Polskie Towarzystwo Okulistyczne
www.pto.com.pl

KLINIKA OCZNA 2020, 122, 1: 1–6
Received: 10.01.2020 Accepted: 26.02.2020

ORIGINAL ARTICLE

Access this article online

Website:
www.klinikaoczna.pl

DOI:
https://doi.org/10.5114/ko.2020.94205

A new platform designed for glaucoma screening: identifying the risk of glaucomatous optic neuropathy using fundus photography with deep learning architecture together with intraocular pressure measurements

Anna Zaleska-Żmijewska[1,2,3], Jacek P. Szaflik[1,2,3], Paweł Borowiecki[3], Katarzyna Pohnke[4], Urszula Romaniuk[4], Izabela Szopa[4], Jacek Pniewski[4], Jerzy Szaflik[3]

[1]Department of Ophthalmology, Second Faculty of Medicine, Medical University of Warsaw, Warsaw, Poland
[2]SPKSO Ophthalmic University Hospital, Warsaw, Poland
[3]Medical Research Centre Brand Med, Warsaw, Poland
[4]Faculty of Physics, University of Warsaw, Warsaw, Poland

ABSTRACT

Aim of the study: To develop a platform designed for glaucoma screening, based on deep learning algorithms, for the diagnosis of glaucomatous optic neuropathy from colour fundus images and intraocular pressure, not requiring medical staff.

Material and methods: A modular platform for glaucoma screening is developed which uses classifiers that independently evaluate the parameters of the visual system. The fundus image classifier is based on trainable mathematical models, while an intraocular pressure classifier is a threshold classifier. Performance analysis is conducted in terms of the statistical parameters: sensitivity, accuracy, precision, and specificity. Glaucoma images were classified by two experts. The cut-off of vertical cup to disc ratio (vCDR) for glaucoma was set at ≥ 0.7. In the training stage 933 healthy and 754 glaucoma images were used. If the intraocular pressure (IOP) was ≥ 24 mmHg in at least one eye the patient was classified in the glaucoma category independently of the fundus image category. During the training stage the following parameters of the image classifier were achieved: sensitivity 0.82 and specificity 0.63.

Results and conclusions: For the test data from two campaigns were used (total 1104 fundus images). For the image classifier, sensitivity 0.73 and specificity 0.83 were obtained for the first campaign, while sensitivity 0.84 and specificity 0.67 were obtained for the second campaign. The final achieved parameters of the platform are: sensitivity 0.79 and specificity 0.67 for the first campaign, sensitivity 0.92 and specificity 0.42 for the second campaign. The results are in accordance with other studies and the platform proved its usability and good performance.

KEY WORDS: glaucoma, artificial intelligence, image classification, screening, deep learning.

INTRODUCTION

Glaucoma is a progressive optic nerve neuropathy, which is one of the leading causes of irreversible blindness worldwide. It affects approximately 70 million people and the number will increase to 112 million in 2040 [1]. Even in advanced stages glaucoma may remain asymptomatic, especially if asymmetric glaucomatous optic nerve damage is present. Thus, there is a great need for screening programmes for glaucoma to diagnose the disease at early levels without changes in visual fields [2, 3]. The main diagnostic tools for glaucoma typically include optic nerve head (ONH) evaluation, detection of visual field defects and elevated intraocular pressure.

Glaucomatous optic neuropathy (GON) is characterised by some typical changes in the appearance of the optic disc: enlarged vertical cup to disc ratio (vCDR), notching and/or thinning of neuroretinal rim, parapapillary atrophy, nasalization of central ONH vessels, disc haemorrhages, asymmetric vCDR [4-6]. These changes can be detected in fundus images and are one of the most important aspects of glaucoma diagnosis. Increased intraocular pressure (IOP) is still considered the strongest risk factor for glaucoma development and progression, but is not necessary for glaucoma diagnosis. Clinical trials confirmed that 30-50% of glaucoma patients have IOP within normal limits [2].

