

Title: Glaucoma patient screening from online retinal fundus images via Artificial Intelligence

Authors: Pablo Franco¹, David Coronado-Gutierrez¹, Carlos Lopez¹, Xavier P. Burgos-Artizzu¹.

Transmural Biotech S. L. Barcelona, Spain.

e-mail: xpburgos@transmuralbiotech.com

Abstract

Objectives: To design and evaluate a novel automated glaucoma classifier from retinal fundus images.

Methods: We designed a novel Artificial Intelligence (AI) automated tool to detect glaucoma from retinal fundus images. We then downloaded publicly available retinal fundus image datasets containing healthy patients and images with verified glaucoma labels. Two thirds of the images were used to train the classifier. The remaining third of the images was used to create several cross-validation evaluation sets with a realistic glaucoma prevalence, to evaluate the classifier's performance in a screening scenario.

Results: 10,658 retinal fundus images from seven different sources were found and downloaded. They were randomly divided into 7,106 for training and 3,551 for validation. Glaucoma prevalence was 24%. Using the validation set, we created 50 random sets of 1,000 images with a 5% glaucoma prevalence. On these sets, the classifier reached a detection rate of 84.1% (CI 95= $\pm$ 1.1%) and 95.8% (CI 95= $\pm$ 0.2%) specificity.

Conclusions: A novel glaucoma classifier from retinal fundus images showed promising results as screening tool on a large cohort of patients. A large clinical study is needed to confirm these results.

Intro

Glaucoma is a neurodegenerative condition that affects the eye, related to increased intraocular pressure¹. When left untreated, patients may experience a gradual loss of the visual field and can even lose their sight completely². Although there exist many types of glaucoma, all are characterized by damage to the optic nerve that can eventually lead to blindness³.

In fact, glaucoma is the second leading cause of blindness worldwide after the cataract, as it affects more than 70 million people⁴. Furthermore, blindness caused by glaucoma is irreversible unlike that caused from cataracts⁵. For this reason, early detection of glaucoma is essential.

Glaucoma can be diagnosed using different techniques. One of the most common is tonometry. This test is done using a device called a tonometer that measures the pressure inside the eye, called intraocular pressure (IOP). The normal value range for IOP is 12-22 mm. Most cases that exceed 22mm are diagnosed with glaucoma⁶. However, some people may have glaucoma at pressures between 12- and 22-mm Hg. IOP is unique to each person, so this method is controversial⁷.

Another common technique is the visual assessment of the optic nerve. The reason is that one of the typical characteristics of glaucoma is that it produces an excavation or depression in the head of the optic nerve⁸. For this visual assessment of the optic nerve two imaging techniques are used: non-invasive retinal fundus images taken by a digital retinal fundus camera and information from a coherence tomography (OCT)⁹.

Retinal fundus imaging is the preferred technique for a first patient screening due to its simplicity, world-wide availability and low cost compared to OCT, which is more costly¹⁰. Prior studies on automated glaucoma detection from retinal fundus images can be divided into two different categories: 1) Those based on retinal clinical indicators and 2) those using image texture analysis techniques.

The first type of works rely on a (manual or automatic) segmentation of the optic disc followed by measurements such as the cup-disc ratio (CDR), the cup-disc area, the ISNT rule, the rim-disc ratio (RDR) or the NRR area^{11 12 13 14}. For example, Wong et al.¹⁵ and Joshi et al.¹⁶ presented different methods to measure CDR, Diaz-Pinto et al.¹⁷ used CDR, ACDR or ISNT indicators or Wong DWK et al.¹³ used CDR and NRR indicators. The main limitation of these

methods is that they require a very accurate disc/cup segmentation and that variations on how the disk is segmented can produce significant changes in performance⁸.

