Original Article

Investigation of the Role of Convolutional Neural Network Architectures in the Diagnosis of Glaucoma using Color Fundus Photography

 Eray Atalay*, Onur Özalp*, Özer Can Devecioğlu**, Hakika Erdoğan***, Türker İnce**,
 Nilgün Yıldırım*

*Eskişehir Osmangazi University Faculty of Medicine, Department of Ophthalmology, Eskişehir, Turkey
**İzmir University of Economics Faculty of Engineering, Department of Electrical and Electronics Engineering, İzmir, Turkey
***Maltepe University Faculty of Medicine, Department of Ophthalmology, İstanbul, Turkey

Abstract

Objectives: To evaluate the performance of convolutional neural network (CNN) architectures to distinguish eyes with glaucoma from normal eyes.

Materials and Methods: A total of 9,950 fundus photographs of 5,388 patients from the database of Eskişehir Osmangazi University Faculty of Medicine Ophthalmology Clinic were labelled as glaucoma, glaucoma suspect, or normal by three different experienced ophthalmologists. The categorized fundus photographs were evaluated using a state-of-the-art two-dimensional CNN and compared with deep residual networks (ResNet) and very deep neural networks (VGG). The accuracy, sensitivity, and specificity of glaucoma detection with the different algorithms were evaluated using a dataset of 238 normal and 320 glaucomatous fundus photographs. For the detection of suspected glaucoma, ResNet-101 architectures were tested with a data set of 170 normal, 170 glaucoma, and 167 glaucoma-suspect fundus photographs.

Results: Accuracy, sensitivity, and specificity in detecting glaucoma were 96.2%, 99.5%, and 93.7% with ResNet-50; 97.4%, 97.8%, and 97.1% with ResNet-101; 98.9%, 100%, and 98.1% with VGG-19, and 99.4%, 100%, and 99% with the 2D CNN, respectively. Accuracy, sensitivity, and specificity values in distinguishing glaucoma suspects from normal eyes were 62%, 68%, and 56% and those for differentiating glaucoma from suspected glaucoma were 92%, 81%, and 97%, respectively. While 55 photographs could be evaluated in 2 seconds with CNN, a clinician spent an average of 24.2 seconds to evaluate a single photograph.

Conclusion: An appropriately designed and trained CNN was able to distinguish glaucoma with high accuracy even with a small number of fundus photographs.

Keywords: Glaucoma, convolutional neural network, artificial intelligence, telemedicine

Introduction

Glaucoma is the most common cause of irreversible blindness.[1] Globally, glaucoma affected 60 million people in 2010 and was projected to affect 80 million people by 2020 and approximately 112 million people in 2040.[2,3] Its average prevalence among people aged 40-80 years is 3.54% worldwide (confidence interval: 2.09%-5.82%), while the prevalence in Turkey has been reported as 1.29% and 2% in different studies (Yıldırım, N., Başmak, H., Kalyoncu, C., Özer, A., Aslantaş, D. Metintaş, S. 2008: Glaucoma prevalence in the population over 40 years of age in the Eskişehir region, 42nd National Congress of the Turkish Ophthalmological Association, Antalya).[3,4] Although glaucoma diagnosis has become more effective with advances in diagnostic technologies, it has been reported that 50-90% of patients do not know that they have glaucoma.[2,5,6,7,8]

Numerous methods such as intraocular pressure (IOP) measurement, optic nerve examination, visual field examination, and retinal nerve fiber analysis are used in the diagnosis of glaucoma. Fundus imaging is often preferred because digital fundus imaging is a noninvasive, cost-effective, and rapid method that is less affected by optic media opacities and is also a practical approach for telemedicine applications. The growing popularity of artificial intelligence applications in recent years further expanded the use of fundus imaging, and fundus photographs have become widely used in the artificial intelligence-assisted diagnosis of many eye diseases.

This study evaluated the performance of convolutional neural networks (CNNs) in differentiating the fundus photographs of eyes with glaucoma from those of normal eyes.

