

Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected Glaucoma

This article was published in the following Dove Press journal:
Clinical Ophthalmology

Miguel Angel Zapata ¹
Dídac Royo-Fibla ¹
Octavi Font¹
José Ignacio Vela^{2,3}
Ivanna Marcantonio^{2,3}
Eduardo Ulises Moya-Sánchez^{4,5}
Abraham Sánchez-Pérez⁵
Darío García-Gasulla ⁴
Ulises Cortés^{4,6}
Eduard Ayguadé^{4,6}
Jesus Labarta^{4,6}

¹Optretina, Barcelona, Spain;

²Ophthalmology Department, Hospital de la Santa Creu i de Sant Pau, Barcelona 08041, Spain; ³Universitat Autònoma de Barcelona (UAB), Campus de la UAB, Barcelona, Spain; ⁴Barcelona Supercomputing Center (BSC), Barcelona, Spain; ⁵Universidad Autónoma de Guadalajara - Postgrado en Ciencias Computacionales, Guadalajara, Mexico; ⁶Universitat Politècnica de Catalunya - BarcelonaTECH, Campus Nord, Barcelona, Spain

Purpose: To assess the performance of deep learning algorithms for different tasks in retinal fundus images: (1) detection of retinal fundus images versus optical coherence tomography (OCT) or other images, (2) evaluation of good quality retinal fundus images, (3) distinction between right eye (OD) and left eye (OS) retinal fundus images, (4) detection of age-related macular degeneration (AMD) and (5) detection of referable glaucomatous optic neuropathy (GON).

Patients and Methods: Five algorithms were designed. Retrospective study from a database of 306,302 images, Optretina's tagged dataset. Three different ophthalmologists, all retinal specialists, classified all images. The dataset was split per patient in a training (80%) and testing (20%) splits. Three different CNN architectures were employed, two of which were custom designed to minimize the number of parameters with minimal impact on its accuracy. Main outcome measure was area under the curve (AUC) with accuracy, sensitivity and specificity.

Results: Determination of retinal fundus image had AUC of 0.979 with an accuracy of 96% (sensitivity 97.7%, specificity 92.4%). Determination of good quality retinal fundus image had AUC of 0.947, accuracy 91.8% (sensitivity 96.9%, specificity 81.8%). Algorithm for OD/OS had AUC 0.989, accuracy 97.4%. AMD had AUC of 0.936, accuracy 86.3% (sensitivity 90.2% specificity 82.5%), GON had AUC of 0.863, accuracy 80.2% (sensitivity 76.8%, specificity 83.8%).

Conclusion: Deep learning algorithms can differentiate a retinal fundus image from other images. Algorithms can evaluate the quality of an image, discriminate between right or left eye and detect the presence of AMD and GON with a high level of accuracy, sensitivity and specificity.

Keywords: artificial intelligence, retinal diseases, screening, retinal fundus image

Introduction

Retinal diseases are on the rise due to the increase of diabetic population and increased longevity. A good example of the effect of age on retinal health is age-related macular degeneration (AMD): in 2020, the number of people with this disease is projected to be 196 million and is expected to grow to 288 million by 2040.¹ Glaucoma is a progressive optic neuropathy, which is the leading cause of blindness in industrialized countries and ocular hypertension is the main risk factor for glaucoma.²

General screening for retinal diseases and glaucoma in the population is ideal, as many pathologies can be detected in the early stages, while other diseases—such as retinal dystrophies, choroidal nevi or epiretinal membranes (ERMs)—can also be diagnosed and are better managed with early detection. Screening for diabetic

Correspondence: Miguel Angel Zapata
Optretina, C/ Las Palmas 11, 08195 Sant
Cugat del Vallès, Barcelona, Spain
Tel +34 655809682
Email mazapata@optretina.com

retinopathy (DR) and AMD is already cost-effective, but there are no published studies to support the case for general screening yet. Retinal fundus images using non-mydiatic cameras (NMC) are currently considered the gold standard for screening DR^{3,4} and are also cost-effective for AMD.^{5,6} Similarly, screening for premature retinopathy is supported by many papers, showing good efficacy and cost-effectiveness.⁷ The detection of disease through retinal fundus images has also been shown to produce competitive results, particularly for neurological⁸ or cardiovascular⁹ diseases. However, general screenings using retinal fundus images have, to the best of our knowledge, never been implemented—even on particularly sensitive segments of the population, such as those over fifty years old, an age at which retinal pathologies increase significantly.

Artificial intelligence (AI) and, more specifically, deep learning, could be helpful in this area. Deep learning is a set of machine learning techniques that have recently undergone a renaissance, achieving breakthrough accuracy in image recognition tasks in different fields from computer vision¹⁰ to the medical sciences, solving problems such as: detecting melanoma,¹¹ detecting cancer metastases from pathology images,¹² and diagnosing pneumonia using x-rays.¹³ We argue that a computer-aided diagnosis system (CADx), whether fully automated or used in conjunction with medical specialists, could substantially reduce the human and economic costs of image analysis. Furthermore, in the last few years, NMC have become significantly cheaper; in the near future, any cell phone could be used as a retinal camera for early screening.