The second type of approaches use image texture analysis techniques, searching retinal fundus images for patterns that determine the presence or not of glaucoma. Machine/Deep Learning is often used to learn from the images what patterns reveal the presence of glaucoma. For example, Chen et al.¹⁸ used an ensemble network named DENet to automatically classify glaucomatous fundus image directly without segmentation. Abbas¹⁹ also used CNN to classify images merging 1200 images achieving specificity of 0.98 and sensitivity of 0.84. Also Orlando et al.²⁰ experimented with the use of CNNs applying techniques like vessels extraction, contrast-limited adaptive histogram equalization (CLAHE) or cropping around the optic nerve head (ONH). Other relevant works (using Deep Learning) are Bajwa et al.²¹, Syna Sreng et al.²² or Chakravarty et al.²³. Finally, the REFUGE challenge²⁴ for glaucoma assessment from retinal fundus images has attracted the development of many strategies based on AI by several research groups. Even though the results of many of these prior studies are promising and already point towards the potential of AI for glaucoma screening of patients, none of these studies had as main objective the evaluation of AI in a real screening scenario. As shown later in the Methods section in Table 1, the prevalence of glaucoma in the public glaucoma datasets show a high variance and are very far from the reported real world-wide glaucoma prevalence of 3.5% for population aged 40-80 years⁴.

The objective of this study is to fill this gap and evaluate a novel glaucoma AI based classifier designed for patient screening. In order to do so, we adjusted test prevalence to a more realistic number according to large studies of glaucoma⁴ and evaluated our novel AI classifier on this test to judge its potential as screening tool.

Materials and Methods

Retinal fundus image datasets

Different public datasets for the evaluation of glaucoma from fundus images are available. To perform this retrospective study, we used datasets Refuge²⁴, Origa²⁵, LAG²⁶, Drishti-GS1²⁷, sjchoi86-HRF²⁸, HRF²⁹ and ODIR³⁰. Figure 1 shows image examples from these datasets. Description of each dataset is as follows:

- **HRF**^{29,31}: This database contains 15 images of healthy patients, 15 images of patients with diabetic retinopathy and 15 images of glaucomatous patients. They were

captured by a Canon CR-1 fundus camera with a field of view of 45 grades with a resolution of 3504×2336 px. Binary gold standard vessel segmentation images generated by clinicians are available for each image.

- **LAG**²⁶: This database contains 4,584 fundus images with 1,711 positive and 3,143 negative glaucoma samples obtained from Beijing Tongren Hospital. Each fundus image is diagnosed by qualified glaucoma specialists, taking the consideration of both morphologic and functional analysis, i.e, intra-ocular pressure, visual field loss and manual optic disc assessment.
- **ODIR**³⁰: This database contains images from 5000 patients collected by Shanggong Medical Technology Co., Ltd. from different hospitals/medical centres in China. In these institutions, fundus images are captured by various cameras in the market, such as Canon, Zeiss and Kowa, resulting into varied image resolutions. They classify patient into eight labels including normal (N), diabetes (D), glaucoma (G), cataract (C), AMD (A), hypertension (H), myopia (M) and other diseases/abnormalities (O) based on both eye images. Glaucoma patient age ranged between 24 and 91 years old.
- **sjchoi86-HRF**²⁸: This dataset contains 601 fundus images divided into 4 groups: normal (300 images), glaucoma (101 images), cataract (100 images) and retina disease (100 images).
- **REFUGE**²⁴: This dataset includes 1200 fundus images with ground truth segmentations and clinical glaucoma labels. They were acquired with two different fundus cameras: Zeiss Visucam 500 (2124×2056 pixels) and Canon CR-2 (1634×1634 pixels).
- **ORIGA(-light)**²⁵: This dataset contains 650 retinal images collected during the Singapore Malay Eye Study (SiMES) population based study, conducted over a 3 year period on adults aged 40 to 80 years. All labels were annotated by trained professionals from Singapore Eye Research Institute.
- **Drishti-GS1**²⁷: this dataset contains 101 retinal images. These images were collected at Aravind eye hospital (India) which correspond to patients between 40-80 years of age with a roughly equal number of males and females. Images have been labelled by 4 clinical experts.

