




Validation of an autonomous artificial intelligence–based diagnostic system for holistic maculopathy screening in a routine occupational health checkup context

Octavi Font¹ · Jordina Torrents-Barrena² · Dídac Royo¹ · Sandra Banderas García^{3,4}  · Javier Zarranz-Ventura^{5,6} · Anniken Bures^{1,7} · Cecilia Salinas^{1,7} · Miguel Ángel Zapata^{1,4}

Received: 4 June 2021 / Revised: 15 March 2022 / Accepted: 31 March 2022

© The Author(s) 2022

Abstract

Purpose This study aims to evaluate the ability of an autonomous artificial intelligence (AI) system for detection of the most common central retinal pathologies in fundus photography.

Methods Retrospective diagnostic test evaluation on a raw dataset of 5918 images (2839 individuals) evaluated with non-mydratric cameras during routine occupational health checkups. Three camera models were employed: Optomed Aurora (field of view — FOV 50°, 88% of the dataset), ZEISS VISUSCOUT 100 (FOV 40°, 9%), and Optomed SmartScope M5 (FOV 40°, 3%). Image acquisition took 2 min per patient. Ground truth for each image of the dataset was determined by 2 masked retina specialists, and disagreements were resolved by a 3rd retina specialist. The specific pathologies considered for evaluation were “diabetic retinopathy” (DR), “Age-related macular degeneration” (AMD), “glaucomatous optic neuropathy” (GON), and “Nevus.” Images with maculopathy signs that did not match the described taxonomy were classified as “Other.”

Results The combination of algorithms to detect any abnormalities had an area under the curve (AUC) of 0.963 with a sensitivity of 92.9% and a specificity of 86.8%. The algorithms individually obtained are as follows: AMD AUC 0.980 (sensitivity 93.8%; specificity 95.7%), DR AUC 0.950 (sensitivity 81.1%; specificity 94.8%), GON AUC 0.889 (sensitivity 53.6% specificity 95.7%), Nevus AUC 0.931 (sensitivity 86.7%; specificity 90.7%).

Conclusion Our holistic AI approach reaches high diagnostic accuracy at simultaneous detection of DR, AMD, and Nevus. The integration of pathology-specific algorithms permits higher sensitivities with minimal impact on its specificity. It also reduces the risk of missing incidental findings. Deep learning may facilitate wider screenings of eye diseases.

Keywords Artificial intelligence · Screening · Retinography · Diabetic retinopathy · Age-related macular degeneration

✉ Sandra Banderas García
sandrabanderasgarcia@gmail.com

¹ Optretina Image Reading Team, Barcelona, Spain

² BCN MedTech, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

³ Facultat de Cirurgia i Ciències Morfològiques, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain

⁴ Ophthalmology Department Hospital Vall d’Hebron, Barcelona, Spain

⁵ Institut Clinic of Ophthalmology (ICOF), Hospital Clinic, Barcelona, Spain

⁶ Institut d’Investigacions Biomediques August Pi I Sunyer (IDIBAPS), Barcelona, Spain

⁷ Instituto de Microcirugía Ocular (IMO), Barcelona, Spain

Key messages

- High diagnostic accuracy in artificial intelligence (AI) system for one pathology in ophthalmology has been achieved but still to be acquired for screening purposes.
- A combination algorithm made of several specific-pathology ones reached high sensitivity and specificity for diabetic retinopathy, age macular degeneration, glaucomatous optic neuropathy and nevus.
- This holistic system could successfully be used in large-population screening programs at working-age.

Introduction

In 2010, 65% of those afflicted by blindness worldwide (32.4 million total) and 76% of those with moderate or severe vision impairment (191 million) had a preventable or treatable cause [1]. The causes vary among regions, but the trend since 1990 shows a decreased incidence due to cataract or refractive errors while age macular degeneration (AMD), glaucoma, and diabetic retinopathy (DR) are on the rise [1]. In developed countries, this trend is even more pronounced: AMD is the leading cause of blindness in people aged 75 years and older [2], whereas DR is the most frequent cause of preventable blindness in the working-age population (adults between 20 and 74 years old) [3]. Future projections do not show any sign of these diseases slowing down either. AMD will affect 288 million people in 2040 [4], glaucoma will impact 111.8 million in 2040 [5], and 191 million people will suffer DR by 2030 [6]. These diseases are treatable with good outcomes if detected early in the course of the disease [7–10], but they often are not symptomatic until late stages of their development. Thus, it is essential to have good screening systems for a timely diagnosis.

Running screening programs at large scale is costly, being the relatively high fixed costs per equipment the main driver of the charge [11]. This is even more noticeable in low-density areas, which are underserved by traditional screening approaches performed in primary-care settings [12]. Additional issues of such programs stem from their limited scope. Primary care physicians, with limited ophthalmological expertise, might often miss abnormalities outside the original screening program or have lower sensitivity than retinal experts [13, 14]. Ophthalmology is also a leading specialty in alternative forms of healthcare delivery. For instance, mobile digital non-mydratic cameras are getting more affordable and have good specificity and sensitivity for DR [15, 16], which has enabled many screening plans in underserved areas [17–20]. There are also many examples of successful telemedicine screening plans in countries such as Australia, USA, India, Singapore, and Spain [21–24].

In parallel to the improvements in imaging and digitalization of healthcare, artificial intelligence (AI) based on

deep learning (DL) [25] represents a breakthrough that has dramatically improved the state-of-the-art in many tasks such as speech recognition, image processing, and text generation, among others [26]. In the field of medicine, DL has been most successful in medical imaging analysis, by enabling the creation of computer-aided diagnosis systems (CADx) with expert-level accuracies. There are many examples in dermatology [27], radiology [28], gastroenterology [29], and ophthalmology [30, 31]. In fact, the first ever US Food and Drug Administration (FDA)-approved autonomous AI is a screening tool for DR [32].