Deep learning has already had success in ophthalmology, demonstrating its ability to detect DR in retinal fundus images.^{14–17} In fact, previous studies report accuracy of over 90% with sensitivity and specificity levels over 90% as well, which coincides with roughly the same level as a human ophthalmology expert. We can find similar outcomes with the detection of referable AMD.¹⁸ However, using such algorithms in real-world situations remains a challenge, as shown by the decreased accuracy (around 80%) achieved in prospective studies.¹⁹ This decrease could be due to inconsistent image quality and other aspects, such as comorbidity. Dataset quality assessment is normally taken for granted, with researchers using trainings and certifications for photographers^{17,18} or manually “cleaning” the datasets;¹⁹ some authors have even used quality filters on algorithms before training the algorithms.²⁰

Optretina is a telemedicine platform which performs general screening for retinal diseases using nonmydiatic cameras and human evaluation by a retinal specialist ophthalmologist.²¹ Our research in artificial intelligence deals with both data acquisition and the diagnostic steps of our telemedicine platform, in a near future we expect AI could assist our specialists in achieving levels of consistency and accuracy beyond unassisted human abilities.

This study aims to evaluate the capacity of AI methods for solving several tasks related to our CADx pipeline. We automatically discard inputs that are not derived from color fundus photography (CFP); next we filter out images that do not pass our quality threshold; we then differentiate the right eye from the left eye; and finally, we proceed to the diagnosis, using the remaining images and our algorithms for AMD and suspected glaucoma.

Methods

Datasets

Optretina has been carrying out a telemedicine screening program since 2013 in optical centers and since 2017 in workplace offices and private companies. This study is based on the tagged dataset from Optretina, composed of 306,302 retinal images. Labelled with diagnosis, laterality and quality. Ophthalmologists—all of whom are retinal specialists—have evaluated all of these images. The dataset is real-life (Not from clinical trials) and anonymized, composed of approximately 80% cases of normal retinas and 20% abnormal ones. Images are acquired using NMC (color fundus and red free) and optical coherence tomography (OCT). The dataset includes images from different types and brands of cameras. [Figure 1](#) shows the distribution of the different cameras used, by percentage. The dataset includes color fundus photography (CFP), macular OCT's, retinal nerve fiber layer OCT (RNFL), and other added non-medical images trying to fool the AI (such as round objects like pizzas or photos of the full moon). Of the 306,302 images, over 250,000 correspond to CFP. In 2017, Optretina published the results of three years of screenings,¹⁷ detailing the percentages of diseases found.

Classification of image type (CFP, OCT, others) and eye of evaluation (right or left) was performed by one ophthalmologist. Labeling for quality assessment, presence of AMD, or glaucomatous optic neuropathy (GON) was performed by two independent retinal specialists, a third ophthalmologist assessed cases where there was

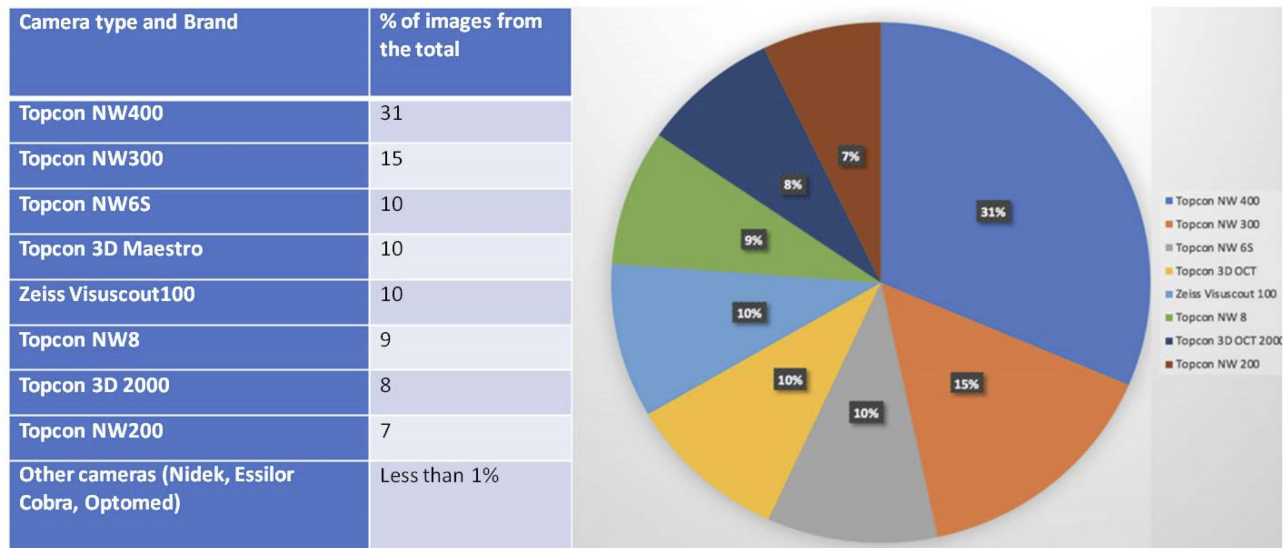


Figure 1 Distribution of nonmydriatic cameras (NMCs) and optical coherence tomography (OCT) used for the dataset. Only central retinal images were included.

