

VALIDATION OF AI ALGORITHMS IN OPHTHALMOLOGICAL POPULATION SCREENING

GABRIEL RIBEIRO DE OLIVEIRA SILVA¹, ÍTALO BEZERRA ALVES², ISABELA LUIZA PACHECO RODRIGUES²,
DOUGLAS WALLACY DA SILVA AMÂNCIO², FABIANO INÁCIO DE SOUZA³

1. Medical student at the Federal University of Jataí, Jataí, Goiás, Brazil.
2. Medical student at the Federal University of Goiás, Goiânia, Goiás, Brazil.
3. Orthopedic Surgeon. PhD in Orthopedics, Traumatology, and Rehabilitation, Goiânia, Goiás, Brazil.

ABSTRACT

The application of artificial intelligence (AI) in ophthalmology has emerged as a promising strategy for the early diagnosis and population screening of visual diseases, mitigating the shortage of specialists and optimizing the management of health resources. The present study aims to analyze the clinical efficacy, robustness, and generalization capacity of AI algorithms applied to the screening and diagnostic support of ophthalmological pathologies. For this purpose, an integrative literature review was conducted through searches in the Google Scholar, PubMed, ResearchGate, and CAPES Periodicals Portal databases, culminating in the selection and analysis of ten studies published between 2018 and 2025. The synthesis of evidence demonstrated that deep learning models focused on the analysis of fundus images present high rates of sensitivity and specificity, especially in the detection of diabetic retinopathy, retinopathy of prematurity, glaucoma, and age-related macular degeneration. Furthermore, it was found that the integration of these technologies significantly expands clinical screening capacity. However, the performance of these algorithms can undergo significant variations when faced with distinct populations or different capture equipment, a technical challenge known as domain shift. Therefore, it is inferred that the safe clinical implementation of AI requires rigorous multicenter external validations and the development of clear ethical-regulatory guidelines, acting as a complement to medical judgment and democratizing access to visual care.

Keywords: Artificial Intelligence, Ophthalmology, Screening, Algorithms, External Validation.

INTRODUCTION

Ophthalmic diseases represent a substantial challenge to global public health, accounting for high rates of morbidity and preventable vision loss. In settings characterized by a shortage of medical specialists and geographical or socioeconomic barriers, timely access to diagnosis and early treatment becomes limited, overburdening healthcare systems and worsening patient prognosis.¹

In this context, the integration of new technologies has driven the development of innovative strategies for population screening. Artificial intelligence (AI), particularly through deep learning algorithms, has emerged as a transformative force in contemporary ophthalmology.² The use of these tools in combination with teleophthalmology has demonstrated a strong capacity to analyze fundus photographs and stratify risks with high accuracy, optimizing the care pathway for highly prevalent

conditions such as diabetic retinopathy, glaucoma, and age-related macular degeneration.^{3,4}

Despite the promising results achieved during the development phase in controlled environments, the scientific literature raises questions regarding the safety and effectiveness of these tools in real-world settings. Algorithm performance often declines when exposed to new populations, different image acquisition protocols, or variations in equipment, constituting a technical challenge known as domain shift.⁵

Therefore, understanding the limitations and accuracy of these technologies is particularly relevant. Assessing the robustness of AI models through external validation processes is essential to ensure that their incorporation into large-scale screening programs occurs safely, reliably, and equitably, without compromising the quality of care provided to patients.^{6,7}

OBJECTIVES

Perinatal hemolytic disease (PHD) is defined as an immunological condition resulting from incompatibility between antigens present on fetal red blood cells and specific maternal antibodies, leading to the destruction of these blood cells either during intrauterine life or in the early neonatal period. This incompatibility occurs predominantly when the mother is previously sensitized to erythrocyte antigens inherited by the fetus from the father.²

The maiGeneral Objective:

- To analyze the robustness, diagnostic performance, and generalizability of artificial intelligence algorithms designed for population screening of ophthalmic diseases, based on scientific evidence compiled from the literature.