CORRESPONDING AUTHOR
Jacek Pniewski, PhD, DSc, Faculty of Physics, University of Warsaw, 5 Pasteura St., 02-093 Warsaw, Poland, e-mail: j.pniewski@uw.edu.pl

There is no single reference standard for establishing the diagnosis of glaucoma. The diagnosis is usually made on the basis of clinical examination combined with several tests for structural and functional optic nerve damage and intraocular pressure measurements [4, 6]. The evaluation of GON may be performed by observing the ONH in fundus photographs and tests using laser scanning imaging techniques such as scanning laser tomography (HRT), scanning laser polarimetry (GDX), and optical coherence tomography (OCT) [7, 8]. In the clinic, retinal fundus photography is a well-established diagnostic tool for eye diseases and the easiest and low-cost test. The interpretation of colour fundus ONH images requires expert knowledge and may differ even among experienced glaucoma specialists. Thus, there is a need for development of automatic methods for GON detection based on fundus images. Several reports have proved the efficacy of machine learning in glaucoma [9-23].

Machine learning is a system of artificial computer intelligence that provides computers with the ability to automatically learn without being explicitly programmed. In ophthalmology machine learning has been used to investigate diabetic retinopathy (DR), age-related macular degeneration (AMD) and glaucoma [10-12]. Recent studies of DR using deep machine learning showed high sensitivity and specificity for the detection of changes typical for DR [11, 12, 15]. The detection of glaucoma may be more challenging than the diagnosis of DR because it relies on estimation of subtle changes in the ONH shape and cupping including the stage of glaucoma and refractive errors of the eye [9, 14].

The aim of this study was to evaluate a simple and cost-effective screening platform designed to use convolutional neural networks for glaucoma detection using colour fundus photography together with non-contact IOP measurements. The existing few commercial systems use only images of the fundus (RetinaLyze) or focus on diabetic retinopathy (Eyenuk), and do not reveal the details of the neural network-based classifiers used for examination.

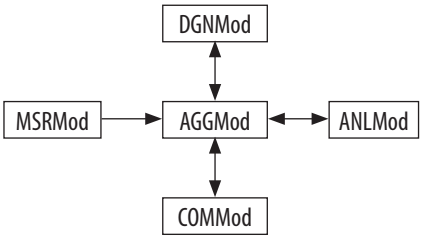

Figure 1. Schematic diagram of the screening platform

Table I. Definitions of statistical parameters

Parameter	Definition
Sensitivity	$\frac{TP}{TP+FN}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Specificity	$\frac{TN}{TN+FP}$

MATERIAL AND METHODS

General description of the platform

The modular platform for glaucoma screening consists of the following modules: aggregating module (AGGMod), measuring module (MSRMod), analytical module (ANLMod), diagnostic module (DGNMod), and communication module (COMMod), as shown in Figure 1.

The AGGMod connects other modules of the platform and provides access to a remote server that collects all data of the examined person and the examination results. An operator registers personal data and medical history of the patient in the server system for future reference.

The MSRMod collects data, at least full-colour fundus images and IOP, from diagnostic devices: a fundus camera and a tonometer. The list of the devices and the number of collected parameters can be extended in future versions of the platform. At least one eye must be imaged, but the platform by default collects data from both eyes.

The ANLMod uses classifiers that independently evaluate the parameters of the visual system: an image of a fundus and an IOP value of an eye. The image classifier is based on trainable mathematical models (deep neural networks, artificial intelligence) and is discussed in detail in the next section. The IOP classifier is a threshold classifier that is based on a threshold value, resulting from the medical knowledge. If the IOP is equal to or higher than this value, the classifier indicates high probability of glaucoma.

The DGNMod issues an initial diagnosis for the patient (glaucoma or healthy) on the basis of the information from the ANLMod. Then, if at least one classifier indicated an increased risk of glaucoma, a message to the patient is sent, using COMMod, as an SMS message or an e-mail, with an indication of a need of a medical appointment. If all classifiers identified healthy eyes, a message is sent with a recommendation for regular eye tests. Otherwise, a message is sent with information about failure during tests. The messages also confirm proper communication with the patient. If the message is not received by the patient he or she is obliged to repeat the test or make a medical appointment.

Performance analysis of a classifier or the whole platform is conducted in terms of the following statistical parameters: sensitivity, accuracy, precision, and specificity, defined in Table I, where TP – true positives, FP – false positives, TN – true negatives, and FN – false negatives, with respect to the unanimous opinion of two qualified ophthalmologists.

The whole information technology system which runs the platform is developed in-house.

