

Use of artificial intelligence in ophthalmology: a narrative review

Thiago Gonçalves dos Santos Martins^I, Paulo Schor^{II}, Luís Guilherme Arneiro Mendes^{III}, Susan Fowler^{IV}, Rufino Silva^V

Universidade Federal de São Paulo (UNIFESP), São Paulo (SP), Brazil, and Faculty of Medicine, University of Coimbra, Coimbra, Portugal

^IMD, PhD. Researcher, Department of Ophthalmology, Universidade Federal de São Paulo (UNIFESP), São Paulo (SP), Brazil; Research Fellow, Department of Ophthalmology, Ludwig Maximilians University (LMU), Munich, Germany; and Doctoral Student, University of Coimbra (UC), Coimbra, Portugal.

<https://orcid.org/0000-0002-3878-8564>

^{II}PhD. Professor, Department of Ophthalmology, Universidade Federal de São Paulo (UNIFESP), São Paulo (SP), Brazil.

<https://orcid.org/0000-0002-3999-4706>

^{III}PhD. Engineer, Association for Innovation and Biomedical Research on Light and Image (AIBILI), Coimbra, Portugal.

<https://orcid.org/0000-0002-3432-3604>

^{IV}RN, PhD. Certified Neuroscience Registered Nurse (CNRN) and Research Fellow of American Heart Association, Department of Ophthalmology, Orlando Health, Orlando, United States; Researcher, Department of Ophthalmology, Walden University, Minneapolis (MN), United States; and Researcher, Department of Ophthalmology, Thomas Edison State University (TESU), Trenton (NJ), United States.

<https://orcid.org/0000-0003-4140-3642>

^VMD, PhD. Fellow of the European Board of Ophthalmology and Professor, Coimbra Institute for Clinical and Biomedical Research (iCBR), Faculty of Medicine, University of Coimbra, Coimbra, Portugal; Fellow, Department of Ophthalmology, Centro Hospitalar e Universitário de Coimbra (CHUC), Coimbra, Portugal; and Researcher, Association for Innovation and Biomedical Research on Light and Image (AIBILI), Coimbra, Portugal.

<https://orcid.org/0000-0001-8676-0833>

KEY WORDS (MeSH terms):

Artificial intelligence.
Glaucoma.
Retinopathy of prematurity.
Ophthalmology.

AUTHORS' KEY WORDS:

Neural network.
Convolutional neural network.
Diabetes.
Macular degeneration.
Deep reinforcement learning.

ABSTRACT

BACKGROUND: Artificial intelligence (AI) deals with development of algorithms that seek to perceive one's environment and perform actions that maximize one's chance of successfully reaching one's pre-determined goals.

OBJECTIVE: To provide an overview of the basic principles of AI and its main studies in the fields of glaucoma, retinopathy of prematurity, age-related macular degeneration and diabetic retinopathy. From this perspective, the limitations and potential challenges that have accompanied the implementation and development of this new technology within ophthalmology are presented.

DESIGN AND SETTING: Narrative review developed by a research group at the Universidade Federal de São Paulo (UNIFESP), São Paulo (SP), Brazil.

METHODS: We searched the literature on the main applications of AI within ophthalmology, using the keywords "artificial intelligence", "diabetic retinopathy", "macular degeneration age-related", "glaucoma" and "retinopathy of prematurity", covering the period from January 1, 2007, to May 3, 2021. We used the MEDLINE database (via PubMed) and the LILACS database (via Virtual Health Library) to identify relevant articles. **RESULTS:** We retrieved 457 references, of which 47 were considered eligible for intensive review and critical analysis.

CONCLUSION: Use of technology, as embodied in AI algorithms, is a way of providing an increasingly accurate service and enhancing scientific research. This forms a source of complement and innovation in relation to the daily skills of ophthalmologists. Thus, AI adds technology to human expertise.

INTRODUCTION

Early diagnosis of eye diseases is of great importance for preventing loss of vision and thereby improving the quality of life. However, the teaching of ophthalmology has declined in medical schools, thus leaving general practitioners with the task and difficulty of identifying ocular pathological conditions.

In addition, there is irregular distribution of ophthalmology specialists in many countries, which makes it difficult for the population to access adequate eye care.¹ Within this scenario, artificial intelligence (AI) may offer an alternative that could increase the population's access to eye care.^{2,3}

Formation of a convolutional layer is the basis of an algorithm. This transforms the input data through application of a set of filters to produce a final response (such as the output). Neural networks set weights on their own for the filters used during the training process. The filters are defined before the training phase but can be optimized during the learning process.

During the learning phase, algorithm performance can be improved. This phase can be supervised when data is assigned during training. It can also be unsupervised and, in this case, the device creates its own input sample.

The training and development phase of an algorithm is generally divided into training, validation and test data sets. These data sets should not be repeated: hence, an image that is in one of the data sets (for example, training) should not be used in any of the other data sets (for example, validation). The data set used during the training phase can be made as subsets and can be optimized through retro-propagation of the information collected.

The data set used in validation is used for selection of parameters and adjustments, and for implementation of training conditions. After the training phase, independent test data is used, captured using different devices, from different populations under different clinical contexts.

When examining the performance results from an algorithm, it is important to evaluate the methodology and the way in which it was developed. For example, an algorithm developed for analysis of fundus retinography may perform poorly if applied to a retinal photograph with a larger field.⁴⁻⁹

OBJECTIVE

The purpose of this article was to provide an overview of the basic principles of AI and its main studies in the fields of glaucoma, retinopathy of prematurity, age-related macular degeneration and diabetic retinopathy. From this perspective, the limitations and potential challenges that have accompanied implementation and development of this new technology within ophthalmology are presented.

