

Deep Learning Application to Detect Glaucoma with a Mixed Training Approach: Public Database and Expert-Labeled Glaucoma Population

Florencia Cellini^a Deborah Caamaño^a Belen Carrasco^b José R. Juberías^{a,b}
Carolina Ossa^a Ramón Bringas^c Francisco de la Fuente^c Pablo Franco^d
David Coronado^d Jose Carlos Pastor^a

^aInstituto de Oftalmobiología Aplicada (IOBA), University of Valladolid, Valladolid, Spain; ^bOphthalmology Department, Hospital Clínico Universitario (HCUV), Valladolid, Spain; ^cOphthalmology Department, Hospital Universitario Río Hortega (HURH), Valladolid, Spain; ^dTransmural Biotech, Barcelona, Spain

Keywords

Deep learning · Glaucoma staging · Automatic detection · Glaucoma screening · ResNet-50 architecture

Abstract

Introduction: Artificial intelligence has real potential for early identification of ocular diseases such as glaucoma. An important challenge is the requirement for large databases properly selected, which are not easily obtained. We used a relatively original strategy: a glaucoma recognition algorithm trained with fundus images from public databases and then tested and retrained with a carefully selected patient database. **Methods:** The study's supervised deep learning method was an adapted version of the ResNet-50 architecture previously trained from 10,658 optic head images (glaucomatous or non-glaucomatous) from seven public databases. A total of 1,158 new images labeled by experts from 616 patients were added. The images were categorized after clinical examination including visual fields in 304 (26%) control images or those with ocular hypertension and 347 (30%) images with early, 290 (25%) with moderate, and 217

(19%) with advanced glaucoma. The initial algorithm was tested using 30% of the selected glaucoma database and then re-trained with 70% of this database and tested again. **Results:** The results in the initial sample showed an area under the curve (AUC) of 76% for all images, and 66% for early, 82% for moderate, and 84% for advanced glaucoma. After retraining the algorithm, the respective AUC results were 82%, 72%, 89%, and 91%. **Conclusion:** Using combined data from public databases and data selected and labeled by experts facilitated improvement of the system's precision and identified interesting possibilities for obtaining tools for automatic screening of glaucomatous eyes more affordably.

© 2023 The Author(s).
Published by S. Karger AG, Basel

Introduction

Open-angle glaucoma (OAG) is a multifactorial chronic optic nerve disease that can produce irreversible blindness. Its estimated global prevalence is 3.5% in the population between 40 and 80 years old and increases

with age [1]. In 2020, the number of people with OAG was an estimated 76 million, and it is expected to increase to 112 million by 2040 [1]. Several population studies have reported that at least 50% of glaucoma cases remain undiagnosed in Europe due to the absence of signs and symptoms. Early diagnosis and appropriate treatment of OAG are crucial, as they usually prevent permanent visual disability [2]. Furthermore, early diagnosis has a relevant impact on the cost of glaucoma management. The economic burden has been calculated to increase fourfold between the treatment of early and advanced OAG [3].

In many cases, glaucoma damage is identified by clinical examination or imaging technologies, although there is a high degree of variability in the design and quality of these studies. Although numerous comparative diagnostic studies have been performed, no evidence indicates which test or combination of tests performs better in early glaucoma diagnosis [4]. The intraocular pressure (IOP) measure is the most used technique. Above 21 mm Hg is considered a risk factor for glaucoma development and progression, although the IOP does not have a Gaussian distribution and 30% of glaucomas do not exhibit high pressure [5]. In moderate-advance stages, depending on the morphology of the papilla, the optic nerve shows typical characteristics of glaucomatous damage: asymmetry of the cup-to-disc ratio, large cup-to-disc ratio for disc size, neuroretinal rim notch or thinning, disc hemorrhages, vessel bayoneting/nasal displacement, peripapillary atrophy, and others [6]. Finally, evaluation of the visual field is also critical since currently the visual fields reflect the existence of functional damage [4].

Ophthalmic imaging technologies are used with the rest of the clinical examinations and their agreement is not always complete. However, they facilitate, in many cases, earlier recognition of functional loss and improve the assessment of structure-function correspondence in early glaucoma [7].