Materials and Methods

The study used 9,950 optic nerve photographs of 5,388 patients obtained as a result of glaucoma prevalence research and stored in the archive of the Eskişehir Osmangazi University Faculty of Medicine Ophthalmology Clinic. The study protocol was planned in accordance with the Declaration of Helsinki and approved by the Eskişehir Osmangazi University Ethics Committee. The photographs to be evaluated in the algorithm were obtained from the hard disk of the nonmydriatic fundus camera (Kowa nonmyd alpha-DIII, Kowa Company Ltd., Tokyo, Japan) in our clinic. All 20° posterior segment photographs in the device archive were anonymized and transferred.

The quality of the photographs was classified and recorded in the database by the relevant physicians (E.A., H.A., O.Ö.) according to the guidelines.[9] Images were considered low quality if more than 50% of the optic disc margins were not visible or haze obscured the optic disc and/or cup margins, medium quality if 0-50% of the optic disc boundaries were not clearly discernible but the cup and its margins were visible and haze did not obscure the optic disc margins or cup; and high quality if all optic disc and cup margins were visible, with or without visibility of the retinal nerve fibers.

Of the total 9,950 photographs, 2,587 medium-quality and 5,970 high-quality photographs comprised the dataset pool, while the 1,393 low-quality photographs were not included. All datasets used in the study were composed of medium- and high-quality photographs.

The clinicians blindly and randomly labeled the fundus photographs as normal, glaucoma suspect, or glaucoma using a computer program (Figure 1). This classification was made according to criteria used in previous studies.[9,10,11,12,13] Glaucoma was diagnosed in the presence of any of the following: vertical cup-to-disc ratio ≥0.9; rim width-to-disc diameter ratio ≤0.05 or presence of localized notching; and any retinal nerve fiber defect corresponding to an area of neuroretinal rim thinning or localized notching. Suspected glaucoma was classified in the presence of any of the following: vertical cup-to-disc ratio of ≥0.7 to <0.9; rim width-to-disc diameter ratio of ≤0.1 to >0.05; retinal nerve fiber defect; and disc hemorrhage. Eyes not exhibiting these characteristics were classified as normal.

Labeling was performed in two stages. In the first stage, the diagnostic codes assigned by two glaucoma specialists (E.A. and H.A.) were entered into the database as the final verdict if there was consensus between them. Disputed photographs were marked for the second stage. In the second stage, these photographs were shown to the third, more senior and experienced expert (N.Y.) and a final verdict was reached by majority vote. In the absence of a majority (e.g., E.A.: suspect, H.A.: normal, N.Y.: glaucoma), the opinion of the third expert was recorded as the final verdict. In addition, the time from the photo being opened in the program to the expert marking an option was automatically calculated by the software. After all fundus photographs were labelled, the CNN stage started.

In this study, we used an adaptive two-dimensional (2D) CNN to combine feature extraction and classification in a single learning body. CNNs are feed-forward artificial neural networks that are inspired by the brain structure and regarded as simple computational models of the mammalian visual cortex.[14,15,16] Therefore, CNNs are mostly used for 2D signals such as pictures and videos. They are generally used by the machine and visual communities as the de facto standard for both understanding and solving many image and video recognition problems. To understand a convolution in the simplest way, it can be thought of as a sliding window function applied in 2D in a matrix. CNNs acquire the basic idea through limited connectivity among multilayer sensors and restricted weight sharing. CNNs are a "restricted" version of multilayer perceptrons, only with subsampling layers added. The fully connected hidden and output layers of CNNs are exactly the same as the layers of multilayer perceptrons. Therefore, a CNN has the characteristic limitations of multilayer perceptrons when learning a complex task. Multilayer perceptrons are truly universal approaches; however, a "better" universal approach is needed for learning tasks with compact configurations.

A conventional CNN architecture uses the convolutional layers for automatic feature learning and extraction from raw or preprocessed data. It then includes multilayer perceptron layers for classification, and a one-dimensional (1D) feature map output is fed to a multilayer perceptron. These two different layers combine to automatically learn optimal features for advanced classification performance. Input data is processed using convolution layers and alternating subsampling layers that produce feature maps that have been filtered and subsampled by the processors (neurons) of the previous layers. The 2D filter cores of the CNN are optimized and trained using an error feedback algorithm (back-propagation). In this study, a 2D CNN classifier was designed and applied to detect glaucoma. The proposed system is shown in Figure 2.

