Screening Multiple Retinal Diseases on Primary Care Population Using Portable Non-Mydriatic Cameras

Dalia Camacho [1] ⓘ, **Daniela Meizner** [2] ⓘ, **Jennifer Enciso** [1] ⓘ, **Hugo Valdez** [3] ⓘ, **Hugo Quiroz-Mercado** [2] ⓘ, **Abdullah Almaatouq** [4] ⓘ, **and Alejandro Noriega** [1,*] ⓘ

1   Prosperia Salud. Mexico City, Mexico; dalia@prosperia.health (D.C.); enciso@prosperia.health (J.E); anc@prosperia.health (A.N.)

2   Retina Department, Asociación para Evitar la Ceguera en México. Mexico City, Mexico; daniela.meiz@gmail.com (D.M.); hugoquiroz@yahoo.com (H.Q.-M.)

3   Retina and Vitreous Department, Fundación Hospital de Nuestra Señora de la Luz. Mexico City, Mexico; valdezfloreshf@gmail.com (H.V.)

4   Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA; amaatouq@mit.edu (A.A.)

*   Correspondence: anc@prosperia.health (A.N.)

Abstract: This study reports on the development and validation of an automated system for screening four of the top causes of irreversible blindness worldwide: diabetic retinopathy, age-related macular degeneration, diabetic macular edema, and pathological myopia. This study is the first in Latin America to develop models for those retinal diseases, and the first study worldwide to demonstrate high accuracy on screening for such diseases with a focus on primary care populations and using images collected with portable, non-mydriatic cameras. Image recognition models were developed using deep multi-task learning and deep transfer learning approaches. For all four diseases, we achieved sensitivities above 90% for cases with above-mild severity levels, and above 83% for all cases including mild severities; all while maintaining specificity levels above 95%, and ROC-AUCs above 98%. Combining single-disease models, we propose a multi-disease referral strategy that achieves sensitivities above 95% for above-mild cases of all four diseases, while maintaining a PPV of 90% for the presence of retinal lesions. These results are relevant to health systems, as they constitute the first study with ideal properties for conducting multi-disease retinal screening campaigns at scale in low-resource environments: high accuracy, focusing on primary care population, and using portable and cost-efficient devices.

Keywords: retinal screening; diabetic retinopathy; age-related macular degeneration; diabetic macular edema; pathological myopia

Contents

1. Introduction

The leading causes of vision impairment and blindness worldwide are uncorrected refractive error, cataract, glaucoma, age-related macular degeneration (AMD), diabetic retinopathy (DR), corneal oppacities, and trachoma [1]. In Latin America these are also the most common ophthalmic problems [1,2], excluding trachoma [3]. Thus, AMD and DR are the two most common retinal diseases in Latin America and worldwide; and detecting them would substantially help decrease cases of avoidable and irreversible blindness.

Through retinal screening AMD and DR can be detected along with other retinal diseases such as pathological myopia (PM), which is a degenerative process that may occur in patients with high myopia and therefore refractive error; and diabetic macular edema (DME), which is considered within the diabetic eye disease [4].

In the past years, the development of deep learning models to automatically detect multiple retinal and ophthalmic diseases has increased [4–24]. Moreover, datasets such as Retinal Fundus Multi-Disease Image Dataset (RFMiD, India) [25] and Ocular Disease Intelligent Recognition (ODIR-2019, China)[26], which contain labels for multiple diseases, have been made public through challenges. Leading to an increase of works focused on detecting multiple retinal diseases [5,15,17,19,20].

Before using the models in real-life scenarios it is important to evaluate performance on a population as similar as possible to the one in which models are going to be used, for example on screening campaigns. Considering specifically retinal screening programs, a larger impact is expected if screening is done outside of eye-clinics or eye hospitals, since patients have higher opportunities to be diagnosed on earlier stages. The results shown by Lin *et al* [13], demonstrate that the performance of retinal disease detection models varies depending whether the data came from eye hospitals, community hospitals or primary care. Therefore, if the aim is to use machine learning models to detect retinal diseases on screening campaigns and reach a broader population, model performance should be evaluated on data coming from general population or from population found on primary care. Furthermore, base population characteristics vary across regions. However, most works have used data from Asian countries such as China, Korea, and India, [4–18]. Only Arenas-Cavalli *et al.* [21] and Argueta-Santillan *et al.* [22] used Latin American data when evaluating performance of models that detected multiple retinal diseases. Arenas-Cavalli *et al* detected DR and DME in Chilean population, achieving an area under the receiver operating characteristic curve (ROC-AUC) of 0.924. While Argueta-Santillan *et al* considered different ophthalmic diseases, including AMD and DR, for which they achieved ROC-AUC values ranging from 0.572 to 0.866 on Mexican data.

Another aspect to consider in screening programs, is fundus camera portability. Portable fundus cameras are easier to transport outside of eye clinics and hospitals, than desktop cameras. Furthermore, most portable cameras have a field of view (FoV) between 40° and 45° leading to images with sufficient area to diagnose the most relevant retinal diseases. However, few works within the context of multiple disease detection work with portable cameras [16,23].

With the aim of reducing cases of avoidable and irreversible blindness due to retinal diseases in Latin America, the purpose of this work was to develop a referral system that used deep learning models to detect the presence of DR and AMD, which are the most prevalent retinal diseases in the region and worldwide. As well as DME and PM, due to their relation with DR and refractive error, respectively. Moreover, the system included an additional model to detect the presence of important retinal lesions (RL) of other retinal diseases. This system was evaluated considering the ideal characteristics for conducting retinal screening campaigns at scale in Latin America on limited-resource environments,

where having a portable and cost-efficient solution focused on primary care population is key. Thus, the performance of each model and of the referral system was evaluated on a subset of images taken with a hand-held portable fundus camera on primary care population from Latin America. The latter conditions reflect those that are expected on a real-life screening campaign.

1.1. Key contributions of this work

Even though, several works have developed deep learning models to detect retinal diseases on fundus images [4–24], other properties, such as population and camera characteristics, number of images, and performance are also important to consider before implementing a model in real-life screening campaigns. In this section, we review previous works aimed to detect multiple retinal diseases with fundus images on adult population, as well as works detecting at least one retinal disease with fundus images on adult Latin American population.