In recent years, artificial intelligence (AI) techniques have been applied to diagnose and treat glaucoma [8], including detecting disease progression [9] to facilitate early detection and stricter control, which would reduce blindness. However, despite the promising performance for many of these algorithms, they are not fully incorporated into clinical practice. The potential of these technologies in glaucoma care is recognized, but some “flaws” must be resolved. In addition, the heterogeneity among studies affects the algorithmic performance [10].

Some challenges are associated with the clinical translation of AI technologies into practice, which were well described by Devalla et al. [8]. First, performance depends on quality training data. The recommendation is

that they be trained using very large databases that include different clinical stages of glaucoma and if they are to be used for screening among the general population, other common ocular pathologies in the real world, such as age-related macular degeneration, diabetic retinopathy, and others should be included [8, 9]. Second, published areas under the curve (AUC) values are considered subjective and do not facilitate an adequate comparison among different series. Third, race and the geographic region affect the prevalence rates of glaucoma, and no database represents the global population [8]. Fourth, most data series are retrospective with well-diagnosed patients; however, there is a lack of prospective information that would provide the true predictive value of the algorithms [8]. Finally, there is the so-called black-box phenomenon that makes clinicians reluctant to adopt these tools. It is not well known how the algorithms discriminate and if the patterns they extract from the databases are related to glaucoma or other demographic parameters [8].

There are multiple strategies for obtaining good performance with AI algorithms, many of which are based on obtaining large databases [11, 12]. However, getting very large accurate databases is difficult. Obviously, public databases can be used, even though they usually have important limitations, such as inadequate information in the selection criteria used to acquire the data, which compromises the results and are difficult to use with confidence. Nevertheless, Chaurasia et al. [10] reported that 38% of AI studies of glaucoma use public database information. Recently, public databases have been used to test improvement in image inpainting algorithms by reducing blurred images, texture distortion, and semantic inaccuracy [13].

For the current study, the main contributions were as follows: we propose to use large public database images to train an AI algorithm and then test it with a protocolized database obtained from different stages of glaucoma well characterized by glaucoma specialists. Once the results were obtained, we retrained the algorithm with new well-characterized data and checked for performance improvement. Upon confirmation, this would allow faster progress in the development of new algorithms using existing data. This approach has not been reported in the peer-reviewed literature.

Materials and Methods

The study goal was to evaluate the ability to improve the performance of a deep learning technique trained on public databases glaucoma by training it with a specifically selected database

of glaucoma patients categorized by different clinical stages. To achieve this objective, the study was divided into two different phases. First, a classifier model trained by using images from public datasets [14] was tested on real practice images (initial sample) from the dataset from the Instituto de Oftalmobiología Aplicada (IOBA) of the University of Valladolid, Valladolid, Spain. Second, the model was retrained with new images from the same IOBA dataset (second sample) and then retested with the same test images (initial sample). All methods described in this section were developed using the Matlab R2018a programming language (MathWorks, Inc., Natick, MA, USA).

Data Collection Design

Regarding the public dataset, to train the deep learning model, a total of 10,658 fundus images from seven different public databases were used: REFUGE [15], Origa [16], LAG [17], Drishti-GS1 [18], sjchoi86 [19], HRF [20], and ODIR [21]. Only images labeled as glaucomatous or healthy eyes were used from each dataset, and images labeled with other ocular pathologies were excluded. The disease stage in these databases has not been considered since many of them did not report the stage [14].

For the IOBA dataset, glaucoma subspecialists selected the participants from the corresponding units of the following institutions: IOBA, Hospital Clínico Universitario, and Hospital Universitario Río Hortega (HURH) all from Valladolid, Spain. Subjects were divided into a control group with IOPs below 21 mm Hg; normal optic disc, visual fields, and nerve fiber layer (RNFL) thickness; an open angle by gonioscopy; and no history of other ocular diseases. The ocular hypertension group included patients with untreated IOPs over 21 mm Hg, normal visual fields, optic disc, and RNFL; an open angle by gonioscopy; no history of other ocular diseases; or use of nonsystemic topical corticosteroids. The glaucomatous eyes, those with IOPs over 21 mm Hg, had repeatable visual field defects by Humphrey analysis using the 24-2 strategy and detectable RNFL loss.

The glaucoma stages were classified based on the simplified Hodapp's classification that defines early glaucoma as the mean deviation (MD) loss of less than 6 decibels (dB), moderate glaucoma as a MD loss ranging from loss of 6 decibels to less than 12 dB, and advanced glaucomatous as a MD loss over 12 dB with data from a visual field examination performed less than 6 months previously [2].

