www.revistaft.com.br

Revista FT | ISSN 1678-0817 | v.30, n.164, 2026 | DOI: 10.69849/zq2jrz97

Inteligência artificial no rastreamento da retinopatia diabética: acurácia diagnóstica, detecção precoce e desafios para implementação na atenção primária à saúde

Artificial intelligence in diabetic retinopathy screening: diagnostic accuracy, early detection, and challenges for implementation in primary health care

Gustavo Beck Patsche[1]

RESUMO

A retinopatia diabética é uma causa evitável de perda visual cujo rastreamento esbarra na escassez de especialistas, sobretudo na atenção primária à saúde (APS). Este artigo analisou, por meio de revisão narrativa, 20 estudos publicados entre 2021 e 2026 sobre o uso da inteligência artificial (IA) no rastreamento da doença, organizados em três eixos: acurácia diagnóstica, detecção precoce e desafios de implementação. Metanálises em cenários reais registraram sensibilidade combinada entre 87% e 94% e especificidade entre 89% e 90%, mas a comparação direta entre sete algoritmos comerciais revelou diferenças significativas de desempenho. Estudos prospectivos em programas populacionais e investigações brasileiras com câmeras portáteis mantiveram sensibilidade acima de 90%, com especificidade entre 54% e 95,4%. Ensaios randomizados demonstraram que a devolutiva imediata elevou a adesão ao encaminhamento de 39,6% para 51,5% em adultos e a conclusão do exame de 22% para 100% em jovens, além de aumentar em cerca de 40% a produtividade do especialista. Entre os desafios para a APS, sobressaem a representatividade dos dados de treinamento, a dependência do equipamento, o impacto dos falsos positivos sobre a rede especializada e a necessidade de avaliações econômicas com parâmetros locais. Conclui-se que a IA é tecnicamente viável para ampliar o rastreamento, mas sua incorporação ao Sistema Único de Saúde depende de validação local e de fluxos de encaminhamento resolutivos.

Palavras-chave: Retinopatia diabética. Inteligência artificial. Programas de rastreamento. Atenção primária à saúde. Aprendizado profundo.

ABSTRACT

Diabetic retinopathy is a preventable cause of vision loss whose screening is hindered by the shortage of specialists, particularly in primary health care (PHC). This article analyzed,

[1] School of Medicine, University of Passo Fundo, Passo Fundo, RS, Brazil – Medical Student. (ORCID: 0000-0003-0333-4519) - E-mail: gustavopatsche@hotmail.com

through a narrative review, 20 studies published between 2021 and 2026 on the use of artificial intelligence (AI) in screening for the disease, organized into three axes: diagnostic accuracy, early detection, and implementation challenges. Meta-analyses in real-world settings reported pooled sensitivity between 87% and 94% and specificity between 89% and 90%, but a head-to-head comparison of seven commercial algorithms revealed significant differences in performance. Prospective studies in population programs and Brazilian investigations with handheld cameras maintained sensitivity above 90%, with specificity ranging from 54% to 95.4%. Randomized trials showed that immediate feedback raised referral adherence from 39.6% to 51.5% in adults and exam completion from 22% to 100% in youth, in addition to increasing specialist productivity by about 40%. Among the challenges for PHC are the representativeness of training data, device dependence, the impact of false positives on specialized services, and the need for economic evaluations using local parameters. AI is technically feasible for expanding screening, but its incorporation into the Brazilian Unified Health System depends on local validation and effective referral pathways.

Keywords: Diabetic retinopathy. Artificial intelligence. Mass screening. Primary health care. Deep learning.

1 INTRODUÇÃO

A retinopatia diabética permanece entre as complicações microvasculares mais frequentes do diabetes mellitus e evolui, em grande parte dos casos, sem sintomas até fases em que a perda visual já se instalou. Na metanálise de Teo *et al.* (2021), a prevalência global entre pessoas com diabetes foi estimada em 22,27%, com 6,17% apresentando formas ameaçadoras à visão. Os mesmos autores projetaram crescimento de 103,12 milhões de adultos afetados em 2020 para 160,50 milhões em 2045, cenário que amplia a pressão sobre os serviços de oftalmologia.

Diante dessa projeção, o rastreamento periódico por retinografia consolidou-se como a estratégia mais difundida para identificar a doença antes do comprometimento visual. A leitura humana das imagens, contudo, exige profissionais treinados e fluxos de laudo que nem sempre acompanham o volume de pacientes. Rajesh *et al.* (2023) lembram que os algoritmos de inteligência artificial (IA) foram desenvolvidos justamente para examinar fotografias do fundo de olho de forma autônoma, reduzindo a dependência da leitura especializada em cada exame.

