

Original Paper

Screening Diabetic Retinopathy Using an Automated Retinal Image Analysis System in Independent and Assistive Use Cases in Mexico: Randomized Controlled Trial

Alejandro Noriega^{1,2}, PhD; Daniela Meizner³, MD; Dalia Camacho^{2,4}, MSc; Jennifer Enciso^{2,5}, PhD; Hugo Quiroz-Mercado³, MD; Virgilio Morales-Canton³, MD; Abdullah Almaatouq⁶, PhD; Alex Pentland¹, PhD

¹MIT Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, United States

²Prosperia Salud, Mexico City, Mexico

³Retina Department, Asociación para Evitar la Ceguera en México, Mexico City, Mexico

⁴Engineering Academic Division, Instituto Tecnológico Autónomo de México, Mexico City, Mexico

⁵Posgrado de Ciencias Bioquímicas, Universidad Nacional Autónoma de México, Mexico City, Mexico

⁶Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, United States

Corresponding Author:

Alejandro Noriega, PhD

Prosperia Salud

58D Secretaria de Marina 1206, Lomas del Chamizal

Mexico City, 05129

Mexico

Phone: 52 617 982 47

Email: noriega@mit.edu

Abstract

Background: The automated screening of patients at risk of developing diabetic retinopathy represents an opportunity to improve their midterm outcome and lower the public expenditure associated with direct and indirect costs of common sight-threatening complications of diabetes.

Objective: This study aimed to develop and evaluate the performance of an automated deep learning–based system to classify retinal fundus images as referable and nonreferable diabetic retinopathy cases, from international and Mexican patients. In particular, we aimed to evaluate the performance of the automated retina image analysis (ARIA) system under an independent scheme (ie, only ARIA screening) and 2 assistive schemes (ie, hybrid ARIA plus ophthalmologist screening), using a web-based platform for remote image analysis to determine and compare the sensibility and specificity of the 3 schemes.

Methods: A randomized controlled experiment was performed where 17 ophthalmologists were asked to classify a series of retinal fundus images under 3 different conditions. The conditions were to (1) screen the fundus image by themselves (solo); (2) screen the fundus image after exposure to the retina image classification of the ARIA system (ARIA answer); and (3) screen the fundus image after exposure to the classification of the ARIA system, as well as its level of confidence and an attention map highlighting the most important areas of interest in the image according to the ARIA system (ARIA explanation). The ophthalmologists' classification in each condition and the result from the ARIA system were compared against a gold standard generated by consulting and aggregating the opinion of 3 retina specialists for each fundus image.

Results: The ARIA system was able to classify referable vs nonreferable cases with an area under the receiver operating characteristic curve of 98%, a sensitivity of 95.1%, and a specificity of 91.5% for international patient cases. There was an area under the receiver operating characteristic curve of 98.3%, a sensitivity of 95.2%, and a specificity of 90% for Mexican patient cases. The ARIA system performance was more successful than the average performance of the 17 ophthalmologists enrolled in the study. Additionally, the results suggest that the ARIA system can be useful as an assistive tool, as sensitivity was significantly higher in the experimental condition where ophthalmologists were exposed to the ARIA system's answer prior to their own classification (93.3%), compared with the sensitivity of the condition where participants assessed the images independently (87.3%; $P=.05$).

Conclusions: These results demonstrate that both independent and assistive use cases of the ARIA system present, for Latin American countries such as Mexico, a substantial opportunity toward expanding the monitoring capacity for the early detection of diabetes-related blindness.

(*JMIR Form Res* 2021;5(8):e25290) doi: [10.2196/25290](https://doi.org/10.2196/25290)

KEYWORDS

diabetic retinopathy; automated diagnosis; retina; fundus image analysis

Introduction

Impact of Diabetes

Diabetes is one of the most challenging health problems in the world, affecting more than 400 million people. Particularly, diabetes threatens the health care systems of low- and middle-income countries, where 80% of the world's diabetic population live [1,2]. Diabetes is a multifactorial and complex disease with a strong genetic component. In this regard, it has been demonstrated that Hispanic/Latino people have a greater susceptibility to develop type II diabetes, as well as diabetes-associated complications, including renal insufficiency and visual impairment [1-4].

In 2015, there were more than 41 million adults diagnosed with diabetes in Latin America and Caribbean countries, making it one of the major causes of premature death and disability in the region [5,6]. Particularly, Mexico ranked sixth among the world's diabetes prevalence in 2015 and second among Latin America, only after Brazil [7,8]. It is estimated that 26 million adults live in Mexico with diabetes or prediabetes, and only half of them have been diagnosed. Diabetes and its related complications are the first cause of disability and the third cause of death in the country, largely impacting productivity, life quality, and the economy [5].

Evolution and Treatment of Diabetic Retinopathy

Diabetic retinopathy (DR) is the most common complication in advanced or uncontrolled diabetic patients and is the leading cause of irreversible vision loss in working-age adults [9,10]. DR is a microvascular complication that emerges in diabetic patients as a consequence of chronic hyperglycemia that contributes to blood vessel damage in the retina, causing a combination of fluid leakage, swelling of the surrounding tissue, blood flow obstruction, and abnormal neovascularization [9,10].

DR progression is slow, gradual, and reversible in its first stage. However, if not treated promptly, it can lead to irreversible blindness. According to the International Clinical Diabetic Retinopathy Severity Scale, the first stage of DR is classified as mild nonproliferative diabetic retinopathy (NPDR), which is characterized by the presence of at least 1 microaneurysm and is highly reversible through blood pressure, cholesterol, and sugar level control. Only very rare cases that present macular edema (swelling of fluid and protein deposits on or under the macula) might require laser photocoagulation or intravitreal injections. Without adequate diabetic control, the disease advances to moderate and severe NPDR stages, which include the presence of hemorrhages, microaneurysms, hard exudates, venous beading, or intraretinal microvascular abnormalities. At these stages, metabolic control is not sufficient

to stop the disease progression, and the patient will require invasive treatments such as photocoagulation and intravitreal antivascular endothelial growth factor agents or corticosteroids. The most advanced stage is proliferative DR and is characterized by neovascularization, preretinal hemorrhages, hemorrhages in the vitreous, traction retinal detachments, or macular edema. Proliferative DR is treated with the more aggressive laser therapy called scatter or pan-retinal photocoagulation; intravitreal injection; and, in some cases, vitreoretinal surgery, which removes scar tissue or blood from the vitreous cavity to repair retinal detachments or treat macular holes [10-13].

To increase early detection and prevent the progression of DR to advanced stages, diabetic patients are recommended to have annual or semiannual retinal screenings beginning at the moment when they are diagnosed with diabetes. However, according to data from the Diabetic Retinopathy Barometer, 27% of people living with diabetes declared that they never discussed eye complications with their doctors before the onset of complications, and only 13% of the diabetic population have visited an ophthalmologist after their diagnosis [4,14]. Through frequent, preventive screenings, 70% of the cases can be captured at the initial stages of the disease and treated with noninvasive strategies such as metabolic control or photocoagulation [15]. Unfortunately, in most developing countries, there is no ophthalmological attention at primary care clinics, and it is only when diabetic patients develop vision attenuation that they are referred to second- and third-level hospitals to be screened, diagnosed, and treated [16]. At this point, significant retinal damage has occurred, and, even with invasive vitreoretinal surgery or photocoagulation, vision cannot be restored.

The limited access to ophthalmologists and retina specialists at primary care clinics, due to financial and staff limitations at national health care institutions, precludes the continuous monitoring of diabetic patients in low- and middle-income countries such as Mexico.

