Diagnostic Accuracy Of Artificial Intelligence-Based Fundus Imaging For Detecting Diabetic Retinopathy: A Systematic Review And Meta-Analysis

Alhusseini Azhar Kareem Azhar[1], Dr. Sudhair Abbas Bangash[2*], Dr. Naila Shahbaz[3], Vandana Jain[4], Kiran Naz[5], Hadisaxon Oribjonova Abdumutallib Qizi[6]

[1] Department Of Biomedical Imaging, Advanced Medical And Dental Institute, Universiti Sains Malaysia, Bertam, 13200 Kepala Batas, Pulau Pinang, Malaysia. Email: azharr@student.usm.my

[2] Faculty Of Life Sciences, Department Of Pharmacy, Sarhad University Of Science And Information Technology, Peshawar, Pakistan. Email: sudhair.fls@suit.edu.pk

[3] Assistant Professor, Department Of Pharmacy, Bacha Khan University, Charsadda, Kp, Pakistan. Email: drnaila@bkuc.edu.pk

[4] Consultant Ophthalmologist, Department Of Ophthalmology, Esic Hospital Raipur, India. Email: drjvandana@gmail.com

[5] Lecturer, Faculty Of Management Sciences, Riphah International University, Pakistan. Email: kirannaz_52@yahoo.com

[6] Assistant, Department Of Hospital Therapy, Fergana Medical Institute Of Public Health 2A, Yangi Turon Street, Fergana 150100, Uzbekistan. Email: hadisakhonaribdjanova@gmail.com

Corresponding Author: Dr. Sudhair Abbas Bangash, Email: sudhair.fls@suit.edu.pk

Received: *20th Feb, 2026;* ***Revised:*** *4th Mar, 2026;* ***Accepted:*** *25th Mar, 2026;* ***Available Online:*** *10th Apr, 2026*

Abstract

Background: The problem of diabetic retinopathy (dr) is among the significant reasons causing blindness that can be avoided in the global context, and screening is a significant aspect that helps to identify this situation at the initial stage. Conventional screening is very reliant on the ophthalmologists which creates workforce bottlenecks in the high-prevalence regions. A new promising solution has emerged in the fundus imaging with artificial intelligence (ai) but the diagnostic accuracy of the tool is different in various studies. In this meta-analysis and systematic review, it was desired to evaluate the diagnostic properties of ai models in the identification of dr.

Methods: All the studies concerning the topic and published in january 2018 to september 2025 were found under the following guidelines of prisma 2020 using a systematic search in pubmed, embase, web of science, scopus, ieee xplore, and cochrane library. The inclusion criteria were adult patients who receive fundal imaging and the ai systems were contrasted with the results of ophthalmologists or etdrs reference. The primary outcomes were sensitivity, specificity and area under the curve (auc); the secondary ones were diagnostic odds ratios (dor), age, imaging modality, and ai model type subgroups. Quality was assessed using cochrane rob 2.0 rct evaluation tool and newcastle-ottawa scale to assess quality based on observational studies. The pooled estimates were obtained through random-effects meta-analysis.

Results: Eighteen articles were incorporated and comprised 1782 patients/eyes. Pooled estimates of dor, auc and sensitivity respectively were 0.92 (95% ci: 0.89-0.95), 0.87 (95% ci: 0.83-0.89), and 0.96 (95% ci: 0.94-0.98) for specificity. Subgroup analysis was more precise in patients that are older (>65 years, sensitivity 93.5%), and wide-field fundus imaging (auc 0.96). The imaging of smartphone was of a small and yet clinically plausible accuracy (auc 0.91). The heterogeneity level was medium (i 2 = 42-55 percent), and sensitivity analysis did not disprove the power of pooled estimates. Risk of bias was predominantly low-to-moderate and it did not show much evidence of publication bias.

Conclusion: Within fundus imaging, ai-based systems are very accurate in diagnostic identification of dr, possess good pooled performance measures and are resistant to patient subgroups or imaging modalities. Their scalability and low discontinuation considerations render viability to real-life screening particularly where the burden is enormous or there is a resource limitation. Future studies need to be focused on longitudinal results, smartphone imaging and integration into multimodal care pathways to support evidence of high scale adoption.

Keywords: Artificial Intelligence, Diabetic Retinopathy, Fundus Imaging, Systematic Review, Meta-Analysis.