Data sets (MSRMod)

The prospective study was conducted between March 2019 and July 2019. Adults (> 18 years) with visual acuity and cen-

tral fixation enough to obtain good quality photography of the fundus were included in the study. The exclusion criteria were: difficulties in obtaining clear fundus images in both eyes due to opacities in optical media or narrow pupil.

The datasets for this study consisted of fundus images obtained from two sources: Ophthalmic Teaching Hospital Glaucoma Outpatients Clinic, which is a clinical base for the Department of Ophthalmology Medical University of Warsaw, and Microsurgery Eye Centre "Laser" in Warsaw.

The study protocol was accepted by the Bioethical Commission of the Medical University of Warsaw and was conducted in accordance with the Declaration of Helsinki.

Before the measurement a short questionnaire was collected including age, family history of glaucoma, systemic diseases that are risk factors for glaucoma (arterial hypertension, systemic hypotony, diabetes mellitus, headaches, Raynaud's syndrome), previous or current treatment for glaucoma, according to indications from the literature [24, 25].

For the preparation of the training set of images for the classifier the following devices were used: Nidek RS 330-Retina Scan Duo (first 196 patients), Haag-Streit DRS (subsequent 200 patients), and Crystalvue NFC-700 (all successive patients). The measurements were carried out in a darkened room with a light intensity not exceeding 1 lux, after 5-10 minutes of adaptation to darkness, in non-mydriatic mode of operation, without pupil dilation. For the operating stage only Crystalvue NFC-700 was used; it makes images of size 4096 × 3072 with 24-bit colour depth, in lossless PNG file compression format. The measurement was centred on the optic disc area, with a 30-degree field of view. No additional correction of illumination or contrast enhancement was applied to the images.

During fundus image acquisition, it is necessary for a device operator to indicate the location of the centre of the optic nerve disc. This procedure allows for the further image preprocessing in ANLMod before the classification of the image and is also used for verification of the quality of the acquired image. The centre of the optic disc was determined within the vascular trunk in eyes with retinal vessels centrally located, while in cases of nasal displacement of the vessels or atypical discs, e.g. oblique or with peripapillary atrophy, every eye was checked individually, looking at the edge of the optic disc area.

We considered two categories of fundus images: healthy or glaucoma. Glaucoma included images classified by two experts as suspected glaucoma or with glaucoma. Graders assessed the specific ONH features in every image. In case of disagreement between the two evaluations performed, both glaucoma experts decided the final classification of the image by consensus. The cut-off of vCDR for the glaucoma category was set at vCDR ≥ 0.7. In spite of increased CDR other features characteristic for GON were considered as specific for this category [4, 6].

After fundus photography a non-contact IOP was measured. For the whole screening project a Topcon CT-800A tonometer was used. The reference cut-off of IOP for healthy eyes was less than 24 mmHg. If the IOP was ≥ 24 mmHg in at least one eye the patient was classified in the glaucoma category independently of the fundus image category.

In preparation of the training and validation dataset, subjects' demographics, such as age and sex, and other ophthalmological findings such as visual field defects, intraocular pressure level, gonioscopic appearance, and OCT measurements, were not considered in the diagnosis of glaucoma.

Fundus image classifier (ANLMod)

Training stage

The data from imaging devices provided at this stage by the MSRMod and passed to the ANLMod are images of different size. Since only the area of the optic nerve is assessed by the classifier the images were cropped using a window of size 900 × 900 pixels centred around a manually pointed centre of the disc, and then scaled using bicubic interpolation to a size of 227 × 227 pixels, as required by the classifier. The 'healthy' set contained 933 images and the 'glaucoma' set contained 754 images. Samples of resulting images are shown in Figure 2. The images presenting increased optic nerve discs are not processed separately. The neural network classifier during the training stage uses only 'healthy/glaucoma' status of the image. In this way, for both increased and normal discs the proper operation should be achieved.

