Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 2s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

AI-Based Diabetic Retinopathy Detection: Comparing Performance with Ophthalmologists

Sandeep kaur[1], Shubhanshi Singh[2], Trisha Tyagi[3], Saumya[4]

[1]Assistant Professor, Chandigarh Group of Colleges, Jhanjeri, Mohali, Punjab, India – 140307,Department of optometry, Chandigarh pharmacy college.

[2]Assistant Professor, Chandigarh Group of Colleges, Jhanjeri, Mohali, Punjab, India – 140307,Department of optometry, Chandigarh pharmacy college.

[3]Assistant Professor, Chandigarh Group of Colleges, Jhanjeri, Mohali, Punjab, India – 140307,Department of optometry, Chandigarh pharmacy college.

[4]Assistant professor, Jaipur national university , Jaipur, Rajasthan, India-302017, Department of optometry.

Cite this paper as: Sandeep kaur, Shubhanshi Singh, Trisha Tyagi, Saumya, (2025) AI-Based Diabetic Retinopathy Detection: Comparing Performance with Ophthalmologists. *Journal of Neonatal Surgery*, 14 (2s), 361-370.

ABSTRACT

Diabetic Retinopathy (DR) is a leading cause of preventable blindness globally. While effective treatments exist, timely detection remains a significant challenge due to limited access to specialists and increasing prevalence of diabetes. Artificial intelligence (AI) systems offer a promising solution for automated DR screening. This review examines current evidence on AI-based DR detection systems and their validation against ophthalmologist diagnosis, with emphasis on clinical implementation challenges. We analyze research published within the last five years (2020-2025), focusing on real-world implementation, performance metrics against the gold standard of expert grading, regulatory approvals, and cost-effectiveness. Recent deep learning algorithms have demonstrated high sensitivity (>90%) and specificity (>85%) for detecting referable DR, comparable to human specialists. However, successful clinical implementation requires addressing challenges including generalizability across diverse populations, integration with existing healthcare workflows, interpretability of AI decisions, and establishing appropriate regulatory frameworks. Cost-effectiveness analyses indicate potential for significant healthcare savings, particularly in underserved regions. We conclude by identifying key gaps and future directions for advancing AI-based DR screening toward widespread clinical adoption.

Keywords: *artificial intelligence, deep learning, diabetic retinopathy, screening, validation, clinical implementation, cost-effectiveness.*

1. INTRODUCTION

Diabetic retinopathy (DR) remains the leading cause of blindness among working-age adults worldwide, with over one-third of the estimated 537 million people with diabetes showing signs of DR.[1] The global burden of DR is projected to increase further as diabetes prevalence rises, particularly in low and middle-income countries where healthcare resources are already strained.[2]

Early detection and timely treatment can prevent up to 95% of vision loss cases.[3] However, multiple barriers impede effective screening, including shortage of eye care specialists, limited access to ophthalmological services in rural and underserved areas, and increasing burden on healthcare systems.

According to estimates from the World Health Organisation, more than half of diabetics do not have the recommended yearly eye exams.[4] These difficulties have increased interest in artificial intelligence (AI)-based automated screening systems. Recent developments in deep learning, a branch of machine learning that uses multi-layer neural networks, have made it possible to create algorithms that can analyse retinal pictures and identify DR with accuracy on par with human specialists.[5]

The present status of AI-based DR detection systems is examined in this thorough review, along with their validation against ophthalmologist diagnoses. We concentrate on works that have been published in the recent five years (2020-2025), highlighting difficulties with clinical implementation as opposed to algorithmic technicalities. The assessment examines cost-effectiveness, regulatory considerations, performance indicators in comparison to the gold standard of expert grading, and real-world implementation experiences. We hope to offer useful insights for physicians, healthcare administrators, and policymakers thinking about using AI-based DR screening systems by combining the most recent research.

More than one-third of the estimated 537 million individuals with diabetes exhibit symptoms of diabetic retinopathy (DR), which continues to be the primary cause of blindness among working-age adults globally.[1] As diabetes prevalence rises, the worldwide burden of DR is expected to grow even more, especially in low- and middle-income nations where healthcare resources are already stretched.

Up to 95% of instances of visual loss can be avoided with early detection and prompt treatment.[3] Effective screening is hampered by a number of factors, such as a lack of eye care professionals, restricted access to ophthalmological services in underserved and rural areas, and mounting strain on healthcare systems. According to projections from the World Health Organisation, more than half of diabetics do not have the needed yearly eye exams.[4]

2. OVERVIEW OF AI TECHNOLOGIES FOR DR DETECTION

2.1 Evolution of AI Approaches

Over the past ten years, there has been a significant evolution in the use of AI for DR detection. Prior to 2016, the majority of early systems used traditional machine learning methods, which necessitated the human extraction of characteristics including exudates, haemorrhages, and microaneurysms.[6] Although these methods showed mediocre performance, they necessitated extensive feature engineering and preprocessing.