METHODS

We searched the literature on the main applications of artificial intelligence within ophthalmology, using the keywords “artificial intelligence”, “diabetic retinopathy”, “macular degeneration age-related”, “glaucoma” and “retinopathy of prematurity”, covering the period from January 1, 2007, to May 3, 2021. We used the Medical Literature Analysis and Retrieval System Online (MEDLINE) database (via PubMed) and the Latin American and Caribbean Literature in Health Sciences (Literatura Latino-Americana e do Caribe em Ciências da Saúde, LILACS) database (via Virtual Health Library) to identify relevant articles.

Through this search, we selected and reviewed articles on the potential automated clinical applications of artificial intelligence technologies and big data analysis. A summary of the articles selected is provided below. The details of the search strategy are shown in **Table 1**.

RESULTS

From the search in the databases, one clinical trial, four meta-analyses, four randomized controlled trials, 47 reviews and four systematic reviews were identified. After screening the titles and abstracts, removing duplicates and screening the citations, 47

studies were considered eligible for critical analysis. The article selection process is detailed in **Figure 1**.

Diabetic retinopathy

Diabetes is the leading cause of blindness in adulthood, affecting more than 415 million people worldwide.^{10,11} Recent studies on the use of AI for monitoring diabetic retinopathy have demonstrated that it has high precision for detecting this disease.^{10,12,13}

In 2018, IDx-DR, which is an AI diagnostic system that autonomously diagnoses patients with diabetic retinopathy (including macular edema), was approved by the United States Food and Drug Administration (FDA) for classifying diabetic retinopathy. This was the first artificial intelligence device approved by that institution.

Ting et al.¹² evaluated the performance of artificial intelligence for screening for diabetic retinopathy, macular degeneration and glaucoma, using 494,661 images of the retina. This algorithm was then tested externally in 11 multiethnic cohorts. For detecting diabetic retinopathy, its sensitivity was 90.5% and specificity was 91.6%. For diabetic retinopathy with the risk of vision loss, its sensitivity was 100% and specificity was 91.1%. For macular degeneration, its sensitivity was 93.2% and specificity was 88.7%. For glaucoma, its sensitivity was 96.4% and specificity was 87.2%.

Tufail et al.¹⁰ evaluated automated screening of patients with diabetic retinopathy by evaluating systems for automatic detection of diabetic retinopathy in comparison with human graders. EyeArt had sensitivity of 94.7% for detecting diabetic retinopathy and 93.8% for referable retinopathy (maculopathy, proliferative or pre-proliferative diabetic retinopathy). Retmaker had sensitivity of 73% for any retinopathy, 85% for referable retinopathy and 97.9% for proliferative retinopathy. AI models were trained to detect microaneurysms, retinal hemorrhages and hard or soft exudates. Retmaker is a system that has been used for screening diabetic retinopathy.¹⁰

Abramoff et al.¹⁴ evaluated an algorithm for automatic detection of diabetic retinopathy, with specificity of 59.45% and sensitivity of 96.8%. Gulshan et al.¹⁵ developed an algorithm for screening diabetic retinopathy using 128,175 images of color fundus retinography, which had specificity of 90.3% and sensitivity of 98.1%, and

Table 1. Details of the search strategy

Database	Search strategies	Papers found
MEDLINE (via PubMed)	("artificial intelligence") and ("diabetic retinopathy")	226
MEDLINE (via PubMed)	("artificial intelligence") and ("macular degeneration age-related")	53
MEDLINE (via PubMed)	("artificial intelligence") AND ("glaucoma")	151
MEDLINE (via PubMed)	("artificial intelligence") and ("retinopathy of prematurity")	26
LILACS (via Biblioteca Virtual em Saúde)	("artificial intelligence") and ("diabetic retinopathy")	188
LILACS (via Biblioteca Virtual em Saúde)	("artificial intelligence") and ("macular degeneration age-related")	90
LILACS (via Biblioteca Virtual em Saúde)	("artificial intelligence") and ("glaucoma")	118
LILACS (via Biblioteca Virtual em Saúde)	("artificial intelligence") and ("retinopathy of prematurity")	21

LILACS = Latin American and Caribbean Literature in Health Sciences (Literatura Latino-Americana e do Caribe em Ciências da Saúde); MEDLINE = Medical Literature Analysis and Retrieval System Online.

reached an area below the receiver operating characteristic (ROC) curve of 0.99 for detecting referable diabetic retinopathy.

Subsequently, Gulshan et al.¹⁶ investigated use of an algorithm in 10 primary care centers for six months, which resulted in sensitivity of 87.2% and specificity of 90.8% for detecting clinically significant macular edema in at least one eye. This follow-up study emphasized the importance of testing an artificial intelligence algorithm in the real world.

Li et al.¹³ developed an artificial intelligence-based model for detecting diabetic retinopathy based on the color of retinography

photographs. Its sensitivity was 97.0% and specificity was 91.4%, as result of using more than 100,000 images. It reached an area below the ROC curve of 0.99 in validation and 0.955 in external validation using an independent multiethnic data set.

Gargeya et al.¹⁷ published a study in which 75,137 fundus retinography photos from diabetic patients were used to train and test an artificial intelligence model. The model had sensitivity of 94% and specificity of 98%.