A transposição desses sistemas do laboratório para a prática, entretanto, envolve etapas distintas. Grzybowski *et al.* (2023) descrevem um percurso que vai do desenvolvimento à validação externa em bancos de imagens independentes, passa pela validação prospectiva e chega à implantação nos serviços, ressaltando que o desempenho obtido em uma fase não garante resultado equivalente na seguinte. Essa distinção pesa quando o cenário de uso é a

atenção primária à saúde (APS), cujas populações, equipamentos e capacidade de encaminhamento diferem dos centros de pesquisa.

No contexto brasileiro, a APS concentra o acompanhamento das pessoas com diabetes no Sistema Único de Saúde (SUS), o que a torna porta de entrada natural para o rastreamento. Penha *et al.* (2023) verificaram, em mutirão realizado em Blumenau (SC), que mais da metade dos participantes fazia ali seu primeiro exame de retina e que 82,5% dependiam exclusivamente do sistema público. O dado evidencia uma lacuna de acesso que tecnologias de leitura automatizada poderiam ajudar a reduzir.

Assim, este artigo tem como objetivo analisar as evidências publicadas entre 2021 e 2026 sobre o uso da IA no rastreamento da retinopatia diabética, com ênfase na acurácia diagnóstica, nos efeitos sobre a detecção precoce e a continuidade do cuidado e nos desafios para sua incorporação à APS.

2 MÉTODO

Trata-se de revisão narrativa da literatura, conduzida em outubro de 2026, com busca nas bases PubMed e PubMed Central, no *Directory of Open Access Journals* e nos portais dos próprios periódicos. Foram combinados os termos *diabetic retinopathy*, *artificial intelligence*, *deep learning*, *screening* e *primary care*. Incluíram-se artigos publicados entre 2021 e 2026, com identificador digital (DOI) ativo, que abordassem desempenho diagnóstico, desfechos assistenciais ou aspectos econômicos e de implementação da IA.

A seleção resultou em 20 estudos, distribuídos entre metanálises de acurácia, estudos diagnósticos prospectivos, ensaios clínicos randomizados, modelagens econômicas, revisões temáticas e um banco público de imagens brasileiro. Deu-se preferência a investigações conduzidas em cenários reais e em países de renda baixa e média, incluindo o Brasil. Os achados foram organizados em três eixos: acurácia diagnóstica, detecção precoce e desfechos do cuidado, e desafios para a implementação na APS. Por se tratar de revisão narrativa, não houve metanálise própria nem avaliação formal do risco de viés.

3 ACURÁCIA DIAGNÓSTICA DOS SISTEMAS DE IA

As sínteses quantitativas disponíveis convergem para níveis elevados de sensibilidade, embora a especificidade varie mais entre os contextos. Wewetzer, Held e Steinhäuser (2021), ao reunirem estudos realizados especificamente na atenção primária, encontraram

sensibilidade combinada de 87% e especificidade de 90%. Considerando prevalência de 10% da doença, os autores estimaram valor preditivo negativo de 98% e positivo de 49%, o que revela boa capacidade de exclusão, mas proporção considerável de encaminhamentos sem confirmação.

Em escopo mais amplo, Joseph *et al.* (2024) analisaram 34 estudos com imagens obtidas em condições reais e obtiveram sensibilidade de 94% e especificidade de 89%, com razão de chances diagnóstica de 128. A metarregressão desses autores associou o desempenho à faixa de renda do país e ao tipo de retinógrafo empregado, achado que desloca a discussão do algoritmo isolado para o conjunto formado por equipamento, operador e população.

A heterogeneidade entre produtos comerciais ficou evidente no estudo de Lee *et al.* (2021), que submeteu sete algoritmos, de cinco empresas, a 311.604 imagens de 23.724 veteranos atendidos em dois sistemas de saúde dos Estados Unidos. As diferenças de desempenho entre os sistemas foram significativas, e os autores defenderam que todo algoritmo seja testado com dados do mundo real antes da adoção clínica. Rajesh *et al.* (2023) acrescentam que comparações diretas desse tipo ainda são escassas, o que dificulta a escolha entre tecnologias.

Entre os estudos prospectivos, o ensaio pivotal de Ipp *et al.* (2021) avaliou o sistema EyeArt em 893 pacientes de 15 serviços de atenção primária e oftalmológica, tendo como referência a graduação de um centro de leitura pela escala do *Early Treatment Diabetic Retinopathy Study*. Sem dilatação pupilar, o sistema alcançou sensibilidade de 95,5% e especificidade de 85,0% para retinopatia mais que leve, e de 95,1% e 89,0%, respectivamente, para formas ameaçadoras à visão, com 87,4% dos olhos passíveis de análise.