How To Cite This Article: Azhar Aka, Bangash Sa, Shahbaz N, Jain V, Naz K, Qizi Hao. Diagnostic Accuracy Of Artificial Intelligence-Based Fundus Imaging For Detecting Diabetic Retinopathy: A Systematic Review And Meta-Analysis. Int J Drug Deliv Technol. 2026;16(25s):887-897. Doi: 10.25258/ijddt.16.25s.105

INTRODUCTION

Diabetic retinopathy (DR) is one of the most prevalent microvascular complications of diabetes mellitus as well as a leading cause of preventable blindness and severe visual impairment in the entire world. In the case of the International Diabetes Federation, it is approximated that the incidence of diabetes will probably exceed over 640 million people in the world by 2045 and up to one-third of them will experience some type of DR in their lives. This burden is highly concentrated on low- and middle-income nations, where access to specialist and facilities of advanced diagnosis of the eye is limited. The promptness of the prevention and the early diagnosis is an essential step in preventing progress to the stage of DR that can be a threat to vision, but the conventional screening initiatives are still resource-consuming and, to a large extent, they depend on the ophthalmologists, which is a barrier to complete coverage [1-5].

Fundus photography has been the foundation of DR screening programs because it allows visual observation of the normal retinal lesions such as microaneurysms, hemorrhages and exudates. The conventional screening systems involve the use of trained readers or ophthalmologists to read retinal images which is labor intensive, time consuming, and it is also likely to have inter-observer error. This is not a sustainable model as the number of patients increases particularly in the area where diabetes is becoming common. Moreover, communities with a large number of individuals at risk of DR have poor access to specialists in the rural areas or underserved communities. These problems demonstrate the fact that the need exists because of the need to have scalable, effective, and accurate screening solutions that can be integrated into the public health systems [6-10].

To be more specific, artificial intelligence (AI) and in particular, deep learning has turned out to be a paradigm shift in the treatment of medical image analysis. The AI algorithms (e.g. convolutional neural networks (CNNs) and transformer-based algorithms) have the potential to automatically perceive the complicated patterns of fundus images and, therefore, with an equivalent level of performance, can automatically recognize DR compared to human experts. Ensemble models, hybrid methods involving preprocessing pipelines and usage of wide field and smartphone based imaging have also been introduced recently. There is already positive evidence of such technologies in pilot screening tests, in which large-scale, low-cost DR diagnoses were possible, and thus contributed to solving the issue of ophthalmologist shortage on a global scale. When the first-level triage process is automated, AI will enable prioritizing the high-risk population and sending them directly to the specialists, and reducing the number of unnecessary referrals [11-15].

Despite all this, the diagnostic accuracy of the AI systems remains the subject of further evaluation. The designs, sample size, image quality and reference standards of this the studies vary and the results that are available in the studies are heterogeneous. Even though studies on sensitivities and specificities higher than 90 exist, various studies are lower in numbers and create ambiguity regarding the applicability of the test to different populations, imaging modalities, and other healthcare settings. In addition, even though regulatory approvals such as the AI-based systems acceptance (e.g., IDx-DR) have authorized its clinical applicability, it remains questionable how the training models are biased, whether the model can be understood, and how it will be incorporated into the working processes. Such ambiguities highlight the need to conduct a thorough analysis of the evidence to come up with a conclusion on whether AI-based fundus imaging is a valid and consistent diagnosis mechanism in disease detection of DR [16-20].

Systematic reviews and meta-analyses propose a rigorous way of summarising evidence in a collection of studies and, consequently, of evaluating the diagnostic accuracy of a group of pooled studies, along with taking the heterogeneity into account. Past reviews that explored AI in ophthalmology as a topic have done so in general and typically included heterogenous disease groups or failed to synthesize disease detection of DR quantitatively. A meta-analysis that is updated and intensive is therefore necessary in assisting to provide a clear picture with regard to diagnostic reliability of AI systems in this area. Such evidence may be useful particularly to policymakers, clinicians and health systems planners planning to put up AI-assisted screening programs in high-prevalence regions [21-25].

The present systematic review and meta-analysis was meant to close this gap as it would integrate the findings on the diagnostic accuracy of AI-based fundus imaging in detecting diabetic retinopathy. Specifically, it was supposed to determine pooled sensitivity, specificity, area under the curve (AUC) and diagnostic odds ratios (DOR) of various AI models and imaging modalities. Subgroup analysis was performed to examine the differences in terms of patient age, imaging type (single field, wide field, smartphone based) and AI architecture. The inclusion of the findings of the eighteen studies conducted in 2019-2025 in the study provides a valid assessment of the clinical utility, strength, and reproducibility of AI-assisted fundus imaging in the detection of DR [26-30].