The data should preferably be validated using different camera systems and populations. Some diabetic retinopathy assessment

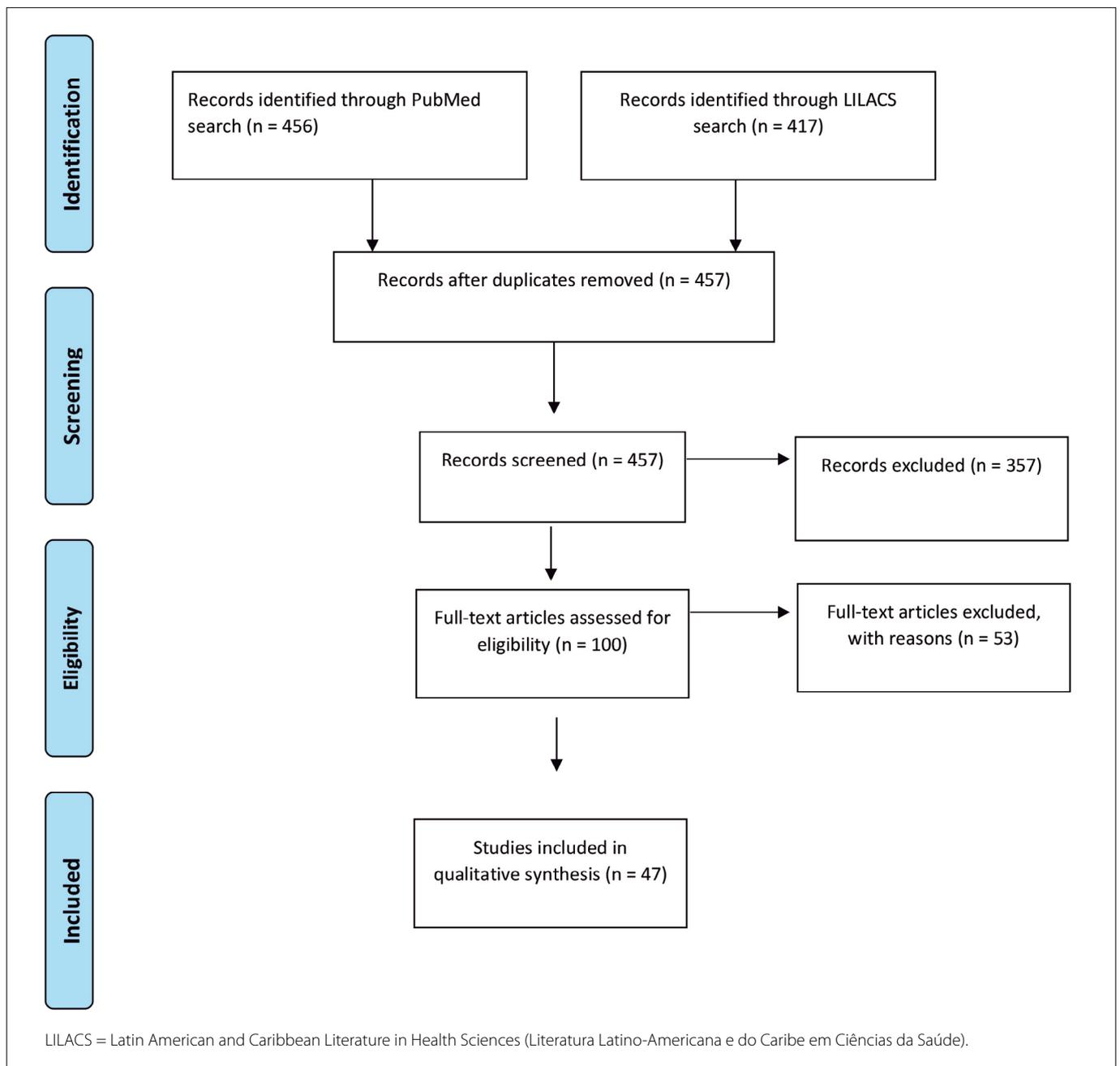


Figure 1. Flow diagram of the study selection process.

systems only evaluate the central 45 degrees, close to the macula, and do not assess diabetic retinopathy lesions that may be occurring on the periphery of the patient's retina.¹⁶

Because of the great variation in the reference standards between different studies, it is difficult to compare the performance of the algorithms. To solve this challenge, algorithms could be tested on an independent data set with a single reference standard.

Age-related macular degeneration

Diabetic retinopathy and age-related macular degeneration (AMD) are the leading causes of blindness among adults over the age of 50 years in the United States. Just like in relation to diabetic retinopathy, development of algorithms for diagnosing and monitoring AMD has therefore been stimulated. AMD cases normally need to be referred to a tertiary-level eye service for clinical evaluation by experts.¹⁸⁻²⁰

Algorithms evaluating color fundus photos

Ting et al.¹² used a database of 72,610 fundus retinography images to classify these patients into intermediate and advanced levels of macular degeneration, with sensitivity of 93.2% and specificity of 88.2%.

Burlina et al.²¹ classified patients using software developed from 130,000 images on 4613 patients and reported that it showed 91.6% accuracy for identifying moderate and advanced macular degeneration. The findings from their study resulted in classification of these patients into four stages, ranging from 1 (without signs of macular degeneration) to 4 (advanced stage).

Grassmann et al.²² tested an algorithm using 120,656 fundus retinography images from 3,654 patients and reported that it showed accuracy of 84.2% for differentiating between early and late macular degeneration and 94.3% for identifying healthy individuals. These investigators used 13 scales based on the Age-Related Eye Disease Study (AREDS), such that stage one showed no signs of degeneration of the macula, stages two to nine represented intermediate disease and stage ten represented the late stage of the disease.

Peng et al.²³ evaluated the severity and risk of progression of macular degeneration using fundus color photography. The performance of the algorithm was compared with that of retinal specialists, and it demonstrated accuracy of 0.94 for detecting large drusen, 0.93 for pigmentary abnormalities and 0.97 for advanced AMD.