Resultados semelhantes apareceram em programas populacionais. Heydon *et al.* (2021) aplicaram o mesmo *software* a 30.405 episódios de rastreamento do programa inglês e obtiveram sensibilidade de 95,7% para retinopatia referenciável, chegando a 100% nos casos de doença não proliferativa moderada a grave e proliferativa. A especificidade, por outro lado, foi de 68% para ausência de retinopatia, e os autores estimaram que o algoritmo poderia reduzir pela metade a carga de leitura dos avaliadores humanos.

Na Tailândia, Ruamviboonsuk *et al.* (2022) integraram um sistema de aprendizado profundo (*deep learning*) ao programa nacional, em nove unidades de atenção primária, com interpretação das imagens em tempo real. Entre 7.651 pacientes analisados, a acurácia para retinopatia ameaçadora à visão foi de 94,7%, com sensibilidade de 91,4% e especificidade de 95,4%. Os autores ponderaram, porém, que fatores socioambientais e fluxos de trabalho locais condicionaram a implantação.

As evidências brasileiras reforçam esse quadro e acrescentam a variável da portabilidade. Malerbi *et al.* (2022) avaliaram uma câmera portátil acoplada a *smartphone* em campanha realizada em Itabuna (BA): dos 824 participantes com diabetes tipo 2, 679 tiveram avaliação completa, e o algoritmo apresentou sensibilidade de 97,8% e especificidade de 61,4% para retinopatia mais que leve, com área sob a curva de 0,89. Todos os falsos negativos corresponderam a retinopatia não proliferativa moderada.

Com o propósito de simplificar o protocolo, Penha *et al.* (2023) testaram a análise de uma única imagem centrada na mácula por olho em 686 pessoas e registraram sensibilidade de 93,6%, especificidade de 71,7% e valor preditivo negativo de 98,0%. Em estudo multicêntrico posterior, Malerbi *et al.* (2024) obtiveram, com o mesmo princípio, sensibilidade de 90,23% e especificidade de 85,06% para retinopatia mais que leve, além de área sob a curva de 0,95 para qualquer grau da doença. Convém registrar que o primeiro autor declarou vínculo de consultoria com a fabricante do equipamento.

Em conjunto, os resultados indicam que a sensibilidade dos sistemas tende a permanecer acima de 90% em diferentes cenários, ao passo que a especificidade oscila conforme o equipamento, o protocolo de captura e a prevalência local. Essa assimetria tem implicações práticas, pois o número de encaminhamentos gerados pela IA precisa ser compatível com a capacidade de absorção da rede especializada.

4 DETECÇÃO PRECOCE E DESFECHOS DO CUIDADO

A acurácia, embora necessária, não assegura por si só que a doença seja tratada a tempo. O benefício clínico depende de que o paciente rastreado conclua o exame, compreenda o resultado e compareça ao serviço especializado quando indicado. Nesse ponto, a contribuição mais concreta da IA parece estar na entrega imediata do laudo no próprio local de atendimento (*point-of-care*), o que encurta o intervalo entre a captura da imagem e a decisão de encaminhar.

O ensaio RAIDERS, conduzido em Ruanda por Mathenge *et al.* (2022), testou essa hipótese ao comparar a devolutiva imediata baseada em IA com a comunicação do resultado após leitura humana, feita de três a cinco dias depois. Entre 275 participantes com indicação de encaminhamento, a adesão foi de 51,5% no grupo intervenção e de 39,6% no grupo controle. Na análise multivariada, a intervenção associou-se a maior aceitação do encaminhamento (*odds ratio* de 1,74), assim como a idade mais avançada, o sexo masculino e a residência em área rural.

Efeito ainda mais pronunciado foi descrito por Wolf *et al.* (2024) no ensaio ACCESS, realizado com crianças e jovens de 8 a 21 anos com diabetes tipo 1 ou tipo 2. A taxa de conclusão do exame em seis meses chegou a 100% no grupo submetido à IA autônoma no próprio ambulatório de diabetes, contra 22% no grupo encaminhado ao oftalmologista. Entre os participantes com resultado alterado, 64% seguiram para avaliação especializada, proporção que não passou de 22% no grupo controle.

Além de ampliar a cobertura, a triagem automatizada pode reorganizar o trabalho do especialista. Abràmoff *et al.* (2023), em ensaio randomizado por conglomerados realizado em Bangladesh, mostraram que os dias de atendimento com uso da IA alcançaram 1,59 atendimentos concluídos por hora de especialista, frente a 1,14 nos dias sem a tecnologia, o que corresponde a ganho de cerca de 40% na produtividade. Na prática, pacientes com resultado negativo deixaram de ocupar vagas do retinólogo.