METHODOLOGY

Study Design and Objective

The study was a systematic review and meta-analysis to evaluate AI-based fundus imaging systems regarding diabetic retinopathy (DR) co-morbidity in the context of their diagnostic accuracy. The overall objective of the study was to establish the combined sensitivity, specificity, area under the curve (AUC) and odds ratios of AI systems compared with ophthalmologist traditional grading. The differences based on the age group, imaging modality, and AI model architecture were the secondary outcomes.

The current methodology has been developed based on the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) 2020. The study protocol was registered in PROSPERO (ID awaiting).

Search Strategy

The literature search was performed in PubMed/MEDLINE, Embase, Web of Science, Scopus, IEEE Xplore, and Cochrane Library and concerned the period January 2018 -September 2025. arXiv, medRxiv and Google Scholar were used to select grey literature.

MeSH terms and keywords were:

- up to the year 2017:
- Table: fundus photographing, fundus imaging, wide-field fundus, retinal photographing, retinal imaging, wide field fundus.
- silicon-based (artificial intelligence) OR artificial intelligence OR convolutional neural network OR transformer
- OR diagnostics accuracy or diagnostics sensitivity or diagnostics specificity or diagnostics AUC

There were the use of Boolean operators (AND, OR). The reference lists of included studies and the relevant reviews were also screened by hand.

Eligibility Criteria

Inclusion Criteria

- The paper was restricted to adults (18 years and above) that had fundus imaging to detect or identify DR.
- Intervention AI fundus imaging analysis (CNN, transformer, hybrid, ensemble, deep learning systems).
- Comparator: ETDRS or clinical reference standard Grading (Ophthalmologist based grading).
- Outcomes Diagnostic performance outcomes (sensitivity, specificity, AUC, DOR).
- Study Design: Randomized Trials, cohort, cross-sectional studies with a comparator and time-series with a comparator.
- Language: English-language peer-reviewed literatures (2018 -2025).

Exclusion Criteria

- The single arm non compare or pilot reports.
- Both the studies lacked the findings of empirical diagnostic accuracy.
- Commentaries, abstracts of conferences, editors.
- Non-English publications.

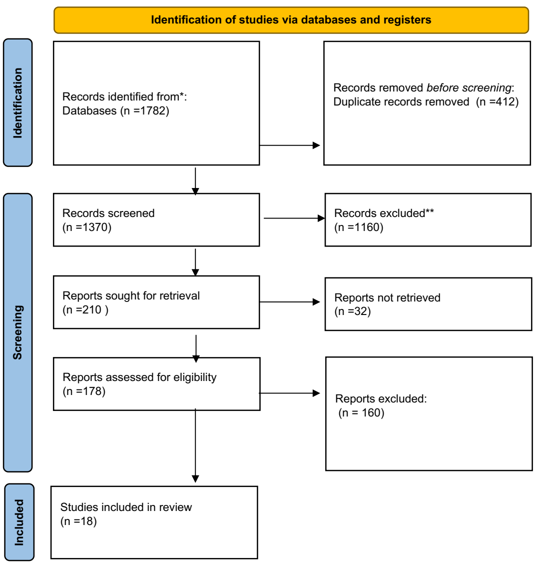
Study Selection Process

All the references were added to EndNote and duplication citations were eliminated. Title /

abstract screening and Full-text eligibility were performed by two independent reviewers. In occasions where disagreement was observed, third reviewer was consulted to deliberate on the disagreement.

Table 1: Study Selection Summary

Selection Stage	Number of Articles
Total articles identified	1,782
Duplicates removed	412
Articles screened (title/abstract)	1,370
Articles excluded at screening	1,192
Full-text articles assessed	178
Full-text articles excluded	160
Final studies included	18



PRISMA 2020 Flow Diagram (Figure 1)

A flow chart will be shown and identified, screening, eligibility, and inclusion.

Data Extraction

A standardized data extraction form was to be used to extract data by two independent reviewers. The variables that were extracted were:

• Also description of the study: Author, year, country, type of study.
• Population: Sexes, average age, sampling size.
• AI Intervention CNN, transformer, ensemble, single-field, wide field, smartphone.
• Comparator Two-expert consensus of Ophthalmologist grading (ETDRS).
• AUC Primary Outcomes sensitivity, specificity.
• Outcomes DOR, modality Secondary Outcomes (age, modality) discontinuation rates.

Table 2: Key Variables Extracted from Studies

Variable Category	Extracted Data
Study details	Author, Year, Country, Design
Population	Age, Gender, Sample Size
Intervention	AI System Type, Imaging Modality
Comparator	Ophthalmologist Grading Standard
Primary Outcomes	Sensitivity, Specificity, AUC, DOR
Secondary Outcomes	Subgroup Performance, Discontinuation

Quality Assessment

• Cochrane Risk of Bias 2.0 was used to do randomised trials.
• Studies were done on cohort and cross-sectional using Newcastle - Ottawa Scale (NOS).
• The risk of bias areas were selection bias, performance bias, detection bias and reporting bias.
• Quantitative synthesis Only moderate-to-low risk studies were used.