Algorithms evaluating optical coherence tomography

Bogunovic et al.²⁴ developed a machine learning method for estimating a risk score and biomarkers associated with progression of macular degeneration in optical coherence tomography (OCT) examinations. In addition, Bogunovic et al.²⁵ evaluated an algorithm for analyzing OCT images in order to predict the best mode of treatment with intravitreal injection. The main predictive

characteristic found in the latter study was the presence of subretinal fluid in the central 3 mm of the macular OCT image.

Schlegl et al.²⁶ developed an AI model for detecting and quantifying intra and subretinal fluid. This algorithm performed well in detecting these lesions in patients with macular degeneration and central retinal vein occlusion. The method had accuracy of 0.94 for detecting intraretinal fluid relating to macular degeneration, diabetic macular edema and retinal vein occlusion. The accuracy for detecting subretinal fluid was 0.92, with superior performance among patients with macular degeneration and retinal vein occlusion, compared with patients with macular edema due to diabetes.²⁷

Venhuizen et al.²⁸ demonstrated the use of AI for classifying the severity of macular degeneration, with sensitivity of 98.2% and specificity of 91.2%. Using a database of 100,000 OCT B-scans (50% from examinations without alteration and 50% from examinations with macular degeneration), Lee et al.²⁹ reported that their algorithm showed accuracy of 87.6% with sensitivity of 84.6% and specificity of 91.5%. That algorithm was developed using images from OCT, diagnoses of macular degeneration provided by a specialist and worst vision of 20/30 in the better eye.

Thus, the main algorithms that have been developed are useful for detecting and segmenting injuries, estimating the risk of progression to advanced stages or evaluating the risk of conversion of dry AMD to an exudative form.

Glaucoma

Glaucoma is an important cause of loss of vision worldwide. In evaluating optic neuropathy, the cup disc needs to be characterized: its size and shape can vary between people. However, defining the cup disc is insufficient for diagnosing glaucoma, due to the large anatomical changes of the optic disc. Examination of the OCT retinal nerve fiber layer thickness and ganglion cell complex can be used for diagnosing glaucoma. Visual field examination is inexpensive and can be used to assess functional loss. However, the sensitivity and specificity of the diagnosis is lower than when a combination of visual field and OCT data is used.

Use of anatomical and functional data together is superior to anatomical data in isolation for diagnosing glaucoma. Artificial intelligence algorithms can combine these factors to aid in making diagnoses. In a study using a database of 125,189 fundus retinography images, Ting et al.¹² reported that their algorithm had sensitivity of 96.4% and specificity of 87.2% for detecting suspected glaucoma, defined as discs with an upper excavation of 0.8 and/or glaucomatous changes. Li et al.³⁰ evaluated a machine learning algorithm for detecting glaucoma based on 48,116 color fundus photographs, with sensitivity of 95.6%, specificity of 92% and an area under the ROC curve of 0.986. The main cause of false negatives in their study was patients with high myopia. Physiological excavation of the optic disc was the most common cause of false

positives in their study. A study using deep learning to classify suspected glaucoma using OCT examinations was developed to differentiate healthy eyes from initial glaucoma.³¹

Kim et al.³² developed a model based on machine learning using three types of records including retinal nerve fiber layer thickness, visual field and ophthalmic clinical data. These investigators reported accuracy of 0.98, sensitivity of 0.983 and specificity of 0.975.

Ahn et al.³³ developed an algorithm that only required retinography data. Use of this limited data resulted in accuracy of 92.2% for identifying glaucoma cases. This model may be convenient with regard to helping with screening of glaucoma cases. Asaoka et al.³⁴ used artificial intelligence to analyze visual fields in patients with pre-perimetric open glaucoma and were able to differentiate them from patients with healthy eyes, with good accuracy (92.6%).

Using 3242 retinography images from eyes with confirmed glaucoma, Shibata et al.³⁵ developed a trained and tested algorithm, from which they reported an excellent area under the ROC curve, of 0.965. This algorithm was trained to detect cup size, optic disc notch, nerve fiber layer atrophy, peripapillary atrophy and optic disc hemorrhage.

Masumoto et al.³⁶ used 1,379 retinography images to detect glaucoma, and found 80.2% specificity and 81.3% sensitivity. The values were higher for severe cases of glaucoma.

Elze et al.³⁷ developed an AI system for identifying patterns of glaucomatous and non-glaucomatous visual field (VF) loss. Through an analysis on 13,231 reliable Humphrey VFs, they identified an ideal solution with 17 prototypes of glaucomatous vision loss. Algorithms show great difficulty in detecting the early stages of glaucoma when patients do not have defects in the visual field. Thus, studies using longitudinal data are needed in order to correctly identify patients who will develop glaucoma.

In patients with severe glaucoma, disease identification by means of algorithms usually has better results. However, caution needs to be exercised due to the great anatomical variability of optic nerves in populations, especially among patients with a high degree of myopia.

Retinopathy of prematurity

Retinopathy of prematurity (ROP), which has a prevalence of 6%-18%, is one of the main causes of loss of vision in childhood worldwide.³⁸ This disease, in its third epidemic, resulted in irreversible blindness in more than 50,000 premature newborns because of a shortage of trained specialists.^{39,40}

Experts usually disagree about the clinical classification of ROP. In the cryotherapy (CRYO)-ROP study, the second examiner disagreed with the first regarding the diagnosis of threshold disease in 12% of the cases.⁴¹ Also, in a multicenter telemedicine study on diagnosing ROP, almost 25% of the tests did not align with one of the three criteria for clinically significant ROP.⁴²

The initial approaches to automated image analysis have been based on quantification of vascular tortuosity and vascular dilation. These systems were developed and validated for wide-angle RetCam images. They were evaluated based on the diagnoses of specialists but did not have any application in the real world because they are only semi-automated, thus requiring manual identification.⁴³