Ainda faltam ensaios com seguimento longo o bastante para medir diretamente a perda visual evitada. Para estimar esse efeito, Channa *et al.* (2023) construíram um modelo de Markov no qual a incidência de perda visual em cinco anos foi de 1.535 por 100 mil pessoas com rastreamento por IA e de 1.625 por 100 mil com exame feito pelo oftalmologista. No modelo, a adesão ao tratamento foi o fator com maior potencial para ampliar o benefício, o que reforça a importância de fluxos de encaminhamento resolutivos.

5 DESAFIOS PARA A IMPLEMENTAÇÃO NA ATENÇÃO PRIMÁRIA

A transferência desses resultados para a APS brasileira esbarra, antes de tudo, na representatividade dos dados usados para treinar e validar os algoritmos. Nakayama *et al.* (2024) disponibilizaram o BRSET, primeiro banco público multirrotulado de retinografias do Brasil e da América Latina, com 16.266 imagens de 8.524 pacientes. Segundo os autores, a iniciativa permite investigar vieses de desempenho entre grupos demográficos, análise que dificilmente seria feita apenas com bases estrangeiras.

A dependência do equipamento constitui um segundo obstáculo. Como a metarregressão de Joseph *et al.* (2024) vinculou a acurácia ao tipo de câmera, a validação de um algoritmo não se estende automaticamente a outro retinógrafo. Os estudos de Malerbi *et al.* (2022) e de Penha *et al.* (2023) mostram que dispositivos portáteis viabilizam o rastreamento fora do consultório oftalmológico, mas também expõem perdas por imagens de qualidade insuficiente: em Itabuna, 17,6% dos participantes não puderam ser plenamente avaliados.

Soma-se a isso um problema operacional ligado à especificidade moderada. Heydon *et al.* (2021) e Malerbi *et al.* (2022) registraram especificidades entre 54% e 68%, e Penha *et al.* (2023) obtiveram valor preditivo positivo de 42,7%. Em municípios com oferta limitada de oftalmologistas, um volume elevado de falsos positivos pode alongar filas e retardar o atendimento dos casos graves, razão pela qual os limiares de encaminhamento precisam ser definidos à luz da capacidade da rede.

No plano econômico, a decisão tampouco se resume à escolha do algoritmo mais preciso. Wang *et al.* (2024), ao simularem 30 anos de rastreamento em 251.535 pessoas com diabetes na China, concluíram que o modelo de maior acurácia não era o mais custo-efetivo e que a IA precisaria atingir sensibilidade mínima de 88,2% e especificidade mínima de 80,4% para gerar economia ou custo-efetividade. Ahmed *et al.* (2025), em análise do rastreamento pediátrico, mostraram que o equilíbrio financeiro depende do volume, sendo alcançado a partir de 241 pacientes por ano em cada serviço.

Essas modelagens, porém, nem sempre incorporam os ganhos assistenciais descritos nos ensaios. Leigh *et al.* (2025) revisaram 18 avaliações econômicas e verificaram que apenas 39% seguiram diretrizes padronizadas de relato e que a maioria desconsiderou a melhora no seguimento após o rastreamento por IA, o que tende a subestimar seu valor. Para o planejamento no SUS, isso significa que estimativas importadas precisam ser recalculadas com parâmetros locais de adesão, custos e prevalência.

Por fim, a implantação envolve questões regulatórias, éticas e de responsabilidade profissional. Rajesh *et al.* (2023) apontam a equidade, os vieses algorítmicos e os aspectos médico-legais como dimensões indissociáveis da incorporação clínica desses sistemas. Ruamviboonsuk *et al.* (2022), por sua vez, mantiveram a revisão das imagens por especialistas como mecanismo de segurança durante a implantação tailandesa, estratégia que pode servir de etapa de transição para serviços que ainda estruturam sua governança sobre a IA.

6 CONSIDERAÇÕES FINAIS

As evidências reunidas mostram que os sistemas de IA para rastreamento da retinopatia diabética alcançam sensibilidade consistente em diferentes populações, com especificidade mais dependente do equipamento, do protocolo de imagem e do contexto assistencial. Os ensaios randomizados acrescentam que a devolutiva imediata aumenta a

conclusão do exame e a adesão ao encaminhamento, efeito que aproxima a tecnologia do objetivo de identificar a doença antes da perda visual.

Para a APS brasileira, os estudos conduzidos com câmeras portáteis em campanhas populacionais indicam viabilidade técnica. A incorporação em larga escala, contudo, ainda requer validação local com os equipamentos efetivamente adquiridos, limiares compatíveis com a oferta de oftalmologistas, avaliações econômicas com parâmetros nacionais e mecanismos de supervisão clínica. Estudos de implementação no SUS, com acompanhamento dos pacientes rastreados até o tratamento, configuram a principal lacuna a ser preenchida.