The proposed 2D CNN classifier is run on Python using the Keras library, which was developed for researchers and developers to make applications of deep learning models as quickly and easily as possible (Keras Deep Learning Library Web Site: https://keras.io/). The architecture of the proposed 2D CNN classifier
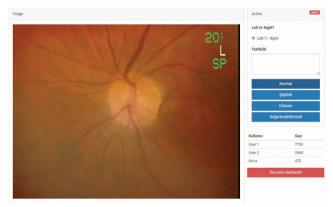

Figure 1. Fundus photograph labeling program

consists of 5 convolutional and two scalar multilayer classifier layers in all experiments to achieve high computational efficiency with competitive performance. The 2D CNN classifier, which is designed with a relatively shallow structure, can be used for real-time glaucoma diagnosis. The number of neurons in the convolutional layers was set to [64 64 32 32 16] and the filter sizes were determined as (15,15), (11,11), (7,7), (3,3), and (3,3). The fully connected multilayer perceptron layer has 512 neurons and the output layer has two neurons for detecting glaucoma. The architecture of the 2D classifier is shown in Figure 3. ReLU was used as a nonlinear activation function in all CNN layers and applied to the subsampling layers with maximum sampling. To train the 2D CNN classifier, a 10-fold cross-validation technique was implemented to improve generalization, thereby preventing over-training of the architecture. The Dropout technique (removing some connections with a certain possibility) was also implemented in the ResNet architecture for the same purpose. The RMSprop algorithm was used in training to optimize the parameters of the designed architecture.

After classification, subdata clusters were created with digital fundus photographs labeled as healthy (n=238)/glaucoma (n=320) and healthy (n=170)/glaucoma (n=170)/glaucoma suspect (n=167). Image size was reduced to 512x512 pixels for 10-fold cross-validation training and testing. To compare the performance of a CNN model with a relatively shallow structure (5 convolutional and 1 hidden scalar layer) that was designed and trained in the most appropriate way using our own data, we also selected the ResNet-50, ResNet-101, and VGG-19 deep structures, which are frequently used in the literature, and trained them using transfer learning with the same data.[17,18,19] Our aim here is to show that in situations where big data is not available (as in most studies in the literature), the glaucoma diagnostic performance of the relatively shallow CNN model that we designed and trained may be similar or better than the performances of the deep structures trained with the transfer learning method proposed in other studies. The deep
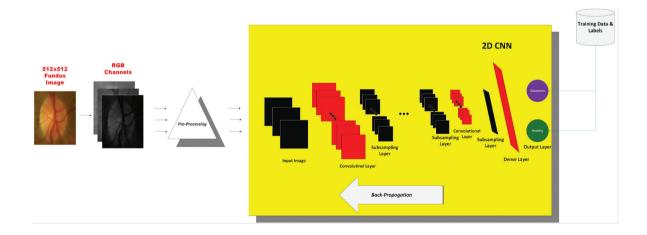

Figure 2. Proposed glaucoma diagnosis system

structures, whose last 3 layers were trained with transfer learning methods using our own data, were previously trained using the ImageNet database, which contains approximately 14 million images.[20] While training the designed architectures, all data were randomly divided into 10 separate clusters for training/validation/generalization. This aimed to train the structure correctly and optimize the generalization performance. Standard performance measurement criteria (accuracy, sensitivity, and specificity) were calculated to compare the performance of the architectures. Heat maps were obtained by upsampling the outputs of the last convolutional layer to the image size.
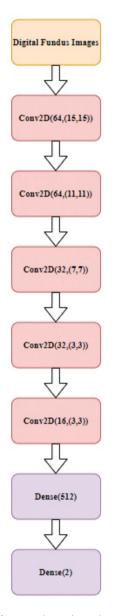

Figure 3. Proposed two-dimensional convolutional neural network architecture (Conv2D [neuron number, filter size) and Dense (neuron number])

To detect suspected glaucoma, two separate datasets (glaucoma suspect/normal and glaucoma suspect/glaucoma) consisting of 170 normal, 170 glaucoma, and 167 glaucoma-suspect fundus photographs were created. The created dataset was divided into 90% and 10% for training and testing, respectively. The transfer learning method was applied with 101-layer ResNet architectures previously trained with the ImageNet dataset.[20]

All experiments reported in this article were run on a 2.2 GHz Intel Core i7-8750H with 8 GB RAM and NVIDIA GeForce GTX 1050Ti graphics card. Both training and test datasets were processed in parallel by a total of 768 CUDA cores.