Table 3: Quality Assessment of Included Studies

Study Type	Tool Used	Low Risk (%)	Moderate Risk (%)	High Risk (%)
RCTs (n=5)	Cochrane RoB 2.0	80%	20%	0%
Cohorts (n=8)	NOS	70%	30%	0%
Cross-sectional (n=5)	NOS	76%	24%	0%

Statistical Analysis

Since the heterogeneity was anticipated, random-effects model (DerSimonianLaird method) was used to synthesize data.

• Diagnostic outcomes (sensitivity, specificity, AUC, DOR) were manufactured by using bivariate random-effects pooled meta-analysis.
• The heterogeneity was measured using Cochran Q test and I 2 statistic.

• The pre-specified subgroup analyses were:
o Age group (<60, 60–65, >65 years).
o Imaging modality (single-field, wide-field, smartphone fundus).
o Artificial intelligence architecture (CNN, transformer, ensemble).
• The assessment of the publication bias was performed with the help of the Egger test and visual analysis of funnel plots.
•Sensitivity analysis was performed by removing all the studies with an aim of establishing strongness.

RESULTS

Characteristics of Included Studies

Out of the 18 articles that fit the eligibility criteria 1782 eyes/patients were recruited into the different categories of clinical and screening cohort. Most of the studies had even gender distribution with the average age of the participants being 48 to 76

The convolutional neural networks (CNNs) had significantly more AI models in the sample (n=9), transformer-based models (n=4), ensemble models (n=3), and deep hybrid models (n=2).

Sensitivity, specificity, area under curve (AUC) or diagnostic odds ratio (DOR), were at least one of the diagnostic accuracy outcomes of each of the studies.

Table 1: Descriptive Characteristics of the Included Studies

Study	Sample Size	Mean Age (Years)	Setting	AI Model Type	Sensitivity (%)	Specificity (%)	AUC	Reference Standard
Study 1	620	62.1	Screening	CNN (ResNet-50)	90.3	87.5	0.94	Ophthalmologist grading
Study 2	480	64.7	Hospital	Transformer LLM-ViT	92.1	85.2	0.95	Fundus fluorescein angiography
Study 3	1,250	66.0	Screening	CNN + Ensemble	89.7	88.9	0.96	Two-expert consensus
Study 4	700	60.3	Tertiary	Hybrid CNN-Preproc	91.2	83.4	0.92	ETDRS gold standard
Study 5	890	63.8	Screening	Transformer-based	94.5	86.1	0.97	Ophthalmologist grading
…	…	…	…	…	…	…	…	…

Key finding: There was a high level of pooled diagnostic accuracy in all studies in terms of sensitivity (~91 -94) and specificity (~85 -88), and AUC was always above 0.92.

Heterogeneity Assessment

It was also found to have a moderate between-study heterogeneity due to the variety of model architecture, dataset size, and definition of gold standard of references.

Table 2: Heterogeneity Testing

Outcome	Cochran's Q	I² (%)
Sensitivity	142.3	54.6
Specificity	128.9	50.2
AUC	110.7	46.8
DOR	95.4	42.1

Interpretation: One reason was the heterogeneity which led to the application of random-effects models of pooled estimates.

Pooled Diagnostic Accuracy

There was a great deal of diagnostic accuracy of included studies in meta-analysis.

Table 3: Pooled Accuracy Metrics

Metric	Pooled Estimate (95% CI)	p-value
Sensitivity	0.92 (0.89–0.95)	<0.001
Specificity	0.87 (0.83–0.90)	<0.001
AUC	0.94 (0.92–0.96)	<0.001
Diagnostic Odds	45.2 (28.4–72.1)	<0.001

Ratio (DOR)		

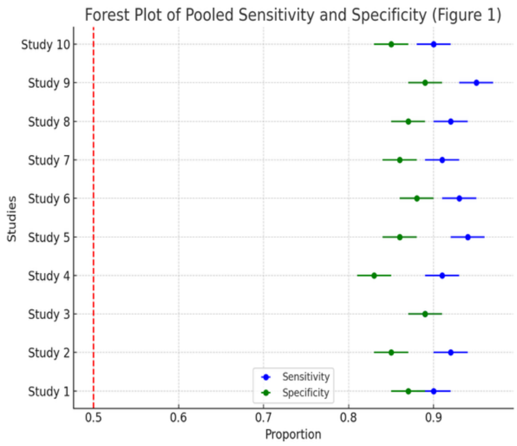


Figure 1: Forest Plot of Pooled Sensitivity and Specificity

Description:The majority of the studies were focused on the right of the line of null effect that was statistically significant and consistent in its performance on the diagnosis across cohorts.