With the advent of deep learning techniques, especially convolutional neural networks (CNNs), which can automatically extract pertinent characteristics from unprocessed picture data, the area saw a paradigm change. This shift was detailed in a thorough analysis by Grzybowski et al.[7] which found that deep learning techniques routinely outperformed conventional techniques on a variety of datasets.

The goal of the latest developments (2020-2025) has been to improve these deep learning architectures and deal with their drawbacks. Important developments include:

1. Ensemble Methods: integrating several models to enhance performance as a whole. When compared to single models, Wang et al.[8] showed that ensemble techniques could lower error rates by 15% to 20%.

2. Attention Mechanisms: These allow algorithms to concentrate on pertinent areas of the image. Li et al[9] have shown that attention mechanism-based approaches perform better when identifying subtle DR traits.

3. Explainable AI: Developing methods to understand and visualize algorithm decision making processes. Zhou et al.[10] implemented gradient-weighted class activation mapping to highlight retinal regions influencing the algorithm's classification decisions.

4. Few-Shot Learning: Developing models requiring fewer labeled examples, addressing the challenge of limited annotated data. Chen et al.[11] demonstrated effective DR grading using just 10% of the training data traditionally required.

5. Multimodal Integration: Combining fundus photography with other imaging modalities such as optical coherence tomography (OCT). Jiang et al.[12] showed improved detection of diabetic macular edema through integrated analysis of fundus and OCT images.

2.2 Current Leading AI Systems

Several AI systems have demonstrated robust performance in DR detection. Notable examples from the last five years include:

IDx-DR: The first FDA-approved autonomous AI system for DR detection. Updated versions have continued to show strong performance, with Abràmoff et al.[13] reporting sensitivity of 95.2% and specificity of 86.7% for detecting more-than-mild DR across diverse populations.

EyeArt: This deep learning system received FDA clearance in 2020. Recent validation by Heydon et al.[14] across 100,000 images from multiple international datasets demonstrated 94.7% sensitivity and 91.3% specificity for referable DR detection.

Google's Deep Learning System: While not commercially deployed as a standalone product, this system has been extensively validated. Sayres et al[15] reported performance exceeding that of general ophthalmologists for DR grading in a multi-center trial.

SELENA+: Developed in Singapore, this system has shown strong performance in Asian populations. Ting et al [5] demonstrated sensitivity of 96.6% and specificity of 93.2% for detecting referable DR in multi-ethnic Asian populations.

ARIAS (Automated Retinal Image Analysis System): Developed by Moorfields Eye Hospital and DeepMind, this system can detect multiple retinal conditions including DR. Faes et al.[16] documented performance matching specialist graders across a range of retinal pathologies.

Many of these systems are now transitioning from research tools to commercially deployed products, with varying degrees of clinical integration and regulatory approval, as will be discussed in later sections.

3. VALIDATION AGAINST OPHTHALMOLOGIST DIAGNOSIS

3.1 Gold Standard for Validation

The gold standard for validating AI systems in DR detection typically involves comparison against expert human grading, often using a consensus of multiple retina specialists or ophthalmologists.

Most validation studies employ established grading systems such as:

- **International Clinical Diabetic Retinopathy (ICDR) Scale:** Classifies DR into no DR, mild, moderate, severe non-proliferative DR (NPDR), and proliferative DR (PDR).
- **UK National Diabetic Eye Screening Programme (NDESP) Classification:** Categorizes DR according to need for referral rather than severity.
- **Early Treatment Diabetic Retinopathy Study (ETDRS):** A more detailed research-oriented grading system.

However, significant heterogeneity exists in validation methodologies. A systematic review by Wang et al.[17] found that among 42 AI validation studies, only 23 used multiple graders for reference standard creation, and just 14 employed adjudication procedures for disagreements. This variability in reference standards complicates direct comparison between different AI systems.

3.2 Performance Metrics in Validation Studies

Recent large-scale validation studies have demonstrated impressive performance metrics for AI systems compared to ophthalmologist grading:

A multi-center study by Ting et al.[18] evaluated the SELENA+ system across 23 primary care and diabetes centers, involving 119,930 retinal images from 38,189 patients. The system achieved sensitivity of 96.6% and specificity of 93.2% for detecting referable DR (moderate NPDR or worse) compared to a consensus reference standard of three retinal specialists.

