

## Systematic Review

# Accuracy of artificial intelligence versus eyecare practitioners in detecting diabetic retinopathy using retinal fundus images: a systematic literature review

Peter K. Mutuma<sup>1\*</sup>, Warren Ayonga<sup>2</sup>, Marera Domnic<sup>2</sup>

<sup>1</sup>Department of Clinical Medicine, School of Health Sciences, Kaimosi Friends University, Vihiga, Kenya

<sup>2</sup>Department of Human Anatomy, School of Health Sciences, Maseno University, Kisumu, Kenya

**Received:** 09 August 2026

**Revised:** 05 October 2026

**Accepted:** 07 October 2026

### \*Correspondence:

Peter K. Mutuma,

E-mail: [pkomzee@gmail.com](mailto:pkomzee@gmail.com)

**Copyright:** © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

Diabetic retinopathy (DR) prevalence rises among low- and middle-income countries where access to appropriate eye care services is lacking. Detecting DR requires retinal fundus imaging, but interpretation conventionally depends on eye-care practitioners, who are affected by shortcomings like inter-observer variability. Artificial Intelligence (AI), a promising technology for automated image examination and interpretation, enables wider, faster screening coverage and consistency. The review compares the diagnostic accuracy of AI and practitioners in detecting DR using retinal images. We conducted a systematic literature review using PRISMA 2020, searching major electronic databases with search terms: diabetic retinopathy, artificial intelligence, eye-care practitioners, and diagnostic accuracy. Eligible studies were published in English, peer-reviewed, and met the inclusion and exclusion criteria. Data extraction, organization, and accuracy were enabled through an Excel data form. The review included 15 of 540 identified studies. AI diagnostic tools like Google's ARDA show equal or higher accuracy than humans in detecting DR, in some instances, in categorizing images. AI brings hope and assurance in prompt identification, management, and referral. Its performance is limited by the tendency to flag images as false positives; however, this depends on image quality, disease severity, datasets used, and overall grading criteria. Human strength remains in clinical judgement, contextualised interpretation of ungradable or low-quality images, and patient-specific management approach. AI implementation to complement care should be steered by local validation through comparison with eye care givers and approval through standardized protocols, quality control, ethical considerations, and data security.

**Keywords:** Diabetic retinopathy, Artificial intelligence, Eye-care practitioners

## INTRODUCTION

Diabetic retinopathy (DR) is a major microvascular complication of diabetes mellitus and an important cause of avoidable visual impairment among working-age adults.<sup>1</sup> DR burden is expected to rise considerably with increasing global prevalence of diabetes.<sup>1</sup> Detection of DR remains challenging due to factors related to the patient, disease, imaging and eye care practitioners, with the burden being greater in low- and middle-income countries.<sup>2</sup> Despite studies showing DR as an important

and growing cause of visual morbidity, many patients are diagnosed late as a result of limited routine eye screening and access to eye care practitioners, as well as the unavailability of fundus cameras and inadequate referral pathways.<sup>1</sup>

Early detection of DR is vital as it ensures timely treatment and management to prevent progression to sight-threatening disease.<sup>3</sup> Retinal fundus imaging is vital for screening as it allows visualization and documentation of lesions and disease severity.<sup>4</sup> Conventionally,

interpretation of fundus images has depended on trained eye-care practitioners, including ophthalmologists, optometrists, ophthalmic clinical officers, ophthalmic nurses, graders and retina specialists.<sup>5</sup> The benefit of human assessment is quick clinical judgment, contextual interpretation and the ability to integrate symptoms, visual acuity, systemic risk factors and ocular comorbidities in the final diagnosis,<sup>6</sup> despite being limited by shortages of trained personnel, inter-observer variability, fatigue, workload, cost and delayed reporting.<sup>7</sup>