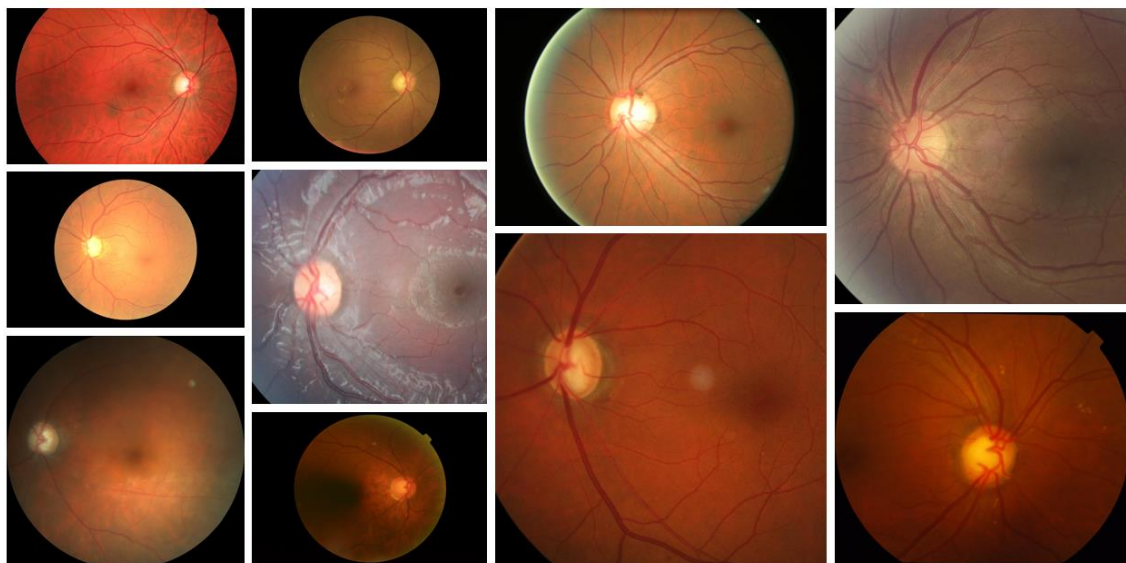


Figure 1. Retinal fundus image examples from the different datasets used.

In each dataset, we downloaded only normal and glaucoma images. Table 1 summarizes the main statistics of each resulting dataset. These datasets have a large variance in terms of glaucoma prevalence, ranging from 10% to 69%. Considering that global glaucoma prevalence for patients from 40 to 80 years is reported at 3.5%⁴, not only there is a high variance, but all datasets are overestimating the prevalence of glaucoma. Reporting results on these datasets, as most prior studies do, makes difficult to judge the potential of the proposed tools for screening.

DATASET	Glaucoma	Normal	Total	Glaucoma prevalence	Origin
LAG	1711	3143	4584	37%	China
ODIR	326	3095	3241	10%	China
HRF	15	15	30	50%	Germany
SJCHOI86-HRF	101	301	402	25%	--
DRISHTI-GS1	70	31	101	69%	India
REFUGE	120	1080	1200	10%	China
ORIGA	168	482	650	26%	Malaysia
TOTAL	2511	8147	10658	24%	

Table 1. Statistics of the different datasets used for the study

To evaluate the potential of the methods proposed for glaucoma screening, while still using data from all these datasets, we designed a specific testing framework which uses a prevalence close to the 3.5% prevalence reported in large studies in population aged 40-80 years⁴. In order to do so, we first merged all datasets, and then partitioned the images into a training set (66%) and a validation set (33%) keeping the prevalence of each original dataset. Then, for the validation set, for each dataset we created 50 random subsets of 1000 images, each containing

5% glaucoma cases. These sets were used to report average performance and confidence intervals.

Image pre-processing

As stated in the Intro, in theory, glaucoma affects mainly the optic disc and the area near it. Therefore, most prior works suggest that only the optic disc part of the image should be used. However, we found that this part of the image can be very small in many images and convey very little signal when the image contrast is not set correctly. We explored different alternatives, to establish what part of the image is better suited to detect glaucoma and found that using the full image reached best results. Please see Supp Figure 1 and Supp Table 1 for more info on these experiments. The only preprocessing performed consists in removing the black borders from the input image, see Supp Figure 1.

Automatic AI glaucoma classifier

A novel glaucoma classifier was developed to classify glaucoma from retinal fundus images. The classifier is a single Deep Convolutional Neural Network³² applied on the entire image (after pre-processing to remove the uninformative black borders, see Supp Figure 1). During the development stage, several possible CNN architectures were benchmarked using our AI online platform. The final chosen model was an adapted version of a ResNet-50 architecture³³.

We relied on Transfer Learning, initially training the CNN on large datasets such as ImageNet Large Scale Visual Recognition Challenge⁶⁰ and Microsoft coco³⁴. Then, the net was adapted to the binary classification task and fully retrained (allowing changes to the entire network) using 66% of all the images downloaded. More specifically, 7,106 images were used for training, with a 24% glaucoma prevalence (1,706 images with glaucoma, 5,400 normal). After training, the net was deployed and applied to all the validation images, outputting full probability score (softmax) for glaucoma.