no agreement, in order to establish at least two consistent gradings. Since Optretina performs screenings and patients in whom pathology is detected are subsequently referred to an ophthalmologist, repetition of images of the same patient are very rare. It is estimated that 0.5% of the dataset images may be of patients already assessed in previous years. The dataset was divided by patient in two splits: training (80%) and testing (20%). 10% of the training split was used for validation purposes during the training process of the models. Table 1 displays the

number of images used for each deep learning model, the number of classes per model, and the number of ophthalmologists who independently labeled the same dataset.

Convolutional Neuronal Networks (CNNs)

The first three CNN models (trained from scratch) were used for different image classification tasks usually applied in data-collection, data-selection, display, and data classification. The first model is very useful for cleaning unlabeled datasets, since it can separate the

Table 1 Number of Images Used for Training the CNNs. Images Were Divided into Two Groups: One for Training the Model and Another Naïve Group for Validating the Training Procedure. Color Fundus photography (CFP), Optical coherence tomography (OCT), Retinal nerve fiber layer (RNFL), Right eye (OD), Left eye (OS), Age related macular degeneration (AMD), referable glaucomatous optic neuropathy (GON).

Model	Classification Task	Images for Training	Data Augmentation from Keras	Number of Classes	Number of Ophthalmologists for Label the Images
1	Image type selection CFP/OCT/RNFL/OTHER	56,396	Shear_range=0.2, zoom_range=0.2	4	1
2	Data Collection Good quality images	150,075	Shear_range=0.2, zoom_range=0.2	2	3
3	Data Display OD/OS	30,119	Shear_range=0.2, zoom_range=0.2	2	1
4	AMD	8832	Zoom_range 0.1 rotation_range 10 horizontal_flip: True	2	3
5	GON	3776	Zoom_range 0.1 rotation_range 15 horizontal_flip: True	2	3

CFP images from OCT or other type of images like red-free image, autofluorescence, angiographic images or other. The second model allows us to select only the CFP with a minimum threshold of image quality. The third model helps us to classify right-eye CFP versus left-eye CFP, which we use in our platform to display the best pair of retinas per patient. The fourth model classifies the AMD CFP versus the normal CFP. Finally, the fifth model allows for the classification of normal CFP versus GON. Table 2 presents the three main CNN. We have also published sample code for these architectures on <https://github.com/octavifs/optretina-cnns>.

Model 1: Image Type Selection CFP/OCT/Other Images

The CNN-1 performs a binary classification, separating CFP from all other image types (OCT/RNFL/other images). The training process was done from scratch, using RGB images resized to 128x128 pixels with geometric data augmentation over the training dataset. It should be noted that the CNN-1 has only three convolution layers, three max pooling layers,

and three dense layers; as a result, the CNN-1 has fewer parameters than very deep models (such as Inception, VGG-19, and RESNET50, among others). Its shallow architecture makes it computationally inexpensive, which is beneficial for increasing the processing throughput of our CADx pipeline. CNN-1, along with the rest of the models, have been implemented using Keras,²² a software library that facilitates the creation of neural networks with access to specialized hardware resources, such as graphics processing units (GPUs), in order to speed up the training and execution of such algorithms.

Model 2: Good Quality Retinal Fundus Image Classification

The second model also implements CNN-1 architecture, but for a different task: to select good quality CFP. The definition of “good quality” was defined by considering the color retinal fundus image, centered on macula with a correct focus; good visualization of the parafoveal vessels; and if it was possible to visualize over two disc diameters around the fovea, the optic

Table 2 CNN Architecture for Algorithms: 1. Color Fundus Photography (CFP) versus Macular Optical Coherence Tomography (OCT), Retinal Nerve Fiber Layer in the OCT and Other images. 2. Good Quality Retinal Fundus image. 3. Right Eye versus Left Eye (OD/OS). 4. AMD. 5. Glaucomatous Optic Neuropathy (GON) Classification. (Notations are Based on Keras²²)