Specific Objectives:

- To classify the deep learning models and the respective ocular diseases mapped in the included studies, outlining the current landscape of technological tools applied to visual health.
- To distinguish the variability in diagnostic accuracy indicators (sensitivity, specificity, and area under the ROC curve) of AI tools according to different diseases, such as diabetic retinopathy, glaucoma, retinopathy of prematurity, and age-related macular degeneration.
- To enumerate the main methodological limitations and ethical and technical challenges reported, with particular emphasis on the phenomenon of domain shift and technical dependence on the quality of the images acquired.
- To select perspectives for future clinical applicability and the public health impact resulting from the integration of these algorithms with primary care and teleophthalmology systems.

METHODS

The present article is characterized as an integrative literature review of a qualitative and exploratory nature, focused on the validation of Artificial Intelligence (AI) algorithms applied to population screening for ophthalmic diseases. Data collection was conducted in May 2026 through a systematic search of Google Scholar, PubMed, ResearchGate, and the CAPES Journal Portal databases, selected for their coverage of high-impact journals and access to recent scientific publications on diagnosis in visual health.

The search strategy employed controlled descriptors and technical terms in Portuguese and English, combined with basic Boolean operators ("Artificial Intelligence" AND "Ophthalmology", "AI algorithms" AND

“Screening” AND “Retina”), in order to ensure the intersection between technology and clinical practice, as well as the grouping of terminological variations such as deep neural networks and machine learning.

The inclusion criteria comprised original articles and reviews published between 2018 and 2025 that specifically addressed the accuracy, sensitivity, and specificity of AI tools in screening settings. After the initial search, titles and abstracts were analyzed to exclude duplicate studies or those without validated data on algorithmic performance. The selected texts were read in full, allowing the synthesis of the main external validation models and evidence regarding the feasibility of these technologies in public healthcare systems and mass screening programs.

RESULTS AND DISCUSSIONS

• Characterization of the Included Studies

Ten studies published between 2018 and 2025 were included, comprising original diagnostic validation studies, observational studies, and literature reviews. The main characteristics of the included studies are presented in Table 1. There was a predominance of research focused on the application and validation of artificial intelligence (AI) algorithms in retinal diseases, with other ophthalmic conditions addressed less frequently.

Table 1: Characterization of the Included Studies.

Author/year	Type of study	Did the study validate the algorithm?	Country/Setting
Pionório et al., 2024	Integrative review	N/A	N/A
Checon et al., 2025	Qualitative/exploratory research	N/A	N/A
Mamprin et al., 2025	Narrative review	N/A	Primary health care
Arenas-Cavalli et al., 2021	Observational clinical validation study	Yes	Chile; national health system
Han et al., 2022	Prospective validation study	Yes	China; hospital and community
Coyner et al., 2022	Diagnostic study with multicenter external validation	Yes	India, Nepal, and Mongolia; telemedicine program
Zapata et al., 2020	Retrospective study	Yes	Spain; teleophthalmology
Kuiava et al., 2021	Development/validation study	Yes	Brazil; teleophthalmology
Chen et al., 2025	Diagnostic study with workflow	Yes	USA and Singapore
Chen et al., 2021	Observational development/validation study	Yes	Taiwan

Source: Pionório et al., 2024; Checon et al., 2025; Mamprin et al., 2025; Arenas-Cavalli et al., 2021; Han et al., 2022; Coyner et al., 2022; Zapata et al., 2020; Kuiava et al., 2021; Chen et al., 2025; Chen et al., 2021.³⁻⁴⁻⁸⁻¹⁴

Table 2: Description and Results of the Studies.