AI has emerged as a promising approach to support DR screening from retinal fundus images, as deep learning and machine learning systems can be trained to identify DR-related retinal changes and classify images.<sup>8,9</sup> Currently, AI algorithms can achieve high diagnostic performance in sensitivity and specificity compared with human graders.<sup>10</sup> AI may therefore improve screening coverage, reduce diagnostic delays, support task shifting and allow rapid triage of patients who require referral services. This is especially relevant in settings where the number of people with diabetes is increasing faster than the availability of eye-care practitioners.<sup>11</sup>

Despite these advantages, the accuracy of AI compared with eye-care practitioners remains an important question.<sup>12</sup> AI systems' performance may vary with image quality, camera type, pupil dilation status, ethnicity, disease severity distribution, co-existing retinal disease, algorithm training data and grading criteria.<sup>13</sup> On the other hand, eye-care practitioners vary in training, experience, access to grading protocols, and exposure to retinal image interpretation.<sup>5</sup> In Kenya, for example, a study evaluating digital retinal camera screening by ophthalmic nurses, ophthalmic clinical officers, and a county ophthalmologist against a retina specialist highlights the importance of assessing diagnostic accuracy across different cadres of eye-care providers, not only among ophthalmologists.<sup>14</sup> Therefore, direct comparison between AI and eye-care practitioners is necessary to determine whether AI can complement or strengthen existing DR screening systems. Therefore, this systematic literature review compared the diagnostic accuracy of AI and eye-care practitioners in detecting DR using retinal fundus images and helped in identifying gaps for safe, ethical and evidence-based integration of AI into ophthalmic eye care screening. The findings may inform clinical practice, digital health policy, training needs, referral models and future research on safe and effective integration of AI into diabetic retinopathy screening programmes.

## METHODS

A systematic review design guided by PRISMA 2020 was used. These methods were applied to ensure accurate identification, selection, appraisal, and synthesis of relevant studies to generate evidence on the comparative diagnostic accuracy of artificial intelligence and eye care practitioners in diabetic retinopathy detection.

## Review design

This study review adopted a systematic literature review design using PRISMA 2020.<sup>15</sup> A systematic literature review is appropriate because it uses a transparent, structured, and reproducible method to identify and synthesize available evidence on a specific research question.

## Search strategy and selection criteria

An extensive literature search was conducted in major electronic databases, including PubMed/MEDLINE, Google Scholar, Scopus, Web of Science, Cochrane Library and Embase, where accessible. Using keywords and Boolean operators related to diabetic retinopathy, artificial intelligence, eye-care practitioners, retinal fundus images, and diagnostic accuracy. The study selection criteria were applied in stages. First, all records identified from the database search will be exported into EndNote, where duplicate records will be removed. Second, titles and abstracts will be screened against the predefined eligibility criteria. Third, all relevant studies underwent full-text review and reasons for exclusion were documented. The final number of included studies was presented using a PRISMA flow diagram.

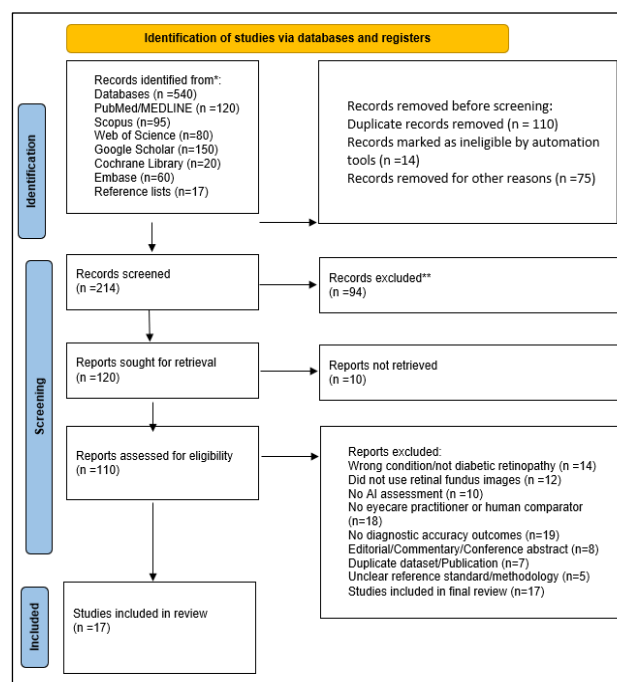


**Figure 1: PRISMA 2020 model for selection of studies.**

Included studies were peer-reviewed full-text articles published in English that evaluated DR detection using retinal fundus images, assessed AI, machine learning, deep learning, or automated retinal image analysis systems, compared AI performance with eyecare practitioners or human graders and reported at least one diagnostic accuracy outcome, such as sensitivity, specificity, accuracy, area under the curve, positive predictive value,