Table 1 shows the number of papers which fulfill the following characteristics: multiple disease detection; usage of portable cameras; whether the data came from general population or primary care; whether more than 1,000 images were used; and if the ROC-AUC for all the models covering retinal diseases was at least 0.95. The number of papers that included specific detection of DR, DME, AMD, and PM is also shown. Furthermore, combinations of characteristics are included to show the number of papers that simultaneously fulfilled them.

Table 1. Number of publications found that have the following characteristics: multiple disease detection; inclusion of DR, DME, AMD, and PM; use of portable cameras; whether the data came from general population or primary care; whether more than 1,000 images were used; and if the ROC-AUC for all the models included was at least 0.95. Works with Latin American data are [21,22,27–34], and papers from data in other parts of the world are [4–20,23,24].

Characteristics present in related work	Latin America	Rest of the World	This work
Total	10	19	-
Multiple disease detection	2	19	✓
Diabetic Retinopathy (DR)	6	17	✓
Age Macular Degeneration (AMD)	1	15	✓
Diabetic Macular Edema (DME)	1	7	✓
Pathological Myopia (PM)	0	10	✓
Portable Camera	2	2	✓
Primary Care	3	5	✓
Images > 1000	4	17	✓
ROC-AUC > 0.95	4	9	✓
DR & DME	1	7	✓
DR & AMD	1	13	✓
DR & DME & AMD	0	4	✓
DR & AMD & PM	0	10	✓
DR & DME & AMD & PM	0	4	✓
ROC-AUC > 0.95 & Multiple disease detection	0	9	✓
ROC-AUC > 0.95 & Images > 1000	1	9	✓
ROC-AUC > 0.95 & Images > 1000 & Primary Care	1	2	✓
ROC-AUC > 0.95 & Images > 1000 & Portable Camera	0	1	✓
ROC-AUC > 0.95 & Images > 1000 & Primary Care & Portable Camera	0	0	✓
Multiple disease detection & Primary care & Portable Camera	0	1	✓
DR & DME & AMD & PM & Primary Care	0	1	✓
DR & DME & AMD & PM & Portable Camera	0	0	✓
DR & DME & AMD & PM & Primary Care & Portable Camera	0	0	✓
DR & DME & AMD & PM & Primary Care & Portable Camera & ROC-AUC > 0.95 & Images > 1000	0	0	✓

In total we identified 21 works, from 2020 on wards, which developed models for detecting multiple diseases in retina using fundus images [4–24], only two of those were evaluated on Latin American data [21,22]. We also found ten works in total, whose aim was retinal disease detection in adult population, that used Latin American data either for training or evaluation [21,22,27–34].

Even though, several works already consider DR, DME, AMD, and PM detection within the same work [7,12–14], none of those use Latin American data on any part of the training and evaluation process. Moreover, they do not use portable fundus cameras for image acquisition and only Lin *et al* [13] evaluated model performance on primary care data.

In terms of camera portability and primary care data, only the work by Zapata *et al* reported on the development of a model for multiple retinal disease detection [23]. Zapata's models detect AMD with a ROC-AUC of 0.936. However, DR, DME, and PM were not considered.

Our work proposes the development of a system to detect multiple retinal diseases including DR, DME, AMD, and PM. The system has been evaluated on data from primary care population, whose images were taken with portable devices. These characteristics correspond to those of a screening setting. Moreover, our test data comes from Latin American population. Hence, our work is framed on a context that had not been previously addressed.

This work is structured as follows. Section 2 describes details on data acquisition, distribution and grading. This section also includes the description of the methods used for model training. Section 3 shows the most important results, Section 4 presents the discussion, and Section 5 states the most important conclusions.

2. Materials and Methods

To generate the models we annotated over 20,000 images from different sources, where 80% of the data came from Mexican population. For the annotation process, we defined a labeling approach, where fundus images were labeled by the presence of retinal lesions instead of doing it by disease and severity scale.

Section 2.1 presents the data sources and its corresponding characteristics. Section 2.2 defines the criteria established to determine presence and severity stages of DR, DME, AMD, and PM. It also provides details on the grading process. Section 2.3 describes the multi-task learning and transfer learning approaches used to train the classification models, as well as the criteria used for the referral system.

2.1. Data

We used 20,017 retinal fundus images from different data sources and cameras.

From Prosperia's screening program to detect diabetic retinopathy, a subset of 8,168 images with sufficient quality from 4,138 patients were used for this project. These images were taken in primary care clinics, pharmacies, and optical shops from October 2020 to January 2022. Most patients were from Mexico City (46%) and Mexico State (41%), 65% of them were female, and the mean age was 50 years old (SD: 16.7). Their images were taken with two different cameras, 5,547 images from 2,570 patients were taken with the Horus 45° auto focus portable fundus non-mydriatic camera from Jedmed (Horus). The other 2,621 images from 1,568 patients were taken with DRS 45° Non-Mydriatic Digital Fundus Camera from Centervue (DRS).

Additionally, we used another 7,902 retinal images from Mexican population found on primary care clinics and on eye-clinics. These images were taken with three different fundus cameras: 1,865 with ZEISS VISUSCOUT 100 Handheld Fundus Camera (Zeiss); 2,581 images with Forus Digital Non-mydriatic Fundus Camera from 3nethra (Forus); and 3,456 with Pictor Plus Fundus camera from Volk optical (Pictor).

Furthermore, a subset of the EyePACS dataset from Kaggle [35] of 3,947 images was used to increase the number of images in the training set.

The distribution of images by disease and severity grades by camera is shown in Table 2. In total there are 20,017 images, 2,465 with some degree of DR, 2,281 with DME, 1,858 with some degree of AMD, and 4,605 with PM.

Table 2. Image distribution across cameras including number of images per disease and severity grade.