Author/year	Disease/Application	AI Model	Main Contribution
Pionório et al., 2024	Focus on diabetic retinopathy and glaucoma	IDx-DR, Eyer, and other systems	Reinforces the role of AI in democratizing diagnosis and providing clinical support
Checon et al., 2025	Focus on diabetic retinopathy, glaucoma, and age-related macular degeneration	Synthesis of studies on deep learning	Highlights the low diversity of databases and the need for local validation
Mamprin et al., 2025	Diabetic retinopathy screening	Deep learning	High diagnostic performance and potential to reduce costs
Arenas-Cavalli et al., 2021	Diabetic retinopathy screening	DART system with convolutional networks	Sensitivity of 94.6%, specificity of 74.3%; feasibility of application in a heterogeneous public healthcare network
Han et al., 2022	Referable diabetic retinopathy, retinal macular disease, and suspected glaucoma	AI-100 (deep learning)	Robustness in real-world settings, but with domain shift across cameras and populations
Coyner et al., 2022	Retinopathy of prematurity screening	Based on gestational age and vascular severity score (VSS)	Sensitivity of 100% in all three settings and an estimated 38–51% reduction in examinations
Zapata et al., 2020	Identification of image quality, laterality, age-related macular degeneration, and suspected glaucoma	5 deep learning algorithms / convolutional neural network	High accuracy for multiple tasks; useful for the screening workflow
Kuiava et al., 2021	Glaucoma, diabetic retinopathy, ocular toxoplasmosis, papilledema, retinal detachment, normal retina	MobileNet and Inception	Sensitivity ~91% and specificity ~98%; useful as a screening support tool
Chen et al., 2025	Diagnosis/classification of age-related macular degeneration	DeepSeeNet and DeepSeeNet+	Increased accuracy for 23/24 clinicians; reduced diagnostic time; reinforces the need for accountability
Chen et al., 2021	Eyelid measurements for ptosis: MRD1, MRD2, and LF	MAIA software and convolutional networks	Demonstrates the application of AI via smartphone for objective and rapid measurement

Source: Pionório et al., 2024; Checon et al., 2025; Mamprin et al., 2025; Arenas-Cavalli et al., 2021; Han et al., 2022; Coyner et al., 2022; Zapata et al., 2020; Kuiava et al., 2021; Chen et al., 2025; Chen et al., 2021.³⁻⁴⁻⁸⁻¹⁴

Overall, there was a predominance of deep learning-based algorithms applied to the analysis of fundus photographs, with a primary focus on diabetic retinopathy, retinopathy of prematurity, age-related macular degeneration (AMD), and suspected glaucoma.⁸⁻¹⁴ Despite the methodological differences observed, the studies converged in assessing the potential of AI to support screening processes and expand access to ophthalmic care, frequently in teleophthalmology and large-scale screening settings.^{4,8-11} In parallel, the included reviews reinforced these general trends, particularly regarding the prospects for incorporating these tools into visual health programs.^{3,4}

• Performance of AI Algorithms

The studies analyzed demonstrated high overall performance for AI algorithms, as evidenced by high sensitivity, specificity, and area under the ROC curve values, reinforcing the potential of these

tools for use in screening programs and clinical decision support.⁹⁻¹⁴

Diabetic retinopathy (DR) was the condition for which this performance was most stable. In a primary care and population-based DR screening context,⁸ deep learning-based systems were described as having accuracy above 90%, accompanied by sensitivity and specificity of approximately 95%, as well as the potential to reduce the costs of manual screening.⁸ In the clinical validation conducted in Chile,⁹ the DART system showed a sensitivity of 94.6%, specificity of 74.3%, and AUC of 0.915 in 1,123 diabetic eye examinations, suggesting applicability within a heterogeneous national healthcare network.⁹ Meanwhile, in a real-world community and hospital setting in China,¹⁰ AI-100 also demonstrated robust performance for referable diabetic retinopathy, with a sensitivity of 95%, specificity of 96.3%, and AUC of 0.954.¹⁰ Thus, the findings demonstrate that AI has high potential for DR screening, particularly in population-based programs and teleophthalmology.