ENGLISH VERSION

1 INTRODUCTION

Diabetic retinopathy remains among the most frequent microvascular complications of diabetes mellitus and, in most cases, progresses without symptoms until vision loss is already established. In the meta-analysis by Teo *et al.* (2021), global prevalence among people with diabetes was estimated at 22.27%, with 6.17% presenting vision-threatening forms. The same authors projected an increase from 103.12 million affected adults in 2020 to 160.50 million in 2045, a scenario that places growing pressure on ophthalmology services.

Given this projection, periodic screening by retinal photography has become the most widespread strategy for identifying the disease before visual impairment. Human grading of images, however, requires trained professionals and reporting workflows that do not always keep pace with patient volume. Rajesh *et al.* (2023) note that artificial intelligence (AI) algorithms were developed precisely to examine fundus photographs autonomously, reducing reliance on specialist reading for each examination.

Moving these systems from the laboratory into practice, however, involves distinct stages. Grzybowski *et al.* (2023) describe a pathway that goes from development to external validation on independent image sets, then to prospective validation and, finally, deployment in health services, stressing that performance achieved in one phase does not guarantee equivalent results in the next. This distinction matters when the setting is primary health care (PHC), whose populations, devices, and referral capacity differ from those of research centers.

In Brazil, PHC concentrates the follow-up of people with diabetes within the Unified Health System (*Sistema Único de Saúde*, SUS), making it a natural entry point for screening. Penha *et al.* (2023) found, during a screening campaign in Blumenau, Santa Catarina, that more than half of the participants were undergoing their first retinal examination and that 82.5% relied exclusively on the public system. These data point to an access gap that automated grading technologies could help to narrow.

This article therefore aims to analyze the evidence published between 2021 and 2026 on the use of AI in diabetic retinopathy screening, with emphasis on diagnostic accuracy, effects on early detection and continuity of care, and the challenges for its incorporation into PHC.

2 METHOD

This is a narrative literature review conducted in October 2026, with searches in PubMed and PubMed Central, the Directory of Open Access Journals, and the journals' own websites. The terms diabetic retinopathy, artificial intelligence, deep learning, screening, and primary care were combined. Articles published between 2021 and 2026, with an active digital object identifier (DOI), addressing diagnostic performance, care outcomes, or economic and implementation aspects of AI were included.

The selection yielded 20 studies, comprising diagnostic accuracy meta-analyses, prospective diagnostic studies, randomized clinical trials, economic models, thematic reviews, and a Brazilian public image dataset. Preference was given to investigations conducted in real-world settings and in low- and middle-income countries, including Brazil. Findings were organized into three axes: diagnostic accuracy, early detection and care outcomes, and challenges for implementation in PHC. As a narrative review, no original meta-analysis or formal risk-of-bias assessment was performed.

3 DIAGNOSTIC ACCURACY OF AI SYSTEMS

Available quantitative syntheses converge on high sensitivity, although specificity varies more across settings. Wewetzer, Held and Steinhäuser (2021), pooling studies conducted specifically in primary care, found a combined sensitivity of 87% and specificity of 90%. Assuming a disease prevalence of 10%, the authors estimated a negative predictive

value of 98% and a positive predictive value of 49%, indicating good rule-out capacity but a considerable proportion of unconfirmed referrals.

With a broader scope, Joseph *et al.* (2024) analyzed 34 studies using images obtained under real-world conditions and found a sensitivity of 94% and specificity of 89%, with a diagnostic odds ratio of 128. Their meta-regression associated performance with country income level and type of fundus camera, a finding that shifts the discussion from the algorithm alone to the combination of device, operator, and population.

Heterogeneity among commercial products became evident in the study by Lee *et al.* (2021), which tested seven algorithms from five companies on 311,604 images from 23,724 veterans seen at two health care systems in the United States. Performance differences between the systems were significant, and the authors argued that every algorithm should be tested on real-world data before clinical adoption. Rajesh *et al.* (2023) add that such head-to-head comparisons remain scarce, which complicates the choice between technologies.

Among prospective studies, the pivotal trial by Ipp *et al.* (2021) evaluated the EyeArt system in 893 patients from 15 primary care and eye care facilities, using reading center grading on the Early Treatment Diabetic Retinopathy Study scale as reference. Without pupil dilation, the system achieved a sensitivity of 95.5% and specificity of 85.0% for more-than-mild retinopathy, and 95.1% and 89.0%, respectively, for vision-threatening forms, with 87.4% of eyes gradable.

Similar results emerged in population programs. Heydon *et al.* (2021) applied the same software to 30,405 screening episodes from the English programme and obtained a sensitivity of 95.7% for referable retinopathy, reaching 100% for moderate to severe non-proliferative and proliferative disease. Specificity, on the other hand, was 68% for absence of retinopathy, and the authors estimated that the algorithm could halve the grading workload of human graders.