Results

Of the 5388 patients included in the study, 3825 were women (71%) and 1563 were men (29%). The mean age of the women was 54.88 ± 10.31 years and that of the men was 58.35 ± 10.82 years. There was no difference between the two groups in terms of age (p>0.05). Of the participants, 643 (11.9%) had diabetes mellitus, 1734 (32.2%) had hypertension, and 505 (9.4%) had coronary heart disease.

Of all the medium- and high-quality photographs labeled by the glaucoma specialists, 416 photos were classified as glaucoma and 342 photos as suspected glaucoma. Diagnostic agreement between E.A. and H.A. was 92%.

Accuracy, sensitivity, and specificity in glaucoma detection were 96.2%, 99.5%, and 93.7% with ResNet-50; 97.4%, 97.8%, and 97.1% with ResNet-101; 98.9%, 100%, and 98.1% with VGG-19, and 99.4%, 100%, and 99% with the 2D CNN. Figure 4 shows sample data of the 2D CNN classifier, while Figure 5 shows heat maps of selected outputs.

ResNet-101 differentiated the fundus photographs of glaucoma-suspect eyes from normal eyes with 62% accuracy, 68% sensitivity, and 56% specificity and from glaucoma eyes with 92% accuracy, 81% sensitivity, and 97% specificity.

In systems trained in diagnosis glaucoma with 10 epochs, the time per epoch was 23.6 s for the 2D CNN, 23.2 s for VGG-19, 35 s for ResNet-50, and 57 s for ResNet-101. While the mean test duration was 2 s per 55 photographs with the CNNs, we determined that clinicians evaluated one photograph in a mean of 24.2 s.

Discussion

In our study, the accuracy rate, sensitivity, and specificity in glaucoma detection were 96.2%, 99.5%, and 93.7% with ResNet-50; 97.4%, 97.8%, and 97.1% with ResNet-101; 98.9%, 100%, and 98.1% with VGG-19; and 99.4%, 100%, and 99% with the 2D CNN, respectively. With ResNet-101, glaucoma suspects were differentiated from normal eyes with an accuracy rate of 62%, sensitivity of 68%, and specificity of 56% and were differentiated from glaucomatous eyes with an

accuracy rate of 92%, sensitivity of 81%, and specificity of 97%. In glaucoma diagnosis, the CNNs tested 55 fundus photographs in 2 seconds, whereas clinicians took approximately 24 seconds on average to evaluate a single fundus photograph.

Imaging with fundus photography is the least costly method for structural evaluation of the optic nerve, but its sensitivity and specificity in detecting glaucoma or suspected glaucoma are not comparable to advanced methods such as optic coherence tomography (OCT). Large numbers of fundus images collected from diabetic retinopathy screening programs have formed a resource for evaluating glaucomatous optic disc changes with deep learning algorithms. Two different studies in which deep learning algorithms were created for glaucoma detection using