In retinopathy of prematurity, the ability to anticipate risk was particularly noteworthy. The model based on gestational age and vascular severity score¹¹ maintained consistent performance across the populations evaluated, with a sensitivity of 100% in all three cohorts, specificity of 63.3% in India, 77.8% in Nepal, and 45.8% in Mongolia, as well as an AUC of 0.919 in 5-fold cross-validation and 0.944 in the validation set. This performance allowed early identification of cases that would progress to require treatment, reinforcing the role of AI not only as a diagnostic tool but also as a predictive resource in a setting in which early detection has a direct impact on visual prognosis.¹¹

Studies evaluating AI for glaucoma or suspected glaucoma showed favorable performance, although it was more heterogeneous than that observed for retinal diseases.^{10,12,13} In a Spanish database containing more than 300,000 images,¹² deep learning models achieved an AUC of 0.863, accuracy of 80.2%, sensitivity of 76.8%, and specificity of 83.8% for suspected glaucomatous neuropathy, maintaining applicability for screening, although with lower performance than for other diseases analyzed in the same study. In contrast, the AI-100 algorithm¹⁰ showed superior performance in a real-world clinical setting, with a sensitivity of 95%, specificity of 94.6%, and AUC of 0.976 for suspected glaucoma, in a validation environment involving different cameras and healthcare settings. Similarly high results were observed with the Brazilian screening software,¹³ whose MobileNet and Inception models achieved sensitivities ranging from 93.33% to 96.55% and specificities above 99% for glaucoma. Taken together, these studies suggest that AI has significant potential for glaucoma screening, although performance varies according to image quality, population characteristics, and database composition.^{10,12,13}

For AMD, the results were also consistent, both for disease detection and for specialist support.^{12,14} In the study conducted in Spain,¹² deep learning models achieved an AUC of 0.936, accuracy of 86.3%, sensitivity of 90.2%, and specificity of 82.5%, indicating effective discrimination between affected and unaffected individuals. A multicenter study evaluated an AI-assisted workflow for the diagnosis and severity classification of AMD.¹⁴ In this study, AI assistance increased the clinicians' mean F1-score from 37.71 to 45.52 and reduced the estimated diagnostic time, although the lower performance observed in the external Singapore cohort suggests potential limitations in generalizability even for promising models. Thus, AMD appears to be a condition in which AI can support both detection and severity stratification, but further refinement is still required for broader use in more diverse populations.^{12,14}

In addition to applications targeting specific diseases, some studies demonstrated the potential of AI for use in a variety of ophthalmic settings.^{13,15} The Brazilian software developed for automated classification of fundus diseases¹³ demonstrated the ability to identify other ophthalmic conditions beyond glaucoma and diabetic retinopathy, including ocular toxoplasmosis, papilledema, retinal

detachment, and normal images, achieving an overall sensitivity above 91% and specificity close to 99% for the MobileNet and Inception architectures. Similarly, a system based on smartphone-acquired photographs for automated measurement of eyelid parameters¹⁵ showed high agreement with assessments performed by specialists, with high correlations for MRD1, MRD2, and levator muscle function. Taken together, these results suggest that AI can be used not only for retinal disease screening but also as a multimodal tool for diagnostic support, telemedicine, and objective standardization of clinical measurements in ophthalmology.^{13,15}

• External Validation and Model Generalizability

External validation emerged as one of the main criteria for assessing the clinical applicability of AI algorithms.^{4,14} The studies analyzed showed that high performance indicators obtained during development do not, by themselves, guarantee equivalent results when models are applied to different populations or exposed to factors such as epidemiological differences, image quality, acquisition protocols, and variations among equipment.^{9,10} In this context, even algorithms with high performance in their training datasets may show reduced performance in external settings, a phenomenon known as domain shift.⁵ Thus, robustness, reliability, and adaptability are as important as diagnostic performance itself,^{4,14} reinforcing the need for independent and multicenter validation before large-scale adoption, particularly in population-based screening programs and public healthcare systems.⁷

The included studies employed different external validation strategies, particularly multicenter validations in independent populations,^{11,14} analyses at different levels of healthcare,¹⁰ and tests performed using different image acquisition equipment.^{9,10} Less frequently, there were also evaluations focused on integrating the models into real-world clinical workflows, not only to assess diagnostic performance but also their practical utility.¹⁴ Taken together, these approaches made it possible to assess the ability of the algorithms to maintain performance and applicability outside the environment in which they were originally developed.^{10,11,14}