Challenges of Diabetic Retinopathy Screening on a Large Scale

In Mexico, DR is a leading cause of irreversible blindness among the working-age population [4,13]. Approximately 30% of the patients diagnosed with diabetes develop DR, and, based on the predictions of diabetes increasing in prevalence, by 2045, there will be 245 million people with DR lesions and 77 million people with vision-threatening DR [17].

One of the main limitations for the establishment of a systematic eye-screening program is the limited availability of ophthalmologists and their unequal distribution around the country. Based on the 2013 registry of society-affiliated

ophthalmologists from the Mexican Society of Ophthalmology, the average number of ophthalmologists per 100,000 people is lower (2.68 per 100,000) than the average among Latin American countries (5.27 per 100,000). There is a particularly worrying distribution in rural areas, with 2 ophthalmologists per 100,000 people [18].

In particular, in low- and middle-income countries such as Costa Rica, Peru, and India, there have been several efforts to implement DR screening programs targeting the limitation of ophthalmologists with mobile screening units integrated with telemedicine [19-21]. In these contexts, 2 key factors were identified for achieving cost-effectiveness of these strategies: (1) accurate identification of the risk population and (2) optimization of the number of people screened per unit of time [21]. Notably, these 2 factors can be improved by leveraging automated retinal image analysis (ARIA) systems such as the one in this study.

ARIA for Diabetic Retinopathy Screening

In recent years, the combination of the development of advanced statistical methods, the greater availability of data, and the substantial increase in computing power has allowed for the application of advanced computational methodologies, including artificial intelligence (AI), in diverse social and medical domains. Among the use of AI for social welfare, AI applications in health care domains are one of the fastest growing sectors, with a compound annual growth rate above 40% during the period between 2014 and 2021 [22]. AI tools have been successfully applied to diagnostics, therapeutics, population health management, administration, and regulation, showing a capacity to augment societies' access to health care and improve the coverage and quality of the services provided.

Ultimately, AI applications in health care present opportunities to improve overall quality of life, patients' prognoses, and optimization of human and financial resources [23]. In particular, ARIA systems have emerged as a promising solution to increase early detection of DR at primary care clinics, particularly, in resource-constrained developing countries, thereby improving health outcomes, avoiding incapacitating complications, and reducing treatment costs.

ARIA systems analyze retinal fundus images by applying techniques such as deep learning (DL) to classify diabetic patients in (1) cases without retinal lesions associated to DR (nonreferable output) and (2) cases that need to undergo examination by an ophthalmologist to confirm diagnosis and define treatment (referable output) [24-28]. As of today, various analysis systems have been developed and implemented on the market in European countries, Canada, and the United States. However, very few have been tested in Latin America and Caribbean countries to evaluate their performance and usability in the particular resource-constrained settings of these countries [29]. To determine qualities of successful implementation in these countries, research must investigate patients' ethnicities, the training of health care personnel, community openness to new technologies, and hospital resources.

Aims and Key Findings of the Study

This study aimed to evaluate the performance of a DL-based ARIA system that classifies retinal fundus images in nonreferable or referable circumstances, based on the presence of DR damage, as well as the potential benefits of its use as an assistive tool for ophthalmic doctors. We also completed a randomized controlled trial where the performance of the ARIA system was compared with the accuracy of 17 ophthalmologists from one of the most reputable ophthalmic hospitals in Mexico, Hospital de la Ceguera, which is part of the "Association to Avoid Blindness in Mexico" (APEC). In particular, the performances of ophthalmologists in 3 experimental conditions were assessed: 1 independent condition, in which the ophthalmologists assessed the images independently from the ARIA system, and 2 assistive conditions, in which either ophthalmologists observed and were influenced by the ARIA system's classification and confidence or an ARIA system-generated, attention heatmap highlighted probable DR lesions in the retina.

The key findings were that the ARIA system developed using a DL strategy was able to classify referable vs nonreferable cases with an area under the receiver operating characteristic curve (AUROC) of 98%, a sensitivity of 95.1%, and a specificity of 91.5% for international patient cases. There was an AUROC of 98.3%, a sensitivity of 95.2%, and a specificity of 90% for Mexican patient cases. For Mexican patient cases, the ARIA system performance was more successful than the average performance of the 17 ophthalmologist participants in the study. Moreover, we found that the ARIA system can be useful as an assistive tool, as we found significant improvement in the specificity in the experimental condition where participants were able to consider the answer of the ARIA system as a second opinion (87.3%), compared with the specificity of the condition where participants assessed the images independently (93.3%; $P=.05$).

Hence, this study aimed to demonstrate the high potential value of the use of ARIA systems, in both independent and assistive schemes, toward the goal of effective mass screening for the early detection of DR in developing countries such as Mexico.

Methods

ARIA System

ARIA System Design

The ARIA system consists of an image preprocessing module and an image analysis module that returns a binary referable and nonreferable DR classification; the level of confidence of that classification; and an attention map that shows, pixel-wise, the indicative features for referable DR according to the model (Figure 1). The models constituting the ARIA system were implemented using the Keras library with the Tensorflow backend [30] in Python 3.5 [31].

Images from all datasets were annotated by ophthalmic specialists for 5-class identification according to the International Clinical Diabetic Retinopathy Severity Scales (ICDRSS) and subsequently labeled as nonreferable or referable

DR [32]. Table 1 describes the classification, and Figure 1A provides a graphical example. The gold standard classification used for the experimental phase of the study was provided by 3 retina specialists, as described in the following subsections.

Figure 1. Deep learning–based automated retinal image analysis system. (A) Example of classified retinal fundus images according to the International Clinical Diabetic Retinopathy Severity Scale used for the training data. (B) Flow chart describing the design of the automated retinal image analysis system; the data used for training, validation, and testing; and the algorithm’s outputs. DR: diabetic retinopathy; NPDR: nonproliferative diabetic retinopathy; PDR: proliferative diabetic retinopathy.