Patients with other ocular pathologies that interfere with image quality or visual field results were excluded. Congenital anomalies of the optic disc and patients with high ametropia (>10 diopters) also were excluded. Fundus images were acquired using three cameras, the Non-Mydriatic Retinal Camera TRC-NW400 (Topcon®, Japan), the Non-Mydriatic Retinal Camera TRC-NW8 (Topcon®), and the FF450 Plus IR 4.5 Fundus Camera (Zeiss®, Germany). Two 45-degree images were obtained from each eye, one centered on the macula and the other on the optic nerve head. All selected images had a clear view of the disc. Only good-quality images were included. Images were stored in an anonymized database.

Using the sample size formula, with a confidence level of 95%, a margin of error of 5%, and considering that the prevalence of glaucoma in Spain is 2.1% [22], a sample size of 375 individuals was determined. However, considering the background described in the literature [23] and to refine the predictive ability of the algorithm as much as possible, we increased the sample size.

Automatic AI Glaucoma Classifier

The supervised deep learning method used in this study was an adapted version of the ResNet-50 architecture [24]. Residual Neural Networks (ResNets) are convolutional neural networks (CNN) with residual learning, which facilitates training and increases accuracy. Specifically, the ResNet-50 is a 50-layer CNN (48 convolutional layers, one MaxPool layer, and one average pool layer). The ResNet-50 was pretrained on large datasets such as ImageNet Large Scale Visual Recognition Challenge [25] and Microsoft Coco [26] during 90 epochs (passes through the dataset) with a 5-epoch warm-up.

The net then was adapted to the binary classification task, and it was trained with labeled glaucoma and control fundus images from public datasets, using 10% of the training set as a validation set. The images were used entirely without any segmentation and resized to $224 \times 224 \times 3$ pixels needed for the ResNet-50 method. This resizing was performed automatically with the bi-cubic interpolation before proceeding to the input image layer.

The net was trained using the Softmax cross-entropy loss function [27] and Adam optimization algorithms [28] during a maximum of 15 epochs, with early stopping if the loss in the validation set did not improve for five consecutive epochs. The learning rate was adjusted using a cosine decay starting at $1e-4$, the batch size was 64 samples, and the weight decay was 0.9. Data augmentation techniques were used during training to improve the robustness of the classifier and avoid overfitting. For example, at each batch, images were randomly flipped, cropped between 0% and 10%, translated 0–10 pixels, and rotated between $[-90; 90]$ degrees.

First Phase

Once the final model was obtained, it was tested on images in the IOBA dataset to evaluate performance on real images from clinical practice. To perform this evaluation, 30% of the IOBA dataset (initial sample) was used, as is usually done in this type of study [29, 30]. An anonymized dataset was delivered to Transmural Biotech S.L. researchers, who only had access to the images and the corresponding glaucoma stage and classified according to visual field results.

Second Phase

The previous model was retrained with newly labeled images from the IOBA dataset to evaluate the performance on the same test set again. To do that, the remaining 70% of the IOBA dataset (second sample) was used with the same CNN configuration as the previous experiment. The final result was retested with the initial sample to compare with the results of the first phase.

Results

The IOBA dataset acquired to retrain and test the deep learning model included 1,158 45-degree fundus photography images centered on the macula and focused on the optic nerve head of 1,158 eyes from 616 patients. The images were distributed as follows: 304 images (26%) from the control group or ocular hypertension, 347 images (30%) from eyes with initial glaucoma, 290 images (25%)

Table 1. The number of control and glaucoma images of the different datasets used to train, test, and retrain the deep learning model

Dataset	Glaucoma	Control	Total images
Training dataset (7 public databases)	2,511	8,147	10,658
Testing dataset (30% IOBA)	269	96	365
Retraining dataset (70% IOBA)	585	208	793
Total	3,365	8,451	11,816

IOBA, Instituto de Oftalmobiología Aplicada.

from eyes with moderate glaucoma, and 217 images (19%) from eyes with advanced glaucoma. Table 1 shows the total images used in this study divided into the testing dataset (initial sample) and the retraining dataset (second sample).