CNN implementation details: The network was trained using Transmural Biotech's AI online platform. The ResNet50 classifier was trained using softmax cross-entropy loss and adam optimizer. 10% of the training set was used as validation set. The first training using ImageNet and Coco images was performed for 90 epochs with a 5-epoch warm-up. Then, the CNN was trained on the training data for a maximum of 15 epochs, early stopping if loss on validation set was not improved for 5 consecutive epochs. Learning rate was adjusted using a cosine decay starting at 1e-4. Batch size was 64 and weight decay was 0.9.

Data augmentation techniques were used during training to improve the classifier robustness and avoid overfitting. Validation images were processed without any alteration nor augmentation tricks.

Statistical analysis

Statistical analysis was performed using Matlab (Mathworks, USA). The final classifier, once trained, was applied to all validation available images, storing the output glaucoma probability score. The scores were used to draw Receiver Operator Characteristic (ROC) curves and compute full Area Under the Curve (AUC) Then, the ROC curves were used to establish the optimal cutoff points in terms of accuracy. Detection rate, false positive rate (FPR), positive and negative predictive values (PPV and NPV) and positive and negative likelihood ratios (LR+ and LR-) were calculated. We repeated this for each cross-validation set, reporting both the average and 95% Confidence Intervals on the sets.

Results

Figure 1 shows the ROC curve for glaucoma detection on the 50 cross-validation sets using a total of 3,551 images. AUC was 91.1% $\pm$ 0.5% and accuracy was 95.3% $\pm$ 0.2. Metric scores for a cut-off maximizing accuracy are shown in Table 2. The detector achieved a 84.1% detection rate at 95.8% $\pm$ 0.2% Specificity. Negative Predictive Value was 99.1% $\pm$ 0.1%.

This scenario would have implied, for a hypothetical population of 1000 people with a prevalence of 5% (50 persons with glaucoma), that of the 50 patients which have glaucoma the CNN would have successfully detected 42, leaving 8 undetected. On the other hand, it would have falsely flagged 39/950 patients without glaucoma disease as having glaucoma.

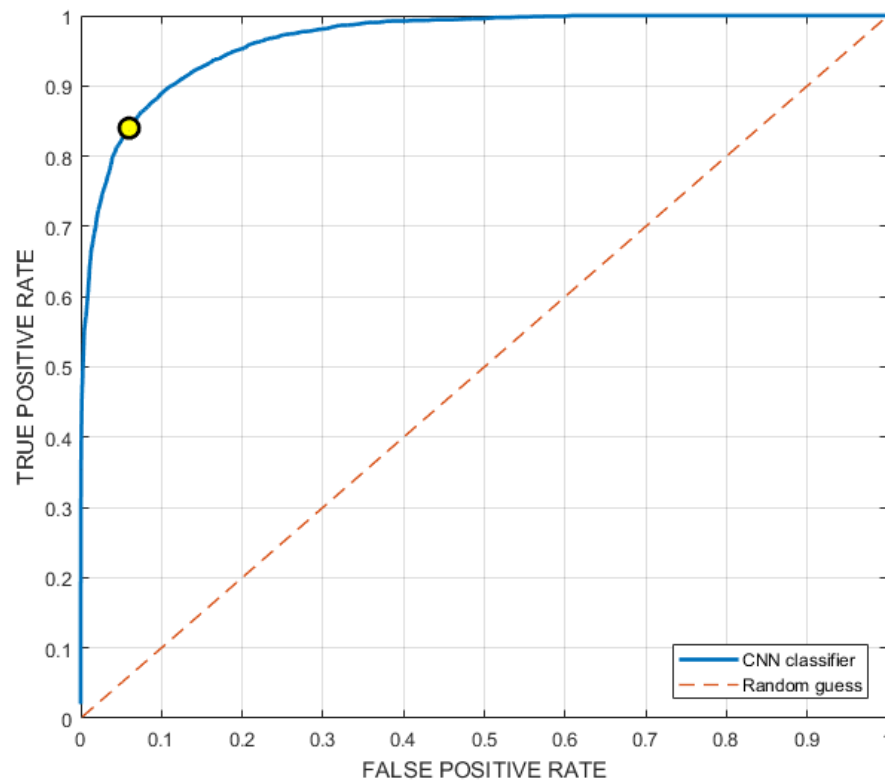


Figure 1. ROC curve for prediction task. The yellow circle marks the optimum cut-off point in terms of accuracy.