While the results of the AI performance presented in many publications are encouraging, there are still questions to be answered regarding their real-world application. The vast majority of publications are limited to retrospective studies taken on datasets captured in populations with prior conditions in a hospital setting [33–35]. Also, most publications limit the scope of their algorithm to just one pathology which, while interesting, is not ideal for screening purposes [36].

This work aims to overcome both limitations, by presenting a retrospective study performed on 2839 patients evaluated by digital fundus images taken with handheld non-mydratic cameras, on a routine checkup performed onsite at work centers. The algorithm evaluated in this study has previously been successful in detecting signs of DR, AMD, glaucoma, and nevus, the most common eye pathologies [24]. The novelty of our proposal is the combination of multiple pathology-specific algorithms to achieve holistic maculopathy detection. Each algorithm is trained to identify individual diseases, and, in conjunction, the final output increases the diagnostic accuracy of the AI system for ocular pathology detection.

Materials and methods

Study population

We present a retrospective diagnostic test evaluation. The dataset consists of 5918 images from a population

Table 1 Breakdown of the study validation datasets, as well as the training datasets for each of the AI algorithms that compose the screening system. (M/F male/female, LE/RE left eye/right eye)

Dataset	<i>n</i> participants (M/F)	Age median (std) (range)	<i>n</i> eyes (LE/RE)	<i>n</i> images	Abnormal (%)	AMD (%)	DR (%)	GON (%)	Nevus (%)	Other (%)
Study validation										
Original	2839 (1786 M/1053 F)	43 (11.52) [5–87]	5483 (2747 LE/2736 RE)	5918	321 (5.42%)	14 (0.23%)	0 (0%)	107 (1.81%)	110 (1.86%)	90 (1.5%)
Original + enriched	3337 (1999 M/1338 F)	46 (15.22) [5–96]	6009 (3013 LE/2996 RE)	6452	855 (13.25%)	398 (6.17%)	150 (2.32%)	107 (1.66%)	110 (1.7%)	90 (1.4%)
Algorithm training										
AMD	– (982 M/1526 F)	75 (11.52) [5–98]	– (1761 LE/1945 RE)	7218	–	4859 (67.31%)	–	–	–	–
DR	71,455 (26,691 M/19,296 F)	66 (11.42) [11–99]	116,501 (60,947 LE/55,554 RE)	139,813	–	–	14,376 (10.28%)	–	–	–
Glaucoma	1206 (619 M/587 F)	54 (15.58) [5–96]	1738 (794 LE/944 RE)	2366	–	–	–	1168 (49.37%)	–	–
GON	18,750 (8,215 M/10,535 F)	42 (11.64) [5–96]	29,352 (15,633 LE/13,719 RE)	30,054	–	–	–	–	4470 (14.87%)	–
Abnormality	31,877 (13,843 M/18,034 F)	49 (17.33) [5–99]	52,791 (27,846 LE/24,945 RE)	53,194	17,433 (32.77%)	–	–	–	–	–

^aThe total number of participants as well as the total number of eyes for the AMD training dataset could not be obtained, since the non-pathological images used for training were not assigned to a patient. The reported numbers (M/F, LE/RE, age) are from the pathological cases, which were referenced to a patient

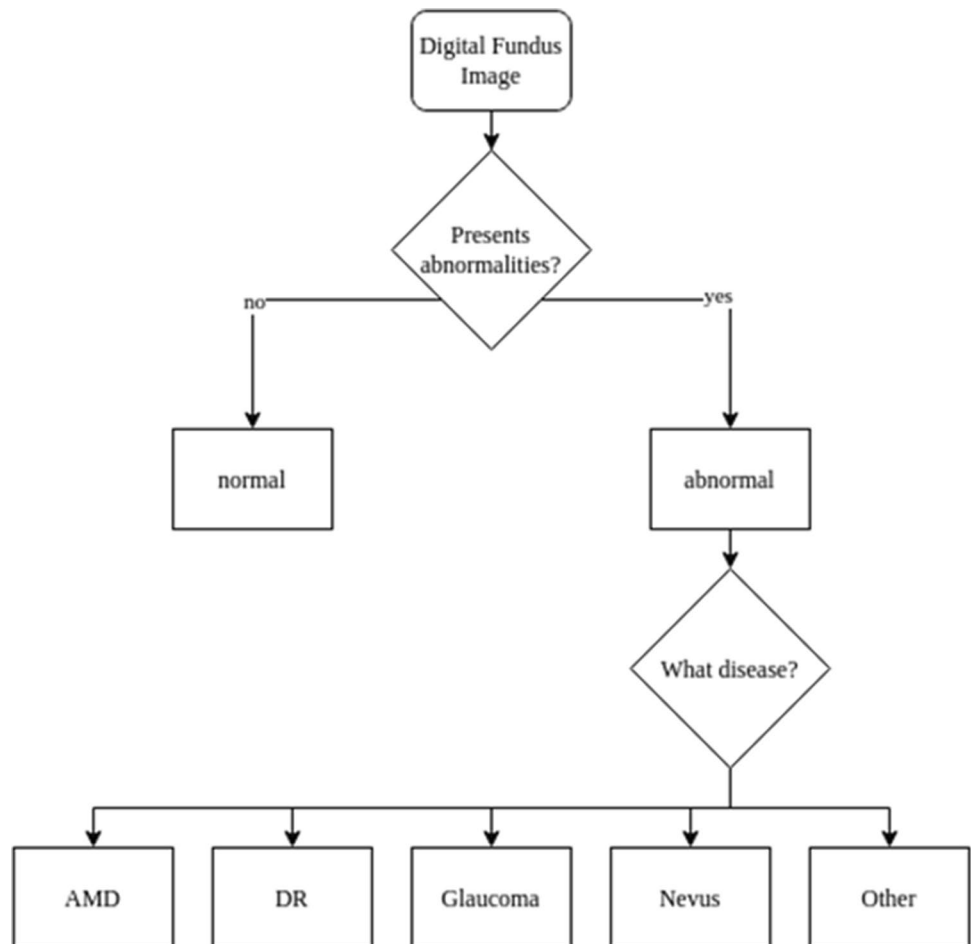
^bThe sex split and age median only accounts for 65% of the dataset. One of the image sources for this dataset did not include sex nor age information per patient