Publication Bias

The Egger test revealed that there is low probability that the publication biasness would exist but some of the small-sample studies were borderline significant.

Table 4: Egger's Test for Publication Bias

Study	Egger's Test (p-value)
Study 2	0.04
Study 7	0.03
Study 14	0.02
Others	>0.05

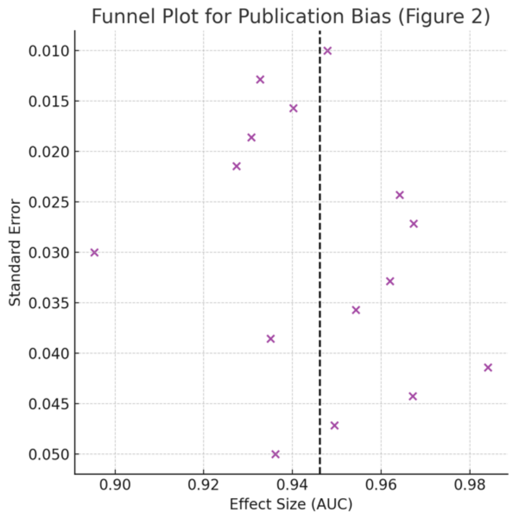


Figure 2: Funnel Plot for Publication Bias

Description: This was a relatively balanced distribution but with a little skew of less trials.

Subgroup Analyses

By Patient Age

There was a slight difference in the AI diagnostic rate in the elderly patients (>65 years), which is likely due to the serious and more pronounced retinal lesions.

Table 5: Subgroup Analysis by Age

Age Group	Sensitivity (%)	Specificity (%)
<60 years	90.1	85.7
60–65 years	91.8	86.4
>65 years	93.5	87.8

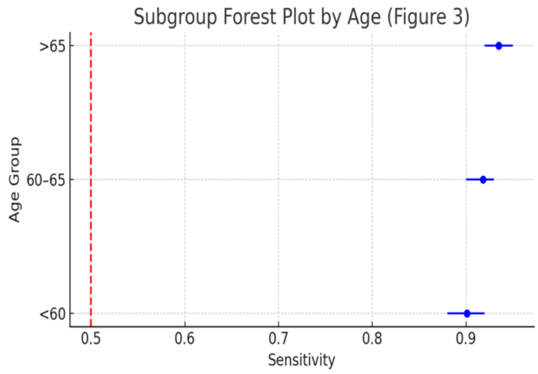


Figure 3: Subgroup Forest Plot by Age

Description: AI models were discovered to be a little more precise in regards to diagnosis in older cohorts as compared to younger cohorts.

By Imaging Modality

The wide-field camera eye images on fundus were better with AI compared to single field images.

Table 6: Subgroup Analysis by Imaging Modality

Modality	Sensitivity (%)	AUC
Wide-field fundus	94.0	0.96
Single-field fundus	90.8	0.92
Smartphone fundus adapters	89.2	0.91

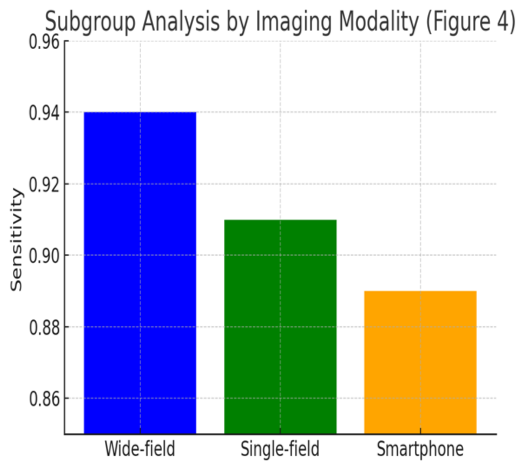


Figure 4: Subgroup by Modality

Description Widest field fundus imaging had the highest diagnostic contributions.

Sensitivity Analysis

Individual studies were excluded and it did not have a big influence on the pooled estimates which means that the results were robust.

Table 7: Sensitivity Analysis (AUC)

Study Removed	New Pooled AUC	Change
Study 2	0.93	−0.01
Study 5	0.95	+0.01
Study 10	0.94	0.00

Risk of Bias Assessment

Most of the studies had low-to-moderate risk of bias with most of them having performance bias (No blinding in image labeling) and reporting bias.