The initial computational approaches for detecting this pathological condition focused on the vascular tortuosity of retinopathy of prematurity-plus (ROP-plus).⁴⁴ Recent work has suggested other possibilities for assessing vessel angles as resources for predictive values for this disease, using linear logistic regression models.⁴⁵

Brown et al. developed and validated a fully automated deep learning system called informatics-retinopathy of prematurity deep learning (i-ROP DL), using a database of 5,511 retinography images obtained by means of a RetCam background camera. This enabled diagnosis of three levels of ROP (plus, pre-plus and normal), with an area under the ROC curve of 0.98 for a positive diagnosis of the disease, in comparison with a reference standard defined by specialists. The i-ROP DL system reached specificity of 94% and sensitivity of 93% for diagnosing ROP-plus and 94% specificity and 100% sensitivity for diagnosing ROP at pre-plus or worse levels.⁴⁶

In an algorithm developed by Redd et al.,⁴⁷ an area under the ROC curve of 0.96 was found for identifying type 1 ROP and 0.91 for clinically significant ROP.⁴⁷ Xiao et al. developed an AI program that quantified the area of neovascularization in patients with ROP.⁴⁸ The AI program reached a higher range of correlation coefficients than that of specialists, for classification of areas with neovascularization. The algorithm works for quantification of key values of oxygen-induced retinopathy images, using deep learning neural networks. Ataer-Cansizoglu et al. developed an AI program with 95% accuracy for analyzing vascularization data.⁴⁹

Current methods for detecting ROP can distinguish between mild and severe cases of ROP but are still unable to identify the stage of the disease.⁵⁰ Campbell et al.⁵¹ demonstrated that automated diagnosis of ROP (i-ROP) had an accuracy of 95%, while the average accuracy of 11 specialists was 87%. Thus, algorithms with performance comparable to that of retinal specialists already exist.

DISCUSSION

Development of algorithms for diagnosing ophthalmic diseases requires many images in order to achieve a classification. When an algorithm is designed, the following need to be considered: the population in which it will be applied, whether it is aligned with current clinical evidence and whether use of the algorithm applies only to diagnosing the disease.⁵²

Because diseases such as glaucoma, macular degeneration, diabetic retinopathy and retinopathy of prematurity have relatively high prevalence, this favors creation of algorithms, given the large

amount of data that has been documented. Rare diseases, with limited data, still present a challenge with regard to development of artificial intelligence programs. Among the topics selected for the present review, retinopathy of prematurity is one for which the fewest algorithms have been developed. This is thought to be due to the lower prevalence of this pathological condition in relation to the others analyzed and the greater difficulty in documenting data among preterm patients. Development of new portable devices that document retinopathy of prematurity may contribute towards development of new algorithms in the future.

Ethical and legal aspects should always be considered by groups that develop algorithms, with the aims of avoiding racial prejudice in healthcare and preserving fundamental rights to protection of personal data.

Currently, there are large databases (big data) of electronic medical records and digital images, which enable recognition of patterns in large volumes of data within a short period of time, thereby reducing errors in diagnostics and therapeutics and creating personalized medicine. In this context, a large database called Intelligent Research in Sight was created to store data on 17,363,018 patients from 7200 ophthalmologists in the United States, in order to improve individual care and public policies.⁵³

The data of some algorithms can be used for three major groups of purposes: classification, segmentation and prediction. In classification, an image will be classified in different categories (presence or absence of disease, for example). This function is explored in disease screening and staging algorithms. In segmentation algorithms, different anatomical structures and lesions of importance

in determining disease biomarkers are outlined. Prediction algorithms, on the other hand, address the relationship of data with future results, and thus help in estimating disease prognosis.

Some AI programs have multiple layers of information input and output, thus enabling a more efficient machine learning process that not only classifies the parameters, but also extracts the results. Most metrics within performance analysis have included calculation of sensitivity, specificity and the area under the ROC curve, which is calculated from sensitivity and specificity values. The closer to 1.0 that the area under the ROC curve reaches, the greater the sensitivity/specificity of the method is. Moreover, it is necessary to evaluate the sensitivity under a fixed specificity. An artificial intelligence system with a good area under the ROC curve may have low sensitivity at a high level of specificity, thus resulting in a high rate of false negatives (Table 2).^{10,12,14-16,21-23,28-30,32-34,36,46,49,51}

It is important to highlight the reliability of the ground truth labels, which in ophthalmological studies are evaluations by specialists, who may nevertheless have divergent opinions. It is important that the sample used for training the algorithms should be specified.

Incorporation of machine learning technology within ophthalmology can improve medical care for the population in regions with limited medical resources, thus reducing some social inequalities.

Future directions, strengths and limitations

Development of enormous longitudinal studies to judge the artificial intelligence systems developed is important for assessing the real security and effectiveness of artificial intelligence systems. Narrative reviews contribute towards providing

Table 2. Comparison of accuracy, sensitivity, specificity and number and type of images analyzed