of 2839 individuals, taken between the 9th of January and the 13th of March of 2020. The median age was 43 years old with a standard deviation of 11.52. From the study population, 1786 (63%) were male and 1053 female (37%) (see Table 1 for a detailed breakdown). Participants of this study were enrolled in a consecutive series during routine occupational health checkups offered by their employer as medical benefits [37]. All ophthalmologic check-ups were performed by a single provider (Optretina, Sant Cugat, Spain). The images were obtained by a trained technician using handheld non-mydratic cameras on the participating center office premises, in a room which had been setup with adequate lighting conditions. The camera models employed were Optomed Aurora (field of view — FOV 50°, 88% of the dataset), ZEISS VISUSCOUT 100 (FOV 40°, 9% of the dataset), and Optomed SmartScope M5 (FOV 40°, 3% of the dataset). Image acquisition took around 2 min per patient. The raw image dataset was included in the study, and no images were discarded due to low resolution or were modified prior to the analysis.

Digital Fundus Image evaluation

The ground truth of the data was evaluated per eye. For patients with multiple captures, an automated quality filtering was employed to select the highest-quality image. Afterward, each image was graded by 2 specialists (intragrader variability kappa of 0.86 and 0.79 respectively) in a 2-tiered approach (Fig. 1). In case of discrepancies, a 3rd retinal specialist reviewed the image (intragrader variability kappa of 0.83). The first step of the labeling process was to classify the image as “normal” or “abnormal,” considering the latter as any digital fundus image showing pathological signs. Abnormal images were further subclassified per pathology. The specific pathologies considered for evaluation were DR (defined as more than mild DR, as per the 2019 revision of the American Academy of Ophthalmology’s Preferred Practice Pattern) [38, 39], AMD (defined as mild or worse), GON (suspicious glaucomatous optic neuropathy was defined by a cup-to-disc ratio of 0.7 or more in the vertical axis and/or other typical changes caused by glaucoma, such as localized notches or retinal nerve fiber layer defects or

Fig. 1 Labeling flowchart. The flowchart depicts the 2-tiered approach followed by all specialists to label the dataset. The ground truth was agreed by at least 2 graders



peripapilar hemorrhages), and Nevus (defined with clinical parameters as an hyperpigmented lesion beneath the retina). Images classified as abnormal (with possible signs of maculopathy) not matching the described taxonomy were classified as other in tier 2.

Dataset enrichment

We aim to assess the effectiveness of our automated screening algorithm on a wide-range general population. Because of the sampling bias of the initial population (working age participants, mostly without known prior pathologies), the prevalence for AMD and DR was far below that reported in the literature for the general population [4, 40]. To balance the data, we enriched the dataset with 384 AMD and 150 DR pathological images to match the prevalence in the general population of our environment [18, 41]. AMD images were obtained from Optretina's image bank (the sample was randomly selected from a cohort of 2212 AMD cases screened from January 2013 to May 2020) [24]. DR images were randomly selected from a series of positive cases detected in the Institut Català de la Salut (ICS) screening program for diabetics (Barcelona, Spain). In both cases, the enriched images were labeled by two expert retinal specialists, following the procedure detailed in Fig. 1. The dataset details are shown in Table 1.

Statistical analysis

The primary outcome of the analysis is the diagnostic accuracy of the AI system, defined by its sensitivity and specificity, versus the ground truth. Since the AI system performs a holistic screening, as well as pathology-specific diagnostic, we calculated the sensitivity and specificity for both. The operating threshold was fixed before the analysis and was not adjusted during the tests. The secondary outcomes are the receiver operating characteristic (ROC) curve and the area under the curve (AUC) index. All reported 95% CIs were obtained by performing a non-parametric bootstrap (1000 samples, with replacement). Study success was defined as reaching a predefined threshold of sensitivity and specificity on our holistic general screening algorithm. The hypotheses of interest were

$$H_0 : p < p_0 \text{ vs } H_A : p \geq p_0$$

where p is the sensitivity or specificity of the AI system. The predefined sensitivity and specificity thresholds were $p_0 = 0.75$ and $p_0 = 0.775$, respectively, benchmark defined by the FDA in their first-approved AI diagnostic system [32]. A one-sided 2.5% type I error binomial test was performed for both null hypotheses.

For the sample size calculation, we estimated a prevalence of retinal abnormalities in an occupational health checkup context of 7.8% with a 95% confidence interval, as per our previous study [42]. With these figures, the total number of participants needed was 2784. Additionally, we also confirmed that the sample size of our enriched dataset was large enough to ensure 80% statistical power ($\beta = 0.2$) on our sensitivity and specificity metrics, given the reported null hypothesis and the levels of pathological prevalence [43].

Training dataset

For algorithm development, macula-centered digital fundus images were retrospectively obtained from Optretina's own image bank (AMD, GON, Nevus, Abnormality) and Institut Català de la Salut and EyePacs (Kaggle). For AMD, Glaucoma, and DR, images were taken from a clinical setting, while Nevus and abnormality images were sourced from screenings, mostly performed with portable cameras. All images were evaluated by at least 1 expert retinologist, following the previously described criteria. The exact breakdown of the training dataset can be found in Table 1.

Individual algorithms

For each dataset, we trained binary classifiers (disease/no disease) using convolutional neural networks (CNNs). This process, with the right training data, allows the CNN to automatically learn features from the images that can be extrapolated successfully outside of the training data. The "AMD" algorithm uses a custom neural network architecture [42] using RGB images of 512×512 pixels [36]. The "DR" algorithm uses an InceptionV3 architecture [44] with inputs of 512×512 . "Glaucoma" uses a ResNet50 [45] with inputs of 224×224 . "Nevus" detection employs an InceptionV3 at 299×299 , and the abnormal images detector another InceptionV3 at 299×299 . The optimization algorithm to train the network was ADAM. We also used batch normalization, as well as using the weights of pretrained ImageNet networks where possible (InceptionV3, ResNet50) to speed up the training.