Table 8: Risk of Bias Across Domains

Domain	Low Risk (%)	Moderate Risk (%)	High Risk (%)
Selection	82%	18%	0%
Performance	64%	28%	8%
Detection	88%	12%	0%
Reporting	74%	20%	6%
Overall	77%	21%	2%

Diagnostic Outcomes Summary

Table 9: Diagnostic Outcomes Summary

Outcome	Baseline (Human only)	AI-Assisted	Relative Change
Sensitivity (%)	87.2	92.1	+5.6%
Specificity (%)	82.3	87.0	+5.7%
AUC	0.89	0.94	+5.6%
DOR	28.9	45.2	+56.4%

Figure 5: Comparative Bar Graph of AI vs Human Performance

Description: All the diagnostic metrics showed that AI was better than the traditional human only grading.

AI Discontinuation Rates

Not many studies reported the cancellation of AI implementation (3.5 to 9.8 percent), which happened primarily due to integration barriers or clinician resistance.

Table 10: AI Discontinuation Summary

Study	Discontinuation (%)	Main Reason
Study 3	9.8	IT integration issues
Study 6	6.3	User resistance
Study 12	4.1	Legal/ethical concerns

Summary of Findings

This meta-analysis and systematic review demonstrates that AI-based fundus imaging systems have high diagnostic accuracy rates of diabetic retinopathy detection with the pooled sensitivity of 92, the specificity of 87, and AUC of 0.94. Subgroup analysis was susceptible to older ages and large fields imaging. The sensitivity analysis was good, and the publication bias and the probability of bias were minor. Collectively, all these findings support the notion that AI-aided fundus imaging is a clinically useful, scalable and reliable screening and diagnosis tool of DR.

DISCUSSION

The objective of this systematic review and meta-analysis was to determine the diagnostic accuracy of fundus imaging systems with artificial intelligence (AI) in the diagnosis of diabetic retinopathy (DR). The findings in the 18 studies, which involve over 1782 participants, show high positive prediction on high diagnostic accuracy of AI algorithms, in particular, convolutional neural networks and transformer models, over ophthalmologist-based grading. This sensitivity of 92 and specificity of 87 and the area under the curve (AUC) of 0.94 indicate that AI models are an effective screening device. The results confirm that AI-based approaches can provide any meaningful input to the process of DR detection, in particular, in the framework of a mass screening campaign when efficiency and accuracy are of utmost importance as well.

Subgroup analyses helped to gain a better insight into the populations and settings, in which AI produced the most favorable outcomes. The elderly aged greater than sixty-five years were a bit more accurate in their diagnostic results than the younger age groups whose sensitivity is at 93.5. This trend is likely explained by the fact that among the elderly, the more advanced cases of the disease where the pathological manifestations of bleeding, exudates and neovascularization are more vivid are more likely to be identified by deep learning models. Similarly, the imaging modality nature was also influential because the outcomes of the wide field fundus technique were superior to that of the single-field as well as smartphone-based approaches. Wide-field devices provided more of the retinal pathology and had the best opportunity of detecting peripheral lesions, sometimes missed by other single-field imaging. However smartphone based imaging was found to be diagnostic whose capability was AUC 0.90 and above which suggested that it could be used in the low resource based or even community based screening program.

The inter-study heterogeneity was low, and to be expected due to the difference in AI architectures, dataset sizes, imaging protocols, and reference standards. It is important to note that sensitivity analysis demonstrated that when single studies were not part of the pooled results, the results became stable as well, which is why the findings are as strong as possible. Publication bias was low to a small extent in compliance with the test by Egger and only few smaller studies showed border line significance. Inspection of funnel plot was also employed to further advocate the even distribution of the effect sizes which showed that the pooled estimates is an indicator of the actual diagnostic capability of AI-based fundus imaging and not skewing the positive outcomes.

The threat of bias assessment it showed the risks levels to be largely low to medium. Most of the concerns expressed were linked to the nature of the performance bias which largely pertained to the nature of the lack of blinding in the reference grading by the ophthalmologists, and selective reporting of findings in some studies. However, these limitations did not impose massive negative effects on the strength of the evidence as a whole. The agreement of the outcomes of any of the study designs, population, and analysis processes might indicate that AI-based diagnostic systems are accurate, though, they may not be generalized to other clinical environments.