CNN-1	AMD-Net	GON-Net
Input (128,128,3) Conv_2D_1(ReLU,3,3,32) Max_Pooling_1(2,2) Conv_2D_2(ReLU,3,3,32) Max_Pooling_2(2,2) Conv_2D_3(ReLU,3,3,64) Max_Pooling_3(2,2) Flatten Dense_1 (ReLU,64) Dropout_1 (0,5) Dense_2 (Sigmoid, 5*/2) We trained only for 50 epochs, batch size of 32 and learning rate of 0.001 for each model.	Input (512,512,3) Conv_2D_1(ReLU,5,5,32) Max_Pooling_1(2,2) Conv_2D_2(ReLU,3,3,32) Conv_2D_3(ReLU,3,3,32) Max_Pooling_2(2,2) Conv_2D_4(ReLU,3,3,32) Conv_2D_5(ReLU,3,3,32) Max_Pooling_3(2,2) Conv_2D_6(ReLU,3,3,32) Conv_2D_7(ReLU,3,3,32) Max_Pooling_4(2,2) Conv_2D_8(ReLU,3,3,32) Conv_2D_9(ReLU,3,3,32) Max_Pooling_5(2,2) Conv_2D_10(ReLU,3,3,64) Conv_2D_11(ReLU,3,3,64) Max_Pooling_6(2,2) Flatten Dense_1(ReLU64) Dropout_1(0,5) Dense_2(ReLU,64) Dropout_2(0,2) Dense_2 (Softmax, 2)	Input (224,224,3) RESNET50 Flatten Dropout_1(0,5) Dense_1(ReLU,64) Dropout_2(0,2) Dense_2 (Softmax, 2)



Figure 2 (A) Examples of good quality images. Retinal fundus images are centered on the macula with correct focus, good visualization of the parafoveal vessels and over two disc diameters around the fovea, and correct observation of the optic disc (at least three-quarters) and the vascular arcades. **(B)** Examples of poor quality images.

disc (at least three-quarters), and the vascular arcades. **Figure 2** shows an example of good and bad quality retinal fundus images. We also trained the same CNN-1 from scratch for a different objective with geometric data augmentation.

Model 3: Right-Eye versus Left-Eye (OD/OS) Classification

The third model is very useful for automatically displaying both CFP images of the same patient in order for the ophthalmologist to label the images. We trained the model from scratch with data augmentation and using CNN-1 architecture in order to classify the left eye versus the right eye. It is important to note that the same simple architecture was useful for classifying images in different tasks. However, while the architecture is the same, each training process teaches the model the features necessary to learn a specific task.

Model 4: Age-Related Macular Degeneration (AMD)

AMD image classification poses special challenges due to the fact that images are almost identical and only small features determine their class.

AMD was classified according to the international classification published in 2013:²³

- Early AMD: the presence of medium drusen within two disc diameters (DD) of the fovea.
- Intermediate AMD: the presence of large drusen or medium drusen with pigmented abnormalities within two DDs from the center.
- Advanced AMD: the presence of geographic atrophy (> 125 microns) or signs of choroidal neovascularization or fibrosis within two DDs from the fovea.

Although the images were classified into the three stages of AMD for further investigations, this first algorithm was binary, and detected AMD retinas (at any stage) from normal ones.

We propose a novel architecture for AMD binary classification which will be referred to hereinafter as AMDNET. The architecture of AMDNET allows us to extract small details and features while also making predictions with relatively few trainable parameters (around 241,000) in comparison with other well-known architectures such as RESNET50²⁴ or InceptionV3²⁵ which have approximately

25 million trainable parameters. The AMDNET architecture is presented in Table 2. We trained AMDNET from scratch with the data specifications presented in Table 1.

Algorithm 5: Referable Glaucomatous Optic Neuropathy (GON)

Referable glaucomatous optic neuropathy (GON) 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 RNFL defects. We modified RESNET50 (23) and training from scratch in order to make a binary GON image classification.

Study of Misclassifications

To study the possible cause of misclassification an experienced retinal specialist and the Optretina medical director (Miguel A. Zapata) reviewed all false positives and negatives in the AMD and the GON algorithm.

Ethics and Data Protection

Given that requesting informed consent for the Optretina evaluation is mandatory, patients are informed of the pros and cons of screening, including the spectrum of diseases that can and cannot be detected by a retinal fundus image. As part of giving consent, patients also agreed to their images being used anonymously for science and research purposes. The dataset was completely anonymized. No Optretina or BSC employee has access to patient data except for the Optretina technical director (Didac Royo), solely for emergency purposes. The keys for reversing the anonymization are physically stored in a bank. The study was approved by the ethics committee of the Hospital Vall d'Hebron in Barcelona, the research followed the tenets of the Declaration of Helsinki.

Statistical Analyses

A confusion matrix was created from the validation set in all five CNNs. Images were divided based on the human and algorithm classification. Receiver operating

characteristics (ROC) curves, area under the curve (AUC), accuracy, and all other calculations were made using SPSS 15. The number of images used for each algorithm depended on the number of good images with the disease in the Optretina database.

Results

Table 3 summarizes the outcomes from the different convolutional neuronal networks (CNNs) in the study. Table 4 reflects the confusion matrix in the validation sets, indicating positives and negatives, comparing human and artificial evaluation. ROC curves are seen in Figure 3.

The first CNN differentiating color fundus images from other medical or non-medical images has an AUC of 0.979, with an accuracy of 96%. This algorithm has a sensitivity over 97% and a specificity of 92%. Artificial intelligence for determining if a retinal fundus image was assessable or not had an accuracy of 92%, compared to human determination and an AUC of 0.947. This CNN also had high sensitivity (97%) and specificity (82%).