The performance of the algorithm was measured by the area under the receiver operating curve (AUC). The reported sensitivity and specificity points have been taken without adjusting the decision threshold (threshold = 0.5). The development datasets were split in an 80/10/10 fashion, where 80% of the data was used for training, 10% for validation (adjusting hyperparameters), and 10% to test the results. This data was split by patient (not image) and is completely independent from the dataset presented for the study validation.

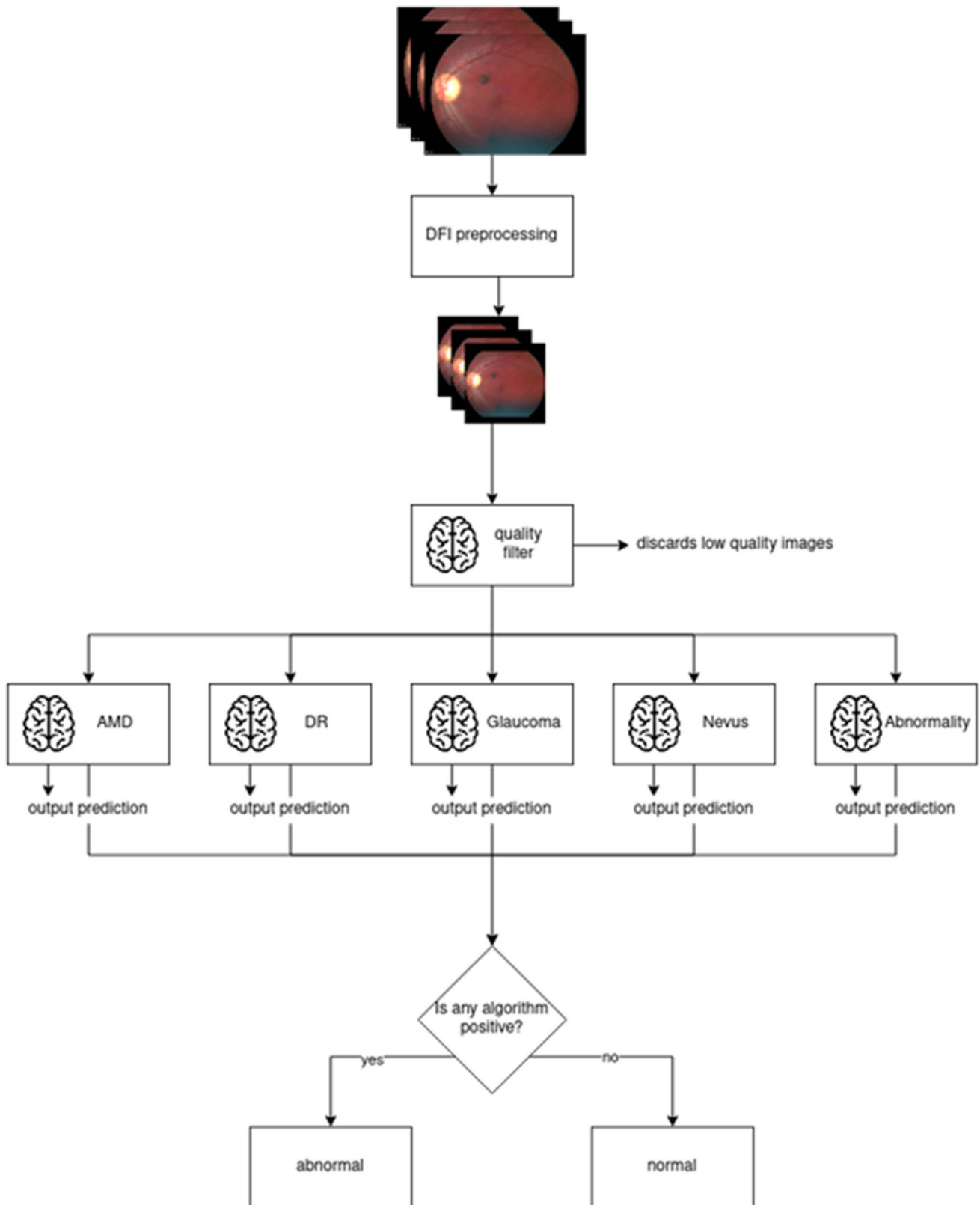


Fig. 2 Algorithm execution flowchart. The predictions are performed at the image level. 5 neural networks process independently each image and in case any algorithm is positive, the screened image is classified as “abnormal”

Screening algorithm

The screening algorithm is a combination of five independently trained neural networks. Four of these neural networks target specific pathologies (AMD, DR, GON, and Nevus), while the fifth one has been trained as an outlier detector, with a training dataset containing images from the aforementioned pathologies as well as other undetermined maculopathies (Fig. 2). Each image evaluated by the system is processed independently by each of the five neural networks and, at a second step, their response is combined in a single output. If an algorithm detects signs of any of the individual pathologies, the screened image is classified as Abnormal. A complete diagram of the AI system architecture is presented in Fig. 2

Results

The combined abnormality algorithm correctly identified 92% of the analyzable images annotated as Abnormal (776/843). The performance of individual disease algorithms is herein described: 99% of AMD images were correctly identified (394/398), 100% of DR images were correctly classified (150/150), 71% of GON images were correctly identified (71%, 75/107), 90% of Nevus images were correctly identified (99/110), and 73% of undetermined maculopathies were correctly classified (65/90). Insufficient quality of the images was observed in 0.23% of the cases (15/6452), which could not be graded. Of these, 80% were labeled as abnormal by our graders (12/15).

The single NN (neural network) abnormality algorithm correctly detected 82% (691/843) of the analyzable images. The percentage of images correctly classified disclosed by pathologies was 98.8% of AMD images (393/398), 94.6% of DR images (142/150), 45.6% of GON images (48/107), 49%

of Nevus images (54/110), and 49% of other maculopathies images (44/90).