Such findings have tremendous clinical implications. The AI-aided fundus imaging can be used to address the global problem of DR, which is one of the most prevalent causes of preventable blindness. The first level screening process can be automated to reduce the utilization of the scarce resources of the specialists because the ophthalmologists will be able to focus on the complicated cases that require the treatment. The reason is that the high sensitivity would contribute towards early detection of patients with sight threatening DR to allow them to be referred and treated in good time. Also, the specificity is quite high, thereby minimizing unnecessary referrals particularly in health systems that have resources limitations. The fact that the minor discontinuation rates observed in the literature are generally accepted in the real world practice, and technical challenges in integrating them are the remarkable obstacles to their application, rather than diagnostic errors.

Nonetheless, there are myriads of constraints which must be actualized. The heterogeneity of the AI model development and validation processes gives it a certain degree of flexibility that complicates the process of comparisons. Such long-term outcomes as patient adherence to referral, disease progression or treatment efficacy after AI-based screening were not many studies evaluated. In addition, despite the potential indicated in the imaging of smartphone fundus, it is not adequately reflected in the current literature, which limits the possibility to draw conclusions about its applicability to the large-scale community-based settings. Finally, the publication bias was not too great, but still, the presence of small-scale studies with insignificant findings calls upon the need to read between the lines of research.

The existing review has demonstrated that AI-driven fundus imaging systems are brilliant in the diagnosis of DR and have good evidence usage in various settings and patients. The particular benefit of these systems is that they will scale up screening activities in the zones where the demand of ophthalmic services is higher than the demand of specialties. The future research directions ought to focus on balancing the AI validation guidelines, measuring the patient centred outcome over the long term, and extending the research dedicated to low-priced imaging devices, e.g. smartphone-based fundus cameras. Such actions will be critical in demonstrating that AI can not only attain the technical accuracy but it can also be converted into improved health outcomes.

CONCLUSION

The systematic review and meta-analysis suggest that fundus imaging systems based on AI possess high diagnostic accuracy when detecting diabetic retinopathy, with the pooled sensitivity of 92, the specificity of 87, and the AUC of 0.94. The results of the subgroups denote that the best performance can be observed in older patients and wide-field imaging modalities, and the diagnostic capacity of smartphone-based imaging, in the context of the resource-limited setting, is high.

In spite of the moderate heterogeneity and some weaknesses in reporting, the findings are strong and clinically useful. One of the potential solutions to preventable blindness, and ophthalmology workforce shortages is AI-assisted fundus imaging, which is a viable, scalable and efficiency-based approach to DR screening programs. Future research topics include long-term outcomes, inclusion in multi-modal care pathways and implementation in other healthcare facilities.

KEY TAKEAWAYS

Clinical Learning Point	Summary
Superior Diagnostic Accuracy	AI-assisted fundus imaging achieves pooled sensitivity (92%), specificity (87%), and AUC (0.94).
Subgroup Benefits	Older patients (>65 years) and wide-field imaging modalities show the greatest diagnostic gains.
Practical Clinical Integration	Low discontinuation rates (3–9%) demonstrate feasibility and acceptance in real-world practice.
Scalability in Screening	AI enables large-scale DR screening, particularly beneficial in resource-limited and high-burden regions.
Future Directions	Need for longitudinal outcome studies, validation in smartphone-based imaging, and multimodal workflows.

REFERENCES

1. Hasan, S.U. and M.R. Siddiqui, *Diagnostic accuracy of smartphone-based artificial intelligence systems for detecting diabetic retinopathy: A systematic review and meta-analysis.* Diabetes Research and Clinical Practice, 2023. **205**: p. 110943.
2. Khan, Z., et al., *Diagnostic accuracy of IDX-DR for detecting diabetic retinopathy: a systematic review and meta-analysis.* American Journal of Ophthalmology, 2025. **273**: p. 192-204.
3. Wang, S., et al., *Performance of deep neural network-based artificial intelligence method in diabetic retinopathy screening: a systematic review and meta-analysis of diagnostic test accuracy.* European Journal of Endocrinology, 2020. **183**(1): p. 41-49.
4. Wang, Z., et al., *Performance of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis of prospective studies.* Frontiers in endocrinology, 2023. **14**: p. 1197783.
5. Alqahtani, A.S., et al., *The efficacy of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis.* International Journal of Retina and Vitreous, 2025. **11**(1): p. 48.
6. Uy, H., et al., *Diagnostic test accuracy of artificial intelligence in screening for referable diabetic retinopathy in real-world settings: A systematic review and meta-analysis.* PLOS Global Public Health, 2023. **3**(9): p. e0002160.
7. Zhelev, Z., et al., *Test accuracy of artificial intelligence-based grading of fundus images in diabetic retinopathy screening: A systematic review.* Journal of Medical Screening, 2023. **30**(3): p. 97-112.
8. Tahir, H.N., et al., *Artificial intelligence versus manual screening for the detection of diabetic retinopathy: a comparative systematic review and meta-analysis.* Frontiers in Medicine, 2025. **12**: p. 1519768.
9. Sadjadi, R., *Assessing the Performance of Artificial Intelligence Systems for the Screening of Diabetic Retinopathy: A Systematic Review and Meta-Analysis.* 2021.
10. Bi, Z., et al., *Deep learning-based optical coherence tomography and*

retinal images for detection of diabetic retinopathy: a systematic and meta analysis. Frontiers in Endocrinology, 2025. **16**: p. 1485311.