The algorithm for the right eye/left eye (OD/OS) had an AUC of 0.989 with an accuracy of 97.4%. CNNs for evaluating diseases had an AUC of 0.936 in the case of age-related macular degeneration (AMD) with an accuracy of 86.3%, and an AUC of 0.863 with an 80% accuracy in the case of glaucomatous optic neuropathy. The algorithm for detecting AMD had a sensitivity of 90.2% and a specificity of 82.5%. In the case of glaucomatous optic neuropathy (GON), sensitivity was 76.8% and specificity was 83.8%.

The first cause of misclassification of images in the AMD and the GON algorithm was the presence of borderline cases. In the case of AMD, false negatives were early AMD that was not seen by the CNN as disease; false positives were cases of small drusen in the macular area that were not tagged as AMD by the retinal specialist. False negatives and positives in suspected glaucoma were borderline cases with approximately 7/10 vertical excavation.

Table 3 Outcomes Calculated Using the Confusion Matrix in Different Algorithms

Model	AUC	Accuracy	Sensitivity	Specificity
Image type	0.979 (0.978–0.981)	0.960 (0.957–0.962)	0.977 (0.976–0.979)	0.924 (0.920–0.929)
Eye	0.989 (0.988–0.991)	0.974 (0.973–0.977)	0.983 (0.981–0.985)	0.966 (0.963–0.970)
Image quality	0.947 (0.945–0.948)	0.918 (0.916–0.919)	0.969 (0.968–0.970)	0.818 (0.814–0.821)
AMD classification	0.936 (0.922–0.946)	0.863 (0.845–0.877)	0.902 (0.880–0.916)	0.825 (0.799–0.849)
GON classification	0.863 (0.827–0.894)	0.803 (0.760–0.836)	0.768 (0.710–0.824)	0.838 (0.784–0.888)

Table 4 Validation Confusion Matrix from the Five CNNs: 1. Color Fundus Photography (CFP) versus Other Images. 2. Good Quality Retinal Fundus Images. 3. Right Eye versus Left Eye (OD/OS). 4. Age Related Macular Degeneration (AMD). 5. Glaucomatous Optic Neuropathy (GON)

Filetype		
Prediction	False	True
	31%	2%
False	3%	64%
	True	
Laterality	False	True
	48%	1%
Prediction	2%	49%
	True	
Quality assessment	False	True
	28%	2%
False	6%	64%
	True	
AMD	False	True
	42%	5%
Prediction	9%	44%
	True	
GON	False	True
	42%	11%
Prediction	8%	39%
	True	

Discussion

For the last three years, artificial intelligence has been considered useful for detecting certain pathologies in retinal fundus images. Studies show spectacular results for the detection of diabetic retinopathy,^{14–17} macular degeneration,^{18,26} and cases of suspected glaucoma.^{15,27,28}

Optretina performs general screening for the main diseases that affect the central retina. The screening is based on retinal fundus images that use mainly table-top non-mydiatic cameras (NMCs), although portable cameras and optical coherence tomography (OCT) are also used. The readings are conducted by specialist retina ophthalmologists who subsequently issue a report on the state of the patient's two central retinas. We think, in a near future artificial intelligence could improve the times and quality of reading, resulting in our readers having better outcomes than human-only readings. In order for this to occur, we must first guarantee the quality of the images. Since the screening is based on retinal fundus images, we consider the first three algorithms to be useful when the photographer (optometrist, general practitioner, or technician)

uploads the images on the telemedicine platform; in this way, the system can detect that there is at least one retinal fundus image for each eye and that the image is of a good enough quality to be evaluated.

Unlike other publications that aim to replace the human reading, we will like to explore future models based on the idea that a hybrid reading—or the use of artificial intelligence (AI) as a diagnostic aid tool can, at this time, be a better solution for both guaranteeing the quality of the readings and decreasing costs and time invested, while also causing fewer ethical or legal problems. However, when working with images from different types of cameras, and even non-retinal fundus image images, studying the image quality is even more important.

In the near future, screening for retinal diseases will most probably be affected by the increasing ease with which images will be able to be produced due to the decreasing cost of NMCs and the use of smartphones and portable digital devices. As other authors have described, these procedures should be autonomous with little requirement for trained personnel.²⁹ It has been well described that the first step necessary for a screening algorithm is the recognition and the exclusion of insufficient image quality.²⁹ This is what our first CNNs dealt with. In our case, algorithms for identification and quality assurance have levels of accuracy that are better than human-only. Retinal landmarks have traditionally been used for recognizing a retinal image and evaluating its quality,^{30,31} but we used two CNNs without any intervention or segmentation, thereby obtaining high levels of AUC for the algorithms.

To guarantee the quality of the images, some authors perform a manual evaluation of their databases;^{16,29} while this system is suitable for retrospective studies, using it in prospective studies is not ideal. Other authors point out the need for training and certifying photographers,¹⁹ which certainly substantially improves the image quality. We believe, however, that the use of quality algorithms may be a more appropriate measure: it has a short feedback circle, even during the image taking process, thus offering a greater guarantee of quality and more cost-effectiveness. Other authors have also used quality algorithms, as was the case with Varadarajan et al,²⁰ who had to exclude 12% of the images of one of their datasets due to poor quality.