		DRS	Forus	Horus	Pictor	Zeiss	EyePACS	Total
Total	-	**2,621**	**2,581**	**5,547**	**3,456**	**1,865**	**3,947**	**20,017**
Retinal Lesions [1]	-	**524 (20.0%)**	**789 (30.6%)**	**1,273 (22.9%)**	**1,194 (34.5%)**	**614 (32.9%)**	**2,533 (64.2%)**	**6,927 (34.6%)**
Diabetic Retinopathy	Mild	34 (1.3%)	34 (1.3%)	76 (1.4%)	61 (1.8%)	22 (1.2%)	220 (5.6%)	447 (2.2%)
	Moderate	108 (4.1%)	261 (10.1%)	219 (3.9%)	334 (9.7%)	202 (10.8%)	532 (13.5%)	1656 (8.3%)
	Severe	30 (1.1%)	21 (0.8%)	29 (0.5%)	23 (0.7%)	15 (0.8%)	13 (0.3%)	131 (0.7%)
	Proliferative	66 (2.5%)	21 (0.8%)	17 (0.3%)	15 (0.4%)	75 (4.0%)	37 (0.9%)	231 (1.2%)
	Total	**238 (9.1%)**	**337 (13.1%)**	**341 (6.1%)**	**433 (12.5%)**	**314 (16.8%)**	**802 (20.3%)**	**2,465 (12.3%)**
Diabetic Macular Edema	Central	64 (2.4%)	152 (5.9%)	144 (2.6%)	260 (7.5%)	97 (5.2%)	422 (10.7%)	1,139 (5.7%)
	Non-central	121 (4.6%)	96 (3.7%)	99 (1.8%)	156 (4.5%)	151 (8.1%)	519 (13.1%)	1142 (5.7%)
	Total	**185 (7.1%)**	**248 (9.6%)**	**243 (4.4%)**	**416 (12.0%)**	**248 (13.3%)**	**941 (23.8%)**	**2,281 (11.4%)**
Age Macular Degeneration	Mild	88 (3.4%)	88 (3.4%)	358 (6.5%)	197 (5.7%)	48 (2.6%)	651 (16.5%)	1430 (7.1%)
	Intermediate	39 (1.5%)	62 (2.4%)	163 (2.9%)	63 (1.8%)	23 (1.2%)	51 (1.3%)	401 (2.0%)
	Advanced	8 (0.3%)	1 (0.0%)	9 (0.2%)	1 (0.0%)	5 (0.3%)	3 (0.1%)	27 (0.1%)
	Total	**135 (5.2%)**	**151 (5.9%)**	**530 (9.6%)**	**261 (7.6%)**	**76 (4.1%)**	**705 (17.9%)**	**1,858 (9.3%)**
Pathological Myopia	1	150 (5.7%)	507 (19.6%)	1068 (19.3%)	965 (27.9%)	255 (13.7%)	808 (20.5%)	3753 (18.7%)
	2	42 (1.6%)	165 (6.4%)	144 (2.6%)	169 (4.9%)	30 (1.6%)	71 (1.8%)	621 (3.1%)
	3	14 (0.5%)	18 (0.7%)	24 (0.4%)	18 (0.5%)	14 (0.8%)	15 (0.4%)	103 (0.5%)
	4	8 (0.3%)	17 (0.7%)	22 (0.4%)	11 (0.3%)	5 (0.3%)	12 (0.3%)	75 (0.4%)
	PLUS	2 (0.1%)	13 (0.5%)	13 (0.2%)	13 (0.4%)	5 (0.3%)	7 (0.2%)	53 (0.3%)
	Total	**216 (8.2%)**	**720 (27.9%)**	**1,271 (22.9%)**	**1,176 (34.0%)**	**309 (16.6%)**	**913 (23.1%)**	**4,605 (23.0%)**

[1] This excludes cases with tessellated fundus only, since having a tessellated fundus only is not pathological, and is unlikely to progress to more severe stages [36].

2.2. Image grading

Here we define the labeling methodology, including the criteria for quality grading, the process of labeling by lesions; the guidelines used to define the main diseases of interest (DR, DME, AMD, and PM); and additional details on the grading process.

2.2.1. Quality

To determine which images had sufficient quality we had two main criteria. The first criterion was the percentage of the image in which the retina was visible. We determined a threshold of 70%, if less than 70% of the retinal image was visible, then the image had insufficient quality. The other criterion, was in terms of definition. An image was classified with insufficient definition if the visible part of the retina was blurry, and it was not possible to determine whether or not a lesion was present. If visibility was less than 70% or the image definition was insufficient, then the image had insufficient quality.

2.2.2. Disease labeling

To have more homogeneous labels in terms of presence and severity of DR, DME, AMD, and PM, we proposed a labeling strategy in which all lesions found on the retina

were annotated. Afterwards, a set of rules was used to determine the presence of retinal diseases as well as the severity grades.

The lesions that were annotated are shown in Table A1 on Appendix A. The severity scale used for DR was the International Clinical Diabetic Retinopathy and Diabetic Macular Edema Disease Severity Scales (ICDR) [37]. For DME we used the presence of hard exudates as a proxy for DME presence [38,39]. The grading for AMD was done according to the criteria of Flaxel *et al* [40]. Finally, to classify pathological myopia we used the categorization by Ohno-Matsui *et al* [41], images with tessellated fundus only were not considered within the pathological myopia model, since in most cases it is a stable condition and only 13.4% progress to more severe stages [36]. Appendix B further describes the grading scales used.

2.2.3. Grading process

Before labeling the presence of lesions, quality was graded according to the criteria defined on Section 2.2.1.

Only images with sufficient quality were labeled by either a retina expert or an ophthalmologist trained by the retina expert. They annotated the presence of lesions in each image according to Table A1 on appendix A. Agreement between labelers was evaluated on a set containing 500 images. At a disease level average percentage of agreement was 90.3% and the average Cohen's kappa was 0.695, which corresponded to a substantial agreement [42]. Since a substantial agreement was achieved, the rest of the images were divided between the ophthalmologist and the retina expert.

Labelers were shown simultaneously the original image and an enhanced version. This was done to ease the detection of lesions. As an example, Figure 1 shows both the original and the enhanced version, microaneurysms are seen more clearly in the latter.
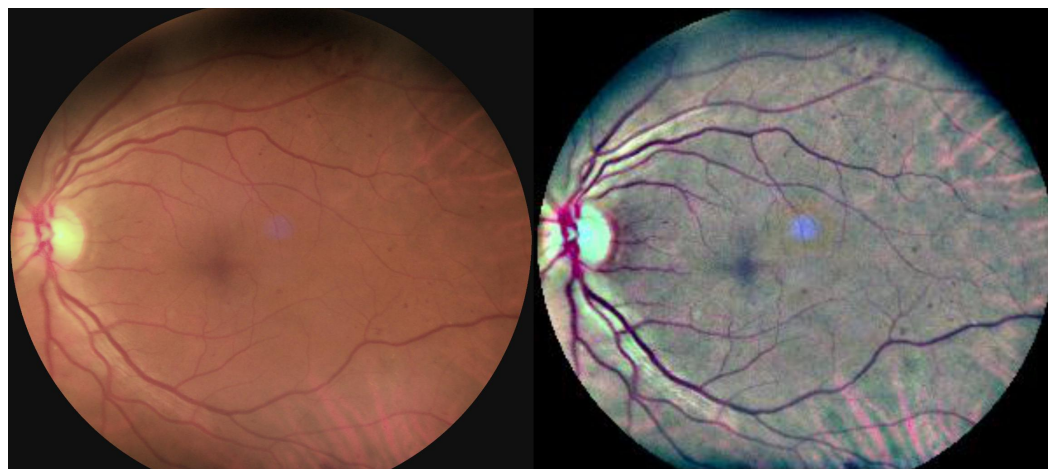

Figure 1. Example of an image shown to labelers which includes the original image and an enhanced version.