The AMD algorithm correctly detected 90% of AMD images (358/398). False positive rates were 62.6% of DR images (94/150), 6.7% of GON images (7/107), 0% of Nevus images (0/110), and 22.5% of other maculopathies images (20/90).

The DR algorithm correctly detected 68.6% DR images (103/150). False positive rates were 4.1% of AMD images (16/398), 0% of GON images (0/107), 0.9% of Nevus images (1/110), and 11.7% of other maculopathies images (10/90).

The GON algorithm correctly detected 58.2% of GON images (62/107). False positive rates were 10.6% of AMD images (42/398), 12.6% of DR images (19/150), 2.9% of 102 Nevus images (3/110), and 8.8% of other maculopathies images (7/90).

The Nevus algorithm correctly detected 88.2% of the Nevus-specific images (90/110). False positive rates were 88% of AMD images (350/398), 100% of DR images (150/150), 2.9% of GON images (3/107), and 63.7% of other maculopathies images (57/90).

Sensitivity and specificity obtained from all algorithms are summarized in Table 2. The threshold for each individual CNN was not adjusted to boost the sensitivity or specificity to a certain operating point. Sensitivity and specificity were calculated per eye, using the best quality image if multiple were available. Discarding eye duplicates had little effect in the metrics in our study cohort (sensitivity: 92.1% duplicates vs 92.8% non-duplicates; specificity: 87.6% duplicates vs 86.8% no duplicates). Enforcing high-quality standards in the pre-processing pipeline, the effect is more noticeable (sensitivity: 92.8% any quality vs 92.6% high quality; specificity: 86.8% any quality vs 89.1% high quality). The dataset, as classified by the automatic quality algorithm, consisted of 53.4% high-quality images ($n = 3444$), 43.6% of acceptable-quality images ($n = 2809$), and 3.0% of low-quality images ($n = 196$).

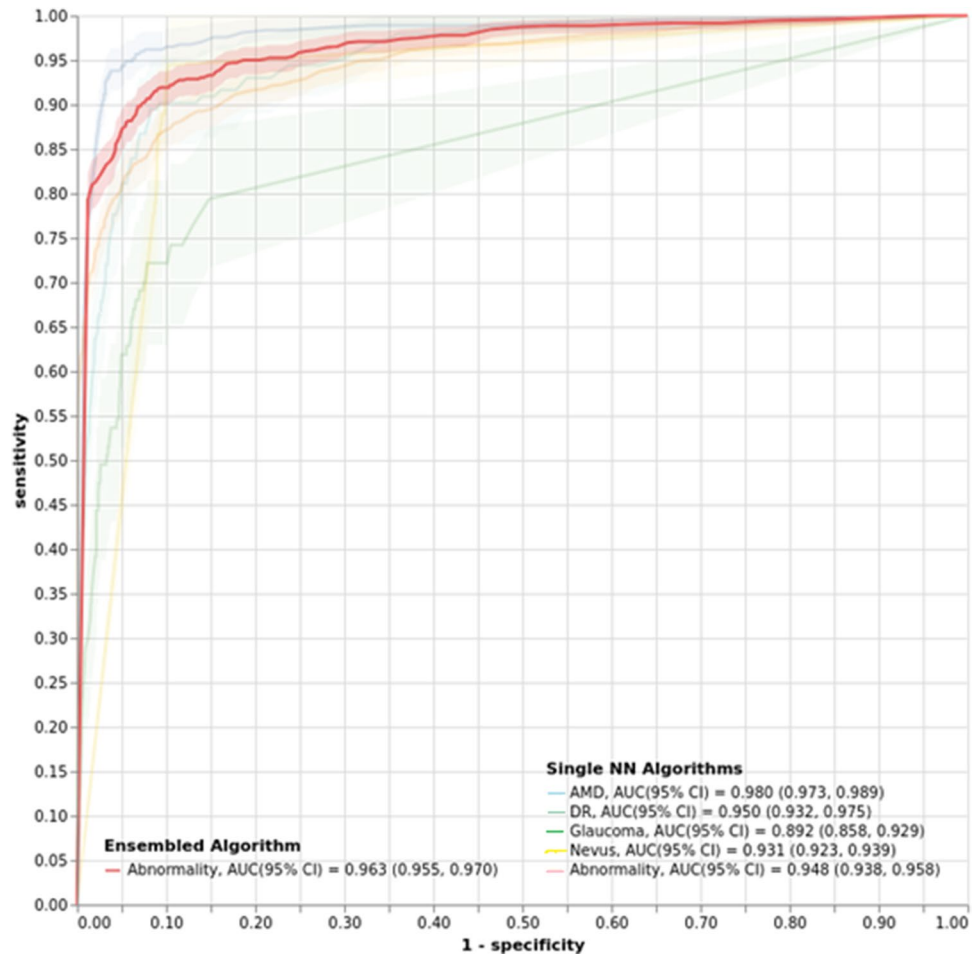
Table 2 Summary of sensitivity, specificity, and AUC aggregated and per individual algorithm

Algorithm	Sensitivity (95% CI) [<i>p</i> -value]	Specificity (95% CI) [<i>p</i> -value]
Combined algorithm:		
Abnormality	92.9% (91.0, 94.6) [<0.001] ^a	86.8% (85.8, 87.7) [<0.001] ^b
Single NN algorithms:		
Abnormality	83.4% (80.6, 85.9)	93.4% (92.7, 94.0)
AMD	93.8% (91.6, 96.3)	95.7% (95.2, 96.2)
DR	81.1% (75.3, 88.1)	94.8% (94.1, 95.4)
Glaucoma	53.6% (43.1, 63.1)	95.7% (95.2, 96.2)
Nevus	86.7% (80.7, 94.0)	90.7% (90.1, 91.5)

^a*p*-value for sensitivity on the combined abnormality algorithm was computed using a one-sided tailed binomial test using a sensitivity of $p = 0.75$ as the null hypothesis

^b*p*-value for specificity on the combined abnormality algorithm was computed using a one-sided tailed binomial test using a sensitivity of $p = 0.775$ as the null hypothesis

Fig. 3 Receiver operating curve for the combined and individual algorithms



To represent the best operation points, we plotted the ROC curve for the combined algorithm and the individual models in Fig. 3. Additionally, we also computed the AUC of the combined algorithm (0.963) and the single NN version (0.948).

