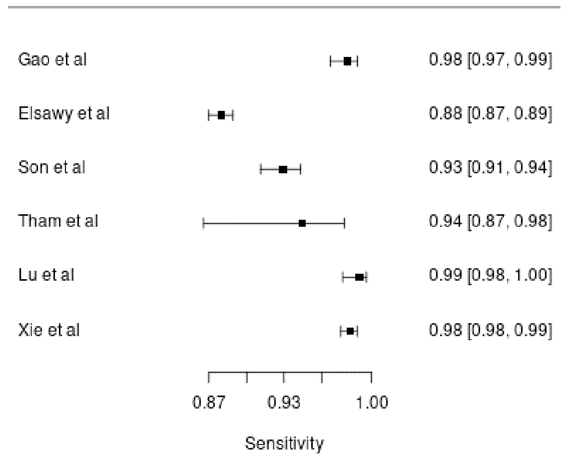

Fig. 2. Forest plot of sensitivity across six included studies.

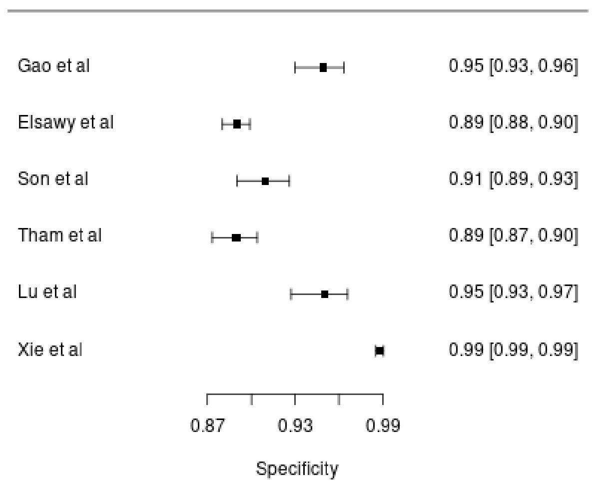

Fig. 3. Forest plot of specificity across six included studies.

- Negative Likelihood Ratio (NLR): The pooled NLR was 0.04, indicating that when the AI predicts absence of cataracts, the probability of actual disease is very low. NLR values <0.1 provide strong evidence to rule out disease.

Summary Receiver Operating Characteristic (SROC) Curve

The hierarchical SROC curve is presented in Figure 4, showing the balance between sensitivity and specificity across studies. The summary point lies in the upper-left quadrant, confirming excellent diagnostic performance. The 95% confidence and predictive regions indicate moderate inter-study heterogeneity.

Discussions

This systematic review and meta-analysis evaluated the diagnostic performance of artificial intelligence (AI) in detecting and grading cataracts using color fundus photographs. Across six eligible studies published between 2022 and 2024, the pooled sensitivity (96%) and specificity (94%) of AI models demonstrate strong diagnostic accuracy, aligning with clinical expectations for screening tools. These findings underscore the growing maturity of AI systems—particularly deep learning-based convolutional neural networks (CNNs)—as viable support tools in cataract screening, especially in low-resource settings where ophthalmologist access is limited.[12,13]

Performance across the included studies was consistently strong, though variation was observed. Lu et al. (2022) and Xie et al. (2022) achieved exceptional sensitivity and specificity values (0.99 and 0.98; 0.95 each), suggesting robust model optimization and high-quality input data.7,8 In contrast, Elsawy et al. (2023), despite utilizing a much larger and diverse dataset (n = 8,681), reported slightly lower performance (sensitivity 0.88; specificity 0.89), likely reflecting real-world challenges such as population heterogeneity and labeling inconsistencies.2 This discrepancy underscores a key point raised in recent ophthalmology literature: while AI models often excel in controlled environments, real-world generalization remains a critical hurdle.[14]

The bivariate meta-analysis confirmed moderate heterogeneity, with τ^2 values of 0.776 (sensitivity) and 0.647 (specificity). A strong correlation (r = 0.765) between logit-transformed sensitivity and specificity suggests that high-performing models tend to maintain balanced diagnostic profiles. The hierarchical summary receiver operating characteristic (HSROC) curve affirmed overall diagnostic