2.3. Model training

Here we describe the model architecture as well as data processing and separation. We also show how we constructed the patient referral decision by using the outputs of DR, DME, AMD, PM, and RL models.

2.3.1. Architectures and training

The models trained for this project were convolutional neural networks, where the convolutional base corresponded to the the Inception V3 architecture [43]. We used a transfer learning approach, the initial weights for the convolutional base were those from ImageNet, these were fine tuned by training a multi-task model that detected the presence of each lesion. The resulting weights from the lesions model were then frozen for the models that detected the presence of each pathology.

The lesions model was a multi-task network that detected the presence of each lesion. For each lesion the model had three dense layers of 64, 32 and one units, where the output unit had a sigmoid activation.

The models for DR, DME, AMD, PM and RL used the resulting convolutional base from the lesions model and only the upper layers were trained. The upper layers of those models had a layer of 2,048 units and the output layer of one unit with sigmoid activation.

Figure 2 shows a diagram of the transfer learning steps and the architectures for the models trained in this work.
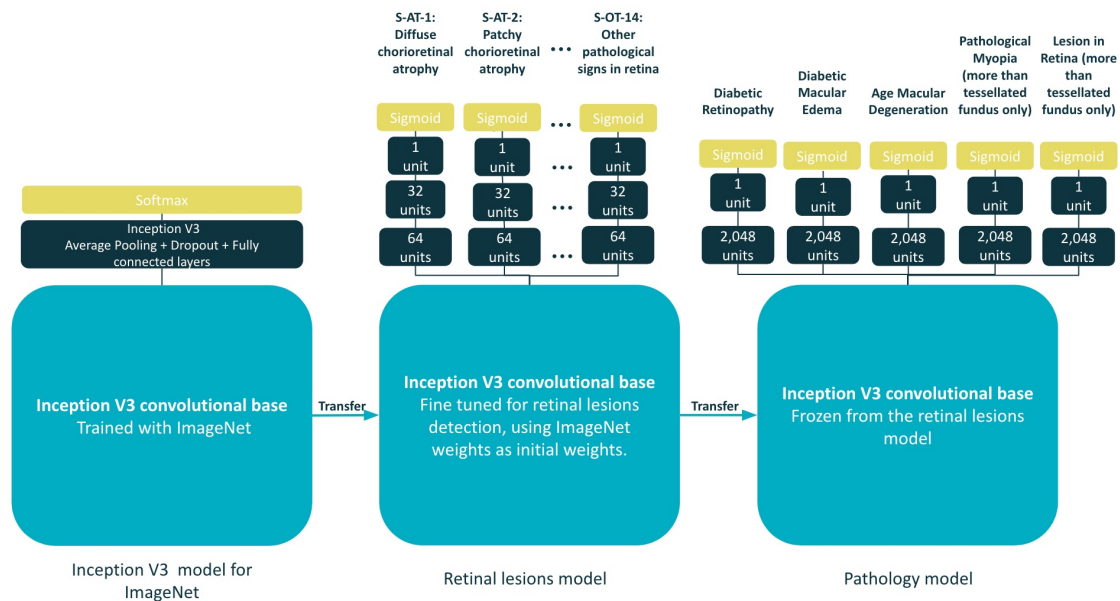

Figure 2. Model architectures for the retinal lesions model and for DR, DME, AMD, PM and RL models. The transfer learning process of weights in convolutional layers is also shown.

2.3.2. Data Preprocessing

The pre-processing consisted on resizing to 299×299, and on image enhancement using the method suggested by Graham [44].

2.3.3. Data Separation

Some considerations were taken when dividing data into train, validation, and test sets. First, to avoid data filtration, all images taken from a unique patient were assigned to the same subset.

Since our goal was to have a good performance mainly on Latin American data, all data from Mexican population was divided as follows: 70% for training, 15% for validation, and 15% for testing. All images from the EyePACS dataset were used for training only. The resulting data separation is shown in Table 3.

Table 3. Number of images from each camera in train, validation, and test sets.

	DRS	Forus	Horus	Pictor	Zeiss	EyePACS
Train	1,844	1,819	3,913	2,423	1,292	3,947
Valdiation	389	361	793	516	275	0
Test	388	401	841	517	298	0

2.3.4. Referral system

In a real-life scenario all models would be used to determine whether a patient needs to be referred to ophthalmic care. We defined a threshold of 0.5 for DR, DME, AMD, and

PM models, and a threshold 0.9 for RL model. If the threshold for any model was surpassed, then the patient would be referred. Figure 3 shows how the reference system works.
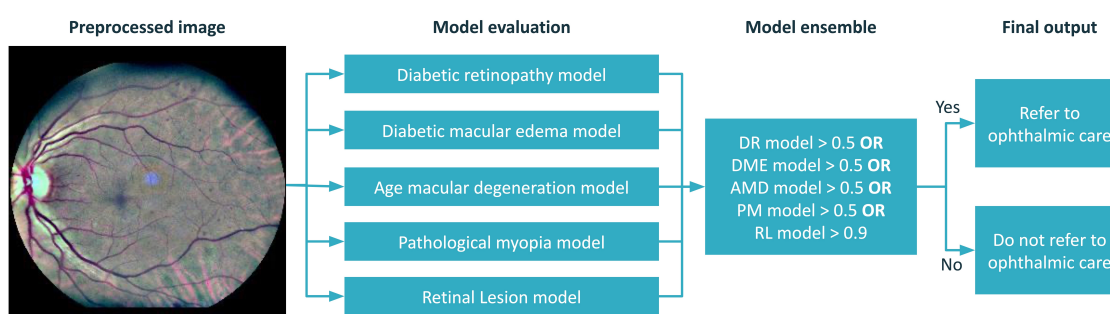

Figure 3. Pathway of the reference system including model evaluation and referral decision.