In Thailand, Ruamviboonsuk *et al.* (2022) integrated a deep learning system into the national programme at nine primary care units, with real-time image interpretation. Among 7,651 patients analyzed, accuracy for vision-threatening retinopathy was 94.7%, with a sensitivity of 91.4% and specificity of 95.4%. The authors cautioned, however, that socio-environmental factors and local workflows shaped the deployment.

Brazilian evidence reinforces this picture and adds the variable of portability. Malerbi *et al.* (2022) evaluated a handheld smartphone-based camera during a campaign in Itabuna, Bahia: of 824 participants with type 2 diabetes, 679 were fully assessed, and the algorithm showed a sensitivity of 97.8% and specificity of 61.4% for more-than-mild retinopathy, with

an area under the curve of 0.89. All false negatives corresponded to moderate non-proliferative retinopathy.

Aiming to simplify the protocol, Penha *et al.* (2023) tested the analysis of a single macula-centered image per eye in 686 people and recorded a sensitivity of 93.6%, specificity of 71.7%, and negative predictive value of 98.0%. In a subsequent multicenter study, Malerbi *et al.* (2024) obtained, with the same approach, a sensitivity of 90.23% and specificity of 85.06% for more-than-mild retinopathy, plus an area under the curve of 0.95 for any degree of disease. It should be noted that the first author disclosed a consulting relationship with the device manufacturer.

Taken together, the results indicate that system sensitivity tends to remain above 90% across settings, whereas specificity fluctuates according to the device, image capture protocol, and local prevalence. This asymmetry has practical implications, since the number of referrals generated by AI must be compatible with the absorptive capacity of specialized services.

4 EARLY DETECTION AND CARE OUTCOMES

Accuracy, although necessary, does not by itself ensure that the disease is treated in time. Clinical benefit depends on the screened patient completing the examination, understanding the result, and attending specialized care when indicated. At this point, the most tangible contribution of AI appears to lie in the immediate delivery of results at the point of care, which shortens the interval between image capture and the referral decision.

The RAIDERS trial, conducted in Rwanda by Mathenge *et al.* (2022), tested this hypothesis by comparing immediate AI-based feedback with communication of results after human grading, performed three to five days later. Among 275 participants requiring referral, adherence was 51.5% in the intervention group and 39.6% in the control group. In multivariate analysis, the intervention was associated with greater referral acceptance (odds ratio of 1.74), as were older age, male sex, and rural residence.

An even more pronounced effect was described by Wolf *et al.* (2024) in the ACCESS trial, carried out with children and young people aged 8 to 21 years with type 1 or type 2 diabetes. The six-month exam completion rate reached 100% in the group receiving autonomous AI at the diabetes clinic itself, compared with 22% in the group referred to an eye care provider. Among participants with an abnormal result, 64% proceeded to specialist evaluation, a proportion that did not exceed 22% in the control group.

Beyond expanding coverage, automated triage can reorganize specialist work. Abràmoff *et al.* (2023), in a cluster-randomized trial in Bangladesh, showed that clinic days using AI achieved 1.59 completed care encounters per specialist hour, compared with 1.14 on days without the technology, corresponding to a productivity gain of about 40%. In practice, patients with negative results no longer occupied retina specialist slots.

Trials with follow-up long enough to directly measure prevented vision loss are still lacking. To estimate this effect, Channa *et al.* (2023) built a Markov model in which the five-year incidence of vision loss was 1,535 per 100,000 people screened by AI and 1,625 per 100,000 with examination by an eye care provider. In the model, treatment adherence was the factor with the greatest potential to enhance the benefit, underscoring the importance of effective referral pathways.

5 CHALLENGES FOR IMPLEMENTATION IN PRIMARY CARE

Transferring these results to Brazilian PHC faces, first and foremost, the issue of representativeness of the data used to train and validate algorithms. Nakayama *et al.* (2024) released BRSET, the first public multilabel retinal image dataset from Brazil and Latin America, with 16,266 images from 8,524 patients. According to the authors, the initiative makes it possible to investigate performance biases across demographic groups, an analysis that would hardly be feasible with foreign datasets alone.

Device dependence represents a second obstacle. Since the meta-regression by Joseph *et al.* (2024) linked accuracy to camera type, validation of an algorithm does not automatically extend to another fundus camera. The studies by Malerbi *et al.* (2022) and Penha *et al.* (2023) show that handheld devices enable screening outside the ophthalmology office, but they also expose losses due to insufficient image quality: in Itabuna, 17.6% of participants could not be fully assessed.

Added to this is an operational problem linked to moderate specificity. Heydon *et al.* (2021) and Malerbi *et al.* (2022) reported specificities between 54% and 68%, and Penha *et al.* (2023) obtained a positive predictive value of 42.7%. In municipalities with a limited supply of ophthalmologists, a high volume of false positives may lengthen waiting lists and delay care for severe cases, which is why referral thresholds need to be set in light of network capacity.