As a classifier the AlexNet convolutional neural network was selected on the basis of preliminary tests [26]. This network is widely used in medical applications [27]. We developed the classifier using the deep learning framework *caffe* [28], on the basis of 30 models of neural network. The models were different in terms of the learning rate (*base_lr*), the learning rate multiplier (*lr_mult*), and the number of layers. For the final selection of the network structure and its parameters two methods were employed: *transfer learning* and *data augmentation*. Transfer learning is a method of solving one problem and applying the solution to a different but related problem. In our case, the initial ability of the AlexNet network to discriminate a number of classes of images was used to discriminate healthy and glaucoma images. Data augmentation is a method of expanding an available set of images using transformations of the original images. In our case, a number of transformations were analysed and used, such as mirror, shift, scale, contrast and intensity change. During the training stage the following statistical parameters were achieved: sensitivity 0.82, accuracy 0.72, precision 0.65, and specificity 0.63. In this calculation, eyes were analysed individually. Statistical tests were used to confirm the hypotheses that certain models are better than others.

The classifier model can be tuned to achieve better performance in terms of statistical parameters, at certain time intervals in the operational stage, when more training images of the fundus are collected. Two or more classifiers can also operate simultaneously, allowing for a smooth change to the updated classifier model.

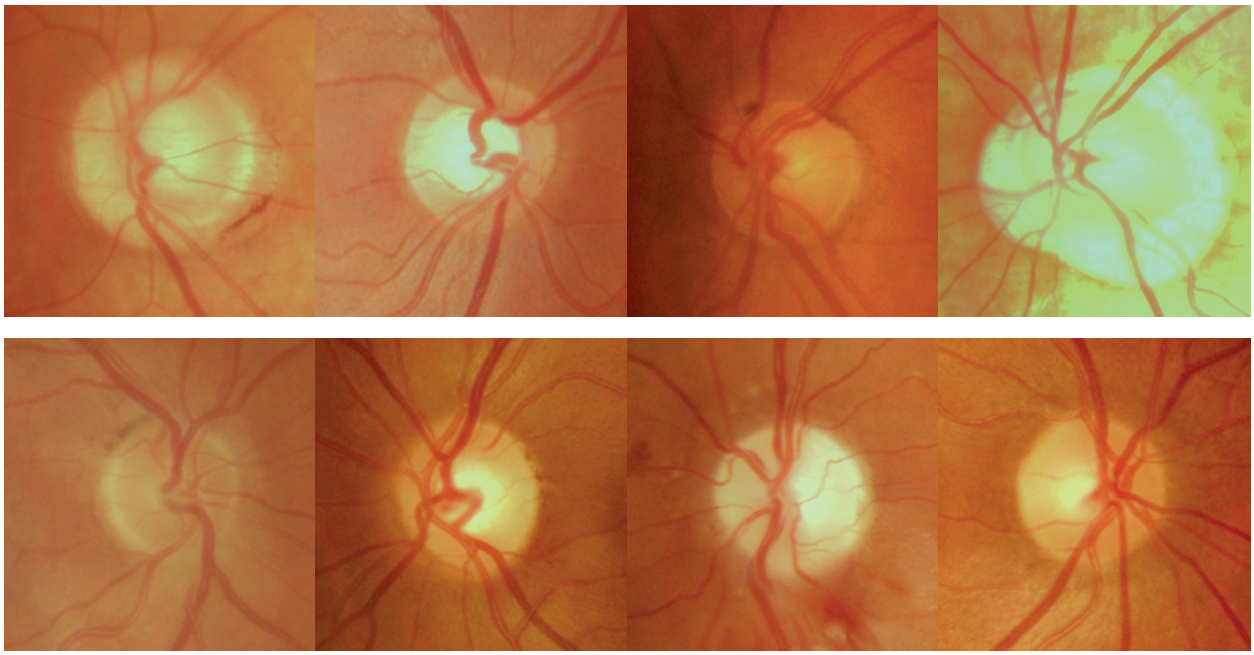

Figure 2. Sample images used to train the classifier: upper row – glaucoma, bottom row – healthy

RESULTS

Performance of the image classifier in the operational stage

For the test of the performance of the platform data from two measurement campaigns were used. In the first one, 752 fundus images were collected from 382 patients, while in the second one 352 images were collected from 202 patients. As a result, values of statistical parameters were achieved as shown in Table II. In this case, the calculation is based only on image data (without IOP) and thus does not represent the overall performance of the platform. Figure 3 depicts a histogram of the age of the patients in the first campaign.