3. Results

All five models were evaluated on different subsets of the test set, according to whether images came from patients in primary care and if they had been taken with a portable camera. The first test set contained all 2,245 test images, the second set contained 1,229 images from patients in primary care, and the final test set contained 841 images from patients in primary care, that were taken with a portable camera.

When evaluating sensitivity, we considered two levels of detection for DR, AMD and PM; a general level that includes all stages[1], and a level that considers cases above mild stages[2].

Table 4 shows the performance for each model on the different test sets, the metrics included are ROC-AUC, accuracy, sensitivity for all cases, sensitivity for cases above mild stages, specificity and positive predictive value (PPV).

---

[1]   We exclude detection of cases that only have tessellated fundus.
[2]   For DR cases above mild stages exclude mild DR; for AMD we exclude cases with mild AMD that do not have retinal pigmentary epithelium changes; and for PM we exclude those cases on stages 1 and 2.

Table 4. Metrics for each model evaluated on different subsets, the first includes all test data, the second includes only data acquired in primary care, finally, the last subset contains images from primary care taken with a portable camera. All images used for evaluation correspond to Mexican data.

	ROC-AUC	Accuracy	Sensitivity including mild	Sensitivity above mild stage	Specificity	PPV
All test data (2,245)						
Retinal Lesions	0.953	92%	83%	90%	95%	88%
Diabetic Retinopathy	0.970	95%	87%	92%	96%	70%
Diabetic Macular Edema	0.985	96%	94%	94%	96%	68%
Age Macular Degeneration	0.970	94%	86%	91%	95%	62%
Pathological Myopia	0.973	95%	80%	91%	96%	58%
Primary Care (1,229)						
Retinal Lesions	0.967	94%	87%	94%	96%	87%
Diabetic Retinopathy	0.973	95%	85%	90%	95%	59%
Diabetic Macular Edema	0.991	96%	94%	94%	96%	53%
Age Macular Degeneration	0.977	94%	86%	94%	95%	65%
Pathological Myopia	0.986	95%	90%	100%	95%	53%
Portable Camera & Primary Care (841)						
Retinal Lesions	0.962	93%	86%	93%	95%	86%
Diabetic Retinopathy	0.983	95%	85%	90%	95%	54%
Diabetic Macular Edema	0.993	96%	96%	96%	96%	46%
Age Macular Degeneration	0.982	95%	85%	100%	96%	70%
Pathological Myopia	0.982	95%	83%	100%	96%	56%

The ROC-AUC surpassed 0.95 for all models in all test sets. The minimum ROC-AUC was 0.953 which corresponded to the detection of RL evaluated on all test data. In general ROC-AUC values for RL were lower than those for specific diseases on all test subsets. In contrast, the best ROC-AUC across all test sets was the one related to DME detection.

The values reported were constrained above a 95% specificity. All sensitivities had values of at least 80% for DR, DME, AMD, PM, and LR for all test subsets. Previously, it had been stated that for DR screening values for sensitivity and specificity should be at least 80% sensitivity and 95% specificity [45]. Therefore, this minimum benchmarks are surpassed for all diseases even considering mild cases. Moreover, for cases that have above mild stages the sensitivity is at least 90%.

Considering the performance on primary care population using portable cameras, we achieved ROC-AUC values over 0.982 for DR, DME, AMD, and PM, and of 0.962 for any retinal disease detection. Given a specificity of at least 95%, sensitivities for each of the five models was at least 83%. Moreover, for cases that have above mild stages, where ophthalmic care becomes more necessary, the sensitivity was at least 90% .

The receiver operating characteristic curves for all test sets are shown in Figure 4, where all possible combinations of sensitivity and (1-specificity) values are shown for DR, DME, AMD, PM and RL on the different test sets.
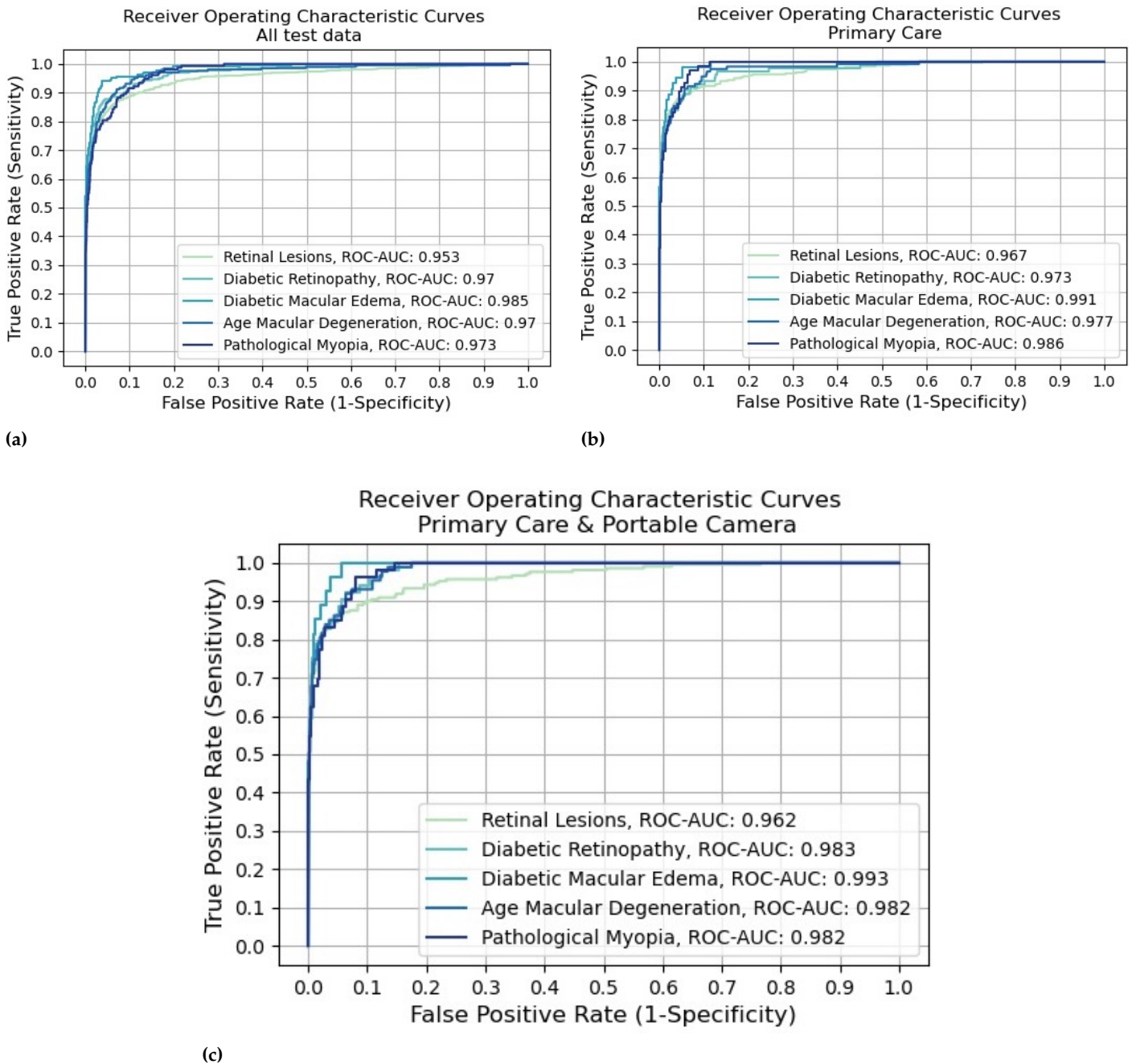
Figure 4. ROC curves for RL, DR, DME, AMD, PM for all test sets. (a) All test data; (b) Primary Care; (c) Portable Camera & Primary Care.