On the economic side, the decision is not simply a matter of choosing the most accurate algorithm either. Wang *et al.* (2024), simulating 30 years of screening for 251,535

people with diabetes in China, concluded that the most accurate model was not the most cost-effective and that AI would need a minimum sensitivity of 88.2% and a minimum specificity of 80.4% to be cost-saving or cost-effective. Ahmed *et al.* (2025), analyzing pediatric screening, showed that financial break-even depends on volume, being reached from 241 patients per year at each site.

These models, however, do not always incorporate the care gains described in the trials. Leigh *et al.* (2025) reviewed 18 economic evaluations and found that only 39% followed standardized reporting guidelines and that most disregarded improved follow-up after AI screening, which tends to underestimate its value. For planning within the SUS, this means that imported estimates must be recalculated with local parameters for adherence, costs, and prevalence.

Finally, deployment involves regulatory, ethical, and professional liability issues. Rajesh *et al.* (2023) identify equity, algorithmic bias, and medicolegal aspects as dimensions inseparable from the clinical incorporation of these systems. Ruamviboonsuk *et al.* (2022), in turn, retained specialist over-reading of images as a safety mechanism during the Thai deployment, a strategy that may serve as a transitional stage for services still building their governance over AI.

6 FINAL CONSIDERATIONS

The evidence gathered shows that AI systems for diabetic retinopathy screening achieve consistent sensitivity across different populations, with specificity more dependent on the device, imaging protocol, and care context. Randomized trials add that immediate feedback increases exam completion and referral adherence, an effect that brings the technology closer to the goal of identifying the disease before vision loss.

For Brazilian PHC, studies conducted with handheld cameras in population campaigns indicate technical feasibility. Large-scale incorporation, however, still requires local validation with the devices actually procured, thresholds compatible with the supply of ophthalmologists, economic evaluations with national parameters, and clinical oversight mechanisms. Implementation studies within the SUS, following screened patients through to treatment, constitute the main gap to be filled.

REFERÊNCIAS

ABRÀMOFF, M. D. *et al.* Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial. **npj Digital Medicine**, London, v. 6, n. 1, art. 184, 2023. DOI: 10.1038/s41746-023-00931-7. Disponível em: https://doi.org/10.1038/s41746-023-00931-7. Acesso em: 30 out. 2026.

AHMED, M. *et al.* Cost-effectiveness of AI for pediatric diabetic eye exams from a health system perspective. **npj Digital Medicine**, London, v. 8, n. 1, art. 3, 2025. DOI: 10.1038/s41746-024-01382-4. Disponível em: https://doi.org/10.1038/s41746-024-01382-4. Acesso em: 30 out. 2026.

CHANNA, R. *et al.* Effectiveness of artificial intelligence screening in preventing vision loss from diabetes: a policy model. **npj Digital Medicine**, London, v. 6, n. 1, art. 53, 2023. DOI: 10.1038/s41746-023-00785-z. Disponível em: https://doi.org/10.1038/s41746-023-00785-z. Acesso em: 30 out. 2026.

GRZYBOWSKI, A. *et al.* Artificial intelligence for diabetic retinopathy screening using color retinal photographs: from development to deployment. **Ophthalmology and Therapy**, [*S. l.*], v. 12, n. 3, p. 1419-1437, 2023. DOI: 10.1007/s40123-023-00691-3. Disponível em: https://doi.org/10.1007/s40123-023-00691-3. Acesso em: 30 out. 2026.

HEYDON, P. *et al.* Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. **British Journal of Ophthalmology**, London, v. 105, n. 5, p. 723-728, 2021. DOI: 10.1136/bjophthalmol-2020-316594. Disponível em: https://doi.org/10.1136/bjophthalmol-2020-316594. Acesso em: 30 out. 2026.

IPP, E. *et al.* Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. **JAMA Network Open**, Chicago, v. 4, n. 11, e2134254, 2021. DOI: 10.1001/jamanetworkopen.2021.34254. Disponível em: https://doi.org/10.1001/jamanetworkopen.2021.34254. Acesso em: 30 out. 2026.

JOSEPH, S. *et al.* Diagnostic accuracy of artificial intelligence-based automated diabetic retinopathy screening in real-world settings: a systematic review and meta-analysis. **American Journal of Ophthalmology**, [*S. l.*], v. 263, p. 214-230, 2024. DOI: 10.1016/j.ajo.2024.02.012. Disponível em: https://doi.org/10.1016/j.ajo.2024.02.012. Acesso em: 30 out. 2026.