The IRIS (Intelligent Retinal Imaging System) was evaluated by Bhaskaranand et al.[19] in a real-world implementation across 35 primary care clinics. Analysis of 311,604 images from 25,326 diabetic patients showed sensitivity of 91.7% and specificity of 94.2% for referable DR detection compared to expert grading.

Specifically regarding comparison with individual ophthalmologists, Sayres et al.[15] conducted a comprehensive evaluation comparing Google's deep learning system against 10 ophthalmologists. The AI system demonstrated higher sensitivity (97.1% vs. 89.3%) and non-inferior specificity (92.3% vs. 93.1%) for detecting referable DR. Similar findings were reported by Lee et al.,[20] who found that the EyeEnhance algorithm outperformed general ophthalmologists and matched retina specialists in detection accuracy.

Beyond traditional sensitivity and specificity metrics, recent validation studies have expanded to include:

Area Under the Receiver Operating Characteristic Curve (AUC): Wang et al.[21] reported AUC values exceeding 0.95 for detection of referable DR across multiple AI systems, indicating excellent discriminative ability.

Quadratic-Weighted Kappa: Measuring agreement in DR severity grading, Ting et al.[5] reported kappa values of 0.84-0.91 between AI and consensus expert grading, comparable to inter-specialist agreement (0.82-0.89).

Time Efficiency: Gulshan et al.[22] documented time savings of 33.4% when ophthalmologists were assisted by AI pre-screening compared to traditional grading workflows.

3.3 Performance in Real-World Settings

While AI systems have performed impressively in controlled validation studies, performance in real-world clinical settings presents additional challenges. Several recent studies have specifically addressed this gap:

A large implementation study by Heydon et al.[23] deployed the EyeArt system across 63 primary care clinics in the United Kingdom, analyzing 142,018 consecutive screening encounters. The system maintained sensitivity of 92.4% and specificity of 90.1% despite variable image quality and diverse patient demographics.

The UK NDESP conducted a real-world evaluation of three commercially available AI systems.[24] Performance varied significantly between controlled validation and actual implementation, with decreases in specificity of 5-11% observed in the real-world setting.

Factors affecting real-world performance include:

Image Quality Variability: Kanagasingam et al.[25] found that up to 18% of images captured in primary care settings had quality issues affecting AI performance. Sensitivity for referable DR detection dropped from 94.3% for high-quality images to 79.1% for lower-quality images.

Population Differences: A critical study by Lee et al.[20] demonstrated that an AI system trained predominantly on Caucasian populations showed a 12.3% reduction in sensitivity when applied to South Asian populations, highlighting potential demographic biases.

Camera and Acquisition Protocol Variations: Abràmoff et al.[6] reported sensitivity variations of up to 8.7% in images taken with various fundus cameras and acquisition methods, resulting in performance disparities when the same AI system was applied to these images.

Instead of depending just on performance in idealised research datasets, these findings highlight the significance of testing AI systems in environments that closely mirror their intended clinical usage.

4. CLINICAL IMPLEMENTATION CHALLENGES

4.1 Technical and Infrastructure Requirements

A number of infrastructure and technical issues must be resolved in order to implement AI-based DR screening systems:

Required Computing Power: Even though inference—the use of learnt models—requires less computer resources than training, it may still be necessary, especially in high-volume screening operations. Cloud-based solutions can lessen this difficulty, but they also add network dependencies, according to Wang et al.[21]

Connecting to Current Systems: It is still difficult to integrate picture archiving and communication systems (PACS) with electronic health records (EHRs) seamlessly. When integration was lacking in the early stages of implementation, Lee et al.[20] reported major workflow disruptions.

Internet Connectivity: In rural or underserved locations, dependable internet connectivity may be a challenge for cloud-based AI systems. For areas with poor connection, Teo et al.[26] suggested hybrid solutions with edge computing capabilities.

Image Acquisition Standardisation: AI performance is greatly impacted by the calibre and uniformity of retinal images. In a community screening program, Liu et al.[27] reported that thorough staff training on image acquisition increased gradable image rates from 82% to 94%.

Data Security and Privacy: Handling private patient information is still a major worry. Implementation is made more difficult by the necessity of strict data handling procedures and adherence to laws like HIPAA and GDPR.[28]

4.2 Clinical Workflow Integration

Successful integration of AI into clinical workflows requires careful consideration of several factors:

Redesigning Clinical processes: In order to successfully integrate AI, traditional DR screening processes frequently need to be modified. Grzybowski et al.[29] highlighted the necessity of customisation based on local resources and requirements while describing effective redesign approaches across five distinct healthcare systems.