First Phase

Shown in Figure 1, the ROC curves obtained by the ResNet-50 CNN algorithm trained on the public database on the initial sample of selected glaucoma patients. The area under the receiver operating characteristic curve obtained was 76% for the entire set, 66% for early glaucoma cases, 82% for moderate cases, and 84% for advanced cases.

Second Phase

Shown in Figure 2, the ROC curves for the same initial sample after retraining the ResNet-50 CNN algorithm with the second sample of glaucoma selected patients. The AUC was 82% for the entire test set, 72% for early cases, 89% for moderate cases, and 91% for advanced cases.

Discussion

In the current study, a relatively original strategy was used to train, test, and improve an AI method that could be useful to detect glaucoma in real patient data. The selected deep learning method was the ResNet-50, which is considered adequate, in that several authors with the same objective have already used it and obtained good results similar to other deep learning techniques at least for detecting glaucomas in an ocular hypertension sample [31]. The state-of-the-art typically uses convolutional layers to represent fundus photographs with one-dimensional visual features [10, 31].

The first approach was to train the method with fundus images from public databases, as other authors did [10]. This strategy allows access to numerous images but obviously is limited because the labeling was uncertain and there was no information about their clinical staging. The resulting model was tested with an IOBA database of fundus images appropriately labeled.

The strategy allowed the detection of the glaucoma stages, with low AUC values in the initial glaucomas and similar values to those of other publications in moderate and advanced results. The results indicated that the proposed deep learning method can differentiate eyes with diagnosed glaucoma (in all stages) more effectively.

The AUC achieved, i.e., 76% for the first experiment on the global sample, is similar to the ranges of other studies. Chaurasia et al. [10] reported a range of AI performance measured in AUC from 79.2% to 99.9% across different studies. Nevertheless, these authors emphasized the heterogeneity across studies and reported many factors affecting the diagnostic performance. Finally, they recommended implementing a standard diagnostic protocol for grading, implementing external data validation, and analysis across different ethnicity groups.

Most studies included in the meta-analysis [10] used private data (67.1%) to train the AI models and evaluate the diagnostic performance, while others resorted to public bases. However, only 13% used external validation to evaluate the model performance according to the study's inclusion and exclusion criteria.

Few studies have used a strategy similar to ours, so it is difficult to make comparisons. Chaurasia et al. [10] used only three studies with a combination of public and private bases. The average AUC was 88.6%, similar to the current results.

One of the differential facts of our work is that we analyzed the results by clinical stages of glaucoma, from initial to advanced cases. Few studies have addressed the performance of AI techniques based on the evolutionary stages of glaucoma [32].

In all analyses, the same result was seen, excellent AUC values and low values in the initial glaucomas. An advantage of the labeled series used in the current study is that the patients were studied exhaustively structurally and functionally and classified according to the visual field alterations. When only advanced glaucoma was detected, the values were higher as in our series where advanced glaucoma was detected in up to 91%.