Method	ACC	SENS	SPEC	PPV	NPV	AUC
CNN	95.3 % (± 0.2 %)	84.1 % (± 1.1 %)	95.8 % (± 0.2 %)	52.4 % (± 1.1 %)	99.1 % (± 0.1 %)	91.1% (± 0.5 %)

Table 2. Metric scores. Values shown as Mean($\pm$ CI 95). ACC=Accuracy. SENS=Sensitivity. SPEC=Specificity. PPV=Positive Predictive Value. NPV=Negative Predictive Value. AUC=Area Under the (ROC) Curve.

For completeness, supp Figure 2 and Supp Table 2 show individual results on each dataset. Furthermore, although a direct comparison on same data was not possible, Supp Table 3 shows a comparison of our results with respect with prior reported results on some of the datasets. Our novel glaucoma classifier reached similar or superior performance compared against prior methods.

Figure 2 shows example correct glaucoma detections, with the class-activation maps³⁵ of the CNN superimposed on top of the image. The CNN appears to be correctly detecting as region of interest the surroundings of optic disc marked in red colour.

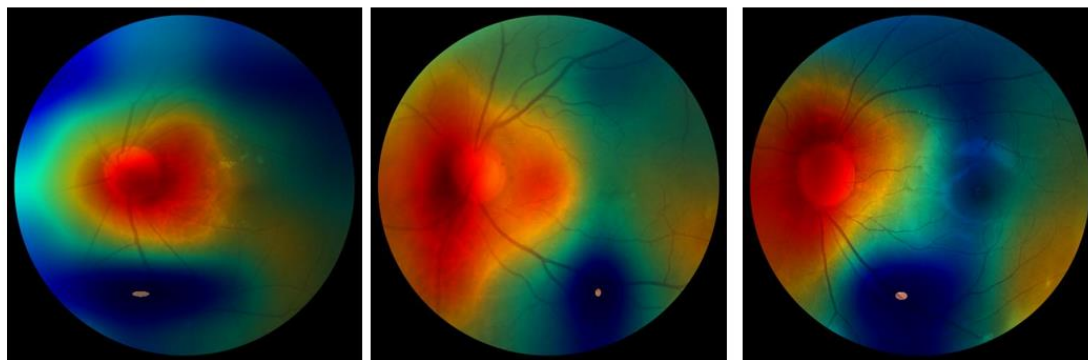


Figure 2. Correct glaucoma detection examples. Original Images are superimposed with the CNN's Class-activation maps³⁵ (red= hot, blue=cold), which provide an estimate of what parts of the image is the CNN model focusing on for classifying the images as glaucoma.

Discussion

Main findings

We designed a novel glaucoma classifier based on fully automated analysis of retinal fundus images. The classifier, using state-of-the-art Deep Learning technology, was designed with a clear focus on screening. Tested on thousands of images from seven different sources, the proposed model was capable of detecting 84.1% of the cases at low (0.9%) false positive rates. The classifier performance showed a low variance (5% in terms of sensitivity) even though the quality, dimension and compression of the images used was varied, as shown by Figure 1.

Negative Predictive Value was 99.1%, making it ideally suited to be used in a screening scenario. Results of this study are promising and suggest that the deep learning-based algorithm can detect glaucoma with high accuracy from retinal fundus images, which are easier to obtain in a large population. This highlights its potential application for glaucoma screening.

Compared to reported results from a varied set of prior approaches, our novel glaucoma classifier showed state-of-the-art performance, always reaching similar or superior performance on all the seven datasets tested, which showed a wide variety in resolution, format, contrast and quality of retinal fundus images. The main advantage of our approach is that it does not require a manual optic disc segmentation and is capable of exploiting information from the entire image. A full analysis of this comparison is shown in Supp Table 3.

Clinical implications

In most cases, a single measurement does not provide enough information for a correct diagnosis of glaucoma³⁶. Therefore, traditional glaucoma screening tools typically consider various aspects in combination. During a regular ophthalmological examination, various tests are performed, such as a visual acuity checking, an intraocular pressure measurement or a preliminary analysis of the cup / disc ratio to detect suspicions of the disease. These tests are integrated into different algorithms that, based on the results, will determine if a further detailed evaluation is necessary³⁷.