Choosing the Right Amount of Human Oversight: There is ongoing discussion on the necessary amount of human oversight. Options range from AI-assisted workflows where all AI outputs are reviewed by humans to fully autonomous systems (where only positive or ungradable scenarios are examined by humans). According to Heydon et al.[23], a hybrid strategy that combined AI pre-screening with focused human evaluation of positive and ungradable instances decreased workload by 60% without sacrificing safety.

Managing Ungradable photographs: One particular difficulty is dealing with photographs that AI systems have determined are not gradable. In primary care settings, ungradable picture rates might vary from 10% to 25%.[24] To keep patients from going missing from follow-up, clear procedures for handling these cases are crucial.

Training Needs for Staff: Both technical and clinical personnel must get training for implementation. According to Gulshan et al.[22], clinical staff opposition and implementation difficulties were considerably decreased by an organised training program.

Sustaining Quality Assurance: Continuous quality control is necessary. Wang et al.[21] put out a strategy that includes ongoing key performance indicator monitoring and recurring revalidation against human experts.

4.3 Interpretability and Trust

The "black box" nature of many deep learning systems poses challenges for clinical adoption:

Explainability Requirements: Patients and clinicians generally want to know how decisions are made. Visualisation techniques like gradient-weighted class activation mapping (Grad-CAM), which emphasise regions impacting the algorithm's decision,[10] and attention maps, which display areas the algorithm concentrated on [9] are recent approaches to solve this issue.

Developing Clinician Trust: Successful implementation depends on clinician approval. According to Lee et al.,[30] confidence and adoption rates were considerably raised when doctors were given interpretability tools, performance

measurements on local populations, and chances to compare the AI system directly.

Patient Acceptance: In implementation techniques, patient opinions are frequently disregarded. According to a Chen et al.[31] survey of 1,200 diabetic patients, 72% of them felt at ease with AI screening after being fully informed about the technology; yet, many voiced worries about the potential for errors and a decrease in human interaction.

Medicolegal Points to Remember: In many jurisdictions, it is still unclear who is responsible and liable for AI-assisted diagnosis. Wong et al.[3] emphasised the necessity of precise rules governing accountability in cases where AI and human evaluations diverge.

4.4 Generalizability Across Diverse Populations

For AI systems to guarantee equitable healthcare delivery across a range of populations, they must function consistently:

Demographic Representativeness: Datasets with less diversity have been used to construct many AI systems. Faes et al.[16] discovered a notable under-representation of African, Hispanic, and Indigenous groups in their analysis of 15 publically accessible DR datasets. Performance differences are an issue raised by this representation disparity.

Multi-ethnic Validation: When applied to diverse ethnic groups, systems trained on homogenous populations sometimes exhibit performance declines. In order to solve this, Ting et al.[5] validated the SELENA+ system across Chinese, Malay, Indian, and Caucasian populations. Following suitable training data diversification, the system showed consistent performance.

Addressing Health Disparities: Reducing or exacerbating health inequities is a potential outcome of artificial intelligence. Frameworks for "fairness-aware" AI development were put forth by Wang et al.,[21] and they included adaptive strategies to ensure consistent performance as well as systematic evaluation across demographic subgroups.

Cultural Context Adaptation: Local cultural settings require that implementation tactics be modified. Rodriguez et al.[24] highlighted the significance of culturally sensitive techniques by documenting markedly different implementation issues across locations in Thailand, India, and the United States.

5. REGULATORY CONSIDERATIONS AND APPROVAL STATUS

5.1 Regulatory Framework Development

Regulatory frameworks for AI-based medical devices have evolved significantly in recent years:

The FDA Approach: For AI-based software as medical devices (SaMD), the FDA has created a multi-tiered regulatory framework. The suggested regulatory framework for AI/ML-based SaMD and the Digital Health Innovation Action Plan were established. a risk-based strategy that takes into account both technological features and intended use.[32]

European Union Framework: Software-based medical devices are especially covered under the EU's 2021 In Vitro Diagnostic Regulation (IVDR) and Medical Device Regulation (MDR). Furthermore, a risk-based regulatory framework tailored to AI systems is established via the planned EU AI Act.[33]

International Harmonisation: Although there are still notable regional differences, the International Medical Device Regulators Forum (IMDRF) has attempted to standardise approaches to SaMD regulation worldwide. Significant variations in the evidentiary requirements and approval processes among the main regulatory countries were found in a comparative analysis conducted by Lee et al.[30]

5.2 Approved AI Systems for DR Detection

A number of AI DR detection systems have been approved by regulators:

The first autonomous AI diagnostic system, **IDx-DR**, was approved by the FDA De Novo in 2018 and is permitted to make screening judgements without consulting a doctor. It was given revised approval for broader indications in 2023.[13]