LEE, A. Y. *et al.* Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. **Diabetes Care**, Alexandria, v. 44, n. 5, p. 1168-1175, 2021. DOI: 10.2337/dc20-1877. Disponível em: https://doi.org/10.2337/dc20-1877. Acesso em: 30 out. 2026.

LEIGH, J. *et al.* Health economic considerations for the implementation of artificial intelligence-enabled diabetic retinopathy screening: a review. **Clinical & Experimental Ophthalmology**, [*S. l.*], 2025. DOI: 10.1111/ceo.70016. Disponível em: https://doi.org/10.1111/ceo.70016. Acesso em: 30 out. 2026.

MALERBI, F. K. *et al.* Automated identification of different severity levels of diabetic retinopathy using a handheld fundus camera and single-image protocol. **Ophthalmology Science**, [*S. l.*], v. 4, n. 4, art. 100481, 2024. DOI: 10.1016/j.xops.2024.100481. Disponível em: https://doi.org/10.1016/j.xops.2024.100481. Acesso em: 30 out. 2026.

MALERBI, F. K. *et al.* Diabetic retinopathy screening using artificial intelligence and handheld smartphone-based retinal camera. **Journal of Diabetes Science and Technology**, Thousand Oaks, v. 16, n. 3, p. 716-723, 2022. DOI: 10.1177/1932296820985567. Disponível em: https://doi.org/10.1177/1932296820985567. Acesso em: 30 out. 2026.

MATHENGE, W. *et al.* Impact of artificial intelligence assessment of diabetic retinopathy on referral service uptake in a low-resource setting: the RAIDERS randomized trial. **Ophthalmology Science**, [*S. l.*], v. 2, n. 4, art. 100168, 2022. DOI: 10.1016/j.xops.2022.100168. Disponível em: https://doi.org/10.1016/j.xops.2022.100168. Acesso em: 30 out. 2026.

NAKAYAMA, L. F. *et al.* BRSET: a Brazilian multilabel ophthalmological dataset of retina fundus photos. **PLOS Digital Health**, San Francisco, v. 3, n. 7, e0000454, 2024. DOI: 10.1371/journal.pdig.0000454. Disponível em: https://doi.org/10.1371/journal.pdig.0000454. Acesso em: 30 out. 2026.

PENHA, F. M. *et al.* Single retinal image for diabetic retinopathy screening: performance of a handheld device with embedded artificial intelligence. **International Journal of Retina and Vitreous**, London, v. 9, art. 41, 2023. DOI: 10.1186/s40942-023-00477-6. Disponível em: https://doi.org/10.1186/s40942-023-00477-6. Acesso em: 30 out. 2026.

RAJESH, A. E. *et al.* Artificial intelligence and diabetic retinopathy: AI framework, prospective studies, head-to-head validation, and cost-effectiveness. **Diabetes Care**, Alexandria, v. 46, n. 10, p. 1728-1739, 2023. DOI: 10.2337/dci23-0032. Disponível em: https://doi.org/10.2337/dci23-0032. Acesso em: 30 out. 2026.

RUAMVIBOONSUK, P. *et al.* Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. **The Lancet Digital Health**, London, v. 4, n. 4, p. e235-e244, 2022. DOI: 10.1016/S2589-7500(22)00017-6. Disponível em: https://doi.org/10.1016/S2589-7500(22)00017-6. Acesso em: 30 out. 2026.

TEO, Z. L. *et al.* Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. **Ophthalmology**, [*S. l.*], v. 128, n. 11, p. 1580-1591, 2021. DOI: 10.1016/j.ophtha.2021.04.027. Disponível em: https://doi.org/10.1016/j.ophtha.2021.04.027. Acesso em: 30 out. 2026.

WANG, Y. *et al.* Economic evaluation for medical artificial intelligence: accuracy vs. cost-effectiveness in a diabetic retinopathy screening case. **npj Digital Medicine**, London, v. 7, art. 43, 2024. DOI: 10.1038/s41746-024-01032-9. Disponível em: https://doi.org/10.1038/s41746-024-01032-9. Acesso em: 30 out. 2026.

WEWETZER, L.; HELD, L. A.; STEINHÄUSER, J. Diagnostic performance of deep-learning-based screening methods for diabetic retinopathy in primary care: a meta-analysis. **PLoS ONE**, San Francisco, v. 16, n. 8, e0255034, 2021. DOI:

10.1371/journal.pone.0255034. Disponível em: https://doi.org/10.1371/journal.pone.0255034. Acesso em: 30 out. 2026.

WOLF, R. M. *et al.* Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: the ACCESS randomized control trial. **Nature Communications**, London, v. 15, art. 421, 2024. DOI: 10.1038/s41467-023-44676-z. Disponível em: https://doi.org/10.1038/s41467-023-44676-z. Acesso em: 30 out. 2026.