Table II. Statistical parameters achieved by the image classifier in the test campaigns

Parameter	Value (campaign 1)	Value (campaign 2)	Reference from training
# of eyes	752	352	
TP	192	236	
TN	560	116	
FP	112	57	
FN	71	44	
Sensitivity	0.73	0.84	0.82
Accuracy	0.80	0.78	0.72
Precision	0.63	0.81	0.65
Specificity	0.83	0.67	0.63

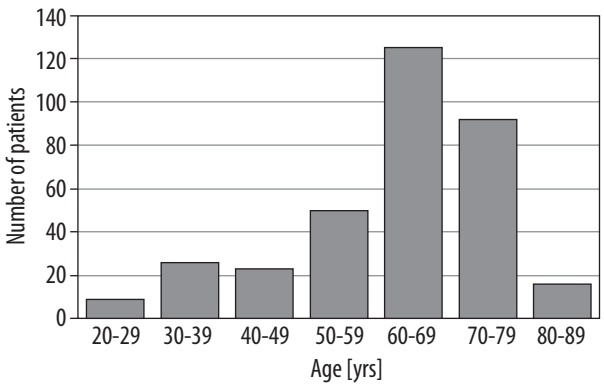

Figure 3. Histogram of the age of the patients in the first test campaign

To summarize, for the first campaign a decrease of sensitivity and precision was observed while the accuracy and specificity were higher than in the training stage. For the second campaign an increase of all parameters was reported.

The final automatic diagnosis is issued on the basis of fundus image(s) and IOP for both eyes, therefore increasing the sensitivity of the platform, because the increased risk of glaucoma is identified for a patient even if only one of all classifiers (one or two fundus images and IOPs) for both eyes returns such an indication. The statistical parameters calculated on the basis of the final diagnosis are shown in Table III.

DISCUSSION

The automatic detection of early-stage glaucoma is still one of the most challenging problems in medical image analysis and a number of test frameworks are available for comparison of different methods [23].

The statistical parameters, particularly sensitivity and specificity, as measures of the performance of a fundus image classifier, determined at the training stage, may not fully predict the operation of a classifier and the whole platform. For training, images are selected which can be clearly assigned to one of the classes: 'glaucoma' or 'healthy'. In real measurements there are several factors that may influence the assignment. They include bad condition of an eye's optical system (e.g. cataracts, haemorrhages), difficulties in operating the measuring device (fuzzy, partial, shifted or very dark images,

very narrow pupil, etc.) and errors in indicating the centre of the optic nerve disc (operator's interpretation). Therefore, the performance of a platform may decrease. Moreover, in an examined group of persons healthy or glaucoma patients may be overrepresented, which leads to change of the statistical parameters. The latter issue may be the reason for the increase of parameters in both campaigns.

For the second campaign the specificity was as low as 0.42. The reason was that in this campaign the patients with advanced glaucoma were overrepresented. The variety of cases included types of images that were not common in the training stage, which caused random assignment in certain cases. These cases will be included in the future versions of the platform.

The above-mentioned actual problems can be partially resolved due to training of the staff that takes the measurements, indicates the centre of the optic nerve disc, and verifies the quality of the image. Additionally, having *a priori* knowledge of a group one can use an additional threshold in ANL-Mod. For example, if the prevalence of glaucoma in specific groups of patients is known, one can use Bayesian inference to control the classifier operation (thresholds), thus maintaining the desired statistical parameters.

Our results are in accordance with the results obtained by other authors, despite using a relatively small number of fundus images for training, as compared, e.g., with the study by Shibata *et al.*, where the training set was selected on the basis of 16,000 images [19]. For example, in the paper by Ahn *et al.* for early-stage glaucoma accuracy in the range 0.73-0.88, depending of the classifier version, was presented [14]. Fu *et al.* reported sensitivity 0.85 and specificity 0.84 [29]. In a review paper by Gómez-Valverde *et al.* the performances of a number of different convolutional neural networks were presented [20]. As a result, the sensitivity in the range 0.79-0.92, and the specificity in the range 0.75-0.91 were achieved. The authors also stressed that the number of images and type of data sets used in the training stage significantly influence the performance of a neural network classifier. It is also worth noting that most of the papers present analyses of specific neural network classifiers, not running operational platforms.