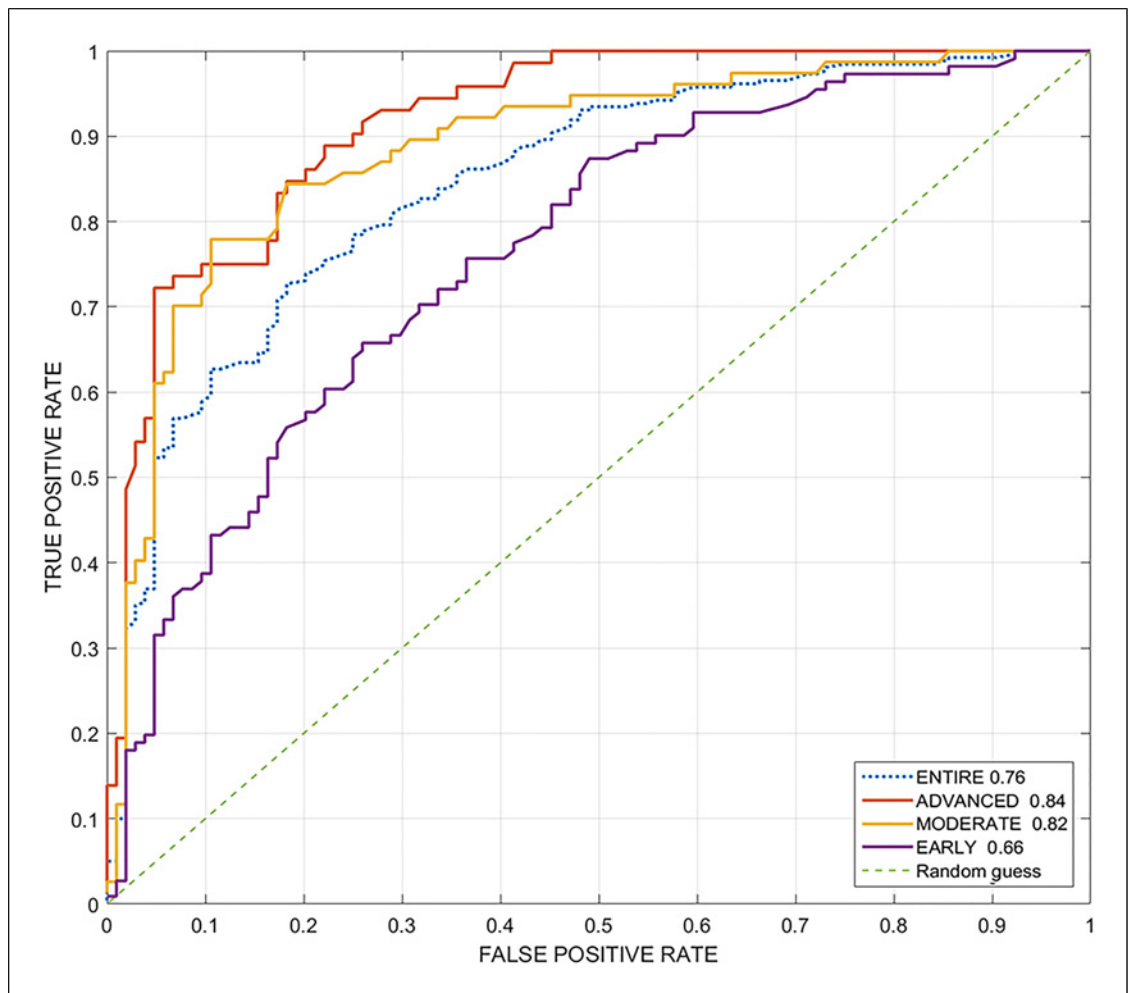


Fig. 1. ROC curve for prediction task for the first experiment. The dashed line shows the global AUC for the entire dataset and the other lines show the AUC for each glaucoma stage.

The second approach of our study was to retrain the model with images from the same population as tested. Improvement in the AUC values between the first and second phases seemed evident, but it must be kept in mind that, as mentioned in the Introduction, the AUC may not be the most appropriate method for comparing different AI strategies. Devalla et al. [8] suggested that the AUC does not allow an adequate comparison among series, so new metrics should be assessed.

In any case, the use of images from the IOBA dataset to retrain the model resulted in an effective improvement in algorithm performance not only in the entire set but also in the different stages in which glaucoma was classified. This could indicate that, with a large number of new appropriately labeled data to retrain the algorithm, it would still obtain better results.

The current study had several limitations. One was the relatively small number of images used to train and test the algorithm compared to other studies, although it can be considered that more than 11,000 images have been used in the current study (public plus selected database). Mursch-Edlmayr et al. [9] affirmed that studies like this may need up to 100,000 images that cover all stages of the disease spectrum to obtain reliable results. Furthermore, these results should be externally validated in a larger multicenter study with prospective data acquired in routine clinical practice. Fan et al. [31] affirmed that the AI-based diagnostic accuracy for glaucoma diagnosis using fundus photography and optical coherence tomography imaging was not better in external datasets compared to the original data source. Different glaucoma definitions between the external datasets also may explain