In the case of multiple acquisition devices and different types of images—red free, color fundus photography (CFP), autofluorescence, OCT, non-medical images—AI could be used as a cascade of filters. First, we could evaluate what kind of image we have, then the eye (in

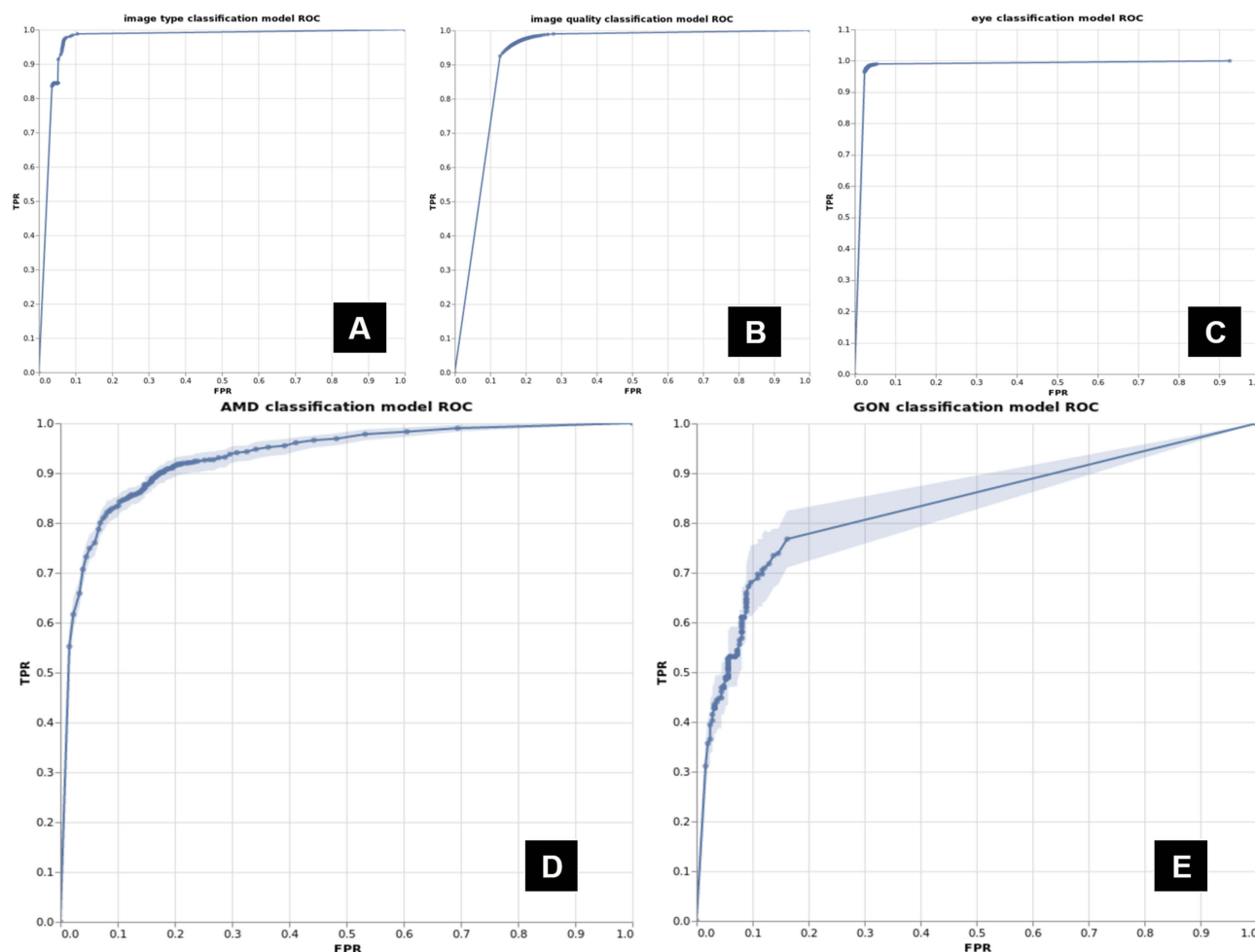


Figure 3 Receiver operating characteristic (ROC) curves from the different CNNs. (A) Image type. (B) Quality assessment. (C) Laterality. (D) AMD. (E) GON.

our model it is important to differentiated laterality, we receive images from different cameras, automatically upload in the platform, not all devices have differentiated file from right or left eye) and if they are of sufficient quality to be assessed; then we could apply CNNs for different pathologies—in our case AMD and GON.