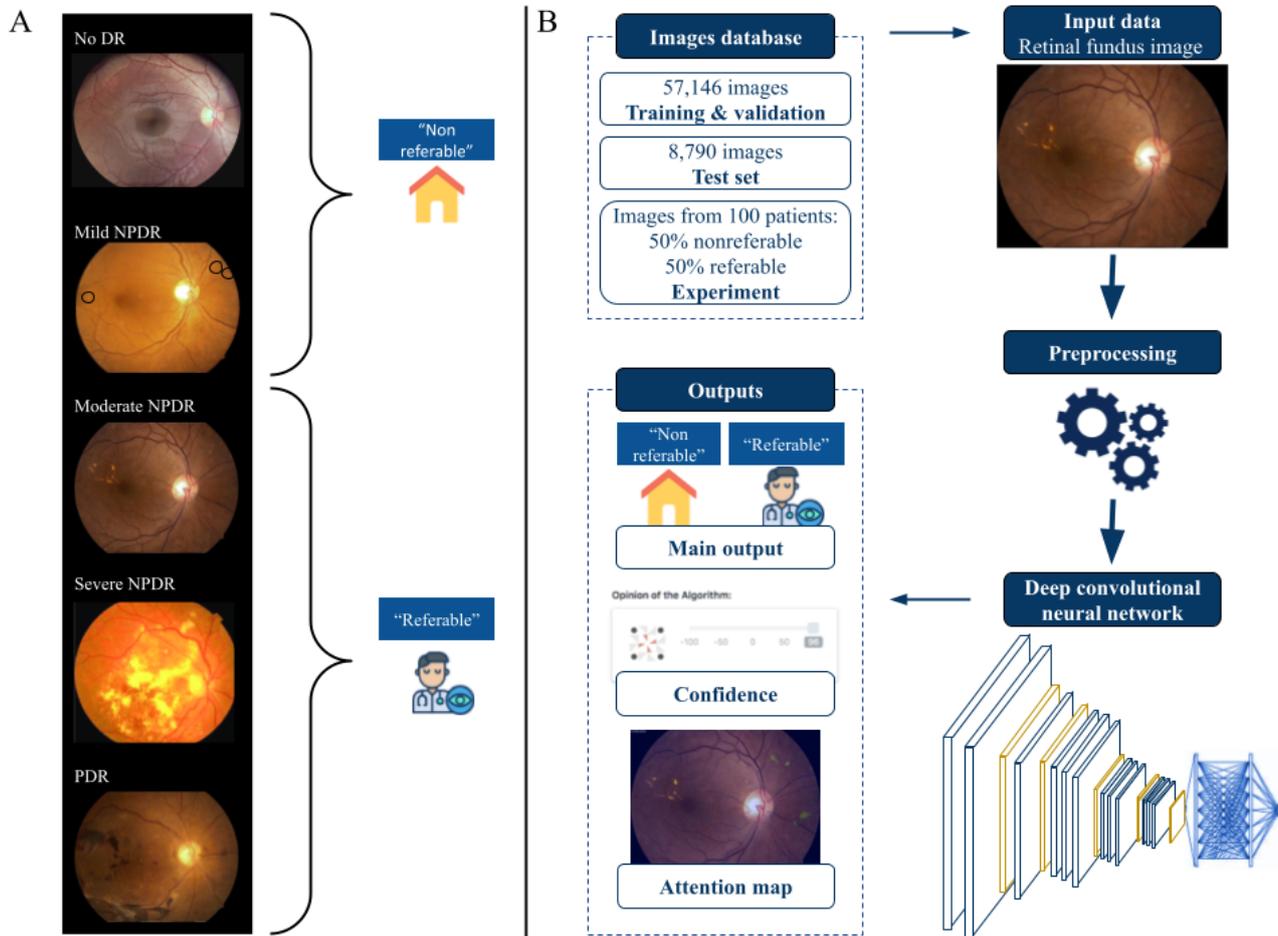


Table 1. International Clinical Diabetic Retinopathy Severity Scale and its classification for the automated retinal image analysis system [32].

ARIA ^a system classification	DR ^b severity scale	Ophthalmoscopy findings
Nonreferable	No apparent retinopathy (no DR)	No abnormalities
	Mild nonproliferative DR (mild DR)	Microaneurysms only
Referable	Moderate nonproliferative DR (moderate DR)	More than just microaneurysms but less than severe nonproliferative diabetic retinopathy
	Severe nonproliferative DR (severe DR)	≥20 intraretinal hemorrhages in each of 4 quadrants, definite venous beading in 2 quadrants, or prominent intraretinal microvascular abnormalities in 1 quadrant. No signs of proliferative retinopathy.
	Proliferative DR	Neovascularization or vitreous/preretinal hemorrhage.

^aARIA: automated retinal image analysis.

^bDR: diabetic retinopathy.

Preprocessing

Before classifying the images and training the algorithms, a preprocessing procedure was applied. The procedure consisted of cropping the background to eliminate noninformative areas,

padding the image to guarantee consistent squared image ratios, resizing the image to 224×224 pixels, and normalizing pixel values to the range 0-1.

Image Classification Model

The model used for image classification consisted of a deep convolutional neural network [33,34]. The network architecture developed for this project consisted of 16 convolutional layers, a dense layer of 1024 neurons, 2 dropout layers to avoid overfitting, and a binary classification layer of a single unit with sigmoid activation. This architecture took the VGG model published by Simonyan and Zisserman [34] as a starting point. Hence, the model output is a value between 0 and 1, which may be interpreted as the confidence of the model regarding a referable DR classification. Lastly, a threshold of 0.5 was used to classify nonreferable (<0.5) and referable (≥ 0.5) DR.

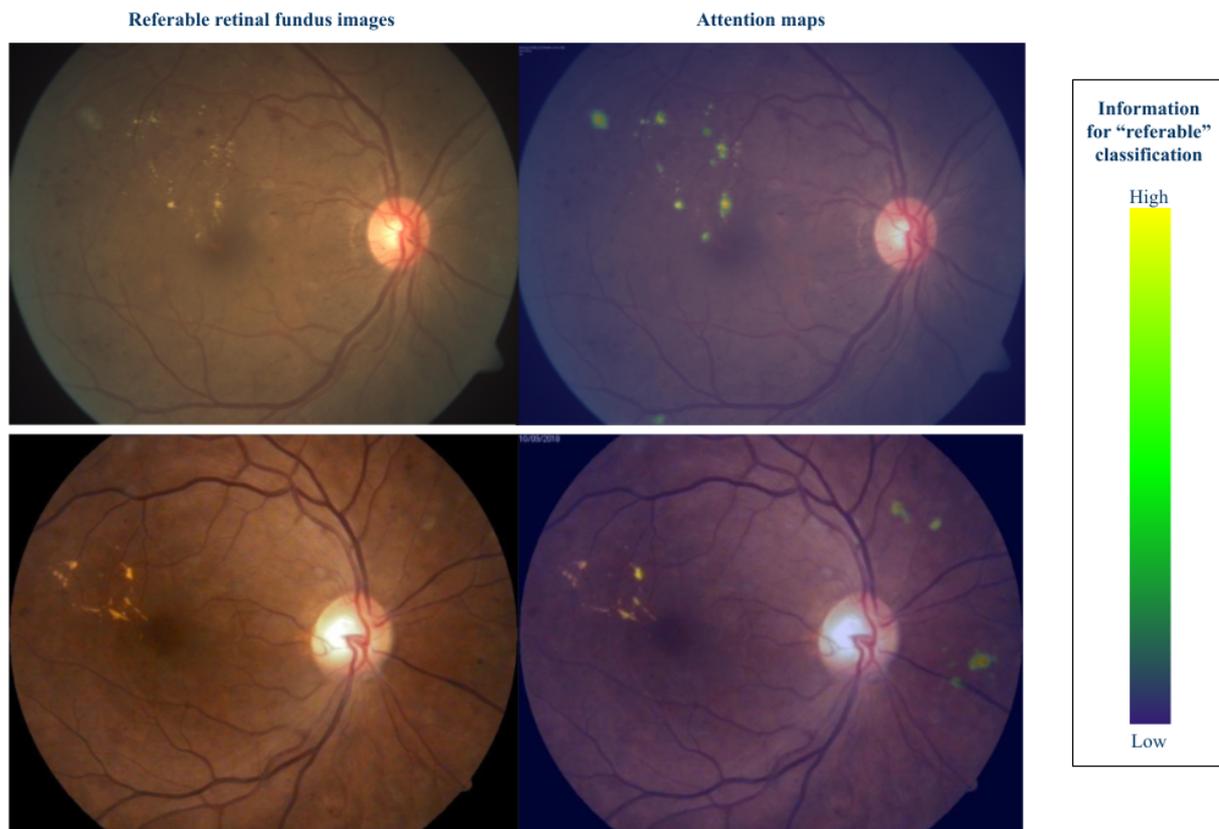
The model was trained on an international dataset, of which most images were taken in primary care clinics in California, United States [35]. The training subset had 57,146 images (16,458/57,146, 28.80% with referable DR; 45,602/57,146, 79.80% gradable), and the evaluation subset had 8,790 images (694/8790, 7.90% with referable DR; 7067/8790, 80.40%

gradable). The training and test subsets followed the same distribution used by Voets and colleagues [36]. Considering real-life scenarios, the training and validation datasets included images from different types of cameras and of different qualities (ie, with artifacts, out of focus, underexposed, or overexposed).