Authors	Pathological condition/ number of images analyzed	Precision
Ting et al. ¹²	Diabetic retinopathy/76,370 images of retinal photographs	Sensitivity of 90.5% and specificity of 91.6%
Tufail et al. ¹⁰	Diabetic retinopathy/20,258 images of retinal photographs	EyeArt (sensitivity of 93.8%) and Retmaker (sensitivity of 97.9%)
Abramoff et al. ¹⁴	Diabetic retinopathy/1,748 images of retinal photographs	Sensitivity of 96.8% and specificity of 59.4%
Gulshan et al. ¹⁵	Diabetic retinopathy/9,963 images of retinal photographs	Sensitivity of 98.1% and specificity of 90.3%
Gulshan et al. ¹⁶	Diabetic retinopathy/103,634 images of retinal photographs	Sensitivity of 87.2% and specificity of 90.8%
Ting et al. ¹²	AMD/72,610 images of retinal photographs	Sensitivity of 93.2% and specificity of 88.2%
Burlina et al. ²¹	AMD/130,000 images of retinal photographs	91.6% accuracy
Grassmann et al. ²²	AMD/120,656 images of retinal photographs	84.2% accuracy
Venhuizen et al. ²⁸	AMD/3,265 images of OCT	Sensitivity of 98.2% and specificity of 91.2%
Peng et al. ²³	AMD/58,402 images of retinal photographs	Accuracy of 97.0%
Lee et al. ²⁹	AMD/48,312 images of OCT	Sensitivity of 84.6% and specificity of 91.5%
Ting et al. ¹²	Glaucoma/125,189 images of retinal photographs	96.4% sensitivity and 87.2% specificity
Li et al. ³⁰	Glaucoma/48,116 images of retinal photographs	Sensitivity of 95.6% and specificity of 92%
Kim et al. ³²	Glaucoma/399 images of visual field	Sensitivity of 98.3% and specificity of 97.5%
Ahn et al. ³³	Glaucoma/1,542 images of retinal photographs	92.2% accuracy
Asaoka et al. ³⁴	Glaucoma/171 images of visual field	92.6% accuracy
Masumoto et al. ³⁶	Glaucoma/982 images of visual field	Sensitivity of 81.3% and specificity of 80.2%
Brown et al. ⁴⁶	ROP/5,511 images of retinal photographs	93% sensitivity and 94% specificity
Ataer-Cansizoglu et al. ⁴⁹	ROP/77 images of retinal photographs	95% accuracy
Campbell et al. ⁵¹	ROP/77 images of retinal photographs	95% accuracy

AMD = age-related macular degeneration; OCT = optical coherence tomography; ROP = retinopathy of prematurity.

updates on the medical knowledge available. These reviews help in formulating new research projects based on interpretation of the results from published studies after non-systematic analysis. New studies that use algorithms that combine analysis on OCT images, fundus retinography and visual fields can be useful for diagnosing and evaluating multiple pathological conditions simultaneously.¹²

Development of algorithms with real-time cloud information analysis is often helpful within the management and monitoring of eye diseases. However, studies need to be carried out with the aim of preventing certain decisions made by algorithms from going beyond certain ethical and moral precepts that have been established by society. Biases in data collection can substantially affect the generalization of the trained model beyond the population in which it was trained.

Thus, further studies are needed, with algorithms developed in different populations. The data used for external validation should come from a geographically distinct population, with validation by independent researchers.⁵⁴ Studies should be developed with protocols that promote transparency in clinical trials, in order to validate the use of algorithms.⁵⁵

The current algorithms have been developed to evaluate two-dimensional images. Incorporation of multimodal images in the training of the algorithms can facilitate identification of three-dimensional ocular pathological conditions.

The articles included in this review generated heterogeneous data because of the diversity in the design of the studies. The main limitation of this review was the lack of tools for methodological assessment of the reviews. In addition, this narrative review does not provide quantitative answers to specific questions about studying artificial intelligence.

CONCLUSION

Use of technology, as embodied in artificial intelligence algorithms, is a way of providing an increasingly accurate service and enhancing scientific research. This forms a source of complement and innovation in relation to the daily skills of ophthalmologists. Thus, artificial intelligence adds technology to human expertise.

REFERENCES

1. Santos Martins TGD, de Azevedo Costa ALF, Schor P. Comment on: "Do We Have Enough Ophthalmologists to Manage Vision-Threatening Diabetic Retinopathy? A Global Perspective." *Eye (Lond)*. 2021;35(2):690-1. PMID: 32346104; <https://doi.org/10.1038/s41433-020-0903-3>.
2. Martins TG, Costa ALF, Helene O, et al. Training of direct ophthalmoscopy using models. *Clin Teach*. 2017;14(6):423-6. PMID: 28401735; <https://doi.org/10.1111/tct.12641>.
3. Rajaraman V, McCarthy J. Father of artificial intelligence. *Reson*. 2014;198-207. <https://doi.org/10.1007/s12045-014-0027-9>.
4. Martins TGD, Costa ALFA. A new way to communicate science in the era of Big Data and citizen science. *Einstein (Sao Paulo)*. 2017;15(4):523. PMID: 29364372; <https://doi.org/10.1590/S1679-45082017CE4280>.
5. Martins TGD, Francisco Kuba MC, Martins TGD. Teaching Ophthalmology for Machines. *Open Ophthalmol J*. 2018;12:127-9. PMID: 30069266; <https://doi.org/10.2174/1874364101812010127>.
6. Martins TGD, Costa ALFA, Martins TGD. Big Data use in medical research. *Einstein (Sao Paulo)*. 2018;16(3):eED4087. PMID: 30231146; <https://doi.org/10.1590/S1679-45082018ED4087>.
7. Lu W, Tong Y, Yu Y, et al. Applications of Artificial Intelligence in Ophthalmology: General Overview. *J Ophthalmol*. 2018;2018:5278196. PMID: 30581604; <https://doi.org/10.1155/2018/5278196>.
8. Meskó B, Hetényi G, Gyórfy Z. Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Serv Res*. 2018;18(1):545. PMID: 30001717; <https://doi.org/10.1186/s12913-018-3359-4>.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44. PMID: 26017442; <https://doi.org/10.1038/nature14539>.
10. Tufail A, Rudisill C, Egan C, et al. Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. *Ophthalmology*. 2017;124(3):343-351. PMID: 28024825; <https://doi.org/10.1016/j.ophtha.2016.11.014>.
11. Balyen L, Ünlü K, Deniz Balyen LS. Outcomes of Intravitreal Triamcinolone Acetonide Injection in Patients With Diabetic Macular Edema. *Van Med J*. 2018;25(1):28-33. <https://doi.org/10.5505/vtd.2018.91300>.
12. Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017;318(22):2211-23. PMID: 29234807; <https://doi.org/10.1001/jama.2017.18152>.
13. Li Z, Keel S, Liu C, et al. An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs. *Diabetes Care*. 2018;41(12):2509-16. PMID: 30275284; <https://doi.org/10.2337/dc18-0147>.
14. Abramoff MD, Folk JC, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol*. 2013;131(3):351-7. PMID: 23494039; <https://doi.org/10.1001/jamaophthalmol.2013.1743>.
15. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316(22):2402-10. PMID: 27898976; <https://doi.org/10.1001/jama.2016.17216>.
16. Gulshan V, Rajan RP, Widner K, et al. Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India. *JAMA Ophthalmol*. 2019;137(9):987-993. PMID: 31194246; <https://doi.org/10.1001/jamaophthalmol.2019.2004>.
17. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*. 2017;124(7):962-9. PMID: 28359545; <https://doi.org/10.1016/j.ophtha.2017.02.008>.