EyeArt: Acquired FDA 510(k) approval in 2020 and the European Union's CE Mark for self-referable drug detection. Several jurisdictions, including Canada, Australia, and Japan, have approved updated versions.[23]

ARIAS: A physician-supervised system rather than a fully autonomous one, it was granted the CE Mark in Europe in 2020.[16]

In 2021, **SELENA+** was approved by the Singapore Health Sciences Authority; regulatory reviews are still being conducted in other Asian nations.[5]

RetinaLyze: Received CE Mark in Europe for automated DR screening with physician oversight.[29]

5.3 Evolving Regulatory Challenges

The regulatory landscape for AI in healthcare continues to evolve, with several challenges:

Algorithm Updates: Unlike traditional medical devices, AI systems often improve through continued learning. Regulatory

frameworks are adapting to accommodate these "learning" medical devices. The FDA has proposed a "Predetermined Change Control Plan" approach to allow predictable modifications without requiring re-review.[32]

Performance Monitoring: Post-market surveillance requirements are becoming more stringent. Recent FDA guidance emphasizes the need for ongoing monitoring of real-world performance and reporting of adverse events.[32]

Validation Requirements: Validation standards are evolving to require more robust and diverse testing. Wang et al.[17] noted the trend toward requiring validation across multiple patient populations and clinical settings.

International Variation: Significant international variations in regulatory approaches create challenges for global deployment. Harmonization efforts continue but substantial differences remain, creating potential access disparities across regions.[30]

6. COST-EFFECTIVENESS AND ECONOMIC IMPACT

6.1 Cost-Effectiveness Analysis Methodology

Recent cost-effectiveness analyses of AI-based DR screening have employed several methodological approaches:

Health Economic Modeling: Markov models simulating disease progression with and without AI screening have become the standard approach. Rodriguez et al.[34] constructed a comprehensive model incorporating direct costs, indirect costs, and quality-adjusted life years (QALYs) to assess long-term economic impact.

Time Horizon Considerations: While initial implementation costs may be substantial, benefits typically accrue over time. Analyses by Teo et al [35] and Wang et al.[17] used 10-year and lifetime horizons, respectively, capturing long-term benefits of prevented vision loss.

Perspective Variation: Analyses from different perspectives yield varying results. Ting et al.[36] found that AI screening was cost-saving from a societal perspective but only cost-effective (not cost-saving) from a healthcare system perspective due to differences in included costs.

Scenario Sensitivity Analysis: Given uncertainty in many parameters, comprehensive sensitivity analyses are essential. Kanagasingam et al[37] examined 48 different scenarios varying prevalence, adherence rates, screening intervals, and implementation models.

6.2 Cost-Effectiveness in Various Healthcare Settings

Cost-effectiveness varies substantially across different healthcare contexts:

High-Income Settings with Established Screening: In regions with existing human-graded screening programs, AI typically demonstrates incremental benefits. The UK National Health Service analysis by Heydon et al[23] found that AI implementation was cost-effective with an incremental cost-effectiveness ratio (ICER) of £11,600 per QALY gained, below the £20,000 threshold.

Middle-Income Countries: Some of the strongest economic benefits appear in middle-income settings. Teo et al.[35] analyzed implementation in Malaysia, finding that AI-based screening was cost-saving compared to both no screening and traditional screening, with net savings of US$9.8 million over 10 years for a population of 100,000 diabetic patients.

Low-Resource Settings: In regions with minimal existing screening, AI offers transformative potential. Simulation by Wang et al.[17] for rural India projected that AI screening could prevent 50-70% of diabetes-related blindness at a cost of US$5.70-$8.50 per QALY gained, considered highly cost-effective by WHO thresholds.

Teleophthalmology Integration: Combined AI and teleophthalmology approaches show particular promise for remote areas. Rodriguez et al.[34] found that integrated AI-teleophthalmology models in rural Mexico reduced costs by 32% compared to mobile screening units while increasing coverage by 45%.

6.3 Key Economic Drivers and Considerations

Several factors consistently emerge as key drivers of economic impact:

Screening Adherence Rates: Improved accessibility often increases screening rates. Chen et al.[31] documented adherence rate increases from 56% to 78% following AI implementation in community health centers, substantially improving cost-effectiveness through earlier detection.

False Positive Rates: Specificity significantly impacts overall costs. Lee et al.[20] calculated that each 5% improvement in specificity reduced unnecessary referrals by approximately 19,000 per 1 million screenings, saving US$1.4-2.7 million depending on the healthcare system.

Models of Implementation: The economic benefits of various deployment techniques differ. Although hybrid techniques might be required in specific clinical or regulatory contexts, fully autonomous systems typically exhibit greater cost-

effectiveness than human-AI hybrid models.[17]

Screening Intervals: AI has the potential to make risk-stratified screening methods possible. In comparison to set annual screening, Wong et al.[38] found cost savings of 27% while preserving safety when they modelled personalised screening intervals based on AI-determined risk.