fundus photographs obtained for diabetic retinopathy screening yielded high rates of 95.6% and 96.4% for sensitivity and 92% and 87.2% for specificity (area under the curve [AUC]: 0.986 and 0.942).[9,21] In another study that evaluated glaucoma from fundus photographs using the ResNet architecture, the AUC value was found to be 0.965.[22] In a study comparing five different ImageNet models used for glaucoma diagnosis, sensitivity and specificity values were found to be 90.6% and 88.2% (AUC: 0.96) with VGG-16, 92.4% and 88.5% (AUC: 0.97) with VGG-19, 92.2% and 87.5% (AUC: 0.97) with InceptionV3, 91.1% and 89.4% (AUC: 0.96) with ResNet50, and 93.5% and 85.8% (AUC: 0.96) with Xception, respectively.[18] The literature data and the results of our study indicate that CNN algorithms
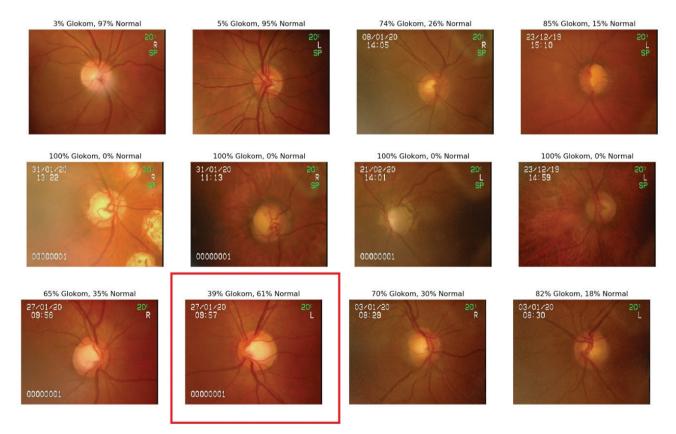

Figure 4. Sample data of the two-dimensional convolutional neural networks classifier (an incorrect result is marked in red)
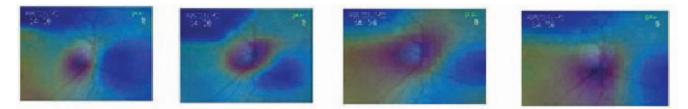

Figure 5. Heat maps of selected outputs

created with fundus photographs are effective in the detection of glaucoma. Although the study comparing different CNN models showed similar sensitivity and specificity in glaucoma detection, in our study we observed higher sensitivity and specificity rates with VGG-19 and the 2D CNN.[18]

The use of OCT and ultra-wide scanning laser ophthalmoscopy images has gained popularity in the evaluation of glaucoma with deep learning. In a study of glaucoma detection using 1399 Optos fundus photographs, it was reported that sensitivity was 81.3% and specificity was 80.2% (AUC: 0.872), with higher values in the detection of patients with advanced glaucoma.[23] In a deep learning study with 2132 OCT images, the sensitivity and specificity for early glaucoma detection were 82.5% and 93.9%, respectively (AUC: 0.937).[24] AUC values ranging from 0.877 to 0.981 were also determined in other studies using different OCT parameters.[25,26,27,28] Confocal scanning laser ophthalmoscopy (CSLO) parameters and scanning laser polarimetry (SLP) parameters have also been used in various artificial intelligence studies conducted with structural evaluations for glaucoma diagnosis. In screens performed with CSLO parameters, the sensitivity ranged from 83% to 92% and the specificity ranged from 80% to 91%,[29,30,31,32,33] while in screenings performed with SLP parameters, the sensitivity was 74% to 77% and specificity was 90% to 92%.[29,34] In the detection of glaucoma by machine learning, there are various data sources for the structural evaluation of the optic disc, such as fundus photograph, OCT, CSLO, and SLP parameters. However, fundus photographs may be preferrable since the use of fundus photographs is the cheapest and easiest method and yields high sensitivity and specificity.

There are also studies in the literature evaluating structural tests together. In an artificial intelligence study using fundus photographs and retinal nerve fiber thickness data obtained from spectral domain OCT, it was reported that glaucoma was diagnosed with 80% specificity and 90% sensitivity.[35] In addition, studies have been conducted to increase the effectiveness of artificial intelligence algorithms by using a combination of structural and functional test data for glaucoma diagnosis. It was reported that fundus photographs and visual field results evaluated together provided greater accuracy (88%) in glaucoma diagnosis than when evaluated separately.[36]

Heat maps, which are formed by upsampling the outputs of the last convolutional layer to the image sizes, can help to understand which regions of an input image affect the output prediction of a CNN. These activation maps are overlayed on the input image to highlight the areas of greatest CNN interaction. Example heat maps from the glaucoma category are shown in Figure 5. They show that alterations around the optic disc, which are also examined in the medical diagnosis of glaucoma, are the areas with which the CNN interacts the most. This suggests that physicians may be able to use machine learning outputs as an aid in the diagnosis of glaucoma or other diseases.