18. Agurto C, Barriga ES, Murray V, et al. Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images. *Invest Ophthalmol Vis Sci.* 2011;52(8):5862-71. PMID: 21666234; <https://doi.org/10.1167/iovs.10-7075>.
19. Zheng Y, Hijazi MH, Coenen F. Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach. *Invest Ophthalmol Vis Sci.* 2012;53(13):8310-8. PMID: 23150624; <https://doi.org/10.1167/iovs.12-9576>.
20. Mookiah MR, Acharya UR, Koh JE, et al. Automated diagnosis of Age-related Macular Degeneration using greyscale features from digital fundus images. *Comput Biol Med.* 2014;53:55-64. PMID: 25127409; <https://doi.org/10.1016/j.combiomed.2014.07.015>.
21. Burlina PM, Joshi N, Pekala M, et al. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol.* 2017;135(11):1170-6. PMID: 28973096; <https://doi.org/10.1001/jamaophthalmol.2017.3782>.
22. Grassmann F, Mengelkamp J, Brandl C, et al. A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. *Ophthalmology.* 2018;125(9):1410-20. PMID: 29653860; <https://doi.org/10.1016/j.ophtha.2018.02.037>.
23. Peng Y, Dharsasi S, Chen Q, et al. DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs. *Ophthalmology.* 2019;126(4):565-75. PMID: 30471319; <https://doi.org/10.1016/j.ophtha.2018.11.015>.
24. Bogunovic H, Montuoro A, Baratsits M, et al. Machine Learning of the Progression of Intermediate Age-Related Macular Degeneration Based on OCT Imaging. *Invest Ophthalmol Vis Sci.* 2017;58(6):BIO141-BIO150. PMID: 28658477; <https://doi.org/10.1167/iovs.17-21789>.
25. Bogunovic H, Waldstein SM, Schlegl T, et al. Prediction of Anti-VEGF Treatment Requirements in Neovascular AMD Using a Machine Learning Approach. *Invest Ophthalmol Vis Sci.* 2017;58(7):3240-8. PMID: 28660277; <https://doi.org/10.1167/iovs.16-21053>.
26. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning. *Ophthalmology.* 2018;125(4):549-58. PMID: 29224926; <https://doi.org/10.1016/j.ophtha.2017.10.031>.
27. Schlanitz FG, Baumann B, Kundi M, et al. Drusen volume development over time and its relevance to the course of age-related macular degeneration. *Br J Ophthalmol.* 2017;101(2):198-203. PMID: 27044341; <https://doi.org/10.1136/bjophthalmol-2016-308422>.
28. Venhuizen FG, van Ginneken B, van Asten F, et al. Automated Staging of Age-Related Macular Degeneration Using Optical Coherence Tomography. *Invest Ophthalmol Vis Sci.* 2017;58(4):2318-28. PMID: 28437528; <https://doi.org/10.1167/iovs.16-20541>.
29. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related Macular Degeneration. *Ophthalmol Retina.* 2017;1(4):322-7. PMID: 30693348; <https://doi.org/10.1016/j.oret.2016.12.009>.
30. Li Z, He Y, Keel S, et al. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology.* 2018;125(8):1199-1206. PMID: 29506863; <https://doi.org/10.1016/j.ophtha.2018.01.023>.
31. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid Deep Learning on Single Wide-field Optical Coherence Tomography Scans Accurately Classifies Glaucoma Suspects. *J Glaucoma.* 2017;26(12):1086-94. PMID: 29045329; <https://doi.org/10.1097/IJG.0000000000000765>.
32. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One.* 2017;12(5):e0177726. PMID: 28542342; <https://doi.org/10.1371/journal.pone.0177726>.
33. Ahn JM, Kim S, Ahn KS, et al. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS One.* 2018;13(11):e0207982. PMID: 30481205; <https://doi.org/10.1371/journal.pone.0207982>.
34. Asaoka R, Murata H, Iwase A, Araie M. Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier. *Ophthalmology.* 2016;123(9):1974-80. PMID: 27395766; <https://doi.org/10.1016/j.ophtha.2016.05.029>.
35. Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep.* 2018;8(1):14665. PMID: 30279554; <https://doi.org/10.1038/s41598-018-33013-w>.
36. Masumoto H, Tabuchi H, Nakakura S, et al. Deep-learning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity. *J Glaucoma.* 2018;27(7):647-52. PMID: 29781835; <https://doi.org/10.1097/IJG.0000000000000988>.
37. Elze T, Pasquale LR, Shen LQ, et al. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *JR Soc Interface.* 2015;12(103):20141118. PMID: 25505132; <https://doi.org/10.1098/rsif.2014.1118>.
38. Fleck BW, Dangata Y. Causes of visual handicap in the Royal Blind School, Edinburgh, 1991-2. *Br J Ophthalmol.* 1994;78(5):421. PMID: 8025088; <https://doi.org/10.1136/bjo.78.5.421-a>.
39. Quinn GE. Retinopathy of prematurity blindness worldwide: phenotypes in the third epidemic. *Eye Brain.* 2016;8:31-6. PMID: 28539799; <https://doi.org/10.2147/EB.S94436>.
40. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *Arch Ophthalmol.* 2007;125(7):875-80. PMID: 17620564; <https://doi.org/10.1001/archophth.125.7.875>.
41. Reynolds JD, Dobson V, Quinn GE, et al. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. *Arch Ophthalmol.* 2002;120(11):1470-6. PMID: 12427059; <https://doi.org/10.1001/archophth.120.11.1470>.
42. Daniel E, Quinn GE, Hildebrand PL, et al. Validated System for Centralized Grading of Retinopathy of Prematurity: Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Study. *JAMA Ophthalmol.* 2015;133(6):675-82. PMID: 25811772; <https://doi.org/10.1001/jamaophthalmol.2015.0460>.