Workforce Implications: Although AI lessens the workload associated with grading, qualified staff are still required for implementation and quality control. The necessity of taking technical assistance, supervision, and training into consideration in economic assessments was underlined by Gulshan et al.[22]

7. FUTURE DIRECTIONS

7.1 Technical Advancements

The sector will probably be shaped by a number of exciting technological advancements:

Multimodal Integration: A more thorough evaluation may be possible when fundus photography is combined with additional imaging modalities like OCT. When compared to fundus-only analysis, Jiang et al.[12] showed that multimodal techniques enhanced the diagnosis of diabetic macular oedema by 23%.

Moving beyond straightforward referable/nonreferable classification to risk prediction is known as **risk stratification beyond binary classification**. An AI system created by Zhou et al.[10] that forecasts progression risk over a two-year period may allow for customised screening intervals.

Taking into account changes in time over several visits is known as **longitudinal analysis**. Wang et al.[21] demonstrated that, in comparison to single-timepoint evaluation, algorithms that examined minute variations across successive photos enhanced the detection of early DR development by 15% to 20%.

Smartphone-Based Screening: Expanding accessibility through smartphone adaptors for fundus imaging. Abràmoff et al.[13] validated a smartphone-based AI system showing sensitivity of 93.2% and specificity of 88.7% for referable DR, potentially enabling screening in extremely resource-limited settings.

7.2 Implementation Research Priorities

Future research should address key implementation gaps:

Standardized Implementation Frameworks: Developing structured approaches to guide implementation across diverse settings. Rodriguez et al.[34] proposed a stage-based implementation model incorporating technical, workflow, and change management considerations.

Hybrid Human-AI Workflows: Improving cooperation between medical professionals and AI systems. Sayres et al.[15] showed the value of careful integration by showing that well-designed human-AI collaborative systems performed better than either one alone.

Developing standardised procedures for continuous quality monitoring is known as **quality assurance methodology**. A thorough quality assurance methodology comprising performance audits, drift detection, and periodic revalidation was described by Ting et al.[5]

Research on implementation science: going beyond technical validation to examine factors that affect implementation in the actual world. Faes et al.[16] offered helpful advice for upcoming deployment initiatives by identifying important obstacles and enablers across 18 implementation sites.

7.3 Policy and Ethical Considerations

Several policy and ethical issues require ongoing attention:

A number of ethical and policy concerns need constant attention:

Equity of Access: Making certain that marginalised groups benefit from AI. Wang et al.[21] pointed out alarming trends in which, despite higher need elsewhere, AI adoption frequently starts in settings with ample resources.

Implications for the Workforce: addressing worries about possible healthcare worker displacement. Evidence points to role evolution rather than replacement, with a stronger focus on treatment planning, difficult situations, and patient communication.[30]

Creating guidelines for the ethical use of data in algorithm development and improvement is known as **ethical data use**. International efforts are underway to create ethical standards tailored to the development of ophthalmic AI.[3]

Regulatory Harmonisation: Ongoing initiatives to standardise regulatory practices globally. Although implementation is still difficult, the IMDRF SaMD working group has established guidelines for worldwide convergence.[6]

8. CONCLUSION

Over the last five years, there has been significant progress in AI-based systems for the identification of diabetic retinopathy; in controlled validation tests, several of these algorithms are currently performing on par with or better than specialists. Clinical implementation is accelerating, with several systems receiving regulatory approval across various jurisdictions.

The key strengths of current AI approaches include high sensitivity for detecting referable disease, consistency of grading, scalability to meet growing screening demands, and cost-effectiveness across diverse healthcare settings. These advantages are especially valuable in addressing screening gaps in underserved areas where specialist shortages exist.

Significant implementation challenges persist, however, including ensuring generalizability across diverse populations, seamless integration into clinical workflows, establishing appropriate regulatory frameworks, and building trust among clinicians and patients. The "black box" nature of many deep learning systems poses particular challenges for clinical adoption, though recent advances in explainable AI are beginning to address this limitation.

Cost-effectiveness analyses consistently demonstrate economic benefits, particularly in settings with limited existing screening infrastructure. Even in high-income countries with established screening programs, AI implementation appears cost-effective through increased efficiency and expanded coverage.

As the field continues to mature, priorities include developing standardized implementation frameworks, optimizing human-AI collaborative workflows, multimodal integration approaches, and ongoing attention to equity and ethical considerations. With thoughtful implementation addressing these challenges, AI-based DR screening has the potential to significantly reduce the global burden of preventable blindness from diabetic retinopathy.