Study Limitations

The deep learning method was able to distinguish glaucomatous eyes from normal eyes with high accuracy, even with a small number of fundus photographs. We predict that more photographs and different optimized learning algorithms will provide even higher sensitivity and specificity values, and that a multi-modelling method combining fundus photographs and OCT or visual field data will increase the effectiveness of artificial intelligence in glaucoma diagnosis. The fundus photographs used in our study were images in the Eskişehir Osmangazi University Faculty of Medicine Eye Clinic archive obtained through glaucoma prevalence research and were not labeled with a diagnosis. For this reason, the artificial intelligence algorithm may not show the same accuracy when used in different races and ethnic groups. An additional limitation of the study was not evaluating false-negative and false-positive rates in the diagnosis of glaucoma.

Conclusion

Glaucoma leads to irreversible blindness, and its early diagnosis and treatment are important. It can be asymptomatic in the early stages, and visual field is not affected until there is between 20% and 50% retinal ganglion cell loss. Therefore, structural evaluation is more important than functional tests. Early glaucoma detection can be achieved by evaluating optical disc images with machine learning algorithms. The development of these methods will allow the creation of telemedicine glaucoma screening programs in primary health care centers and enable effective triage. Artificial intelligence-based algorithms should not be regarded as a system that can replace physicians, but as a tool that aids physicians in diagnosis and follow-up.

References

1. Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, Jonas JB, Keeffe J, Leasher J, Naidoo K, Pesudovs K, Resnikoff S, Taylor HR; Vision Loss Expert Group. Causes of vision loss worldwide, 1990-2010: a systematic analysis. Lancet Glob Health. 2013;1:e339-e349.

2. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol. 2006;90:262-267.

3. 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;121:2081-2090.

4. Yavaş GF, Küsbeci T, Şanlı M, Toprak D, Ermiş SS, İnan ÜÜ, Öztürk F. Risk Factors for Primary Open-Angle Glaucoma in Western Turkey. Turk J Ophthalmol. 2013;43:87-90.

5. Bourne RR, Jonas JB, Flaxman SR, Keeffe J, Leasher J, Naidoo K, Parodi MB, Pesudovs K, Price H, White RA, Wong TY, Resnikoff S, Taylor HR; Vision Loss Expert Group of the Global Burden of Disease Study. Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990-2010. Br J Ophthalmol. 2014;98:629-638.

6. Varma R, Ying-Lai M, Francis BA, Nguyen BB, Deneen J, Wilson MR, Azen SP; Los Angeles Latino Eye Study Group. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study. Ophthalmology. 2004;111:1439-1448.

7. Baskaran M, Foo RC, Cheng CY, Narayanaswamy AK, Zheng YF, Wu R, Saw SM, Foster PJ, Wong TY, Aung T. The Prevalence and Types of Glaucoma in an Urban Chinese Population: The Singapore Chinese Eye Study. JAMA Ophthalmol. 2015;133:874-880.

8. Gupta P, Zhao D, Guallar E, Ko F, Boland MV, Friedman DS. Prevalence of Glaucoma in the United States: The 2005-2008 National Health and Nutrition Examination Survey. Invest Ophthalmol Vis Sci. 2016;57:2905-2913.

9. 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;125:1199-1206.

10. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol. 2002;86:238-242.

11. Iwase A, Suzuki Y, Araie M, Yamamoto T, Abe H, Shirato S, Kuwayama Y, Mishima HK, Shimizu H, Tomita G, Inoue Y, Kitazawa Y; Tajimi Study Group, Japan Glaucoma Society. The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. Ophthalmology. 2004;111:1641-1648.

12. He M, Foster PJ, Ge J, Huang W, Zheng Y, Friedman DS, Lee PS, Khaw PT. Prevalence and clinical characteristics of glaucoma in adult Chinese: a population-based study in Liwan District, Guangzhou. Invest Ophthalmol Vis Sci. 2006;47(7):2782-2788.

13. Topouzis F, Wilson MR, Harris A, Anastasopoulos E, Yu F, Mavroudis L, Pappas T, Koskosas A, Coleman AL. Prevalence of open-angle glaucoma in Greece: the Thessaloniki Eye Study. Am J Ophthalmol. 2007;144:511-519.