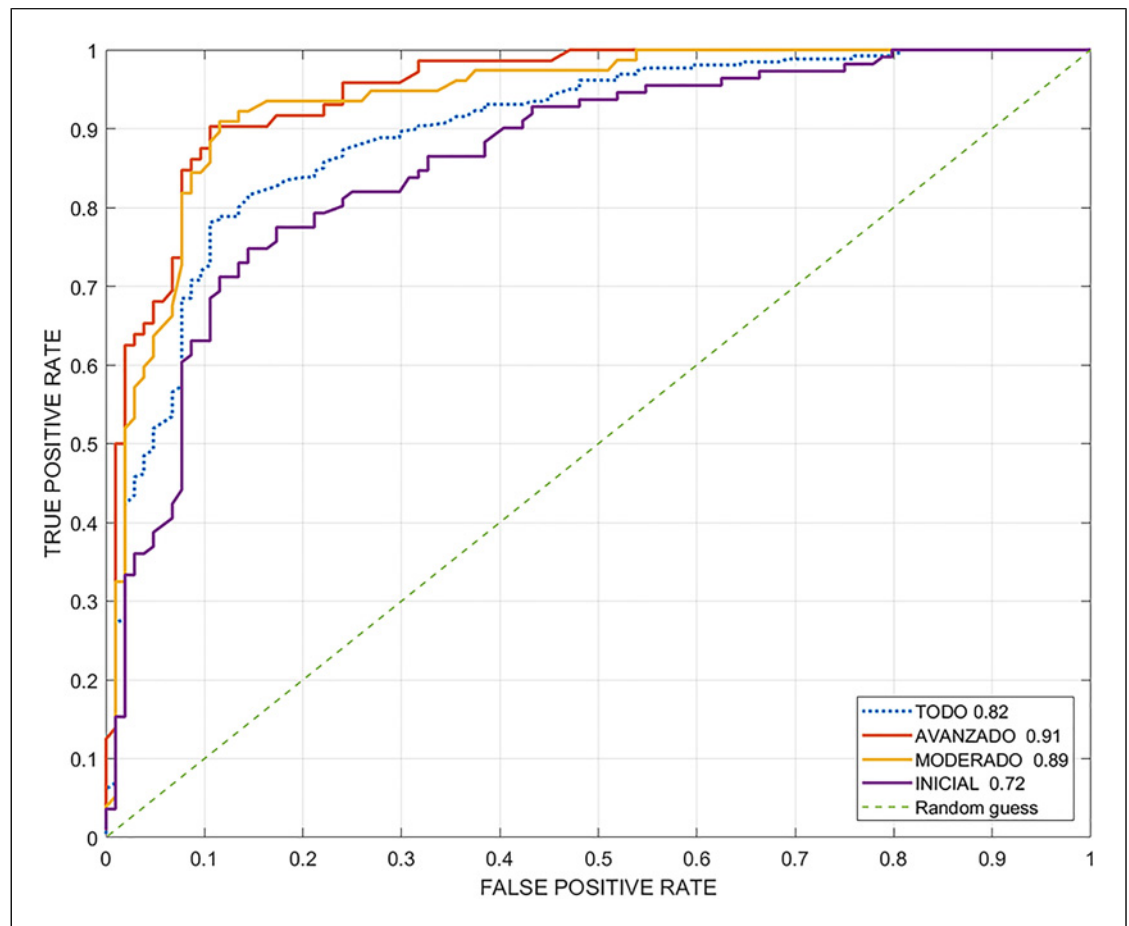


Fig. 2. ROC curve for prediction task for the second experiment. The dashed line shows the global AUC for the entire dataset and the other lines show the AUC for each glaucoma stage.

the differences in performance (e.g., visual field vs. expert photography review, cup-to-disc ratio, etc.) [31].

Another limitation was that the clinical variables of the patients, such as age, ethnicity, etc., have not been considered due to the current database being limited to patients from a Spanish region. This limitation should be considered if a generalized use of this type of algorithm is intended. Finally, another limitation, mentioned by Mursch-Edlmayr et al. [9], was the lack of knowledge of whether the accuracy results obtained can be generalizable to populations of patients in the real world where prevalent comorbidities exist, such as cataract and ocular surface disease, which negatively affect image quality.

The main strength of the study was that the current method was trained with large public databases and also was tested on data obtained under a clear protocol and, therefore, perfectly labeled. We believe that the use of different databases worldwide gives more robustness to the algorithm compared

to others that have used only one database. However, the use of public databases has the uncertainty of how the outcome of the pathology was acquired, making it difficult to judge the potential of the proposed tools to detect glaucoma. Therefore, the use of data, specifically selected and reviewed by glaucoma experts, to test the model provides confidence that the proposed method will work in a real population of glaucoma. According to Chaurasia et al. [10] only 5% of the AI studies in glaucoma used public and private databases together. Therefore, we consider that this combination of datasets is a reasonable strategy that deserves further investigation.