In contrast to results shown in Table 4, Table 5 shows the expected results of the referral decision, where the sensitivities for each disease of interest are shown, as well as the sensitivity of retinal diseases not specifically considered within this work. The overall specificity and PPV values are also shown.

Table 5. Results of all models working simultaneously with the given thresholds

	Threshold	Sensitivity including mild	Sensitivity above mild stage	Specificity	PPV
All test data (2,245)					
Retinal lesions	0.9	80%	88%	96%	90%
Diabetic Retinopathy	0.5	90%	95%	-	-
Diabetic Macular Edema	0.5	93%	93%	-	-
Age Macular Degeneration	0.5	83%	90%	-	-
Pathological Myopia	0.5	68%	91%	-	-
Other retinal diseases	-	80%	80%	-	-
Primary Care (1,229)					
Retinal lesions	0.9	81%	91%	98%	92%
Diabetic Retinopathy	0.5	92%	97%	-	-
Diabetic Macular Edema	0.5	96%	96%	-	-
Age Macular Degeneration	0.5	85%	94%	-	-
Pathological Myopia	0.5	65%	89%	-	-
Other retinal diseases	-	81%	81%	-	-
Portable Camera & Primary Care (841)					
Retinal lesions	0.9	82%	93%	97%	90%
Diabetic Retinopathy	0.5	94%	95%	-	-
Diabetic Macular Edema	0.5	100%	100%	-	-
Age Macular Degeneration	0.5	87%	100%	-	-
Pathological Myopia	0.5	66%	100%	-	-
Other retinal diseases	-	74%	74%	-	-

For all test sets PPV was at least 90% and specificity was 96%. Sensitivity of all retinal lesions was 80% and for those cases above mild stages sensitivity was at least 88%. The least sensitivity was for PM that included stage 2, where sensitivity was as low as 66%, however considering cases in stage 3 and above, sensitivity ranged from 89% to 100%. Thus, the most severe cases are detected correctly. For all DR cases sensitivity was at least 90% even when including mild cases, and was at least 95% for cases in moderate, severe or proliferative stages. Sensitivity of DME was also above 90%. Sensitivity of AMD was above 80% for all cases and above 90% for intermediate and advanced stages and for those in mild stages that have RPE changes.

Evaluating on the subset where data came from primary care population and portable cameras, we achieved a specificity of 97% and a PPV of 90%, which means that 90% of the cases that would be referred to ophthalmic care would have a retinal disease. Also, the sensitivity for the four main diseases for cases above mild stages was at least 95%, hence the models are properly detecting cases above mild stages. However, those cases with mild stages are not neglected, since detection for all cases of DR and AMD have a sensitivity of 87%. Sensitivity is lower on detecting mild cases of pathological myopia, however this is also related to the difficulty to differentiate mild cases of diffuse chorioretinal atrophy with tessellated fundus only.

4. Discussion

In this work we developed models to detect DR, DME, AMD and PM as well as the general presence of retinal diseases using primarily Latin American data and we evaluated them on different subsets, including one that considers only data from Latin American population on primary care, whose images were acquired with portable non-mydriatic fundus cameras. This work is the first to consider those four retinal diseases in a Latin American context, furthermore, it is the first to detect them on data from primary care population taken with portable fundus cameras. These can be seen on Table 1.

Moreover, all models have a ROC-AUC above 0.95 on all test sets. These models can be used for retinal disease detection in a real-life scenario for all outputs including DR,

DME, AMD, PM, and RL. Since, the resulting sensitivities and specificities surpass the minimum benchmarks of 80% sensitivity and 95% specificity established in [45].

The performance obtained on primary care data taken with portable cameras for DR, DME, AMD, and PM had ROC-AUC values of at least 0.982, achieving a sensitivity of at least 83% and specificity values of at least 95%. Furthermore, excluding mild cases, the models achieved a sensitivity of at least 90%.

On a real-life scenario all models would interact jointly to refer patients to ophthalmic care, evaluating the resulting referrals, we established thresholds of 0.5 for the four diseases and 0.9 for general retinal lesions. With these thresholds, a PPV of 90% would be achieved, this PPV value means that 90% of patients that arrive to ophthalmic care have a retinal disease. The performance of referral in terms of sensitivity guarantees that 95% of patients with cases of moderate, severe, and proliferative DR would be referred. Also, all cases of DME, of mild AMD with EPR changes, or of intermediate and advanced AMD, and of PM over stage 3 would be referred to ophthalmic care. Additionally, 74% of patients that have other retinal diseases would be properly detected. Hence, using this referral strategy would lead to retinal screenings with high sensitivity on cases above mild stages (above 95%) for DR, DME, AMD, and PM, an overall high specificity (97%) and an overall high PPV (90%).