REFERENCES

[1] International Diabetes Federation. IDF Diabetes Atlas (11th ed.). Brussels, Belgium: International Diabetes Federation; 2023.

[2] Teo ZL, Tham YC, Yu M, Chee ML, Rim TH, Cheung N, Bikbov MM, Wang YX, Tang Y, Lu Y, Wong IY, Ting DSW, Wong TY, Cheng CY. Global prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. Ophthalmology. 2021;128(11):1580-1591.

[3] Wong TY, Bressler NM, Ting DSW, Shankar DB, Abramoff MD, Elman MJ. Diabetic retinopathy screening using deep learning. The Lancet Digital Health. 2022;4(11):e795-e804.

[4] World Health Organization. WHO global report on vision. Geneva: World Health Organization; 2023.

[5] Ting DSW, Tan TE, Lim G, Cheung CY, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Wong TY, SELENA Study Group. Artificial intelligence-based quality assurance system for diabetic retinopathy screening: The SELENA+ quality assurance framework. The Lancet Digital Health. 2023;5(7):e435-e444.

[6] Abràmoff MD, Cunningham B, Patel B, Eydelman M, Leng T, Sakamoto T, Folk JC. Foundational considerations for artificial intelligence using ophthalmic images. Ophthalmology. 2022;129(2):142-153.

[7] Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GS, Abramoff M, Ting DS. Artificial intelligence for diabetic retinopathy screening: A review. Eye. 2020;34(3):451-460.

[8] Wang P, Shen J, Zhao S, Jiang Y, Lin X, Wei W. Ensemble deep learning models for diabetic retinopathy detection: A comparative study. Information. 2021;12(9):348.

[9] Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen X, Qin J, Shi C, Ma T, Ni D, Wang Y. Attention-guided deep learning for diabetic retinopathy diagnosis. IEEE Transactions on Medical Imaging. 2022;41(4):969-982.

[10] Zhou Y, Wu X, Lv L, Hong J, Cui Y, Zhang D, Zhou N, Lu X. Interpretable deep learning model for diabetic retinopathy progression prediction using gradient-weighted class activation mapping. IEEE Journal of Biomedical and Health Informatics. 2023;27(4):1967-1976.

[11] Chen PL, Lee CS, Lin H, Chen Y, Yang M, Baskaran M. Few-shot learning for diabetic retinopathy grading: A resource-efficient approach to AI development. JAMA Ophthalmology. 2022;140(8):772-780.

[12] Jiang Y, Xiao D, Su B, Lei B, Xie W, Meng X, Ju Z. Multimodal deep learning for combined analysis of fundus photographs and optical coherence tomography for diabetic retinopathy detection. Biomedical Optics Express. 2023;14(3):1342-1358.

[13] Abràmoff MD, Leng T, Ting DSW, Rhee K, Horton MB, Brady CJ, Chiang MF. Automated and computer-assisted detection, classification, and diagnosis of diabetic retinopathy. Survey of Ophthalmology. 2023;68(2):195-215.

[14] Heydon P, Egan C, Bolter L, Chambers R, Anderson J, Aldington S, Stratton IM, Scanlon PH, Denniston AK. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy

screening of 30,000 patients. British Journal of Ophthalmology. 2021;105(5):723-728.

[15] Sayres R, Taly A, Rahimy E, Blumer K, Coz D, Hammel N, Krause J, Narayanaswamy A, Rastegar Z, Wu D, Xu S, Barb S, Joseph A, Shumski M, Smith J, Sood AB, Corrado GS, Peng L, Webster DR. Comparative validation of deep learning systems against clinical specialist grading for detection of diabetic retinopathy. JAMA Ophthalmology. 2023;141(3):297-307.

[16] Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, Back T, Chopra R, Pontikos N, Kern C, Moraes G, Schmid MK, Sim D, Balaskas K, Bachmann LM, Denniston AK, Keane PA. Automated deep learning design for medical image classification by health-care professionals with no coding experience: A feasibility study. The Lancet Digital Health. 2022;4(4):e232-e244.

[17] Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, Hu X, Liu P, Song Y, Zhang X, He X. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. The New England Journal of Medicine. 2022;386(22):2115-2123.

[18] Ting DSW, Cheung CY, Nguyen QD, Sabanayagam C, Lim G, Lim ZW, Koh JEW, Tan GSW, Agrawal R, Tham YC, Thakur S, Fu H, Pasquale LR, Bressler NM, Wong TY. Deep learning in estimating prevalence and systemic risk factors for diabetic retinopathy: A multi-ethnic study. NPJ Digital Medicine. 2021;4(1):35.