Discussion

The proposed study has exceeded the expectations of the holistic solution proposed for screening the central retina's diseases. The reported diagnostic accuracy levels are similar to other algorithms already available on the market and higher than those required by the FDA for approval [32]. However, the proposed system presents the additional benefit that several pathologies are simultaneously screened together with those that cause greater visual loss among industrialized countries. These are also the main causes of preventable, non-reversible blindness, which are experiencing more growth in the world [46].

It is widely demonstrated that early detection of these diseases (mainly AMD, DR, and GON) and their early

treatment, if necessary, can prevent visual loss in a very high percentage of patients [47–49]. Simultaneous screening for multiple pathologies of the retina has previously been contemplated in some publications, both associated with human reading and artificial intelligence, mainly combining the detection of AMD and DR [14, 50], and also glaucoma [51]. To date, all artificial intelligence studies using fundus images for other diseases than diabetes have been carried out only in existing databases, with no clinical validation studies performed prospectively.

Currently, screening programs in most countries focus on DR, probably for cost-effectiveness reasons. Our study population is relatively young, and a priori, healthy, despite previous studies that report alterations in fundus images in almost 8% of cases in this type of population [37]. Although these alterations do not usually represent serious or urgent cases, any pathology in this population, young and working, can have significant socio-economic repercussions. The incorporation of artificial intelligence and the simultaneous screening of several diseases can make these early detection systems more cost-effective. Despite the fact that the objective of this study is not an economic evaluation, the use of

automatic detection software can reduce previously reported costs, lower than 10 euros per patient [37].

One of the most important causes of loss of effectiveness of AI is related to the quality of the images [52]. In our series, we have had 3% of fundus photographs of low quality, a figure significantly lower than other published series such as Abramoff et al. (above 8%) [32] or Liu et al. (16%) [53]. We would like to note that our numbers are obtained on real conditions, with portable cameras and, generally, under certainly strict timeframes. This difference can be due to different reasons. While portable cameras traditionally tended to be of lower quality than desktop cameras, the technical advances in recent years have improved the quality of the images, and current cameras like Aurora are of equivalent quality. Moreover, screening has been carried out in relatively young people, in which ocular media opacities are much less common and tend to have more dilated pupils in scotopic conditions. One of the most interesting points is the study of comorbidities; to date, it is also one of the limits set by AI. With this type of approach, lesions can be detected by independent algorithms that combine together in a holistic diagnosis. This results in a more robust system, less likely to miss incidental findings, with higher overall sensitivity, while only penalizing slightly on specificity. The use of AI for DR screening is already being implemented successfully in some countries. However, this pathology-specific approach carries the risk of ignoring other possible findings, since the neural networks employed are not designed for it. We do believe that it would be beneficial if those IA systems are combined and set up with a more holistic approach, to minimize the risk of ignoring these incidental findings. The combination of multiple algorithms also makes it easier to deploy improvements on the system. We can tackle algorithms per pathology, and any improvements in the individual models will benefit in the final output. We have already achieved gains in multiple retrainings of the DR algorithm, and we believe that the rest of algorithms could be similarly improved in the near future. GON algorithm alone shows weaker performance than the rest of single algorithms, making the single and combined algorithms not reliable for screening of this pathology.

This study's limits are those determined by carrying out a retrospective study, those related to the population studied (in this case, younger and with a lower rate of pathology than the general population), and the limits derived from the pathology studied. To compensate for possible biases, the database has been enriched with a presence similar to the population of AMD and DR. It would be convenient, in the future, to introduce other highly prevalent pathologies in the population, such as the presence of epiretinal membranes or macular signs associated with high myopia. Another area that we want to study further is in the image capture workflow, to offer not only an automated way of

screening, but a better screening workflow with hybrid systems. We believe that by integrating the image acquisition process with an online platform for automated data collection, it is possible to instrument the whole process and guide the technician through, with the additional benefit that the images are automatically assigned to the right patient and checked for adequate quality prior to running any subsequent diagnostic analysis.

In conclusion, the use of an autonomous AI-based diagnostic system based on fundus images for holistic maculopathy screening in a routine occupational health checkup context seems effective, with high levels of sensitivity and specificity that improves further those achieved by specific algorithms. The application of these systems could allow more extensive screening programs with greater detection of pathology in working-age patients.

Author contribution All the listed authors meet the criteria for the authorship according to the ICMJE guidelines.

All persons who meet the authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the Graefe's Archive for Clinical and Experimental Ophthalmology. O. Font, MA. Zapata, and D. Royo made a substantial contribution to the present work in the design of the study, analysis and interpretation of data, manuscript drafting, and approval of the version of the manuscript to be published. MA. Zapata, A. Bures, and C. Salinas had contributed to acquisition of data, conception and design of the study, critical revision for important intellectual content, and approval of the version to be published. O. Font, J. Zarranz-Ventura, and S. Banderas had an important participation in the conception of the study, drafting of the paper, and approval of the present version of the manuscript. Edition of the manuscript was taken by MA. Zapata and S. Banderas. J. Torrents and J. Zarranz-Ventura's contribution to the present work included furthermore critical revision for important intellectual content and approval of the present version. All the authors agree to be accountable for all aspects of the work in ensuring its accuracy and integrity.

Funding Open Access Funding provided by Universitat Autònoma de Barcelona.

Data availability Any information not included in the present manuscript is available upon request by contacting the corresponding author.

Code availability The AI system described in this study is copyrighted by Optretina S.L. and consequently the underlying source code is not available.