Table III. Statistical parameters achieved in the test campaigns

Parameter	Value (campaign 1)	Value (campaign 2)
# of patients	381	202
TP	89	137
TN	179	22
FP	88	30
FN	23	12
Sensitivity	0.79	0.92
Accuracy	0.71	0.79
Precision	0.50	0.82
Specificity	0.67	0.42

The image and IOP classifiers need single-digit seconds to issue an initial diagnosis, depending on the performance of the computer used in examination. The time required to examine a single patient is about 5 to 10 minutes, while the final message is sent within the next 5 minutes.

CONCLUSIONS

Our first platform proved its usability and good performance in the real screening projects. This study demonstrates that deep learning techniques combined with fundus photography and IOP measurements are an effective tool for a glaucoma screening programme distinguishing between normal and glaucoma patients.

Although AlexNet was successfully employed in our platform in future versions other architectures that can handle bigger input images than 227 × 227 should be used, and the focus should be set to improve the image classifier that can reliably evaluate images from different devices, both professional and simple ones, such as mobile phones equipped with additional optics.

DISCLOSURE

The authors declare no conflict of interest.

References

1. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 2014; 121: 2081-2090.
2. Tielsch JM, Katz J, Singh K, et al. A population-based evaluation of glaucoma screening: The Baltimore Eye Survey. Am J Epidemiol 1991; 134: 1102-1110.
3. Maul EA, Jampel HD. Glaucoma screening in the real world. Ophthalmology 2010; 117: 1665-1666.
4. Fingeret M, Medeiros FA, Susanna R, et al. Five rules to evaluate the optic disc and retinal nerve fiber layer for glaucoma. J Am Optom Assoc 2005; 76: 661-668.
5. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. JAMA 2014; 311: 1901-1911.
6. Prum BE, Rosenberg LF, Gedde SJ, et al. Primary open-angle glaucoma preferred practice pattern® guidelines. Ophthalmology 2016; 1: P41-111.
7. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. J Glaucoma 2017; 26: 1086-1094.
8. Tatham AJ, Medeiros FA. Detecting structural progression in glaucoma with optical coherence tomography. Ophthalmology 2017; 124: S57-65.
9. Chan K, Lee TW, Sample PA, et al. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. IEEE Trans Biomed Eng 2002; 49: 963-974.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436-444.

11. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316: 2402-2410.
12. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016; 57: 5200-5206.
13. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. PLoS One 2017; 12: e0177726.
14. Ahn JM, Kim S, Ahn K-S, et al. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. PLoS One 2018; 13: e0207982.
15. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018; 172: 1122-1131.e9.
16. An G, Omodaka K, Tsuda S, et al. Comparison of machine-learning classification models for glaucoma management. J Health Eng 2018; 2018: 6874765.
17. Christopher M, Belghith A, Bowd C, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep 2018; 8: 1-13.
18. Li Z, He Y, Keel S, et al. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology 2018; 125: 1199-1206.
19. Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep 2018; 8: 1-9.
20. Gómez-Valverde JJ, Antón A, Fatti G, et al. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomed Opt Express 2019; 10: 892-913.
21. Phene S, Dunn RC, Hammel N, et al. Deep learning and glaucoma specialists: the relative importance of optic disc features to predict glaucoma referral in fundus photographs. Ophthalmology 2019; 126: 1627-1639.
22. Asaoka R, Murata H, Hirasawa K, et al. 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.
23. Orlando JI, Fu H, Barbosa 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.
24. Kośny J, Jurowski P. Wybrane czynniki ryzyka jaskry i ich rola w progresji choroby. Część I – czynniki ogólne oraz głównie miejscowe i mechaniczne. Klin Oczna 2018; 3: 159-163.
25. Kośny J, Jurowski P. Wybrane czynniki ryzyka jaskry i ich rola w progresji choroby. Część II – czynniki głównie ogólnoustrojowe i naczyniowe. Klin Oczna 2018; 3: 164-167.
26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, et al. (eds.). Advances in Neural Information Processing Systems 25. Curran Associates, Inc. 2012; 1097-1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (accessed 31 Dec 2019).
27. Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. J Digit Imaging 2017; 30: 234-243.
28. Caffe/Deep Learning Framework. https://caffe.berkeleyvision.org/ (accessed 8 Feb 2020).
29. Fu H, Cheng J, Xu Y, et al. Disc-aware ensemble network for glaucoma screening from fundus image. IEEE Trans Med Imaging 2018; 37: 2493-2501.