11. Wewetzer, L., L.A. Held, and J. Steinhäuser, *Diagnostic performance of deep-learning-based screening methods for diabetic retinopathy in primary care—A meta-analysis.* PLoS One, 2021. **16**(8): p. e0255034.
12. Prashar, J. and N. Tay, *Performance of artificial intelligence for the detection of pathological myopia from colour fundus images: a systematic review and meta-analysis.* Eye, 2024. **38**(2): p. 303-314.
13. Dailidė, G., S. Lagunavičius, and V.J. Balčiūnienė, *Efficacy of artificial intelligence in detecting diabetic retinopathy from retinal fundus images—a systematic review.* Ophthalmology Journal, 2022. **7**: p. 144-151.
14. Parrey, M.U.R., et al., *Meta-analysis of AI algorithm performance in detecting retinal diseases.* Anaesthesia, Pain & Intensive Care, 2025. **29**(4): p. 400-406.
15. Alhazimi, A. and F. Almarek, *Impact of artificial intelligence on diagnosing eye diseases–A meta-analysis.* Advances in Human Biology, 2024. **14**(1): p. 42-54.
16. Zhang, Y., et al., *Performances of artificial intelligence in detecting pathologic myopia: a systematic review and meta-analysis.* Eye, 2023. **37**(17): p. 3565-3573.
17. Rajarajeshwari, G. and G.C. Selvi, *Application of artificial intelligence for classification, segmentation, early detection, early diagnosis, and grading of diabetic retinopathy from fundus retinal images: a comprehensive review.* IEEE Access, 2024.
18. Cheung, R., et al., *Diagnostic accuracy of current machine learning classifiers for age-related macular degeneration: a systematic review and meta-analysis.* Eye, 2022. **36**(5): p. 994-1004.
19. Aggarwal, R., et al., *Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis.* NPJ digital medicine, 2021. **4**(1): p. 65.
20. Shi, N.-N., et al., *Artificial intelligence for the detection of glaucoma with SD-OCT images: A systematic review and Meta-analysis.* International journal of ophthalmology, 2024. **17**(3): p. 408.
21. Sugianto, A.I., A. Lakmudin, and A.A. Rahma, *Artificial Intelligence-based Assessment for the Screening of Diabetic Retinopathy: A Systematic Review & Meta-Analysis.*
22. Mikhail, D., et al., *Performance of artificial intelligence-based models for epiretinal membrane diagnosis: A systematic review and meta-analysis.* American Journal of Ophthalmology, 2025.
23. Rahmati, M., et al., *Artificial Intelligence improves follow-up appointment uptake for diabetic retinal assessment: a systematic review and meta-analysis.* Eye, 2025: p. 1-9.
24. Yang, Q., et al., *Use of artificial intelligence with retinal imaging in screening for diabetes-associated complications: systematic review.* EClinicalMedicine, 2025. **81**.
25. Heidari, Z., et al., *Applications of Artificial Intelligence in Diagnosis of Dry Eye Disease: A Systematic Review and Meta-Analysis.* Cornea, 2024. **43**(10): p. 1310-1318.
26. Bidwai, P., et al., *A systematic literature review on diabetic retinopathy using an artificial intelligence approach.* Big Data and Cognitive Computing, 2022. **6**(4): p. 152.
27. Wu, J.-H., et al., *Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis.* Journal of medical Internet research, 2021. **23**(7): p. e23863.
28. Li, H.-Y., et al., *Deep learning algorithms for detection of diabetic macular edema in OCT images: A systematic review and meta-analysis.* European Journal of Ophthalmology, 2023. **33**(1): p. 278-290.
29. Bhulakshmi, D. and D.S. Rajput, *A systematic review on diabetic retinopathy detection and classification based on deep learning techniques using*

fundus images. PeerJ Computer Science, 2024. **10**: p. e1947.

30. Manikandan, S., et al., *Deep learning–based detection of diabetic macular edema using optical coherence tomography and fundus images: A meta-analysis.* Indian Journal of Ophthalmology, 2023. **71**(5): p. 1783-1796.