Declarations

Ethics approval Optretina is a company supervised by the ethics committee of the Vall d'Hebron Hospital in Barcelona, who approved this clinical trial (Study code PR(OPT)370/2019).

Consent to participate Patient's informed consent was exempted because of the retrospective nature of the study, using fully anonymized retinal images.

Conflict of interest MA. Zapata is the founding member and medical director of Optretina. D. Royo is founding partner and technical director of the company. O. Font is employed in Optretina and A. Bures and C. Salinas work as funduscopy reader in the company. The authors declare no other direct financial conflicts of interest as individuals.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bourne RAA, Stevens GA, White RA et al (2013) Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob Health* 1:339–349. [https://doi.org/10.1016/S2214-109X\(13\)70113-X](https://doi.org/10.1016/S2214-109X(13)70113-X)
- Klaver CCW, Wolfs RC, Vinberling JR et al (1998) Age-specific prevalence and causes of blindness and visual impairment in an older population: the Rotterdam study. *Arch Ophthalmol* 116:623–658. <https://doi.org/10.1001/archophth.116.5.653>
- Mohamed Q, Gillies MC, Wong TY (2017) Management of diabetic retinopathy. *JAMA* 298:902–916. <https://doi.org/10.1001/jama.298.8.902>
- Wong WL, Su X, Li X et al (2014) Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health* 2:e106–e116. [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1)
- Tham YC, Li X, Wong TY et al (2014) Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 121:2081–2090. <https://doi.org/10.1016/j.ophtha.2014.05.013>
- Ting DSW, Ming Cheung GCM, Wong TY (2016) Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. *Clin Exp Ophthalmol* 44:260–277. <https://doi.org/10.1111/ceo.12696>
- Ferris FL (1993) How effective are treatments for diabetic retinopathy? *JAMA-J Am Med Assoc* 269:1–129
- Lim LS, Mitchell P, Seddon JM et al (2012) Age-related macular degeneration. *Lancet* 379:1728–1738. [https://doi.org/10.1016/S0140-6736\(12\)60282-7](https://doi.org/10.1016/S0140-6736(12)60282-7)
- Hamilton WS (2007) New treatments for age-related macular degeneration. *Lancet* 370:1–1480. [https://doi.org/10.1016/S0140-6736\(07\)61627-4](https://doi.org/10.1016/S0140-6736(07)61627-4)
- Jonas JB, Aung T, Bourne RR et al (2017) Glaucoma. *Lancet* 390:2183–2193. [https://doi.org/10.1016/S0140-6736\(17\)31469-1](https://doi.org/10.1016/S0140-6736(17)31469-1)
- Lairson DR, Pugh JA, Kapadia AS et al (1992) Cost-effectiveness of alternative methods for diabetic retinopathy screening. *Diabetes Care* 15:77–1369. <https://doi.org/10.2337/diacare.15.10.1369>
- Lee SJ, McCarty SC, Taylor HR, Keeffe JE (2008) Costs of mobile screening for diabetic retinopathy: a practical framework for rural populations. *Aust J rural Health* 9:186–192. <https://doi.org/10.1046/j.1038-5282.2001.00356.x>
- Farley TF, Mandava N, Prall FR, Carsky C (2008) Accuracy of primary care clinicians in screening for diabetic retinopathy using single-image retinal photography. *Ann Fam Med* 6:34–428. <https://doi.org/10.1370/afm.857>
- Chan CKW, Gangwani RA, McGhee SM et al (2015) Cost-effectiveness of screening for intermediate age-related macular degeneration during diabetic retinopathy screening. *Ophthalmology* 122:85–2278. <https://doi.org/10.1016/j.ophtha.2015.06.050>
- Ahmed J, Ward TP, Bursell SE et al (2006) The sensitivity and specificity of nonmydriatic digital stereoscopic retinal imaging in detecting diabetic retinopathy. *Diabetes Care* 29:9–2205. <https://doi.org/10.2337/dc06-0295>
- Massin P, Erginay A, Mehidi AB et al (2003) Evaluation of a new non-mydriatic digital camera for detection of diabetic retinopathy. *Diabet Med* 20:635–641. <https://doi.org/10.1046/j.1464-5491.2003.01002.x>
- Levy J, Lifshitz T, Goldfarb D, Knyazer B, Belfair N (2011) Screening for diabetic retinopathy with a mobile non-mydriatic digital fundus camera in Southern Israel. *Isr Med Assoc J* 13:40–137
- Romero-Aroca P, Sagarra-Alamo R, Traveset-Maeso A et al (2010) The non-mydriatic camera as a screening method in diabetics in Tarragona. *Arch Soc Esp Oftalmol* 85:232–238. <https://doi.org/10.1016/j.oftal.2010.09.005>
- Spurling GKP, Deborah AA, Hayman NE, Hansar N, Cooney AM, Jackson CL (2010) Retinal photography for diabetic retinopathy screening in Indigenous primary health care: the Inala experience. *Aust N Z J Public Health* 34(SUPPL):30–33. <https://doi.org/10.1111/j.1753-6405.2010.00549.x>
- Beynat JA, Charles A, Astruc K et al (2009) Screening for diabetic retinopathy in a rural French population with a mobile non-mydriatic camera. *Diabetes Metab* 35:49–56. <https://doi.org/10.1016/j.diabet.2008.07.002>
- Kumar S, Tay-Kearney ML, Chaves F, Constable IJ, Yogesan K (2006) Remote ophthalmology services: Cost comparison of telemedicine and alternative service delivery options. *J Telemed Telecare* 12:19–22. <https://doi.org/10.1258/135763306775321399>
- Sharma M, Chakrabarty AS, Rathi P, Sharma RC, Pratibha G (2011) An integrated, mobile service for diabetic retinopathy in rural India. *Community Eye Health* 24:17–18
- Nguyen HV, Tan GSW, Tapp RJ et al (2016) Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology* 123:2571–2580. <https://doi.org/10.1016/j.ophtha.2016.08.021>
- Zapata MA, Arcos G, Fonossola A et al (2017) Telemedicine for a general screening of retinal disease using nonmydriatic fundus cameras in optometry centers: three-year results. *Telemed J E Health* 23:30–36. <https://doi.org/10.1089/tmj.2016.0020>
- Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Inform Process Syst* 25:1–9
- Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:18–115. <https://doi.org/10.1038/nature21056>
- Lakhani P, Sundaram B (2017) Deep Learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284:574–582. <https://doi.org/10.1148/radiol.2017162326>
- Urban G, Tripathi P, Alkayali T et al (2018) Deep learning localizes and identifies polyps in real time with 96% accuracy