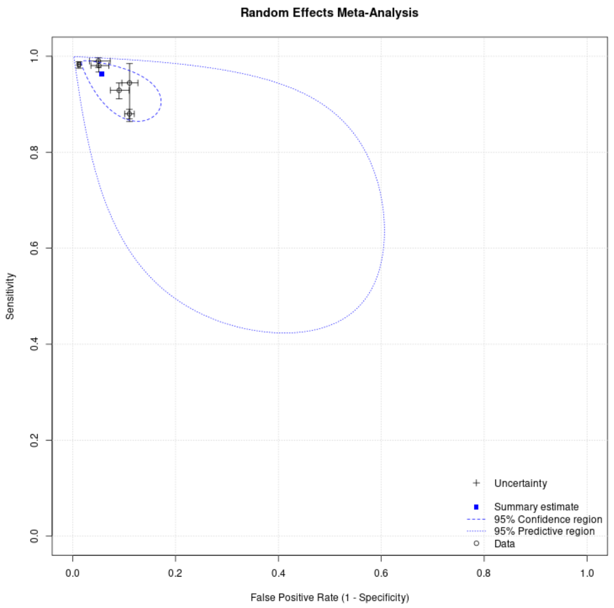

Fig. 4. HSROC curve from random-effects meta-analysis, showing uncertainty regions and summary estimates.

strength, with a summary point located in the upper-left quadrant and tight confidence bounds indicating consistency. However, predictive region spread points to ongoing challenges, such as image quality differences and varying fundus camera protocols across studies.

Importantly, the review highlights that most included studies utilized advanced CNNs like ResNet-50, DenseNet1[21], and Inception-v3, which offer robust feature extraction across diverse imaging conditions. As noted in recent technical reviews, these architectures remain foundational in ophthalmic AI, although emerging transformer-based models may offer improved context modeling and generalizability.15 The growing interest in attention-based models suggests future directions may involve shifting from traditional CNNs to hybrid or transformer architectures.

Moreover, the exclusive use of color fundus photographs offers significant advantages—non-invasiveness, accessibility, and compatibility with portable devices. Yet, studies have shown that diagnostic confidence may be further improved when combined with modalities like OCT, which enhance posterior segment visualization and reduce false negatives in early cataract or comorbid pathology detection.[16]

From an implementation standpoint, several limitations persist. Most studies were retrospective and set in academic environments, with limited demographic diversity. Few models provided visual interpretability outputs (e.g., saliency maps), a limitation echoed in broader AI surgical literature.17 Additionally, model integration with electronic health records (EHRs), clinician workflow, and regulatory frameworks remains limited. These gaps, along with data privacy and ethical concerns, must be addressed through

Table 2. Data Extraction

Study	Year	Model	Grading System	True positive	False negative	False positive	True negative	Total sampel	Sensitivity	Specificity	Weight sensitivity	Weight specificity
Gao et al[9]	2024	Dual-Stream Cataract Evaluation Network	Age-Related Eye Disease Study grading system	657	13	34	636	1340	0.98	0.95	0.49	0.475
Elsawy et al[2]	2023	DeepOpacityNet	Wisconsin Cataract Grading System	3819	521	478	3863	8681	0.88	0.89	0.43	0.44
Son et al[10]	2022	ResNet18, WideResNet50-2, and ResNext50	LOCS III	916	70	89	897	1972	0.93	0.91	0.46	0.45
Tham et al[11]	2022	Convolutional neural network (CNN)	Wisconsin and AREDS grading systems	68	4	179	1441	1692	0.95	0.89	0.04	0.85
Lu et al[8]	2022	R-CNN and ResNet	LOCS III	495	5	25	475	1000	0.99	0.95	0.49	0.47
Xie et al[7]	2022	DenseNet121	N/A	1554	27	82	6732	8395	0.98	0.98	0.18	0.8

robust clinical trials and policy innovation.

Finally, the diagnostic odds ratio (88.5), positive likelihood ratio (16.85), and negative likelihood ratio (0.04) collectively suggest that AI systems not only perform well statistically but also possess strong clinical utility in both ruling in and ruling out cataract cases. In public health contexts, particularly in underserved or high-volume regions, AI can extend diagnostic reach through mobile fundus cameras and teleophthalmology—enhancing early detection and reducing preventable blindness.[18]

Limitation

A prevalent technical constraint is the absence of uniformity in the outcome metrics and validation methods reported, complicating direct comparisons across studies and leaving some uncertainty about whether the findings of AI-driven techniques necessitate further validation by ophthalmologists, despite demonstrating high AUC, specificity, and sensitivity. Nevertheless, we cannot overlook the occurrence of technical inaccuracies, and no technology surpasses the value of experience.

Conclusion

Artificial intelligence demonstrates excellent diagnostic accuracy for detecting and grading cataracts using color fundus photographs, with pooled sensitivity and specificity exceeding 90%. These results support the implementation of AI-assisted tools in clinical and community-based eye screening programs. Future research should focus on prospective validation, algorithm transparency, and integration into clinical workflows to maximize impact and ensure equitable access.

Authors' contribution

PKP (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Validation; Visualization; Writing – original draft; Writing – review & editing)

IAY (Conceptualization; Formal analysis; Investigation; Methodology; Supervision; Writing – original draft; Writing – review & editing)

ASG (Conceptualization; Formal analysis; Investigation; Methodology; Supervision; Writing – original draft; Writing – review & editing)

Conflict of interest

None to declare.