Among the strategies employed, multicenter validations provided the most robust evidence of generalizability, demonstrating maintenance of performance even in the presence of epidemiological, operational, and technological differences.^{10,11} However, the reduction in performance observed in some external cohorts indicates that generalizability remains a relevant challenge for some of the models evaluated.¹⁴ These findings reinforce that external validation should be understood as a fundamental requirement for the safe incorporation of AI into ophthalmic screening programs.^{4,14,16}

• Clinical Applicability and Public Health Impact

The included studies demonstrated that the main clinical contribution of AI to ophthalmology is related to expanding screening capacity and improving healthcare workflows.^{3,4,8} In practical terms, the algorithms showed potential for the early identification of individuals at higher risk of ocular disease, thereby facilitating timely referral and reducing the burden on referral services.⁹⁻¹¹ This potential was observed particularly in screening programs for highly prevalent diseases, in which AI was frequently proposed as a complementary tool for primary care and teleophthalmology.^{4,8,9,17}

Another recurring aspect was the rationalization of healthcare resources.^{4,8} The studies suggest that incorporating AI systems may contribute to reducing unnecessary examinations and improving the allocation of the specialized workforce.^{8,11} In some settings, a potential reduction in healthcare workload was observed without compromising diagnostic sensitivity, positioning these technologies as tools to support the management of screening programs.^{10,11,14,18,19}

The relevance to public health was particularly evident in settings characterized by a shortage of ophthalmologists and difficulties in accessing specialized services.^{3,4,11} In this regard, the integration of AI, teleophthalmology, and digital platforms was identified as a strategy capable of expanding diagnostic coverage and strengthening prevention programs, particularly in regions with geographical barriers to access.^{3,4,9} In vulnerable populations and healthcare systems with limited resources, these algorithms emerge as a concrete means of expanding ophthalmic care.^{1,4,8,11,17}

• Methodological Limitations and Ethical-Technical Challenges

Despite the promising results of the included studies, several methodological limitations were identified and should be considered when interpreting the findings. The main limitation was the heterogeneity among the studies, which differed in terms of the diseases evaluated, population characteristics, methodological structures, equipment used, and reported performance metrics.^{3,4} In addition, the algorithms were developed for different purposes, such as screening, severity classification, risk prediction, and clinical decision support, which makes direct comparisons difficult and limits the standardization of the available evidence.^{3,4,20}

Another critical issue was the dependence of the algorithms on the quality of the data used.^{9,10,11} The studies highlighted that factors such as image quality, differences among equipment, acquisition protocols, and specific characteristics of the populations evaluated can significantly influence model performance.^{9-11,14} This variability favors the occurrence of domain shift, demonstrating that the reliability of the algorithms remains directly related to the quality and representativeness of the data used in their training and application.^{10,14,21}

Regarding ethical and regulatory aspects, the studies raised concerns related to model interpretability, the limited diversity of certain training datasets, the risk of algorithmic bias, and the definition of clinical responsibilities in cases of diagnostic error.^{4,14} Issues related to data governance, system transparency, and the need for continuous professional oversight were also highlighted.^{4,14,22,23} Therefore, the safe implementation of these technologies depends not only on the technical performance of the algorithms but also on regulatory and monitoring mechanisms that ensure their safety, reliability, and equitable use across different clinical application settings.^{4,14,24}

• Perspectivas futuras

The future perspectives identified in the studies converge toward the consolidation of AI as a tool integrated into ophthalmic healthcare workflows,^{2-4,14} which requires advances related to algorithm robustness, generalizability, and validation.^{4,14} In this regard, priorities include conducting prospective, multicenter studies in more diverse populations, as well as increasing the representativeness of the datasets used for training and validation.^{4,11,14} These measures are essential to increase system reliability and reduce limitations related to domain shift and selection bias, particularly when these models are intended for incorporation into different healthcare settings.^{4,6,10,14}