Attention Heatmaps

Attention heatmaps were developed to show lesion areas in the image by highlighting each pixel according to their importance to a referable DR classification, according to the model. These heatmaps were obtained by applying one of the most effective methods for building saliency maps on images, the layer-wise relevance propagation method, with an alpha-beta rule [37,38]. In essence, the layer-wise relevance propagation method redistributed the output value throughout the layers until the input layer (input image) was reached. Figure 2 shows examples of fundus images and the heatmaps generated using the methodology described.

Figure 2. Attention heatmaps for 2 referable images. Green and yellow colors indicate regions in the image that provide information to the algorithm to classify the image as referable.



Study Populations

We had 17 ophthalmologists from the Mexican ophthalmic hospital participating in the experimental study, and 3 retina specialists from the same institution participated in the generation of the gold standard. The 17 ophthalmologists evaluated 45° macula-centered fundus images from 100 Mexican patients, where 50% (50/100) had nonreferable DR and 50% (50/100) had referable DR levels. Each ophthalmologist evaluated 45 retinal images, in order for each image to be

evaluated more than once. The ophthalmologists were retina specialization resident students, where 3 residents were in their second year, 12 were in their third year, and 2 were in their fourth year of residency.

Experimental Design

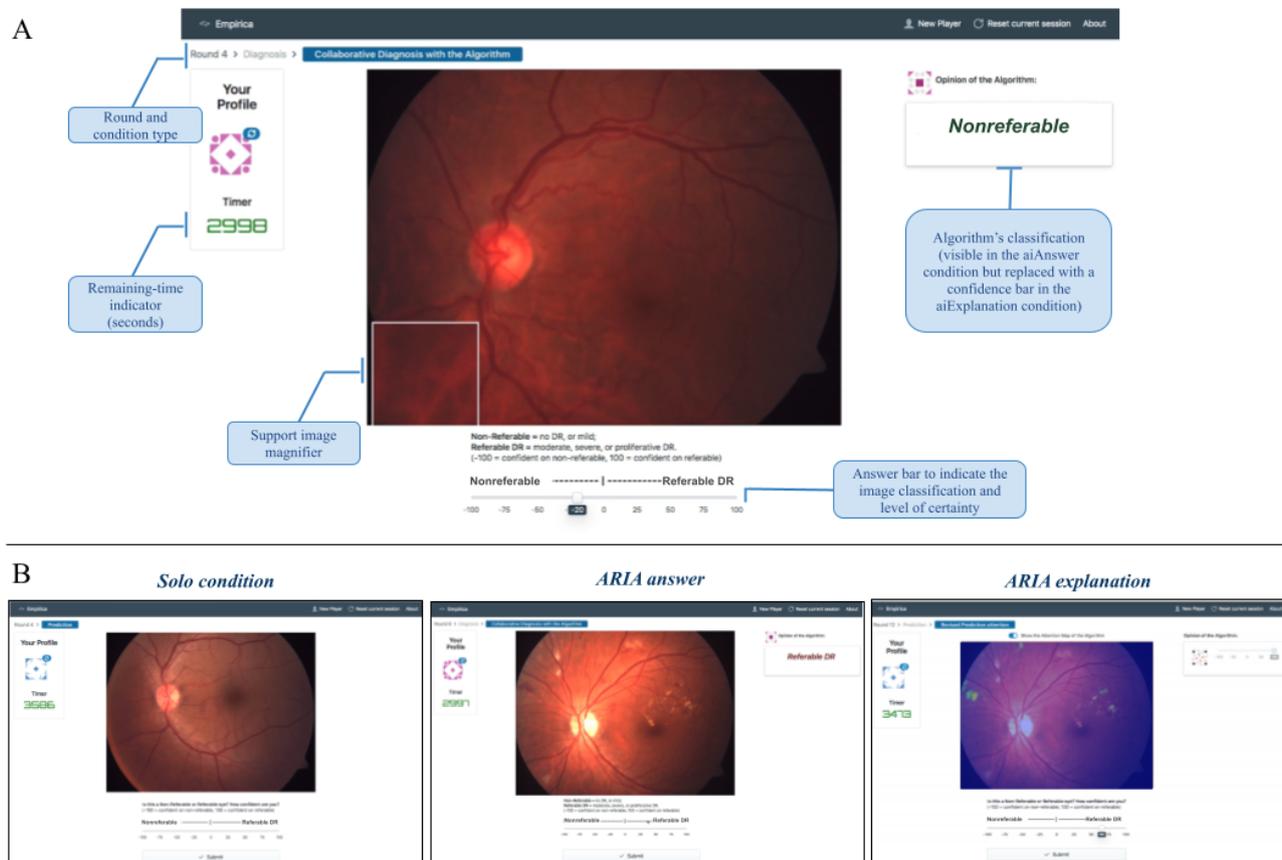
Overview of Study Design

We conducted a randomized controlled experiment to assess the performance of the ARIA system in comparison with

ophthalmic doctors from the Mexican ophthalmic hospital and to evaluate the potential benefits of using the system as an assistive tool for doctors. To achieve this, a web-based experiment platform was developed where ophthalmologists evaluated fundus retinal images under 3 different

conditions—solo, ARIA answer, and ARIA explanation—described below. The platform was developed based on the Empirica framework [39]. Figure 3 displays the main screens of the web platform used in this experiment.

Figure 3. Web-platform design for patient-case classification. (A) Visual indicators and components of the classification window. (B) Visualization of the 3 experimental conditions. ARIA: automated retinal image analysis; DR: diabetic retinopathy.



Gold Standard and Image Quality

To generate a gold standard, the fundus images of all patient cases used in the experiment were graded by 3 retina specialists of the ophthalmic hospital, and a majority rule was used (ie, if there was a disagreement in the nonreferable/referable label, the label selected by 2 of 3 experts was considered the gold standard). We used the same web-based platform described in Figure 3 for image grading. The retina specialists also graded the image quality, and images graded as bad quality were not considered for the experiment. From the remaining images, 50 images from patients with referable DR and 50 images from patients with nonreferable DR were selected at random to be used for the study. According to the ICDRSS, the selected images had the following distribution: 49 with no apparent retinopathy, 1 with mild DR, 33 with moderate DR, 12 with severe nonproliferative DR, and 5 with proliferative DR. Since these images were taken at an ophthalmic hospital, most patients with DR were under treatment and therefore had more advanced DR stages (moderate, severe, and proliferative DR).

Experimental Conditions

The experiment followed a within-subjects design, where each ophthalmologist evaluated 45 randomly selected fundus images

(from 45 different patients), 15 for each of the 3 treatment conditions: solo, ARIA answer, and ARIA explanation. The ophthalmologists were first asked to evaluate 15 fundus retinal images in the solo condition, followed by 30 images that randomly alternated between the ARIA answer and the ARIA explanation conditions. The 15 images in each condition subset were randomly selected for each participant without replacement from all images available for the experiment, generating a rough balance in the proportion of referable and nonreferable images across conditions. In particular, the average proportion of referable images was 49.8% (127/255) for the solo condition, 52.5% (134/255) for the ARIA answer condition, and 46.7% (119/255) for the ARIA explanation condition. In addition, Multimedia Appendix 1 reports the average number of observations of each ICDRSS class for each treatment condition.

In the solo condition, participants responded to the task in isolation, without any exposure to the ARIA system. In contrast, in the ARIA answer condition, participants were exposed to the binary answer of the ARIA system (ie, nonreferable or referable), as a second opinion, and then asked to submit their postexposure answer. The ARIA explanation condition was identical to the ARIA answer condition, with the exception that

participants were shown not only the binary answer of the ARIA system but also its level of confidence and attention heatmap.