negative predictive value, or kappa agreement and used an acceptable reference standard such as retina specialist grading, ophthalmologist diagnosis, Ophthalmic clinical officer diagnosis, expert panel grading, or internationally recognized diabetic retinopathy grading criteria

Studies excluded were those that did not focus on diabetic retinopathy, did not use retinal fundus images, assessed other retinal diseases without reporting separate DR results, lacked comparison with human graders or eyecare practitioners, failed to report diagnostic accuracy outcomes, had unclear methodology, lacked a defined reference standard, or were editorials, opinion papers, letters to the editor, conference abstracts without full text, commentaries, narrative reviews, or duplicate publications without additional relevant.

### **Data extraction procedure**

Data were extracted from the included studies using a standardized data extraction form. The form was designed and developed in Microsoft Excel before data extraction to ensure consistency and completeness. The literature search and screening process was guided by the PRISMA framework, with all eligible studies independently reviewed to extract relevant information, including author name, year of publication, country or setting, study title, study design, study population or image dataset, sample size, type of retinal fundus images used, diagnostic method assessed, comparator group, reference standard, sensitivity, specificity, accuracy, area under the curve, agreement measures, key findings, limitations, and conclusions. The PRISMA process was used, where identified records were screened, eligibility assessed, and final inclusion determined, showing the number of records identified from databases, duplicates removed, records screened by title and abstract, full-text articles assessed, studies excluded with reasons, and studies finally included in the review. Data searching and collection were standardised and extracted by two individuals and evaluated independently and all discrepancies were resolved by consensus via the contribution of a third reviewer. The raw data were then cleaned, organized and summarized to allow for comparison of the accuracy of AI and eye care practitioners in detecting diabetic retinopathy through interpretation of retinal images.

## **RESULTS**

A total of 540 studies were identified, with 15 studies included in the final analysis. The reviewed literature shows that diabetic retinopathy remains a major and increasing public health problem, especially as the global burden of diabetes continues to rise. The problem is more serious in low- and middle-income countries, where late diagnosis, inadequate routine screening, shortage of trained eye-care practitioners, scarce fundus cameras, and weak referral systems contribute to delayed diagnosis and treatment. Retinal fundus imaging is identified as a key screening tool as it allows visualization and documentation

of retinal lesions. Conventionally, fundus image interpretation has depended on eye-care practitioners such as ophthalmologists, optometrists, ophthalmic clinical officers, ophthalmic nurses, retinal graders, and retina specialists. Human interpretation remains valuable as it applies clinical judgment, patient symptoms, visual acuity, systemic risk factors and other ocular conditions. This is, however, affected by shortages of trained personnel, inter-observer variability, fatigue, workload, cost, and delayed reporting.

The study findings also demonstrate that artificial intelligence, especially deep learning and machine-learning systems, shows strong potential to support diabetic retinopathy screening by classifying retinal fundus images as normal, non-referable, referable, or sight-threatening disease. The current evidence suggests that AI algorithms can achieve high diagnostic performance in terms of sensitivity and specificity when compared with human graders or reference standards. The findings indicate that AI has potential to improve screening coverage, reduce diagnostic delays, support task-shifting, and assist in rapid triage and referral of patients. Despite this, the diagnostic accuracy may vary depending on image quality, type of camera, pupil dilation, population features, disease severity, co-existing retinal conditions, training datasets, and grading criteria. Similarly, the performance of eye-care practitioners may vary depending on professional cadre, level of training, experience, and access to standardized grading protocols. Hence, direct comparison between AI and eye-care practitioners is necessary before AI can be safely integrated into diabetic retinopathy screening programmes.