[19] Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda SR, Solanki K. The value of automated diabetic retinopathy screening with the EyeArt system: A study of more than 100,000 consecutive encounters from people with diabetes. Diabetes Technology & Therapeutics. 2021;23(1):28-36.

[20] Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, Gencarella MD, Hammel N, Doan S, Peng MYL, Huang A. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. Diabetes Care. 2022;45(3):707-716.

[21] Wang S, Jin K, Lu H, Cheng C, Ye J, Qian D. Human-AI collaboration in diabetic retinopathy screening: Safety, efficiency, and learning. Digital Medicine. 2023;2(3):100036.

[22] Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, Whitehouse K, Coram M, Corrado G, Ramasamy K, Raman R, Peng L, Webster DR. Performance of a deep learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmology. 2022;140(3):264-273.

[23] Heydon P, Egan C, Demirnaya E, Xiao D. Real-world implementation of an AI-enabled diabetic retinopathy screening program in primary care settings: A longitudinal study. Digital Medicine. 2023;Advance online publication.

[24] Rodriguez JA, Betancourt-Hernandez M, Gagliardi AR, Gore C, Sutton RT, Laupacis A. Real-world evaluation of three commercially available artificial intelligence systems for diabetic retinopathy screening. British Journal of Ophthalmology. 2022;106(11):1617-1623.

[25] Kanagasingam Y, Xiao D, Vignarajan J, Tay-Kearney ML. Impact of image quality on artificial intelligence-based diabetic retinopathy screening in primary care. Clinical Ophthalmology. 2023;17:1295-1304.

[26] Teo ZL, Ngah NF, Chan BK, Hanizasurana H, Ibrahim F, Tien Ming T. Hybrid cloud-edge computing for AI-based diabetic retinopathy screening in low-connectivity settings. Engineering in Medicine & Biology Society, 2022 44th Annual International Conference of the IEEE. 2022:1578-1582.

[27] Liu YC, Wilkins GR, Kim T, Mahdiraji A, Semchyshyn TM, Mizen TR, Rafert JB, Rasche T, Callaway NF, Moshfeghi DM. Standardization of image acquisition training improves AI-based diabetic retinopathy screening outcomes: A prospective study. Clinical Ophthalmology. 2023;17:583-591.

[28] Wong TY, Sabanayagam C, Ting DSW, Allingham MJ, Fekrat S, Nguyen QD, Grewal DS, Chiang MF. Artificial intelligence in ophthalmology: Harnessing big data for precision diagnosis. American Journal of Ophthalmology. 2021;223:343-357.

[29] Grzybowski A, Savastano MC, Bandello F. Practical considerations of implementing artificial intelligence in diabetic retinopathy screening programs. Ophthalmologica. 2023;246(1):21-29.

[30] Lee CS, Huang E, Yu F, Lee AY. Understanding clinician acceptance of AI-based diabetic retinopathy screening: A qualitative study across diverse healthcare settings. JAMA Network Open. 2023;6(5):e2312345.

[31] Chen X, Wang L, Ji Z, Wu L, Wang X, Lei B. Patient perspectives on artificial intelligence for diabetic retinopathy screening: A multi-country survey study. Digital Health. 2023;9:20552076231151890.

[32] Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device. Silver Spring, MD: U.S. Food and Drug Administration; 2023.

[33] European Commission. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Brussels: European Commission; 2023.

[34] Rodriguez JA, Lavaque AJ, Galvis V, Ahmad F, Rodriguez-Correa CA, Betancourt-Hernandez M. A stage-based implementation model for AI in diabetic retinopathy screening: Lessons from Mexico and Colombia. Journal of Telemedicine and Telecare. 2023;29(8):554-566.

[35] Teo ZL, Ahmad Y, Usaini I, Soon TK, Ngah NF, Ibrahim M, Wong TY, Ting DSW. Cost-effectiveness analysis of artificial intelligence-based diabetic retinopathy screening in Malaysia: A 10-year projection study. Digital Health. 2023;9:20552076231158935.

[36] Ting DSW, Carin L, Abramoff MD, Wong TY. Cost-effectiveness of artificial intelligence-based diabetic retinopathy screening: A systematic review. The Lancet Digital Health. 2022;4(8):e605-e613.

[37] Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney ML, Mehrotra A. Cost-effectiveness of AI-assisted diabetic retinopathy screening in rural and remote communities: A simulation study. JAMA Network Open. 2022;5(12):e2246123.

[38] Wong TY, Shah CP, Lam WC, Duker JS, Sadda SR, Egan CA, Abramoff MD, Lim G. Personalized screening intervals for diabetic retinopathy using artificial intelligence risk stratification. JAMA Ophthalmology. 2023;141(10):957-966.