Finally, after completing all the classification tasks, the ophthalmologists were asked to submit an optional feedback survey about their experience.

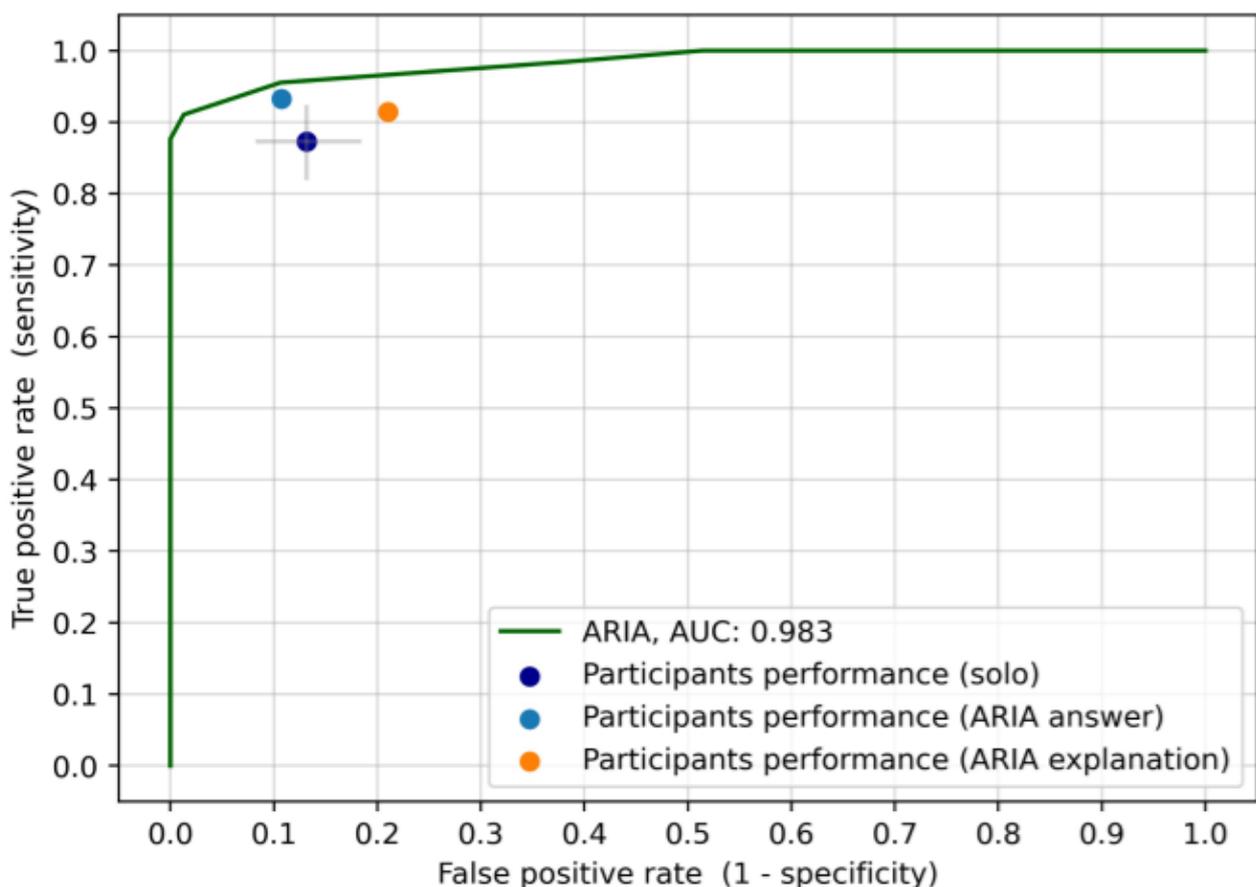
The study was reviewed and approved by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology, and all participants provided explicit consent prior to their participation.

Results

ARIA's Independent Performance

The ARIA system was first tested in a large dataset of international cases. It achieved an out-of-sample area under the receiver operating characteristic curve (AUROC) of 98% (Multimedia Appendix 1). In particular, using a given acceptance threshold, the ARIA system achieved a sensitivity of 95.1% and a specificity of 91.5%. Most importantly, the ARIA system also displayed high accuracy classifying images from patients from the Mexican ophthalmic hospital, where it had an AUROC of 98.3%, a sensitivity of 95.2%, and specificity of 90% (Figure 4).

Figure 4. Receiver operating characteristic curve of the ARIA system compared with the ophthalmologist's accuracy under the 3 experimental conditions (solo, ARIA answer, and ARIA explanation). Grey lines indicate 95% CIs for the solo condition. ARIA: automated retinal image analysis; AUC: area under the curve.



ARIA's Assistive Performance

Figure 4 shows the sensitivity and false positive rate (false positive rate = 1 – specificity) for each condition—solo, ARIA answer, and ARIA explanation—and compares them with the receiver operating characteristic curve of the ARIA system. The average sensitivity in the solo condition across the 17 participants was 87.3%, and the average specificity was 86.8%. In comparison, the average sensitivity and specificity across the 17 participants for the ARIA answer condition were 93.3% and 89.3%, respectively, and the average sensitivity and specificity across participants for the ARIA explanation condition were 91.5% and 79%, respectively.

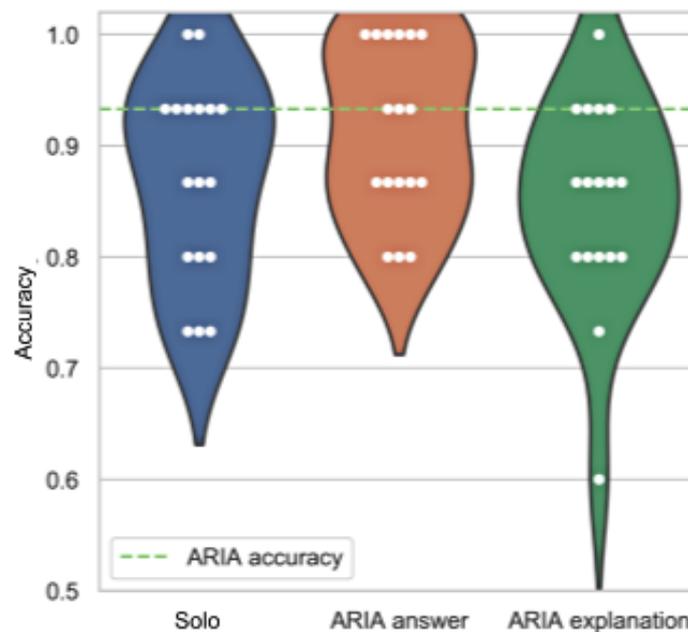
The joint analysis of the ARIA system performance for Mexican patients, compared with the 3 experimental conditions involving ophthalmologist assessments, showed that the ARIA system is more accurate than the average accuracy of participants under any of the exposure conditions. In particular, the ARIA system increased sensitivity from 87.3% to 93.3% ($P=.05$; vertical movement between the dark blue dot and the green line in Figure 4) while maintaining participants' specificity at 86.8%. Compared with the solo condition, the ARIA system also increased specificity to 100% while maintaining participants' average sensitivity at 87.3% (horizontal movement from the dark blue dot leftwards to the green line in Figure 4).

Most interestingly, Figure 4 shows that exposure to the ARIA system was able to improve the performance of human experts, particularly, in the ARIA answer condition, which significantly improved the sensitivity and specificity compared with the solo condition (distance between dark blue and light blue dots in Figure 4). However, performance in the ARIA explanation condition had mixed results, showing improved sensitivity but worse specificity (distance between dark blue and orange dots in Figure 4).