Funding

No external funding was received.

References

1. WHO. Blindness and vision impairment prevention. Https://WwwWhoInt/News-Room/Fact-Sheets/Detail/Blindness-and-Visual-Impairment. 2019;(8 October 2019).
2. Elsawy A, L Keenan TD, Chen Q, Thavikulwat AT, Bhandari S, Cheer Quek T, et al. A deep network DeepOpacityNet for detection of cataracts from color fundus photographs Plain language summary. 2023; Available from: https://doi.org/10.5281/
3. Müller S, Jain M, Sachdeva B, Shah PN, Holz FG, Finger RP, et al. Artificial Intelligence in Cataract Surgery: A Systematic Review. Vol. 13, Translational Vision Science and Technology. Association for Research in Vision and Ophthalmology Inc.; 2024.
4. Shih KC, Hung KW, Lau KP, Yip WM, Lee A, Fong A, et al. Deep Learning Automated Diagnosis and Grading of Cataracts using Colour Fundus Images: The Fundus Cataract-AI Project. Invest Ophthalmol Vis Sci. 2024 Jun 17;65(7):5950.
5. Prashar J, Tay N. Performance of artificial intelligence for the detection of pathological myopia from colour fundus images: a systematic review and meta-analysis. Eye (Basingstoke). 2024;38(2).
6. Wu X, Xu D, Ma T, Li ZH, Ye Z, Wang F, et al. Artificial Intelligence Model for Antiinterference Cataract Automatic Diagnosis: A Diagnostic Accuracy Study. Front Cell Dev Biol. 2022;10.
7. Xie H, Li Z, Wu C, Zhao Y, Lin C, Wang Z, et al. Deep learning for detecting visually impaired cataracts using fundus images. Front Cell Dev Biol. 2023;11.
8. Lu Q, Wei L, He W, Zhang K, Wang J, Zhang Y, et al. Lens Opacities Classification System III-based artificial intelligence program for automatic cataract grading. J Cataract Refract Surg. 2022 May 1;48(5):528–34.
9. Gao W, Shao L, Li F, Dong L, Zhang C, Deng Z, et al. Fundus photograph-based cataract evaluation network using deep learning. Front Phys. 2023;11.
10. Son KY, Ko J, Kim E, Lee SY, Kim MJ, Han J, et al. Deep Learning-Based Cataract Detection and Grading from Slit-Lamp and Retro-Illumination Photographs: Model Development and Validation Study. Ophthalmology Science. 2022 Jun 1;2(2).
11. Tham YC, Goh JHL, Anees A, Lei X, Rim TH, Chee ML, et al. Detecting visually significant cataract using retinal photograph-based deep learning. Nat Aging. 2022;2(3).
12. Ramya N, Hemavathi D. Artificial Intelligence in Cataract Diagnosis and Management With Its Future Directions. Responsible AI for Digital Health and Medical Analytics. 2025;189–210.
13. Khan I, Akbar W, Soomro A, Hussain T, Khalil I, Khan MN, et al. Enhancing Ocular Health Precision: Cataract Detection Using Fundus Images and ResNet-50. IECE Transactions on Intelligent Systematics [Internet]. 2024 Oct 29;1(3):145–60. Available from: https://www.iece.org/article/tis/51/91
14. Rampat R, Debellemanière G, Gatinel D, Ting DSJ. Artificial intelligence applications in cataract and refractive surgeries. Curr Opin Ophthalmol [Internet]. 2024;35(6). Available from: https://journals.lww.com/co-ophthalmology/fulltext/2024/11000/artificial_intelligence_applications_in_cataract.8.aspx
15. Sakunthala D, Gireesh N. Automated Glaucoma Detection Using Vision and Swin Transformers: Advancing Ophthalmic AI. 2025.
16. Carretero MM, Parejo A de los RS, Pérez CM, Amores RR, Gómez ÁP, Pérez MB. Comparative Analysis of Diagnostic Accuracy in Retinal Imaging: Fundus Photography Versus Optical Coherence Tomography. Science and Technology: Developments and Applications Vol 2 [Internet]. 2025 Jan 22;163–88. Available from: https://stm.bookpi.org/STDA-V2/article/view/16909
17. Ahuja AS, Paredes AA, Eisel MLS, Kodwani S, Wagner I V., Miller DD, et al. Applications of Artificial Intelligence in Cataract Surgery: A Review. Vol. 18, Clinical Ophthalmology. Dove Medical Press Ltd; 2024. p. 2969–75.
18. Bressler I, Dollberg D, Aviv R, Margalit D, Harris A, Siesky B, et al. Non-contact optical blood pressure biometry using AI analysis of fundus imaging. doi:101101/2025010625320084.