14. Cireşan DC, Meier U, Gambardella LM, Schmidhuber J. Deep, big, simple neural nets for handwritten digit recognition. Neural Comput. 2010;22:3207-3220.

15. Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. International conference on artificial neural networks: Springer, 2010.

16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017;60:84-90.

17. Gómez-Valverde JJ, Antón A, Fatti G, Liefers B, Herranz A, Santos A, Sánchez CI, Ledesma-Carbayo MJ. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomed Opt Express. 2019;10:892-913.

18. Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A. CNNs for automatic glaucoma assessment using fundus images: an extensive validation. Biomed Eng Online. 2019;18:29.

19. Serener A, Serte S. Transfer learning for early and advanced glaucoma detection with convolutional neural networks. 2019 Medical Technologies Congress (TIPTEKNO): IEEE, 2019.

20. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition: Ieee, 2009.

21. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng CY, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY. 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;318:2211-2223.

22. Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, Asaoka R. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep. 2018;8:14665.

23. Masumoto H, Tabuchi H, Nakakura S, Ishitobi N, Miki M, Enno H. Deep-learning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity. J Glaucoma. 2018;27:647-652.

24. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, Kanamoto T, Ikeda Y, Mori K, Iwase A, Shoji N, Inoue K, Yamagami J, Araie M. Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images. Am J Ophthalmol. 2019;198:136-145.

25. Burgansky-Eliash Z, Wollstein G, Chu T, Ramsey JD, Glymour C, Noecker RJ, Ishikawa H, Schuman JS. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. Invest Ophthalmol Vis Sci. 2005;46:4147-4152.

26. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. PLoS One. 2017;12:e0177726.

27. Barella KA, Costa VP, Gonçalves Vidotti V, Silva FR, Dias M, Gomi ES. Glaucoma Diagnostic Accuracy of Machine Learning Classifiers Using Retinal Nerve Fiber Layer and Optic Nerve Data from SD-OCT. J Ophthalmol. 2013;2013:789129.

28. Christopher M, Belghith A, Weinreb RN, Bowd C, Goldbaum MH, Saunders LJ, Medeiros FA, Zangwill LM. Retinal Nerve Fiber Layer Features Identified by Unsupervised Machine Learning on Optical Coherence Tomography Scans Predict Glaucoma Progression. Invest Ophthalmol Vis Sci. 2018;59:2748-2756.

29. Bowd C, Zangwill LM, Medeiros FA, Hao J, Chan K, Lee TW, Sejnowski TJ, Goldbaum MH, Sample PA, Crowston JG, Weinreb RN. Confocal scanning laser ophthalmoscopy classifiers and stereophotograph evaluation for prediction of visual field abnormalities in glaucoma-suspect eyes. Invest Ophthalmol Vis Sci. 2004;45:2255-2262.

30. Townsend KA, Wollstein G, Danks D, Sung KR, Ishikawa H, Kagemann L, Gabriele ML, Schuman JS. Heidelberg Retina Tomograph 3 machine learning classifiers for glaucoma detection. Br J Ophthalmol. 2008;92:814-818.

31. Zangwill LM, Chan K, Bowd C, Hao J, Lee TW, Weinreb RN, Sejnowski TJ, Goldbaum MH. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. Invest Ophthalmol Vis Sci. 2004;45:3144-3151.

32. Uchida H, Brigatti L, Caprioli J. Detection of structural damage from glaucoma with confocal laser image analysis. Invest Ophthalmol Vis Sci. 1996;37:2393-2401.

33. Adler W, Peters A, Lausen B. Comparison of classifiers applied to confocal scanning laser ophthalmoscopy data. Methods Inf Med. 2008;47:38-46.

34. Weinreb RN, Zangwill L, Berry CC, Bathija R, Sample PA. Detection of glaucoma with scanning laser polarimetry. Arch Ophthalmol. 1998;116:1583-1589.

35. Medeiros FA, Jammal AA, Thompson AC. From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs. Ophthalmology. 2019;126:513-521.

36. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. Am J Ophthalmol. 1996;121:511-521.