Another recurring trend is the integration of artificial intelligence, teleophthalmology, and established clinical workflows.^{3,4,14} The studies suggest that the future of these technologies is less related to replacing specialists and more to developing tools capable of supporting decision-making, optimizing screening programs, and expanding access to ophthalmic care.^{4,8,14,25} Thus, aspects such as algorithm interpretability, continuous performance monitoring, and safe integration into healthcare systems are likely to play a central role in consolidating AI as a complementary strategy for the prevention, detection, and management of ophthalmic conditions.^{4,14,26,27}

CONCLUSION

The review showed that artificial intelligence algorithms, particularly those based on deep learning applied to the analysis of ophthalmic images, demonstrate high performance for the screening and diagnostic support of various ocular diseases. Studies conducted on diabetic retinopathy, retinopathy of prematurity, age-related macular degeneration, and glaucoma generally demonstrated high sensitivity, specificity, and area under the ROC curve values. Among these diseases, the population-based screening model and its application in primary care for the detection of diabetic retinopathy achieved the best results among the diseases analyzed, with high rates of accuracy, specificity, and sensitivity. This analysis reinforces the potential of these tools to expand screening capacity, optimize healthcare workflows, and facilitate the early detection of conditions associated with preventable vision loss.

However, the findings also demonstrate that performance obtained exclusively in development datasets does not guarantee the clinical applicability of the models, since limiting factors such as image quality, diversity, dataset size, and differences among the technological equipment used represent relevant challenges for the safe implementation of these technologies. In this regard, external validation proved to be a fundamental step in assessing the robustness, reliability, and generalizability of algorithms across different populations, epidemiological contexts, and operational conditions.

In addition to its high diagnostic potential, AI proved to be a tool with great potential to contribute to expanding access to ophthalmic care, particularly when integrated with ophthalmic teleconsultation and large-scale screening programs, potentially benefiting populations with limited access to specialists. However, its effective incorporation into healthcare systems requires addressing methodological, ethical, and regulatory issues related to model transparency, bias mitigation, clinical supervision, regulation, and data security.

Therefore, artificial intelligence represents a promising and potentially transformative tool for contemporary ophthalmology. However, the future of AI in the specialty does not point toward replacing clinical judgment, but rather toward its use as a tool to support greater diagnostic efficiency.

REFERENCES

1. Kawaguchi A, Sharafeldin N, Sundaram A, Campbell S, Tennant M, Rudnisky C, Weis E, Damji KF. Tele-Ophthalmology for Age-Related Macular Degeneration and Diabetic Retinopathy Screening: A Systematic Review and Meta-Analysis. *Telemed J E Health*. 2018; 24(4): 301-308. doi: 10.1089/tmj.2017.0100.
2. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Adv Ophthalmol Pract Res*. 2022; 2(3): 100078. doi: 10.1016/j.aopr.2022.100078.
3. Pionório VAL, Silva DF da, Pionório JL. Avanços recentes da IA na oftalmologia - revisão de literatura. *Braz J Health Rev [Internet]*. 2024 May 15 [Citado em 07 Jun 2026]; 7(3): 1-19. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/69712>.
4. Checon MES, Grill HM, Wendt MG, Nunes GT, Ávila GC de, Ferreira NM, Carvalho GSC, Andreazza IC, Rocha FLA da. Inteligência artificial na oftalmologia: Diagnóstico preciso, acesso ampliado e novos desafios. *JMBR [Internet]*. 2025 Aug 9 [Citado em 07 Jun 2026]; 2(4): 771-779. Disponível em: <https://journalmbr.com.br/index.php/jmbr/article/view/831>.
5. Guan H, Liu M. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Trans Biomed Eng*. 2022; 69(3): 1173-1185. doi: 10.1109/TBME.2021.3117407.
6. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020; 370: m3210. doi: 10.1136/bmj.m3210.
7. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, Golub RM, Denniston AK, Shetty S, Moher D, Bossuyt PM, Darzi A, Ashrafian H. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med*. 2025; 31(10): 3283-3289. doi: 10.1038/s41591-025-03953-8.
8. Mamprinj J, Prates NA, Mascarenhas Júnior OMR, Rossoni LT, Massote G de B. Inteligência artificial na triagem da retinopatia diabética na atenção primária: acurácia diagnóstica e custo-efetividade. *REASE [Internet]*. 2025 Nov 10 [Citado em 07 Jun 2026]; 11(11): 2177-2187. Disponível em: <https://periodicorease.pro.br/rease/article/view/22147>. doi: 10.51891/rease.v11i11.22147.