Several studies have evaluated the effectiveness of current glaucoma screening tools. R Khandeka et al.³⁸ carried out a study on 3324 patients using ocular pressure and visual inspection of retinal fundus changes as parameters. Using only ocular pressure, the reported sensitivity was 49.7% at 95.6% specificity. Using only disc changes, sensitivity was 48.4% at 97.9% specificity. Sensitivity improved to 67.3% at 96.5% specificity when combining both. They concluded that these parameters were not accurate enough for a trustworthy glaucoma screening. Also in 2013 S-Farhad Mohammadi et al.³⁹ evaluated a glaucoma screening algorithm which included examination of the optic disc for vertical cupping, asymmetry, tonometry and automated perimetry over a population of 124 patients (prevalence 4.8%). They reported a Positive Predictive Value of 30%.

Other studies have analysed the performance of screening using other techniques as the Heidelberg Retinal Tomography [HRT]. Healey et al.⁴⁰ reported 46% of sensitivity and 91% of specificity, Saito et al.⁴¹ reported a sensitivity of 39% and 96% of specificity. The results from these 2 studies predict that a screening program of 10 000 individuals from a general population with a 2% prevalence of OAG would detect only 80–90 out of the 200 affected, and falsely detect glaucoma on 390–880 normal individuals.

In conclusion, results of the different studios suggest that screening to detect all individuals with glaucoma remains an unsolved challenge⁴². In this context, the method proposed in this study, which is not invasive, automated, capable of providing results in seconds and achieving a sensitivity of 84.1% at 95.8% specificity, could, individually or as an additional step in the screening algorithm, add additional information in order to obtain a more accurate early glaucoma diagnosis.

The proposed methods will be integrated into an online platform, making them available to any professional.

Strengths and limitations

This study has several strengths. We designed a novel glaucoma classifier fully automated, which is capable to extract information from the full retinal fundus image. The novel classifier was evaluated on a large set of retinal fundus images online which were collected from different centers, showing a wide variety in terms of image resolution, format and quality, mimicking a real clinical scenario. The classifier showed state-of-the-art results when compared directly with prior reported results from a large variety of studies. Furthermore, we combined combine all available retinal fundus image datasets while adjusting the prevalence of glaucoma in order to evaluate our classifier in a scenario realistic from a clinical point of view. Finally, to further imitate realistic conditions, we did not filter the images by imposing our own quality constraints and used all online images available directly.

We acknowledge two main limitations. Firstly, the study was conducted on online images and therefore we weren't able to neither double-check clinical outcomes of the images nor verify the role image acquisition had on performance (a known Deep Learning issue), since the specific image acquisition details from each source was not provided. Secondly, each source had varied glaucoma prevalence and even though we tried to mitigate this issue by adjusting prevalence manually and using images acquired by several different centers, these results cannot substitute a proper, large clinical validation study, which should and will be pursued.

As future work, information provided by AI methods could be combined with all other information about the patient (presence of a positive family history, medical history, ethnicity, age, gender). This information was not available in the public datasets used. Additional evaluation using these variables in order to detect relations would be very interesting.

Conclusions

A novel glaucoma classifier from retinal fundus images showed promising results as screening tool on a large cohort of patients. The new tool could help in the first screening of patients, improving screening accuracy compared to visual inspection by the oftalmologists. A large clinical study is needed to confirm these results.

References

1. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: A review. *JAMA - J Am Med Assoc.* 2014;311(18):1901-1911. doi:10.1001/jama.2014.3192