The previous metrics correspond to the evaluation on the subset containing primary care data taken with a portable camera, therefore, the models can be used in retinal screening programs with the prior characteristics. Thus, the models developed in this work can be used as a tool to detect retinal diseases and help prevent avoidable blindness.

Future work may include efforts to detect other retinal diseases such as epirretinal membranes, retinal detachment, and macular holes. In terms of DR, AMD, and PM future work may include differentiation between severity stages, which may be useful to guide patients' care.

5. Conclusions

In this work we developed models that detect DR, DME, AMD, and PM with sensitivity over 90% for cases above mild severity stages and of at least 83% for all cases, while maintaining a specificity of at least 95% evaluated on Latin American population in primary care, whose images were taken with a portable camera. No previous works developed and evaluated models for those four diseases simultaneously for those conditions.

Moreover, we evaluated models jointly, as they would be used on a real life-scenario, and we were able to define thresholds that guarantee a PPV of 90% in terms of presence of retinal disease and a specificity of 97% achieving high sensitivities of at least 95% for the four main retinal diseases on cases above mild severity stage.

Given the previous results these models can be used for screening programs on primary care population using portable cameras to help detect retinal diseases and prevent avoidable and irreversible blindness due to the most common retinal diseases.

References

1. Flaxman, S.R.; Bourne, R.R.; Resnikoff, S.; Ackland, P.; Braithwaite, T.; Cicinelli, M.V.; Das, A.; Jonas, J.B.; Keeffe, J.; Kempen, J.H.; et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health* **2017**, *5*, e1221–e1234.
2. de Salud, S., 2020.
3. Mitchell, C. Paho/WHO: Mexico eliminates trachoma, leading infectious cause of blindness, 2017.
4. Sarki, R.; Ahmed, K.; Wang, H.; Zhang, Y. Automated detection of mild and multi-class diabetic eye diseases using deep learning. *Health Information Science and Systems* **2020**, *8*, 1–9.
5. Dabholkar, A.J.; Shah, A.; Mehta, P. Deep Ensemble Model for Retinal Diseases Detection and Classification.
6. Hong, J.; Liu, X.; Guo, Y.; Gu, H.; Gu, L.; Xu, J.; Lu, Y.; Sun, X.; Ye, Z.; Liu, J.; et al. A novel hierarchical deep learning framework for diagnosing multiple visual impairment diseases in the clinical environment. *Frontiers in Medicine* **2021**, *8*.
7. Huang, J.H.; Yang, C.H.H.; Liu, F.; Tian, M.; Liu, Y.C.; Wu, T.W.; Lin, I.; Wang, K.; Morikawa, H.; Chang, H.; et al. Deepopht: medical report generation for retinal images via deep models and visual explanation. Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021, pp. 2442–2452.

8. Kim, K.M.; Heo, T.Y.; Kim, A.; Kim, J.; Han, K.J.; Yun, J.; Min, J.K. Development of a fundus image-based deep learning diagnostic tool for various retinal diseases. *Journal of Personalized Medicine* **2021**, *11*, 321.

9. Lee, J.; Lee, J.; Cho, S.; Song, J.; Lee, M.; Kim, S.H.; Lee, J.Y.; Shin, D.H.; Kim, J.M.; Bae, J.H.; et al. Development of decision support software for deep learning-based automated retinal disease screening using relatively limited fundus photograph data. *Electronics* **2021**, *10*, 163.

10. Li, F.; Wang, Y.; Xu, T.; Dong, L.; Yan, L.; Jiang, M.; Zhang, X.; Jiang, H.; Wu, Z.; Zou, H. Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. *Eye* **2021**, pp. 1–9.

11. Li, B.; Chen, H.; Zhang, B.; Yuan, M.; Jin, X.; Lei, B.; Xu, J.; Gu, W.; Wong, D.C.S.; He, X.; et al. Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography. *British Journal of Ophthalmology* **2021**.

12. Li, X.; Zhou, Y.; Wang, J.; Lin, H.; Zhao, J.; Ding, D.; Yu, W.; Chen, Y. Multi-Modal Multi-Instance Learning for Retinal Disease Recognition. *arXiv preprint arXiv:2109.12307* **2021**.

13. Lin, D.; Xiong, J.; Liu, C.; Zhao, L.; Li, Z.; Yu, S.; Wu, X.; Ge, Z.; Hu, X.; Wang, B.; et al. Application of Comprehensive Artificial intelligence Retinal Expert (CARE) system: a national real-world evidence study. *The Lancet Digital Health* **2021**, *3*, e486–e495.

14. Cen, L.P.; Ji, J.; Lin, J.W.; Ju, S.T.; Lin, H.J.; Li, T.P.; Wang, Y.; Yang, J.F.; Liu, Y.F.; Tan, S.; et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nature communications* **2021**, *12*, 1–13.

15. Luo, X.; Li, J.; Chen, M.; Yang, X.; Li, X. Ophthalmic disease detection via deep learning with a novel mixture loss function. *IEEE Journal of Biomedical and Health Informatics* **2021**, *25*, 3332–3339.

16. Son, J.; Shin, J.Y.; Kim, H.D.; Jung, K.H.; Park, K.H.; Park, S.J. Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. *Ophthalmology* **2020**, *127*, 85–94.

17. Li, N.; Li, T.; Hu, C.; Wang, K.; Kang, H. A benchmark of ocular disease intelligent recognition: one shot for multi-disease detection. International Symposium on Benchmarking, Measuring and Optimization. Springer, 2020, pp. 177–193.

18. Nazir, T.; Nawaz, M.; Rashid, J.; Mahum, R.; Masood, M.; Mehmood, A.; Ali, F.; Kim, J.; Kwon, H.Y.; Hussain, A. Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model. *Sensors* **2021**, *21*, 5283.

19. Wang, J.; Yang, L.; Huo, Z.; He, W.; Luo, J. Multi-label classification of fundus images with efficientnet. *IEEE Access* **2020**, *8*, 212499–212508.

20. Müller, D.; Soto-Rey, I.; Kramer, F. Multi-disease detection in retinal imaging based on ensembling heterogeneous deep learning models. In *German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal*; IOS Press, 2021; pp. 23–31.

21. Arenas-Cavalli, J.T.; Abarca, I.; Rojas-Contreras, M.; Bernuy, F.; Donoso, R. Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. *Eye* **2022**, *36*, 78–85.