Figure 5 provides more detail on the effect that exposure to information of the ARIA system had on the performance of

ophthalmologists. In particular, it shows that the accuracy (% of correct answers) of the 17 experts consistently improved in the ARIA answer condition, shifting the distribution upwards and decreasing the variance across participants. For example, while only 2 participants had a perfect score in the solo condition, up to 6 participants had a perfect score in the ARIA answer condition. However, the ARIA explanation condition had mixed beneficial and detrimental effects on participants' accuracy and increased the variance of performance across participants compared with the solo condition.

Figure 5. Influence of the ARIA system on the ophthalmologists' decisions: ophthalmologists' performance after exposure to the ARIA answer or the ARIA explanation condition outputs. ARIA: automated retinal image analysis.



Discussion

Principal Findings

The number of people living with diabetes by 2045 is projected to reach 700 million people worldwide [7,40]. This means that routine eye screening might prevent vision loss in approximately 230 million patients. Just in Mexico, the prevention of DR would implicate savings of up to US \$10 million for the 3 main public health care institutions [41]. The development of ARIA systems represents a possible solution to the increasing demand of eye screenings in health care systems, particularly, in limited-resource settings. However, it has been shown that acceptance of the human factors involved in the field processes are critical for the effective implementation of screening systems [42,43].

In this study, we successfully developed and evaluated a DL-based ARIA system to determine its performance as an independent decision-making system, as well as a supportive tool for health care professionals. As an independent decision-making tool, the ARIA system outperformed the average ophthalmologist participant in the experiment. On the other hand, as a supportive tool, the ARIA system exerted a strong influence on the opinion of human participants. However,

its effect depended on the output's format, highlighting the importance of a well-designed platform that has been user-tested with final users.

ARIA's Independent Performance

The DL-based ARIA system presented in this work was evaluated with a subset of retinal images from international patient cases and an image set of patients from a Mexican ophthalmic hospital. In both datasets, the ARIA system outperformed the average sensitivity and specificity of 17 ophthalmology residents of retina specialty.

The sensitivities (95.1% and 95.2% for the international and Mexican datasets, respectively) are comparable to those reported for 7 other automated DR screening systems assessed in a systematic review, whose sensitivity values were between 87% and 95% [44]. On the other hand, the specificities reached by our ARIA system (91.5% and 90% for the international and Mexican datasets, respectively) were higher than the average specificity values of between 49% and 69% reported by Nørgaard and Grauslund [44]. Also, our system's sensitivity and specificity were comparable with those reported for commercial DR screening technologies with DL features, whose sensitivity and specificity values were 85%-99.3% and 68.8%-97.9%, respectively [45]. Compared with these

commercial DR screening technologies, our ARIA system has one of the best balances between sensitivity and specificity, with both measurements above 90%.

ARIA Assistive Performance

Besides the sensitivity and specificity assessment, the ARIA system evaluation included 2 hybrid decision schemes, either assistive or a combination of human and AI. The experimental design was developed to reflect that in real-world applications, results of an automated system are reviewed and confirmed by health care professionals to choose the most adequate therapeutic protocol for each patient. In these assistive evaluations, we confirmed the existence of significant synergies derived from the interaction among the human and AI dyads.

The ARIA output's influence on ophthalmologists' overall precision depended on its format. A simplified output (ie, nonreferable or referable classification) resulted in the most successful sensitivity and specificity for ophthalmologists' inputs. On the other hand, a more complex output (ie, with a confidence bar and attention map) partially improved ophthalmologists' decisions, increasing their sensitivity but also increasing the incidence of false positive classifications.

These results are coherent with some of the ophthalmologists' feedback submitted after the classification tasks, where some expressed that even when attention heatmaps were useful, the bar showing the confidence of the ARIA system was confusing.

Limitations

Future pilot studies with a larger number of patients and ophthalmologists will be useful to confirm the ARIA system's accuracy. Also, future studies might include direct ophthalmoscopy by retina specialists as the gold standard, in order to avoid errors related to image quality.

Additional experiments with alternative platform designs might be useful to generate a suitable screening tool that optimizes patient evaluations and referrals in 2 stages. In the first stage, an ARIA system might be useful to identify patients with a higher probability of developing DR. In the second stage, ophthalmologists would be able to evaluate the retinal images of high-risk patients, in combination with the ARIA system output, to make a first decision about the disease stage and treatment, sending referrals to retina specialists only for patients with an advanced disease.

Conclusions

The results of this study demonstrate a substantial opportunity for Latin American countries such as Mexico toward developing efficient mass screening systems for early detection of diabetes-related blindness, considering the short supply of ophthalmologists in their public health care system.

The web-based platform developed for this study was designed for the implementation of the ARIA system as an automatic screening tool and as a telemedicine platform to confirm or reject the ARIA system's output with assessment of an ophthalmologist or retina specialist. The platform was useful for this study and can be easily adapted for future studies that include the collection of additional information about other eye diseases detectable by image analysis (ie, glaucoma, age-related macular degeneration, or coat disease).

The conclusion of these results suggests the proposed ARIA system is valuable in an independent or assistive condition and can be useful to increase and improve DR diagnosis, as well as other ophthalmic diseases in the future. However, special attention to the design of an explanatory platform is required for successful implementation of the system.

Acknowledgments

The authors gratefully thank the retina specialists and the ophthalmologists from the APEC Hospital involved in this study for their evaluations of the retina fundus images. This project was carried out thanks to the fellowships received by individual members of the team, including fellowships of the Massachusetts Institute of Technology and the National Council of Science and Technology (CONACYT).

Authors' Contributions

AN conceived and designed the experiments, analyzed data, and contributed to the discussion and review of the paper. DC trained the models, performed the experiments, analyzed data, and contributed to the discussion of the paper. DM contributed to the experimental design, image classification, and discussion of the paper. JE contributed with data analysis and paper writing, including the discussion. HQM, VMC, AA, and AP contributed to various aspects of the paper, including experimental design, machine learning strategies, medical feedback, image evaluations, and the discussion.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Average number of observations according to the International Clinical Diabetic Retinopathy Severity Scales classification, for each set of 15 retina images in each treatment condition: solo, ARIA answer, and ARIA explanation.

[\[DOC File, 58 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).

[\[PDF File \(Adobe PDF File\), 407 KB-Multimedia Appendix 2\]](#)