9. Arenas-Cavalli JT, Abarca I, Rojas-Contreras M, Bernuy F, Donoso R. Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. *Eye (Lond)*. 2022; 36(1): 78-85. doi: 10.1038/s41433-020-01366-0.
10. Han R, Cheng G, Zhang B, Yang J, Yuan M, Yang D, Wu J, Liu J, Zhao C, Chen Y, Xu Y. Validating automated eye disease screening AI algorithm in community and in-hospital scenarios. *Front Public Health*. 2022; 10: 944967. doi: 10.3389/fpubh.2022.944967.
11. Coyner AS, Oh MA, Shah PK, Singh P, Ostmo S, Valikodath NG, Cole E, Al-Khaled T, Bajimaya S, K C S, Chuluunbat T, Munkhuu B, Subramanian P, Venkatapathy N, Jonas KE, Hallak JA, Chan RVP, Chiang MF, Kalpathy-Cramer J, Campbell JP. External Validation of a Retinopathy of Prematurity Screening Model Using Artificial Intelligence in 3 Low- and Middle-Income Populations. *JAMA Ophthalmol*. 2022; 140(8): 791-798. doi: 10.1001/jamaophthalmol.2022.2135.
12. Zapata MA, Royo-Fibla D, Font O, Vela JL, Marcantonio I, Moya-Sánchez EU, Sánchez-Pérez A, García-Gasulla D, Cortés U, Ayguadé E, Labarta J. Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected Glaucoma. *Clin Ophthalmol*. 2020; 14: 419-429. doi: 10.2147/OPTH.S235751.
13. Kuiava VA, Kuiava EL, Chielle EO, Syllós R. Desenvolvimento de sistema estruturado com inteligência artificial para apoio no diagnóstico de patologias oftalmológicas mais relevantes. *Clin Biomed Res [Internet]*. 2021 Jun 28 [Citado em 07 Jun 2026]; 41(1): 27-32. Disponível em: <https://seer.ufrgs.br/index.php/hcpa/article/view/109565>.
14. Chen Q, Keenan TDL, Agron E, Allot A, Guan E, Duong B, Elsayy A, Hou B, Xue C, Bhandari S, Broadhead G, Cousineau-Krieger C, Davis E, Gensheimer WG, Golshani CA, Grasic D, Gupta S, Haddock L, Konstantinou E, Lamba T, Maiberger M, Mantopoulos D, Mehta MC, Elnahry AG, Al-Nawafli M, Oshinsky A, Powell BE, Purt B, Shin S, Stiefel H, Thavikulwat AT, Wroblewski KJ, Tham YC, Cheung CMG, Cheng CY, Chew EY, Hribar MR, Chiang MF, Lu Z. AI Workflow, External Validation, and Development in Eye Disease Diagnosis. *JAMA Netw Open*. 2025; 8(7): e2517204. doi: 10.1001/jamanetworkopen.2025.17204.
15. Chen HC, Tzeng SS, Hsiao YC, Chen RF, Hung EC, Lee OK. Smartphone-Based Artificial Intelligence-Assisted Prediction for Eyelid Measurements: Algorithm Development and Observational Validation Study. *JMIR Mhealth Uhealth*. 2021; 9(10): e32444. doi: 10.2196/32444.
16. Alqahtani AS, Alshareef WM, Aljadani HT, Hawsawi WO, Shaheen MH. The efficacy of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis. *Int J Retina Vitreous*. 2025; 11(1): 48. doi: 10.1186/s40942-025-00670-9.
17. Kalogeropoulos D, Kalogeropoulos C, Stefanidou M, Neofytou M. The role of tele-ophthalmology in diabetic retinopathy screening. *J Optom*. 2020 Oct-Dec;13(4):262-268. doi: 10.1016/j.optom.2019.12.004. Epub 2020 Jan 14. PMID: 31948924; PMCID: PMC7520530.
18. El Arab RA, Al Moosa OA. Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare. *NPJ Digit Med*. 2025; 8(1): 548. doi: 10.1038/s41746-025-01722-y.
19. Abramoff MD, Whitestone N, Patnaik JL, Rich E, Ahmed M, Husain L, Hassan MY, Tanjil MSH, Weitzman D, Dai T, Wagner BD, Cherwek DH, Congdon N, Islam K. Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial. *NPJ Digit Med*. 2023; 6(1): 184. doi: 10.1038/s41746-023-00931-7.
20. Nittas V, Daniore P, Landers C, Gille F, Amann J, Hubbs S, Puhan MA, Vayena E, Blasimme A. Beyond high hopes: A scoping review of the 2019-2021 scientific discourse on machine learning in medical imaging. *PLOS Digit Health*. 2023; 2(1): e0000189. doi: 10.1371/journal.pdig.0000189.
21. Zhou K, Liu Z, Qiao Y, Xiang T, Loy CC. Domain Generalization: A Survey. *IEEE Trans Pattern Anal Mach Intell*. 2023; 45(4): 4396-4415. doi: 10.1109/TPAMI.2022.3195549.
22. Benzinger L, Ursin F, Balke WT, Kacprowski T, Salloch S. Should Artificial Intelligence be used to support clinical ethical decision-making? A systematic review of reasons. *BMC Med Ethics*. 2023; 24(1): 48. doi: 10.1186/s12910-023-00929-6.
23. Jung J, Lee H, Jung H, Kim H. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon*. 2023; 9(5): e16110. doi: 10.1016/j.heliyon.2023.e16110.
24. Kerasidou A. Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. *J Oral Biol Craniofac Res*. 2021; 11(4): 612-614. doi: 10.1016/j.jobcr.2021.09.004.
25. Ouanes K, Farhah N. Effectiveness of Artificial Intelligence (AI) in Clinical Decision Support Systems and Care Delivery. *J Med Syst*. 2024; 48(1): 74. doi: 10.1007/s10916-024-02098-4.
26. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019; 25(1): 44-56. doi: 10.1038/s41591-018-0300-7.
27. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus*. 2023; 15(10): e46454. doi: 10.7759/cureus.46454.