The findings, therefore, illustrate the important role of AI in complementing rather than replacing eye-care practitioners in the detection of DR. This is because human practitioners still retain the strengths of superior clinical judgement, patient empathy, the ability to tailor a patient-specific treatment plan, the ability to offer contextualised interpretation of ungradable or low-quality images and the ability to decide on the referral of patients in need of specialised care. We can comfortably conclude that AI offers quick, efficient and consistent results, while eye-care practitioners provide superior clinical judgment, contextualised interpretations, patient empathy and the ability to make quick decisions on patient referral to a specialist. The major gaps that are explicitly portrayed in the reviews are the direct comparison and local validation of the accuracy of AI systems. The comparison and validation should be tailored to ensure a sustainable quality assurance protocol, proper ethical considerations, and induction of eye-care practitioners in accurate AI systems use before AI-assisted DR screening is implemented for routine purposes, particularly in low-resource settings.

## **DISCUSSION**

Diabetic retinopathy is a significant cause of avoidable visual impairment and, if left unmanaged, can ultimately

cause total blindness.<sup>1</sup> It is estimated that one-third of all diabetic patients worldwide have some form of the disease, with about 35% of individuals with diabetes in Sub-Saharan Africa found to have diabetic retinopathy.<sup>16</sup> This is expected to rise due to the increasing prevalence of diabetes mellitus (DM); the number of people with DM is projected to reach 1.31 billion by 2050, up from 529 million in 2021 globally.<sup>1</sup> The situation is worse in low- and middle-income countries where the population has a shortage of well-trained eye-care personnel, limited resources, delayed diagnosis and weak referral pathways.<sup>1</sup> These limitations, coupled with the rising burden of DR, call for prompt, regular and accurate screening for retinal diseases.<sup>1</sup>

Diabetic retinopathy can remain asymptomatic in its early stages and progress from mild to sight-threatening disease; hence, routine screening is essential to identify changes before irreversible visual loss occurs.<sup>4</sup> Fundus imaging plays a key role in this process, allowing visualization, documentation, grading, and follow-up of retinal lesions and disease severity.<sup>4</sup> The effectiveness of any DR screening programme depends not only on the availability of imaging technology but also on accurate interpretation of retinal images and functional referral systems.<sup>3</sup> Conventionally, interpretation of retinal images has depended on trained eye-care practitioners.<sup>5</sup> as they can integrate retinal findings with patient symptoms, systemic risk factors and co-existing eye disease, making human judgment particularly valuable in complex cases, poor-quality images, atypical retinal lesions, and patients with multiple ocular or systemic comorbidities.<sup>6</sup> However, the literature also highlights important limitations of human-based screening, including a shortage of trained personnel, inter-observer variability, fatigue, high workload, cost, and delayed reporting.<sup>7</sup> These challenges are especially significant in resource-limited settings.

The integration of AI tools in medical care has proven an important factor in supporting DR screening, as deep learning and machine learning systems can be trained to detect retinal abnormalities associated with DR and classify retinal images.<sup>9</sup> Evidence from the literature suggests that AI can achieve high diagnostic performance, particularly in sensitivity and specificity, when compared with human interpreters or reference standards.<sup>10</sup> This shows that AI has the potential to improve screening coverage, reduce reporting delays, support task shifting, and assist in the rapid triage of patients, especially in settings where the diabetic population is increasing faster than the availability of resources.<sup>11</sup>

The diagnostic accuracy of AI cannot be assumed to be uniform across all populations and clinical settings; the reviews portray the AI system's accuracy as one that is affected by image quality, pupil status, retinal pigmentation, ethnicity, DR severity, other comorbidities, standardisation of training data and the grading method used.<sup>13</sup> Training and validation of AI is done using data collected elsewhere from populations with dissimilar

characteristics from the local population; hence, they lack validation in low- and middle-income countries where the population has different characteristics altogether.<sup>10</sup> The performance of AI models trained mainly on high-quality images from well-resourced settings may not translate to images from a health centre, outreach screening programmes, or those taken with poorer-quality cameras. This justifies the need for validation in local populations to ensure standardisation, accuracy, and efficiency before full adoption of AI as a complementary tool in routine screening.<sup>17</sup>