The use of AI could decrease the costs of screening tools in the near future, allowing mass screening for retinal disease in the general population and not just in known diabetics. The incidence of AMD and glaucoma is currently increasing around the world. These two diseases are the second and third most frequent causes for visual impairment in Europe,³² which is why we based our research on them. Other publications on automated detection of referable AMD with accuracies of 90% can be found in the literature.^{18,26} With our CNNs, we have achieved 86% accuracy, because we have included all types of AMD, including the early stages; at this point, one of the most important

reasons for misclassification is difficulty in differentiating between early stages of AMD and normal retinas. We believe that it is worthwhile for future research to study the classification of AMD stages using AI in CFP, as other authors demonstrate using OCT.³³ For the purposes of our current research, we prefer to identify any stage of AMD in order to create future algorithms for classification. We believe that although early AMD will not threaten visual acuity in the short term, it should still be studied and monitored by the ophthalmologist. In a recent review, Schmidt-Erfurth et al indicate the limited role of CFP for advanced stages of AMD compared to OCT screening,²⁹ but for a general screening platform it may make sense to start with color fundus photography due to cost considerations. Also AI has detected patterns that are invisible to the human eye in CFP such as gender, age, and smoking status. Detection of advanced AMD in retinal fundus images should therefore be studied.

Screening for glaucoma is much more complicated than doing so for other ophthalmological diseases. Authors have had the best outcomes with a combination of the slit lamp, OCT retinal nerve fiber layer (RNFL), and visual field tests, but this combination is not the modality of choice when we want to consider devices that are low cost, widely available, and easy to use.²⁹ In the latter case, we are limited to detecting suspected glaucoma by visualizing only the optic disc, similar to Ting et al, who obtained a 94% accuracy for possible glaucoma in a dataset of 125,000 diabetic patients.¹⁵ In our case, with only 4720 images in the suspected glaucoma algorithm, we achieved levels of accuracy of over 80%. Further research with a better-defined glaucoma database would be desirable in order to reach the maximum potential for AI in screening for this disease.

We believe the kind of technology discussed in this paper could be used not only for prospective purposes, but also in retrospective studies, for classification in non-tagged big datasets such as hospitals or ophthalmological clinics. We use AI for classification, but one of the most exciting types of future implementation of CNNs could be the prediction of disease progression or clinical data.²⁹ In the near future, application of big data will require a well-tagged database, and AI could be very helpful in this arena—for massive retrospective classifications of image datasets

The accuracy of CNNs appears to be related primarily to two factors. The first factor is the similarity of classes, and the second is the number of well-classified images. Regarding the first factor, it is necessary to explore the use of deep learning for fine-grained image classification techniques, given that it is easy for CNNs to differentiate a horse from a car but the complexity increases when considering the classification of similar images. Concerning the second factor, it is easier to obtain high levels of accuracy if we have large and well-classified datasets. Although we used data augmentation techniques in order to improve validation accuracy, having a large dataset could be very important for obtaining a robust and more general prediction model. We believe false positives could be avoided with larger and better-tagged datasets. In our case, the main cause of false positives and negatives were borderline cases—images with which even retinal specialists have low interobserver agreement.

From an engineering point of view, we have shown that it is possible to propose an effective CNN architecture with high validation accuracy and fewer trainable parameters. We would like to explore the combination of

different types of architecture in a future study in order to try to predict other retinopathies.

One of the limitations of this study is the inclusion of multiples images, from multiple devices and photographers, some of which only include 40°, 45°, and 50° of the central retina. However, this same limitation could also be a strength for the resulting algorithms, which could be more reliable for use in the real world without having to worry about camera brands or types. The only significant barrier would be the assessable AI filter.

From an ethical and legal point of view, AI is far from being reliable for general screening for retinal diseases. AI can detect DR, AMD, or suspected glaucoma perfectly, but we have to see what would happen—and who is responsible in the event the algorithm fails—when the system evaluates an image with uncommon findings, such as dystrophies, tumors, or minor infections. In our opinion, this problem could be solved with bigger datasets that include rare diseases. In the end, the more we teach the AI system, the more it learns. Meanwhile, AI could be used to ameliorate the cost-effectiveness of using hybrid systems supervised by a medical expert, raising doctor's performance beyond unaided human-only abilities, also would be desirable for further investigations evaluate AI-assisted interpretation in other non-retinal specialist personnel such as general ophthalmologists or primary care physicians.

Deep learning opens an exciting chapter for the future. AI algorithms seem to be reliable for determining whether or not an image is CFP. Those algorithms are useful for evaluating the quality of the retinal fundus image and also for determining DR, AMD, and suspected glaucoma. As discussed earlier, the next steps will be to continue “feeding the beast,” providing more and more classified images to improve accuracy and creating new CNNs for other diseases. One of the most exciting challenges ahead is designing new networks for making the normal/abnormal algorithm over 90% of accuracy. Apart from technical and ethical limitations, medical doctors must also establish how to incorporate the algorithm properly into our clinics and offices in order to control it and ensure a good use of AI technology in medicine.

Disclosure

MAZ is founder and Medical director in Optretina. DRF is founder and CTO in Optretina. OF is employee in Optretina. The authors report no other conflicts of interest in this work.