References

1. Zhang X, Saaddine JB, Chou C, Cotch MF, Cheng YJ, Geiss LS, et al. Prevalence of diabetic retinopathy in the United States, 2005-2008. *JAMA* 2010 Aug 11;304(6):649-656 [[FREE Full text](#)] [doi: [10.1001/jama.2010.1111](https://doi.org/10.1001/jama.2010.1111)] [Medline: [20699456](https://pubmed.ncbi.nlm.nih.gov/20699456/)]
2. SIGMA Type 2 Diabetes Consortium T, Williams AL, Jacobs SBR, Moreno-Macías H, Huerta-Chagoya A, Churchhouse C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 2014 Feb 06;506(7486):97-101 [[FREE Full text](#)] [doi: [10.1038/nature12828](https://doi.org/10.1038/nature12828)] [Medline: [24390345](https://pubmed.ncbi.nlm.nih.gov/24390345/)]
3. Caballero AE. Understanding the Hispanic/Latino patient. *Am J Med* 2011 Oct;124(10 Suppl):S10-S15. [doi: [10.1016/j.amjmed.2011.07.018](https://doi.org/10.1016/j.amjmed.2011.07.018)] [Medline: [21939793](https://pubmed.ncbi.nlm.nih.gov/21939793/)]
4. Secretaría de Salud. 2016. Encuesta Nacional de Salud y Nutrición de Medio Camino 2016 (Ensanut MC 2016). Cuernavaca: INSP; 2016. URL: <https://www.gob.mx/cms/uploads/attachment/file/209093/ENSANUT.pdf> [accessed 2019-11-08]
5. Institute for Health Metrics and Evaluation. 2017. Country profiles (Mexico). URL: <http://www.healthdata.org/results/country-profiles> [accessed 2019-08-24]
6. Barcelo A, Arredondo A, Gordillo-Tobar A, Segovia J, Qiang A. The cost of diabetes in Latin America and the Caribbean in 2015: Evidence for decision and policy makers. *Journal of Global Health* 2017 Dec;7(2):020410 [[FREE Full text](#)] [doi: [10.7189/jogh.07.020410](https://doi.org/10.7189/jogh.07.020410)] [Medline: [29163935](https://pubmed.ncbi.nlm.nih.gov/29163935/)]
7. International Diabetes Federation. 2019. IDF Diabetes Atlas Ninth edition. 2019. URL: https://diabetesatlas.org/upload/resources/material/20200302_133352_2406-IDF-ATLAS-SPAN-BOOK.pdf [accessed 2019-11-15]
8. Gómez EJ. Political party ambitions and type-2 diabetes policy in Brazil and Mexico. *Health Econ Policy Law* 2020 Apr;15(2):261-276. [doi: [10.1017/S1744133118000415](https://doi.org/10.1017/S1744133118000415)] [Medline: [30394254](https://pubmed.ncbi.nlm.nih.gov/30394254/)]
9. Deshpande A, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. *Phys Ther* 2008 Nov;88(11):1254-1264 [[FREE Full text](#)] [doi: [10.2522/ptj.20080020](https://doi.org/10.2522/ptj.20080020)] [Medline: [18801858](https://pubmed.ncbi.nlm.nih.gov/18801858/)]
10. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *The Lancet* 2010 Jul;376(9735):124-136. [doi: [10.1016/S0140-6736\(09\)62124-3](https://doi.org/10.1016/S0140-6736(09)62124-3)]
11. Secretaría de Salud. Diagnóstico y Tratamiento de Retinopatía Diabética. Evidencias y Recomendaciones. Catálogo Maestro. Guías de Práctica Clínica. México: CENETEC; 2011. URL: http://www.cenetec.salud.gob.mx/descargas/gpc/CatalogoMaestro/171_GPC_RETINOPATIA_DIABETICA/Imss_171RR.pdf [accessed 2019-07-19]
12. Jaime CL. RETINOPATÍA DIABÉTICA DESDE LA PREVENCIÓN. INTEGRAR LA PESQUISA EN LOS CENTROS DE DIABETES. *Revista Médica Clínica Las Condes* 2016 Mar;27(2):195-203. [doi: [10.1016/j.rmcl.2016.04.009](https://doi.org/10.1016/j.rmcl.2016.04.009)]
13. Barría VF, Martínez CF. Clinical Practice Guide for Diabetic Retinopathy for Latin America for Ophthalmologists and Healthcare Professionals. 2011. URL: <https://www.iapb.org/wp-content/uploads/2011-Clinical-Practice-Guide-for-DR-for-Latin-America.pdf> [accessed 2020-08-21]
14. Cavan D, Makaroff L, da Rocha Fernandes J, Sylvanowicz M, Ackland P, Conlon J, et al. The Diabetic Retinopathy Barometer Study: Global perspectives on access to and experiences of diabetic retinopathy screening and treatment. *Diabetes Res Clin Pract* 2017 Jul;129:16-24. [doi: [10.1016/j.diabres.2017.03.023](https://doi.org/10.1016/j.diabres.2017.03.023)] [Medline: [28499163](https://pubmed.ncbi.nlm.nih.gov/28499163/)]
15. Jiménez Báez M, Márquez González H, Bárcenas Contreras R, Morales Montoya C, García Espinosa L. EARLY DIAGNOSIS OF DIABETIC RETINOPATHY IN PRIMARY CARE. *Colombia Médica* 2015 Jan 01;46(01):14-18 [[FREE Full text](#)] [doi: [10.25100/cm.v46i1.1681](https://doi.org/10.25100/cm.v46i1.1681)]
16. Carrillo-Alarcón L, Ávila-Pozos R, López LE, Cruz-Castillo R, Ocampo-Torres M, Alcalde-Rabanal J. Projection of Diabetic Patients Retinopathy in Hidalgo State-México, through 2030. *EC Ophthalmology* 2017;5(2):73-80.
17. Thomas RL, Halim S, Gurudas S, Sivaprasad S, Owens D. IDF Diabetes Atlas: A review of studies utilising retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018. *Diabetes Res Clin Pract* 2019 Nov;157:107840. [doi: [10.1016/j.diabres.2019.107840](https://doi.org/10.1016/j.diabres.2019.107840)] [Medline: [31733978](https://pubmed.ncbi.nlm.nih.gov/31733978/)]
18. Hong H, Mújica OJ, Anaya J, Lansingh VC, López E, Silva JC. The Challenge of Universal Eye Health in Latin America: distributive inequality of ophthalmologists in 14 countries. *BMJ Open* 2016 Nov 18;6(11):e012819 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2016-012819](https://doi.org/10.1136/bmjopen-2016-012819)] [Medline: [27864248](https://pubmed.ncbi.nlm.nih.gov/27864248/)]
19. Martínez J, Hernandez-Bogantes E, Wu L. Diabetic retinopathy screening using single-field digital fundus photography at a district level in Costa Rica: a pilot study. *Int Ophthalmol* 2011 Apr;31(2):83-88. [doi: [10.1007/s10792-010-9413-9](https://doi.org/10.1007/s10792-010-9413-9)] [Medline: [21274593](https://pubmed.ncbi.nlm.nih.gov/21274593/)]
20. Salamanca O, Geary A, Suárez N, Benavent S, Gonzalez M. Implementation of a diabetic retinopathy referral network, Peru. *Bull World Health Organ* 2018 Oct 01;96(10):674-681 [[FREE Full text](#)] [doi: [10.2471/BLT.18.212613](https://doi.org/10.2471/BLT.18.212613)] [Medline: [30455515](https://pubmed.ncbi.nlm.nih.gov/30455515/)]