22. Argueta-Santillan, M.; Campos-Castolo, E.M.; Ángel Méndez-Lucero, M.; Lima-Sánchez, D.N.; Fabricio Urbina-González, J.; Cerón-Solís, O.; Alayola-Sansores, A.; Fajardo-Dolci, G. Use of artificial intelligence to evaluate the detection of retinal alterations as a screening test in Mexican patients. *International Journal of Combinatorial Optimization Problems & Informatics* **2021**, *12*.

23. Zapata, M.A.; Royo-Fibla, D.; Font, O.; Vela, J.I.; Marcantonio, I.; Moya-Sánchez, E.U.; Sánchez-Pérez, A.; Garcia-Gasulla, D.; Cortés, U.; Ayguadé, E.; et al. Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma. *Clinical Ophthalmology (Auckland, NZ)* **2020**, *14*, 419.

24. Triwijoyo, B.K.; Sabarguna, B.S.; Budiharto, W.; Abdurachman, E. Deep learning approach for classification of eye diseases based on color fundus images. In *Diabetes and Fundus OCT*; Elsevier, 2020; pp. 25–57.

25. Pachade, S.; Porwal, P.; Thulkar, D.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Giancardo, L.; Quellec, G.; Mériaudeau, F. Retinal fundus multi-disease image dataset (rfmid): a dataset for multi-disease detection research. *Data* **2021**, *6*, 14.

26. Ocular Disease Intelligent Recognition ODIR-5K.

27. Gonzalez-Briceno, G.; Sanchez, A.; Ortega-Cisneros, S.; Contreras, M.S.G.; Diaz, G.A.P.; Moya-Sanchez, E.U. Artificial intelligence-based referral system for patients with diabetic retinopathy. *Computer* **2020**, *53*, 77–87.

28. Rogers, T.W.; Gonzalez-Bueno, J.; Garcia Franco, R.; Lopez Star, E.; Mendez Marin, D.; Vassallo, J.; Lansingh, V.; Trikha, S.; Jaccard, N. Evaluation of an AI system for the detection of diabetic retinopathy from images captured with a handheld portable fundus camera: the MAILOR AI study. *Eye* **2021**, *35*, 632–638.

29. Valdez-Rodríguez, J.E.; Felipe-Riverón, E.M.; Calvo, H. Optic Disc Preprocessing for Reliable Glaucoma Detection in Small Datasets. *Mathematics* **2021**, *9*, 2237.

30. Alves, S.S.A.; Matos, A.G.; Almeida, J.S.; Benevides, C.A.; Cunha, C.C.H.; Santiago, R.V.C.; Pereira, R.F.; Reboucas Filho, P.P. A new strategy for the detection of diabetic retinopathy using a smartphone app and machine learning methods embedded on cloud computer. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2020, pp. 542–545.

31. Parra, R.; Ojeda, V.; Vázquez Noguera, J.L.; García-Torres, M.; Mello-Román, J.C.; Villalba, C.; Facon, J.; Divina, F.; Cardozo, O.; Castillo, V.E.; et al. A Trust-Based Methodology to Evaluate Deep Learning Models for Automatic Diagnosis of Ocular Toxoplasmosis from Fundus Images. *Diagnostics* **2021**, *11*, 1951.

32. Abeyrathna, D.; Subramaniam, M.; Chundi, P.; Hasanreisoglu, M.; Halim, M.S.; Ozdal, P.C.; Nguyen, Q. Directed Fine Tuning Using Feature Clustering for Instance Segmentation of Toxoplasmosis Fundus Images. 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE, 2020, pp. 767–772.

33. Carrillo, J.; Bautista, L.; Villamizar, J.; Rueda, J.; Sanchez, M.; et al. Glaucoma detection using fundus images of the eye. 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA). IEEE, 2019, pp. 1–4.

34. Noriega, A.; Meizner, D.; Camacho, D.; Enciso, J.; Quiroz-Mercado, H.; Morales-Canton, V.; Almaatouq, A.; Pentland, A.; et al. Screening Diabetic Retinopathy Using an Automated Retinal Image Analysis System in Independent and Assistive Use Cases in Mexico: Randomized Controlled Trial. *JMIR formative research* **2021**, *5*, e25290.

35. Diabetic retinopathy detection, 2015.

36. Ohno-Matsui, K. Myopic chorioretinal atrophy. In *Pathologic myopia*; Springer, 2014; pp. 187–209.

37. Wilkinson, C.P.; Ferris, F.L.; Klein, R.E.; Lee, P.P.; Agardh, C.D.; Davis, M.; Dills, D.; Kampik, A.; Pararajasegaram, R.; Verdaguer, J.T. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* **2003**, *110*, 1677–82.

38. Wong, T.Y.; Sun, J.; Kawasaki, R.; Ruamviboonsuk, P.; Gupta, N.; Lansingh, V.C.; Maia, M.; Mathenge, W.; Moreker, S.; Muqit, M.M.K.; et al. Guidelines on Diabetic Eye Care: The International Council of Ophthalmology Recommendations for Screening, Follow-up, Referral, and Treatment Based on Resource Settings. *Ophthalmology* **2018**, *125*, 1608–1622.

39. Litvin, T.V.; Ozawa, G.Y.; Bresnick, G.H.; Cuadros, J.A.; Muller, M.S.; Elsner, A.E.; Gast, T.J. Utility of hard exudates for the screening of macular edema. *Optom Vis Sci* **2014**, *91*, 370–375.

40. Flaxel, C.J.; Adelman, R.A.; Bailey, S.T.; Fawzi, A.; Lim, J.I.; Vemulakonda, G.A.; Ying, G.S. Age-Related Macular Degeneration Preferred Practice Pattern®. *Ophthalmology* **2020**, *127*, P1–P65.

41. Ohno-Matsui, K.; Kawasaki, R.; Jonas, J.B.; Cheung, C.M.; Saw, S.M.; Verhoeven, V.J.; Klaver, C.C.; Moriyama, M.; Shinohara, K.; Kawasaki, Y.; et al. International photographic classification and grading system for myopic maculopathy. *Am J Ophthalmol* **2015**, *159*, 877–883.

42. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *biometrics* **1977**, pp. 159–174.

43. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

44. Graham, B. Kaggle diabetic retinopathy detection competition report. *University of Warwick* **2015**.

45. Scanlon, P.H. The English national screening programme for diabetic retinopathy 2003–2016. *Acta diabetologica* **2017**, *54*, 515–525.