References

- Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Glob Health*. 2014;2(2):e106–e116. doi:10.1016/S2214-109X(13)70145-1
- Leggio GM, Bucolo C, Platania CB, Salomone S, Drago F. Current drug treatments targeting dopamine D3 receptor. *Pharmacol Ther*. 2016;165:164–177. doi:10.1016/j.pharmthera.2016.06.007
- Surendran TS, Raman R. Teleophthalmology in diabetic retinopathy. *J Diabetes Sci Technol*. 2014;8(2):262–266. doi:10.1177/1932296814522806
- Mansberger SL, Gleitsmann K, Gardiner S, et al. Comparing the effectiveness of telemedicine and traditional surveillance in providing diabetic retinopathy screening examinations: a randomized controlled trial. *Telemed J E Health*. 2013;19(12):942–948. doi:10.1089/tmj.2012.0313
- Chan CK, Gangwani RA, McGhee SM, Lian J, Wong DS. Cost-effectiveness of screening for intermediate age-related macular degeneration during diabetic retinopathy screening. *Ophthalmology*. 2015;122(11):2278–2285. doi:10.1016/j.ophtha.2015.06.050
- De Bats F, Vannier Nitenberg C, Fantino B, Denis P, Kodjikian L. Age-related macular degeneration screening using a nonmydriatic digital color fundus camera and telemedicine. *Ophthalmologica*. 2014;231(3):172–176. doi:10.1159/000356695
- Fierson WM; American Academy of Pediatrics Section on Ophthalmology; American Academy of Ophthalmology; American Association for Pediatric Ophthalmology and Strabismus; American Association of Certified Orthoptists. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics*. 2013;131(1):189–195.
- Pérez MA, Bruce BB, Newman NJ, Bioussé V. The use of retinal photography in nonophthalmic settings and its potential for neurology. *Neurologist*. 2012;18(6):350–355. doi:10.1097/NRL.0b013e318272f7d7
- Frost S, Brown M, Stirling V, Vignarajan J, Prentice D, Kanagasingam Y. Utility of ward-based retinal photography in stroke patients. *J Stroke Cerebrovasc Dis*. 2017;26(3):600–607. doi:10.1016/j.jstrokecerebrovasdis.2016.11.112
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;1097–1105.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118. doi:10.1038/nature21056
- Liu Y, Gadepalli K, Norouzi M, et al. Detecting cancer metastases on gigapixel pathology images. *arXiv*. 2018;1806.07064v1.
- Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv*. 2017.
- Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124(7):962–969. doi:10.1016/j.ophtha.2017.02.008
- Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318(22):2211–2223. doi:10.1001/jama.2017.18152
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410. doi:10.1001/jama.2016.17216
- Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57:5200–5206. doi:10.1167/iovs.16-19964
- Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol*. 2017;135(11):1170–1176. doi:10.1001/jamaophthalmol.2017.3782
- Verbraak FD, Abramoff MD, Bausch GCF, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. *Diabetes Care*. 2019;42(4):651–656. doi:10.2337/dc18-0148
- Varadarajan AV, Poplin R, Blumer K, et al. Deep learning for predicting refractive error from retinal fundus images. *Invest Ophthalmol Vis Sci*. 2018;59(7):2861–2868. doi:10.1167/iovs.18-23887
- Zapata MA, Arcos G, Fonollosa A, et al. Telemedicine for a general screening of retinal disease using nonmydriatic fundus cameras in optometry centers: three-year results. *Telemed J E Health*. 2017;23(1):30–36. doi:10.1089/tmj.2016.0020
- Chollet F. Keras; 2015. Available from: <https://github.com/fchollet/keras>. Accessed January 29, 2020.
- Ferris FL 3rd, Wilkinson CP, Bird A, et al.; Beckman Initiative for Macular Research Classification Committee. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013;120(4):844–851. doi:10.1016/j.ophtha.2012.10.036
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2016: 770–778.
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: IEEE conference on computer vision and pattern recognition. *arXiv*. 2016.
- Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med*. 2017;82:80–86. doi:10.1016/j.compbimed.2017.01.018
- Kim SJ, Cho KJ, Oh S, Liu B. Development of machine learning models for diagnosis of glaucoma. *PLoS One*. 2017;12(5):e0177726. doi:10.1371/journal.pone.0177726
- Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125(8):1199–1206. doi:10.1016/j.ophtha.2018.01.023
- Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. *Prog Retin Eye Res*. 2018;67:1–29. doi:10.1016/j.preteyeres.2018.07.004
- Moccia S, Banali R, Martini C, et al. Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images. *Magma*. 2019;32(2):187–195. doi:10.1007/s10334-018-0718-4
- Molina-Casado JM, Carmona EJ, García-Feijóo J. Fast detection of the main anatomical structures in digital retinal images based on intra- and inter-structure relational knowledge. *Comput Methods Programs Biomed*. 2017;149:55–68. doi:10.1016/j.cmpb.2017.06.022
- Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob Health*. 2013;1(6):e339–e349. doi:10.1016/S2214-109X(13)70113-X
- Keremany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172(5):1122–1131.e9. doi:10.1016/j.cell.2018.02.010

Clinical Ophthalmology

Dovepress

Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed

Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-ophthalmology-journal>