- in screening colonoscopy. *Gastroenterology* 155:1069–1078.e8. <https://doi.org/10.1053/j.gastro.2018.06.037>
30. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316:2402–2410. <https://doi.org/10.1001/jama.2016.17216>
 31. De-Fauw J, Ledsam JR, Romera-Paredes B et al (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 24:1342–1350. <https://doi.org/10.1038/s41591-018-0107-6>
 32. Abramoff MD, Lavin PT, Birch M, Shah N, Folk CJ (2018) Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *Npj Digital Medicine* 1:39. <https://doi.org/10.1038/s41746-018-0040-6>
 33. Burlina PM, Joshi N, Pekala M et al (2017) Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks". *JAMA Ophthalmol* 135:6–1170. <https://doi.org/10.1001/jamaophthalmol.2017.3782>
 34. Brown JM, Campbell JP, Beers A et al (2018) Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol* 136:10–803. <https://doi.org/10.1001/jamaophthalmol.2018.1934>
 35. Ahn JM, Kim S, Ahn KS et al (2018) A deep learning model for the detection of both advanced and early glaucoma using fundus photography". *PLoS ONE* 13:1–8. <https://doi.org/10.1371/journal.pone.0207982>
 36. Chew EY, Schachat AP (2015) Should we add screening of age-related macular degeneration to current screening programs for diabetic retinopathy? *Ophthalmology* 122:6–2155. <https://doi.org/10.1016/j.ophtha.2015.08.007>
 37. Zapata MA, Martín R, Garcia-Arumí C, et al (2020) Remote screening of retinal and optic disc diseases using handheld non-mydiatic cameras in programmed routine occupational health checkups onsite at work centers. *Graefes Arch Clin Exp Ophthalmol* 1-9. <https://doi.org/10.1007/s00417-020-04860-z>
 38. Wilkinson CP, Ferris FL 3rd, Klein RE et al (2003) Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 110:1677–1682. [https://doi.org/10.1016/S0161-6420\(03\)00475-5](https://doi.org/10.1016/S0161-6420(03)00475-5)
 39. Flaxel CJ, Adelman RA, Bailey ST et al (2020) Diabetic retinopathy preferred practice pattern. *Ophthalmology* 127:66–145. <https://doi.org/10.1016/j.ophtha.2019.09.025>
 40. Yau JWY, Rogers SL, Kawasaki R et al (2012) Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 35:556–564. <https://doi.org/10.2337/dc11-1909>
 41. Spanish Eyes Epidemiological (SEE) Study Group (2011) Prevalence of age-related macular degeneration in Spain. *Br J Ophthalmol* 95:931–936. <https://doi.org/10.1136/bjo.2010.187773>
 42. Zapata MA, Royo-Fibla D, Font O et al (2020) Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma. *Clin Ophthalmol* 4:419–429. <https://doi.org/10.2147/OPTH.S235751>
 43. Bujang MA, Adnan TH (2016) Requirements for minimum sample size for sensitivity and specificity analysis. *J Clin Diagn Res* 10:YE01–YE06. <https://doi.org/10.7860/JCDR/2016/18129.8744>
 44. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. <https://doi.org/10.1109/CVPR.2016.308>
 45. Wu S, Zhong S, Liu Y (2018) Deep residual learning for image steganalysis. *Multimed Tools Appl* 77:10437–10453. <https://doi.org/10.1007/s11042-017-4440-4>
 46. GBD (2019) Blindness and Vision Impairment Collaborators, Vision Loss Expert Group of the Global Burden of Disease Study (2021) Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the right to sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health* 9:e144–e160. [https://doi.org/10.1016/S2214-109X\(20\)30489-7](https://doi.org/10.1016/S2214-109X(20)30489-7)
 47. AREDS2-HOME Study Research Group, Chew EY, Clemons TE et al (2014) Randomized trial of a Home Monitoring System for Early Detection of Choroidal Neovascularization Home Monitoring of the Eye (HOME) study. *Ophthalmology* 21:535–544. <https://doi.org/10.1016/j.ophtha.2013.10.027>
 48. Ferris FL (1994) Results of 20 years of research on the treatment of diabetic retinopathy. *Prev Med* 23:740–742. <https://doi.org/10.1006/pmed.1994.1127>
 49. Phu J, Agar A, Wang H, Masselos K, Kalloniatis M (2020) Management of open-angle glaucoma by primary eye-care practitioners: toward a personalised medicine approach. *Clin Exp Optom*. <https://doi.org/10.1111/cxo.13114>
 50. González-Gonzalo C, Sánchez-Gutiérrez V, Hernández-Martínez P et al (2019) Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. *Acta Ophthalmol* 98:368–377. <https://doi.org/10.1111/aos.14306>
 51. Ting DSW, Cheung CYL, Lim G et al (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318:2211–2223. <https://doi.org/10.1001/jama.2017.18152>
 52. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H (2018) Artificial intelligence in retina. *Prog Retin Eye Res* 67:1–29. <https://doi.org/10.1016/j.preteyeres.2018.07.004>
 53. Liu J, Gibson E, Ramchal S et al (2021) Diabetic retinopathy screening with automated retinal image analysis in a primary care setting improves adherence to ophthalmic care. *Ophthalmol Retina* 5:71–77. <https://doi.org/10.1016/j.oret.2020.06.016>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.