In the future, we will feed the database with more perfectly labeled patients to verify how far the learning capacity of these classifiers extends. If the results are adequate, the algorithm will be subjected to external validation with a real-world nonselected population.

We must also consider the constant evolution of AI techniques. It is accepted [33] that deep CNNs can effectively

improve the performance of single-image super-resolution reconstruction. Nevertheless, deep CNNs will lead to an important increase in the size of parameters, limiting its application. Thus, other procedures are tested.

In conclusion, the current results support further research to continue to develop automated AI methods to achieve better results in detecting glaucoma in fundus images, in unselected populations, thus contributing to earlier diagnosis of this challenging disease.

Statement of Ethics

The study follows the Organic Law 15/99 on the Protection of Personal Data (LOPD) and Law 14/2007 of Biomedical Research. The protocol was approved by the Research Ethics Committee of the IOBA and Hospital Clínico Universitario de Valladolid (HCUV), with the code PI 21-2278 and registered on clinicaltrials.gov NCT04972695. The entry into the study was voluntary, and the patients provided written informed consent.

Conflict of Interest Statement

Collaboration with sponsor employees related to the content of the manuscript. The co-author Florencia Cellini during her fellowship in glaucoma received small financial support for 10 months for the collection of data and to cover the cost of retinographies. The financial support was provided by the company Transmural Biotech S.L. There is no other relationship with the company.

References

- Tham Y, Li X, Wong T, Quigley H, Aung T, Cheng C. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081–90.
- European Glaucoma Society Terminology and Guidelines for Glaucoma. *Br J Ophthalmol*. 2021;105(Suppl 1):1–169.
- Lee P, Walt J, Doyle J, Kotak S, Evans S, Budenz D, et al. A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch Ophthalmol*. 2006;124(1):12–9.
- McCann P, Hogg R, Wright D, Pose-Bazarra S, Chakravarthy U, Peto T, et al. Glaucoma in the northern Ireland cohort for the longitudinal study of ageing (NICOLA): cohort profile, prevalence, awareness and associations. *Br J Ophthalmol*. 2020;104(11):1492–9.
- Häkkinen M, Ekström C. Distribution of intraocular pressure in a Swedish population. *Ups J Med Sci*. 2022;127:127.
- Gandhi M, Dubey S. Evaluation of the optic nerve head in glaucoma. *J Curr Glaucoma Pract*. 2013;7(3):106–14.
- Lin S, Singh K, Jampel H, Hodapp E, Smith S, Francis B, et al. Optic nerve head and retinal nerve fiber layer analysis: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2007;114(10):1937–49.
- Devalla S, Liang Z, Pham T, Boote C, Strouthidis N, Thiery A, et al. Glaucoma management in the era of artificial intelligence. *Br J Ophthalmol*. 2020;104(3):301–11.
- Mursch-Edlmayr A, Ng W, Diniz-Filho A, Sousa D, Arnold L, Schlenker M, et al. Artificial intelligence algorithms to diagnose glaucoma and detect glaucoma progression: translation to clinical practice. *Transl Vis Sci Technol*. 2020;9(2):55.
- Chaurasia A, Greatbatch C, Hewitt A. Diagnostic accuracy of artificial intelligence in glaucoma screening and clinical practice. *J Glaucoma*. 2022;31(5):285–99.
- Li Z, He Y, Keel S, Meng W, Chang R, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125(8):1199–206.
- Medeiros F, Jammal A, Thompson A. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology*. 2019;126(4):513–21.
- Chen Y, Xia R, Yang K, Zou K. High resolution image inpainting via DR-GAN and contextual attention. *Multimed Tools Appl*. 2023;1–21.
- Franco P, Coronado-Gutierrez D, Lopez C, Burgos-Artizzu X. Glaucoma patient screening from online retinal fundus images via artificial intelligence. *medRxiv*. 2021:2021–2.
- Orlando J, Fu H, Barbosa Breda J, Van Keer K, Bathula D, Diaz-Pinto A, et al. REFUGE Challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal*. 2020;59(101570):101570.

Funding Sources

This work was funded by Transmural, which provided the necessary financial support for the collection of data and the cost of retinographies. There has been no relationship beyond this except for the two co-authors who are currently employees. The company has had no involvement in the collection and preparation of data or the preparation of the manuscript.