2. ResearchGate.
https://www.researchgate.net/publication/336359605_REFUGE_Challenge_A_Unified_Framework_for_Evaluating_Automated_Methods_for_Glaucoma_Assessment_from_Fundus_Photos/link/5da63e06299bf1c1e4c37326/download. Accessed July 26, 2020.
3. Types of Glaucoma | National Eye Institute. <https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/glaucoma/types-glaucoma>. Accessed July 27, 2020.
4. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090. doi:10.1016/j.optha.2014.05.013
5. Kingman S. Glaucoma is second leading cause of blindness globally. *Bull World Health Organ*. 2004;82(11):887-888. doi:/S0042-96862004001100019
6. Five Common Glaucoma Tests | Glaucoma Research Foundation.
<https://www.glaucoma.org/glaucoma/diagnostic-tests.php>. Accessed July 17, 2020.
7. High Eye Pressure and Glaucoma | Glaucoma Research Foundation.
<https://www.glaucoma.org/gleams/high-eye-pressure-and-glaucoma.php>. Accessed July 27, 2020.
8. Díaz Pinto AY. Machine Learning for Glaucoma Assessment using Fundus Images. June 2019. doi:10.4995/Thesis/10251/124351
9. Sharma P, Sample PA, Zangwill LM, Schuman JS. Diagnostic Tools for Glaucoma Detection and Management. *Surv Ophthalmol*. 2008;53(6 SUPPL.):S17. doi:10.1016/j.survophthal.2008.08.003
10. Chalakkal RJ, Abdulla WH, Hong SC. Fundus retinal image analyses for screening and diagnosing diabetic retinopathy, macular edema, and glaucoma disorders. In: *Diabetes and Fundus OCT*. Elsevier; 2020:59-111. doi:10.1016/b978-0-12-817440-1.00003-6
11. Kumar JRH, Seelamantula CS, Kamath YS, Jampala R. Rim-to-Disc Ratio Outperforms Cup-to-Disc Ratio for Glaucoma Prescreening. *Sci Rep*. 2019;9(1). doi:10.1038/s41598-019-43385-2
12. Fernandez-Granero MA, Sarmiento A, Sanchez-Morillo D, Jiménez S, Alemany P, Fondón I. Automatic CDR Estimation for Early Glaucoma Diagnosis. *J Healthc Eng*. 2017;2017.

doi:10.1155/2017/5953621

13. Das P, Nirmala SR, Medhi JP. Diagnosis of glaucoma using CDR and NRR area in retina images. *Netw Model Anal Heal Informatics Bioinforma*. 2016;5(1). doi:10.1007/s13721-015-0110-5
14. Verdú-Monedero R, Morales-Sánchez J, Berenguer-Vidal R, Sellés-Navarro I, Palazón-Cabanes A. Automatic Measurement of ISNT and CDR on Retinal Images by Means of a Fast and Efficient Method Based on Mathematical Morphology and Active Contours. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 11487 LNCS. Springer Verlag; 2019:361-370. doi:10.1007/978-3-030-19651-6_35
15. Wong DWK, Liu J, Lim JH, et al. Level-set based automatic cup-to-disc ratio determination using retinal fundus images in argali. In: *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology."* Vol 2008. IEEE Computer Society; 2008:2266-2269. doi:10.1109/iembs.2008.4649648
16. Joshi GD, Sivaswamy J, Krishnadas SR. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Trans Med Imaging*. 2011;30(6):1192-1205. doi:10.1109/TMI.2011.2106509
17. Diaz A, Morales S, Naranjo V, Alcoceryz P, Lanzagortayz A. Glaucoma diagnosis by means of optic cup feature analysis in color fundus images. In: *European Signal Processing Conference*. Vol 2016-Novem. European Signal Processing Conference, EUSIPCO; 2016:2055-2059. doi:10.1109/EUSIPCO.2016.7760610
18. Fu H, Cheng J, Xu Y, Liu J. Glaucoma Detection Based on Deep Learning Network in Fundus Image. In: *Advances in Computer Vision and Pattern Recognition*. Springer London; 2019:119-137. doi:10.1007/978-3-030-13969-8_6
19. Abbas Q. Glaucoma-Deep: Detection of Glaucoma Eye Disease on Retinal Fundus Images using Deep Learning. *Int J Adv Comput Sci Appl*. 2017;8(6). doi:10.14569/ijacsa.2017.080606
20. Orlando JJ, Prokofyeva E, del Fresno M, Blaschko MB. Convolutional neural network transfer for automated glaucoma identification. In: Romero E, Lepore N, Brieva J, Larrabide I, eds. *12th International Symposium on Medical Information Processing and*