The comparison between AI systems and eye-care practitioners is particularly important, as both exhibit a synergistic relationship. As stated earlier, AI has enormous advantages despite several other shortcomings.<sup>6</sup> On the other hand, human caregivers have uncontested abilities in contextualising their clinical reasoning, expertise and accuracy, which may differ depending on the level of training, experience, availability of grading guidelines and patient workload.<sup>5</sup>

Kenya, like other low-resourced settings, needs to adopt AI systems as a complement to eye care practitioners to strengthen the existing DR screening services, ensuring accessibility and improved screening accuracy while maintaining patient safety.<sup>1</sup> AI screening studies that have evaluated eye-care practitioners like ophthalmic nurses, ophthalmic clinical officers, and an ophthalmologist against a retina specialist have highlighted the importance of assessing the diagnostic accuracy of different eye-care providers at different levels.<sup>1</sup> Screening and diagnosis of DR involves ophthalmologists, nurses, clinical officers, optometrists and other eye-trained personnel; hence, comparing AI with all other eye-care practitioners involved in screening DR is more practical in ensuring an accurate picture of the happenings across the board is well validated for real-world implementation, rather than comparing AI with a single cadre.<sup>14</sup>

The use of AI as a tool to complement practitioners in detecting DR is highly justifiable and supported by the majority of the literature reviewed. This complementation will appropriately ensure prompt identification of cases, grading and quick referral of complications to a specialist while preserving the crucial role of human clinical judgment. Ultimately, eye-care practitioners and retina specialists are relieved to focus more on complex cases and management of sight-threatening disease. Successful implementation requires clear guidelines on quality assurance, staff training, ethical adherence, data security and local population involvement and AI validation.

Therefore, the systematic review comparing the diagnostic accuracy of AI and eye-care practitioners in detecting DR from retinal fundus images is justified in synthesizing available evidence, identifying performance differences, assessing methodological quality and guiding safe integration of AI into diabetic retinopathy screening.

The review had several strengths, including the use of fundus photography, which provides a standardized diagnostic modality that reflects the increased use of AI and telemedicine in patient management. Utilization of comprehensive search engines such as Scopus, PubMed/MEDLINE, among others, with predefined eligibility criteria and PRISMA 2020 reporting increases the likelihood of identifying crucial and relevant studies, reducing bias and dependence on a single database. However, the study was limited by the relatively small number of studies specifically evaluating AI detection for diabetic retinopathy in low- and middle-income countries like Kenya. Different studies tend to differ significantly due to various underlying structures and tools used, hence making direct comparison of data difficult.

The study included different eye care practitioners who have different training and experience in interpreting the findings, which makes comparison with AI unequal. Selection bias in the studies included, whereby some studies may use highly selected datasets containing clear, high-quality images, unlike others that may have poorer images, hence high differences in accuracy estimates. Lastly, there is limited data from local settings, hence reliance on datasets from countries or ethnic populations that differ from the Kenyan ethnic setup, hence restricting generalizability of findings.

## CONCLUSION

The existing literature review shows that AI has solid potential to complement eye-care practitioners in screening and detection of diabetic retinopathy, especially in resource-limited settings where available speciality screening capability is way below the demand for the service. Implementation should be guided by results from local validation, quality assurance, ethical safeguards, training, data protection and clear referral pathways.

## Recommendations

Artificial intelligence should be incorporated into the screening of diabetic retinopathy as a supportive tool, as it has shown promising potential in its screening and detection using retinal fundus images. The system, however, needs validation using retinal fundus images from the targeted population, especially in low-resource settings. Eyecare practitioners should receive continuous training in diabetic retinopathy grading, retinal image interpretation and use of AI outputs. Future research should compare AI with different cadres of eye-care practitioners under real-world conditions and report standardized diagnostic accuracy outcomes.