Author Contributions

Florencia Cellini and Deborah Caamaño: conceptualization, methodology, investigation, resources, data curation, writing – original draft. Belen Carrasco, José Ramón Juberías, Carolina Ossa, Ramón Bringas, and Francisco de la Fuente: resources, data curation, and investigation. Pablo Franco and David Coronado-Gutierrez: conceptualization, methodology, software, validation, writing – original draft, funding acquisition. J. Carlos Pastor: conceptualization, methodology, writing – review and editing, supervision, and project administration.

Data Availability Statement

The data in this study were obtained with the financial support of Transmural Biotech S.L where restrictions may apply. All data generated or analyzed during this study are included in this article. Further inquiries can be directed to the corresponding author.

- 16 Zhang Z, Yin F, Liu J, Wong W, Tan NM, Lee BH, et al. ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research. *Annu Int Conf IEEE Eng Med Biol Soc.* 2010;2010:3065–8.
- 17 Li L, Xu M, Liu H, Li Y, Wang X, Jiang L, et al. A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Trans Med Imaging.* 2020;39(2):413–24.
- 18 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. *JSM Biomed Imaging Data Papers.* 2015;2(1):1004.
- 19 Sjchoi. Retina Dataset. [Online]. 2016. Available from: https://github.com/cvblab/retina_dataset (accessed 26 August 2020).
- 20 Budai A, Bock R, Maier A, Hornegger J, Michelson G. Robust vessel segmentation in fundus images. *Int J Biomed Imaging.* 2013; 2013:154860.
- 21 National Institute of Health Data science at Peking University (NIHDS-PKU), Institute of Artificial Intelligence at Peking University (IAI-PKU). ODIR Dataset. grand-challenge.org. [Online]. 2019. Available from: <https://odir2019.grand-challenge.org/dataset/> (accessed 26 August 2020).
- 22 Anton A, Andrada M, Mujica V, Calle M, Portela J, Mayo A. Prevalence of primary open-angle glaucoma in a Spanish population: the Segovia study. *J Glaucoma.* 2004;13(5):371–6.
- 23 Zhao R, Chen X, Liu X, Chen Z, Guo F, Li S. Direct cup-to-disc ratio estimation for glaucoma screening via semi-supervised learning. 2020 Apr;24(4):1104–13. *IEEE J Biomed Health Inform.* 2020;24(4):1104–13.
- 24 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Comput Soc; 2016. Vol. 2016; p. 770–8.
- 25 Qian G, Zhang L, Zhang Q. Fast conceptor classifier in pre-trained neural networks for visual recognition. In: Cong F, Leung A, Wei Q, editors. *Advances in neural networks.* Cham: Springer International Publishing AG; 2017. p. 290–8.
- 26 Lin T, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in context. In: Fleet D, editor. *Computer vision—ECCV 2014.* Cham: Springer International Publishing; 2014. p. 740–55.
- 27 Goodfellow I, Bengio Y, Courville A. Softmax units for multinoulli output distributions. *Deep Learning.* 2018:180–4.
- 28 Kingma D, Ba J. arXiv.org. [Online]. 2014. Available from: <http://arxiv.org/abs/1412.6980>.
- 29 Raghavendra U, Fujita H, Bhandary S, Gudigar A, Tan J, Acharya U. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Inf Sci.* 2018;441(441):41–9.
- 30 Girard M, Chin K, Devalla S, Aung T, Jonas J, Wang Y. Deep learning can exploit 3D structural information of the optic nerve head to provide a glaucoma diagnostic power superior to that of retinal nerve fibre layer thickness. *Invest Ophthalmol Vis Sci.* 2018; 59(9):4081.
- 31 Fan R, Alipour K, Bowd C, Christopher M, Brye N, Proudfoot J, et al. Detecting glaucoma from fundus photographs using deep learning without convolutions: transformer for improved generalization. *Ophthalmol Sci.* 2023;3(1):100233.
- 32 Yousefi S. Clinical applications of artificial intelligence in glaucoma. *J Ophthalmic Vis Res.* 2023;18(1):97–112.
- 33 Chen Y, Xia R, Yang K, Zou K. MFFN: image super-resolution via multi-level features fusion network. *Vis Comput.* 2023:1–16.