43. Wittenberg LA, Jonsson NJ, Chan RV, Chiang MF. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus*. 2012;49(1):11-9; quiz 10, 20. PMID: 21366159; <https://doi.org/10.3928/01913913-20110222-01>.
44. Capowski JJ, Kylstra JA, Freedman SF. A numeric index based on spatial frequency for the tortuosity of retinal vessels and its application to plus disease in retinopathy of prematurity. *Retina*. 1995;15(6):490-500. PMID: 8747443; <https://doi.org/10.1097/00006982-199515060-00006>.
45. Oloumi F, Rangayyan RM, Ells AL. Quantification of the changes in the openness of the major temporal arcade in retinal fundus images of preterm infants with plus disease. *Invest Ophthalmol Vis Sci*. 2014;55(10):6728-35. PMID: 25168897; <https://doi.org/10.1167/iovs.13-13640>.
46. Brown JM, Campbell JP, Beers A, et al. Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. *JAMA Ophthalmol*. 2018;136(7):803-10. PMID: 29801159; <https://doi.org/10.1001/jamaophthalmol.2018.1934>.
47. Redd TK, Campbell JP, Brown JM, et al. Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol*. 2018;bjophthalmol-2018-313156. PMID: 30470715; <https://doi.org/10.1136/bjophthalmol-2018-313156>.
48. Xiao S, Bucher F, Wu Y, et al. Fully automated, deep learning segmentation of oxygen-induced retinopathy images. *JCI Insight*. 2017;2(24):e97585. PMID: 29263301; <https://doi.org/10.1172/jci.insight.97585>.
49. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, et al. i-ROP Research Consortium. Computer-Based Image Analysis for Plus Disease Diagnosis in Retinopathy of Prematurity: Performance of the "i-ROP" System and Image Features Associated With Expert Diagnosis. *Transl Vis Sci Technol*. 2015;4(6):5. PMID: 26644965; <https://doi.org/10.1167/tvst.4.6.5>.
50. Wang J, Ju R, Chen Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine*. 2018;35:361-368. PMID: 30166272; <https://doi.org/10.1016/j.ebiom.2018.08.033>.
51. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Imaging and Informatics in ROP (i-ROP) Research Consortium. Expert Diagnosis of Plus Disease in Retinopathy of Prematurity From Computer-Based Image Analysis. *JAMA Ophthalmol*. 2016;134(6):651-7. PMID: 27077667; <https://doi.org/10.1001/jamaophthalmol.2016.0611>.
52. Abràmoff MD, Tobey D, Char DS. Lessons Learned About Autonomous AI: Finding a Safe, Efficacious, and Ethical Path Through the Development Process. *Am J Ophthalmol*. 2020;214:134-42. PMID: 32171769; <https://doi.org/10.1016/j.ajo.2020.02.022>.
53. Chiang MF, Sommer A, Rich WL, Lum F, Parke DW 2nd. The 2016 American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight) Database: Characteristics and Methods. *Ophthalmology*. 2018;125(8):1143-1148. PMID: 29342435; <https://doi.org/10.1016/j.ophtha.2017.12.001>.
54. Faes L, Liu X, Wagner SK, et al. A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. *Transl Vis Sci Technol*. 2020;9(2):7. PMID: 32704413; <https://doi.org/10.1167/tvst.9.2.7>.
55. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351-63. PMID: 32908284; <https://doi.org/10.1038/s41591-020-1037-7>.

Authors' contributions: Martins TGS: contributed actively to the discussion of the study results and reviewed and approved the final version; Schor P: contributed actively to the discussion of the study results and reviewed and approved the final version; Mendes LGA: contributed actively to the discussion of the study results and reviewed and approved the final version; Fowler S: contributed actively to the discussion of the study results and reviewed and approved the final version; and Silva R: contributed actively to the discussion of the study results and reviewed and approved the final version

Sources of funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), under finance code 001

Conflicts of interest: None of the authors had any conflict of interest with this submission

Date of first submission: September 4, 2021

Last received: January 13, 2022

Accepted: February 22, 2022

Address for correspondence:

Thiago Gonçalves dos Santos Martins
Departamento de Oftalmologia, Universidade Federal de São Paulo (UNIFESP)
R. Botucatu, 821
Vila Clementino — São Paulo (SP) — Brasil
CEP 04023-062
Tel. (+55 21) 2571-2248
E-mail thiagogsmartins@yahoo.com.br