Analysis. Vol 10160. SPIE; 2017:101600U. doi:10.1117/12.2255740

21. Naseer Bajwa M, Malik MI, Siddiqui SA, et al. Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. 2019. doi:10.1186/s12911-019-0842-8
22. Sreng S, Maneerat N, Hamamoto K, Win KY. Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images. *Appl Sci*. 2020;10(14):4916. doi:10.3390/app10144916
23. Chakravarty A, Sivaswamy J. *Glaucoma Classification with a Fusion of Segmentation and Image-Based Features*.; 2016. doi:10.0/Linux-x86_64
24. Orlando JI, Fu H, Barbossa Breda J, et al. REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal*. 2020;59:101570. doi:10.1016/j.media.2019.101570
25. Zhang Z, Yin FS, Liu J, et al. ORIGA-light : An online retinal fundus image database for glaucoma analysis and research. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*. IEEE Computer Society; 2010:3065-3068. doi:10.1109/IEMBS.2010.5626137
26. Li L, Xu M, Wang X, Jiang L, Liu H. Attention Based Glaucoma Detection: A Large-scale Database and CNN Model. March 2019:1-11. <http://arxiv.org/abs/1903.10831>.
27. Sivaswamy J, Chakravarty A, Datt Joshi G, Abbas Syed T. *A Comprehensive Retinal Image Dataset for the Assessment of Glaucoma from the Optic Nerve Head Analysis*. Vol 2.; 2015.
28. cvblab/retina_dataset: Retina dataset containing 1) normal 2) cataract 3) glaucoma 4) retina disease. https://github.com/cvblab/retina_dataset. Accessed August 26, 2020.
29. High-Resolution Fundus (HRF) Image Database. <https://www5.cs.fau.de/research/data/fundus-images/>. Accessed August 27, 2020.
30. Dataset-数据集 - ODIR-2019 - Grand Challenge. <https://odir2019.grand-challenge.org/dataset/>. Accessed August 27, 2020.
31. Budai A, Odstrcilik J, Kollar R, et al. *A Public Database for the Evaluation of Fundus Image Segmentation Algorithms*.

32. (PDF) Convolutional Networks for Images, Speech, and Time-Series.
https://www.researchgate.net/publication/216792820_Convolutional_Networks_for_images_Speech_and_Time-Series. Accessed September 11, 2020.
33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol 2016-Decem. IEEE Computer Society; 2016:770-778.
doi:10.1109/CVPR.2016.90
34. Microsoft COCO: Common Objects in Context | Request PDF.
https://www.researchgate.net/publication/262049707_Microsoft_COCO_Common_Objects_in_Context. Accessed September 11, 2020.
35. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int J Comput Vis*. 2016;128(2):336-359. doi:10.1007/s11263-019-01228-7
36. Glaucoma Screenings: Challenges and Failures. <http://www.uniteforsight.org/health-screenings/glaucoma-screening>. Accessed September 8, 2020.
37. Glaucoma Screening - EyeWiki. https://eyewiki.aao.org/Glaucoma_Screening. Accessed September 6, 2020.
38. (PDF) Oman Eye Study 2005: validity of screening tests used in the glaucoma survey.
https://www.researchgate.net/publication/23930295_Oman_Eye_Study_2005_validity_of_screening_tests_used_in_the_glaucoma_survey. Accessed September 6, 2020.
39. Mohammadi S-F, Mirhadi S, Mehrjardi HZ, et al. An algorithm for glaucoma screening in clinical settings and its preliminary performance profile. *J Ophthalmic Vis Res*. 2013;8(4):314-320. <http://www.ncbi.nlm.nih.gov/pubmed/24653818>.
40. Healey PR, Lee AJ, Aung T, Wong TY, Mitchell P. Diagnostic accuracy of the Heidelberg Retina Tomograph for Glaucoma: A population-based assessment. *Ophthalmology*. 2010;117(9):1667-1673. doi:10.1016/j.ophtha.2010.07.001
41. Saito H, Tsutsumi T, Araie M, Tomidokoro A, Iwase A. Sensitivity and Specificity of the Heidelberg Retina Tomograph II Version 3.0 in a Population-based Study: The Tajimi Study. *Ophthalmology*. 2009;116(10):1854-1861. doi:10.1016/j.ophtha.2009.03.048
42. Maul EA, Jampel HD. Glaucoma screening in the real world. *Ophthalmology*.

2010;117(9):1665-1666. doi:10.1016/j.ophtha.2009.11.001

Funding: This work in its entirety was supported by Transmural Biotech SL.

Competing interests: All authors are Transmural Biotech employees.

Data and materials availability: All data associated with this study is available in the main text and comes from public datasets. The developed glaucoma detection method will be made available through a novel platform.