21. Rachapelle S, Legood R, Alavi Y, Lindfield R, Sharma T, Kuper H, et al. The cost-utility of telemedicine to screen for diabetic retinopathy in India. *Ophthalmology* 2013 Mar;120(3):566-573. [doi: [10.1016/j.ophtha.2012.09.002](https://doi.org/10.1016/j.ophtha.2012.09.002)] [Medline: [23211635](https://pubmed.ncbi.nlm.nih.gov/23211635/)]
22. 2016. From \$600 M to \$6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare. 2016. URL: <https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare/> [accessed 2019-12-01]
23. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019 Jan;25(1):30-36. [doi: [10.1038/s41591-018-0307-0](https://doi.org/10.1038/s41591-018-0307-0)] [Medline: [30617336](https://pubmed.ncbi.nlm.nih.gov/30617336/)]
24. Arenas-Cavalli J, Ríos S, Pola M, Donoso R. A Web-based Platform for Automated Diabetic Retinopathy Screening. *Procedia Computer Science* 2015;60:557-563 [FREE Full text] [doi: [10.1016/j.procs.2015.08.179](https://doi.org/10.1016/j.procs.2015.08.179)]
25. Yang Y, Li T, Li W, Wu H, Fan W, Zhang W. Lesion detection and Grading of Diabetic Retinopathy via Two-stages Deep Convolutional Neural Networks. In: *Lecture Notes in Computer Science*. Switzerland: Springer Nature; 2017 Presented at: 20th Medical Image Computing and Computer Assisted Intervention - MICCAI 2017; September 11-13; Quebec City, Canada p. 533-540 URL: https://doi.org/10.1007/978-3-319-66179-7_61 [doi: [10.1007/978-3-319-66179-7_61](https://doi.org/10.1007/978-3-319-66179-7_61)]
26. Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda S, et al. Automated Diabetic Retinopathy Screening and Monitoring Using Retinal Fundus Image Analysis. *Journal of Diabetes Science and Technology* 2016 Feb 16;10(2):254-261 [FREE Full text] [doi: [10.1177/1932296816628546](https://doi.org/10.1177/1932296816628546)] [Medline: [26888972](https://pubmed.ncbi.nlm.nih.gov/26888972/)]
27. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016 Dec 13;316(22):2402-2410. [doi: [10.1001/jama.2016.17216](https://doi.org/10.1001/jama.2016.17216)] [Medline: [27898976](https://pubmed.ncbi.nlm.nih.gov/27898976/)]
28. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, et al. Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. *Ophthalmology* 2017 Mar;124(3):343-351 [FREE Full text] [doi: [10.1016/j.ophtha.2016.11.014](https://doi.org/10.1016/j.ophtha.2016.11.014)] [Medline: [28024825](https://pubmed.ncbi.nlm.nih.gov/28024825/)]
29. Dutz M, Almeida R, Packard T. *The Jobs of Tomorrow: Technology, Productivity, and Prosperity in Latin America and the Caribbean*. Washington DC: World Bank Group; 2018:1-71.
30. Chollet F, others. Keras. 2015. URL: <https://keras.io> [accessed 2018-12-01]
31. Van RG, Drake JF. Python Tutorial Release 2. 2001. URL: <https://docs.python.org/2.0/tut/tut.html> [accessed 2018-12-01]
32. Wilkinson C, Ferris FL, Klein RE, Lee PP, Agardh CD, Davis M, Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003 Sep;110(9):1677-1682. [doi: [10.1016/S0161-6420\(03\)00475-5](https://doi.org/10.1016/S0161-6420(03)00475-5)] [Medline: [13129861](https://pubmed.ncbi.nlm.nih.gov/13129861/)]
33. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1989 Dec;1(4):541-551. [doi: [10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541)]
34. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv* 2015;1409.1556v6:1-14 [FREE Full text]
35. Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *Journal of Diabetes Science and Technology* 2009 May 01;3(3):509-516 [FREE Full text] [doi: [10.1177/193229680900300315](https://doi.org/10.1177/193229680900300315)] [Medline: [20144289](https://pubmed.ncbi.nlm.nih.gov/20144289/)]
36. Voets M, Møllersen K, Bongo L. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PLoS One* 2019;14(6):e0217541 [FREE Full text] [doi: [10.1371/journal.pone.0217541](https://doi.org/10.1371/journal.pone.0217541)] [Medline: [31170223](https://pubmed.ncbi.nlm.nih.gov/31170223/)]
37. Samek W, Montavon G, Binder A, La-puschkin S, Müller KR. Interpreting the Predictions of Complex ML Models by Layer-wise Relevance Propagation. *ArXiv* 2016;1611.08191v1:1-5 [FREE Full text]
38. Bach S, Binder A, Montavon G, Klauschen F, Müller K, Samek W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS One* 2015 Jul;10(7):e0130140 [FREE Full text] [doi: [10.1371/journal.pone.0130140](https://doi.org/10.1371/journal.pone.0130140)] [Medline: [26161953](https://pubmed.ncbi.nlm.nih.gov/26161953/)]
39. Almaatouq A, Becker J, Houghton JP, Paton N, Watts DJ, Whiting ME. Empirica: a virtual lab for high-throughput macro-level experiments. *Behavior Research Methods* 2021 Mar 29;1-14 [FREE Full text] [doi: [10.3758/s13428-020-01535-9](https://doi.org/10.3758/s13428-020-01535-9)] [Medline: [33782900](https://pubmed.ncbi.nlm.nih.gov/33782900/)]
40. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9 edition. *Diabetes Res Clin Pract* 2019 Nov;157:107843 [FREE Full text] [doi: [10.1016/j.diabres.2019.107843](https://doi.org/10.1016/j.diabres.2019.107843)] [Medline: [31518657](https://pubmed.ncbi.nlm.nih.gov/31518657/)]
41. Barquera S, Campos-Nonato I, Aguilar-Salinas C, Lopez-Ridaura R, Arredondo A, Rivera-Dommarco J. Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy. *Global Health* 2013 Feb 02;9:3 [FREE Full text] [doi: [10.1186/1744-8603-9-3](https://doi.org/10.1186/1744-8603-9-3)] [Medline: [23374611](https://pubmed.ncbi.nlm.nih.gov/23374611/)]
42. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit Med* 2019;2:25 [FREE Full text] [doi: [10.1038/s41746-019-0099-8](https://doi.org/10.1038/s41746-019-0099-8)] [Medline: [31304372](https://pubmed.ncbi.nlm.nih.gov/31304372/)]

43. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, et al. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. : Association for Computing Machinery; 2020 Apr Presented at: 2020 CHI Conference on Human Factors in Computing Systems (CHI '20); April 25-30; Honolulu p. 1-12 URL: <https://doi.org/10.1145/3313831.3376718> [doi: [10.1145/3313831.3376718](https://doi.org/10.1145/3313831.3376718)]
44. Nørgaard MF, Grauslund J. Automated Screening for Diabetic Retinopathy - A Systematic Review. *Ophthalmic Res* 2018;60(1):9-17 [FREE Full text] [doi: [10.1159/000486284](https://doi.org/10.1159/000486284)] [Medline: [29339646](https://pubmed.ncbi.nlm.nih.gov/29339646/)]
45. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan G, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. *Eye (Lond)* 2020 Mar;34(3):451-460 [FREE Full text] [doi: [10.1038/s41433-019-0566-0](https://doi.org/10.1038/s41433-019-0566-0)] [Medline: [31488886](https://pubmed.ncbi.nlm.nih.gov/31488886/)]

Abbreviations

AI: artificial intelligence

ARIA: automated retinal image analysis

AUROC: area under the receiver operating characteristic curve

DL: deep learning

DR: diabetic retinopathy

ICDRSS: International Clinical Diabetic Retinopathy Severity Scales

NPDR: nonproliferative diabetic retinopathy

Edited by G Eysenbach; submitted 18.11.20; peer-reviewed by G Lim, Y Li; comments to author 24.02.21; revised version received 12.04.21; accepted 19.05.21; published 26.08.21

Please cite as:

*Noriega A, Meizner D, Camacho D, Enciso J, Quiroz-Mercado H, Morales-Canton V, Almaatouq A, Pentland A
Screening Diabetic Retinopathy Using an Automated Retinal Image Analysis System in Independent and Assistive Use Cases in Mexico:
Randomized Controlled Trial*

JMIR Form Res 2021;5(8):e25290

URL: <https://formative.jmir.org/2021/8/e25290>

doi: [10.2196/25290](https://doi.org/10.2196/25290)

PMID:

©Alejandro Noriega, Daniela Meizner, Dalia Camacho, Jennifer Enciso, Hugo Quiroz-Mercado, Virgilio Morales-Canton, Abdullah Almaatouq, Alex Pentland. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.