*Funding: No funding sources*

*Conflict of interest: None declared*

*Ethical approval: The study was approved by the Institutional Ethics Committee*

## REFERENCES

1. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. 2023;402(10397):203-34.
2. Aftab O, Khan H, VanderBeek BL, Scoles D, Kim B, Tsui, JC. Evaluation of ChatGPT-4 in detecting referable diabetic retinopathy using single fundus images. *AJO Int*. 2025;2(2):100111.
3. Dar MA, Maqbool M, Ara I, Qadrie, Z. Preserving sight: Managing and preventing diabetic retinopathy. *Open Health*. 2023;4(1):20230019.
4. Tahir MZ, Nasir M, Zhang S. Advances in retinal microaneurysms detection, segmentation and datasets for the diagnosis of diabetic retinopathy: a systematic literature review. *Multimedia Tool Appl*. 2024;83(30):74897-935.
5. Cheung R. The key to retinal imaging. University of New South Wales (Australia). 2024.
6. Sevgi M, Antaki F, Khan AZ, Ong AY, Merle DA, Hu K, et al. Complementary Human-AI Clinical Reasoning in Ophthalmology. *arXiv preprint* 2025;2510:22414.
7. Dow ER, Khan NC, Chen KM, Mishra K, Perera C, Narala R, et al. AI-Human Hybrid Workflow Enhances Teleophthalmology for the Detection of Diabetic Retinopathy. *Ophthalmol Sci*. 2023;3(4):100330.
8. Huang X, Wang H, She C, Feng J, Liu X, Hu X, et al. Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy. *Front Endocrinol (Lausanne)*. 2022;13:946915.
9. Mukherjee N, Sengupta S. Application of deep learning approaches for classification of diabetic retinopathy stages from fundus retinal images: a survey. *Multimed Tools App*. 2024;83(14):43115-75.
10. Joseph S, Selvaraj J, Mani I, Kumaragurupari T, Shang X, Mudgil P, et al. Diagnostic Accuracy of Artificial Intelligence-Based Automated Diabetic Retinopathy Screening in Real-World Settings: A Systematic Review and Meta-Analysis. *Am J Ophthalmol*. 2024;263:214-30.
11. Ripatti AR. Screening Diabetic Retinopathy and Age-Related Macular Degeneration with Artificial Intelligence. *New Innov Eye Care*. 2025.
12. Scanzer AC, Shorter E, Kinnaird C, Valikodath N, Al-Khaled T, Cole E, et al. Optometrist's perspectives of Artificial Intelligence in eye care. *J Optom*. 2022;15(1):S91-7.
13. Park SH, Choi J, Byeon JS. Key Principles of Clinical Validation, Device Approval, and Insurance Coverage Decisions of Artificial Intelligence. *Korean J Radiol*. 2021;22(3):442-53.
14. Ong'ang'o JR, Mashedi O, Gichangi M, Kiplimo R, Nyamori J, Alwenya K. The diagnostic accuracy of diabetes retinopathy screening by ophthalmic clinical officers, ophthalmic nurses and county ophthalmologists against a retina specialist in 2

selected county referral hospitals, Kenya. *Front Ophthalmol* (Lausanne). 2023;3:1082205.

15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
16. Wondmeneh TG, Mohammed JA. Prevalence of diabetic retinopathy and its associated risk factors among adults in Ethiopia: a systematic review and meta-analysis. *Sci Rep*. 2024;14(1):28266.
17. Shrinkhal, Shukla A, Shukla M, Garg P, Shukla R, Singh SP. Artificial Intelligence-Assisted Diagnosis of Retinal and Neuro-Ophthalmic Diseases: a Comparative Evaluation of ChatGPT, Google

Gemini and An Ophthalmologist. *Maedica (Bucur)*. 2025;20(4):765-70.

**Cite this article as:** Mutuma PK, Ayonga W, Domnic M. Accuracy of artificial intelligence versus eyecare practitioners in detecting diabetic retinopathy using retinal fundus images: a systematic literature review. *Int J Community Med Public Health* 2026;13:6883-8.