MAILING ADDRESS

GABRIEL RIBEIRO DE OLIVEIRA SILVA
BR 364, KM 195, No. 3800, Jataí, GO, Brazil
E-mail: oliveira.gabriel@discente.ufj.edu.br

EDITORIAL AND REVIEW

Chief editors:

Waldemar Naves do Amaral - <http://lattes.cnpq.br/4092560599116579> - <https://orcid.org/0000-0002-0824-1138>
Nílzio Antônio da Silva - <http://lattes.cnpq.br/1780564621664455> - <https://orcid.org/0000-0002-6133-0498>

Authors:

Gabriel Ribeiro de Oliveira Silva - <http://lattes.cnpq.br/9082313490110126> - <https://orcid.org/0009-0004-2846-762X>

Ítalo Bezerra Alves - <http://lattes.cnpq.br/5900325370274094> - <https://orcid.org/0009-0000-2393-3135>

Isabela Luiza Pacheco Rodrigues - <http://lattes.cnpq.br/7700244285082906> - <https://orcid.org/0009-0009-3936-0007>

Douglas Wallacy da Silva Amâncio - <http://lattes.cnpq.br/3261605374734058> - <https://orcid.org/0009-0003-2058-308X>

Fabiano Inácio de Souza - <http://lattes.cnpq.br/8305270225089632> - <https://orcid.org/0000-0001-5862-4674>

Library Review: Izabella Goulart

Spell Check: Dario Alvares

Translation: Soledad Montalbetti

Received: 24/08/26. Accepted: 27/08/26. Published in: 29/09/2026.