

# A Deep Learning Algorithm to Quantify Neuroretinal Rim Loss From Optic Disc Photographs



ATALIE C. THOMPSON, ALESSANDRO A. JAMMAL, AND FELIPE A. MEDEIROS

- **PURPOSE:** To train a deep learning (DL) algorithm that quantifies glaucomatous neuroretinal damage on fundus photographs using the minimum rim width relative to Bruch membrane opening (BMO-MRW) from spectral-domain optical coherence tomography (SDOCT).
- **DESIGN:** Cross-sectional study.
- **METHODS:** A total of 9282 pairs of optic disc photographs and SDOCT optic nerve head scans from 927 eyes of 490 subjects were randomly divided into the validation plus training (80%) and test sets (20%). A DL convolutional neural network was trained to predict the SDOCT BMO-MRW global and sector values when evaluating optic disc photographs. The predictions of the DL network were compared to the actual SDOCT measurements. The area under the receiver operating curve (AUC) was used to evaluate the ability of the network to discriminate glaucomatous visual field loss from normal eyes.
- **RESULTS:** The DL predictions of global BMO-MRW from all optic disc photographs in the test set (mean  $\pm$  standard deviation [SD]:  $228.8 \pm 63.1 \mu\text{m}$ ) were highly correlated with the observed values from SDOCT (mean  $\pm$  SD:  $226.0 \pm 73.8 \mu\text{m}$ ) (Pearson's  $r = 0.88$ ;  $R^2 = 77\%$ ;  $P < .001$ ), with mean absolute error of the predictions of  $27.8 \mu\text{m}$ . The AUCs for discriminating glaucomatous from healthy eyes with the DL predictions and actual SDOCT global BMO-MRW measurements were 0.945 (95% confidence interval [CI]: 0.874–0.980) and 0.933 (95% CI: 0.856–0.975), respectively ( $P = .587$ ).
- **CONCLUSIONS:** A DL network can be trained to quantify the amount of neuroretinal damage on optic disc photographs using SDOCT BMO-MRW as a reference. This algorithm showed high accuracy for glaucoma detection, and may potentially eliminate the need for human gradings of disc photographs. (Am J Ophthalmol 2019;201:9–18. © 2019 Elsevier Inc. All rights reserved.)

**G**LAUCOMA IS THE LEADING CAUSE OF IRREVERSIBLE blindness worldwide, and will affect approximately 80 million people by 2020.<sup>1</sup> Nevertheless, several population-based surveys have suggested that the majority of patients with glaucoma are unaware that they have the disease.<sup>2,3</sup> The reasons for this are likely multifactorial, including poor public knowledge about glaucoma<sup>3,4</sup> and the fact that glaucoma symptoms may be minimal until the later stages.<sup>5</sup> Public health interventions that can effectively and inexpensively screen populations for glaucoma are thus needed so that patients can be diagnosed and treated earlier in the disease course, before they suffer irreversible visual dysfunction.

Over the past two decades, nonmydriatic fundus photographs have been widely and effectively employed to screen for diabetic retinopathy via teleophthalmology.<sup>6,7</sup> However, the use of optic disc photographs for glaucoma screening has been limited by the dependence on human graders. Manual review of disc photographs is not only time-consuming, but also highly subjective. When used in a screening setting, monoscopic fundus photograph gradings have also shown poor sensitivity for glaucoma.<sup>8–13</sup> Numerous studies have documented that gradings of optic disc photographs have only modest reproducibility<sup>14,15</sup> and fair interrater reliability,<sup>8,15,16</sup> even among expert graders with fellowship training in glaucoma.<sup>16</sup>

Nevertheless, interest in the potential use of fundus photographs for glaucoma screening<sup>17,18</sup> has been renewed in light of recent publications demonstrating the ability of deep learning algorithms to provide accurate classification of retinal diseases on fundus photographs.<sup>18–21</sup> Most recently, Li and associates<sup>17</sup> showed that a deep learning neural network algorithm could be trained to detect optic nerves suspicious, or “referable,” for glaucoma on fundus photographs. While their work is enlightening, the true accuracy of their algorithm is limited by their reliance on the use of subjectively graded fundus photographs for the reference standard. If a neural network is trained using subjectively graded photographs, then the neural network will be trained to mimic the original errors in discrimination made by the human graders. For example, optic nerves with physiologically enlarged cups may be misdiagnosed with glaucoma, whereas true glaucomatous damage on small optic discs may be underappreciated.<sup>17</sup>

When ophthalmologists in clinical practice have difficulty determining whether an optic nerve's appearance is

AJO.com

Supplemental Material available at [AJO.com](http://AJO.com).

Accepted for publication Jan 17, 2019.

From the Vision, Imaging and Performance (VIP) Laboratory, Duke Eye Center and Department of Ophthalmology, Duke University, Durham, North Carolina, USA.

Inquiries to Felipe A. Medeiros, Duke Eye Center, Department of Ophthalmology, Duke University, 2351 Erwin Rd, Durham, NC 27705 USA; e-mail: [felipe.medeiros@duke.edu](mailto:felipe.medeiros@duke.edu)

glaucomatous, they turn to objective metrics such as spectral-domain optical coherence tomography (SDOCT) to provide quantitative structural information regarding the optic disc and surrounding tissue. We propose that such quantitative data derived from structural measurements of the optic nerve head may provide a better reference standard for development of neural networks than qualitative gradings of the optic disc on fundus photographs. The minimum rim width relative to the Bruch membrane opening (BMO-MRW) is a fairly new parameter that has been introduced for the evaluation of the neuroretinal rim on SDOCT. The BMO-MRW is defined by the minimum distance from the internal limiting membrane (ILM) to the inner opening of the BMO within the SDOCT scans averaged around the disc ([Supplemental Figure](#); supplemental material available at [AJO.com](#)).<sup>22,23</sup> Several studies have suggested that BMO-MRW is at least as accurate as<sup>24–26</sup> and may be more sensitive than peripapillary retinal nerve fiber layer (RNFL) and other rim-based parameters<sup>22,24,27</sup> for diagnosing glaucoma. BMO-MRW has shown a very strong correlation with visual field loss in glaucoma.<sup>24,25,27,28</sup> Moreover, BMO-MRW may be particularly sensitive for the identification of early glaucoma<sup>22,29</sup> and glaucoma suspects,<sup>24</sup> making it a potentially useful metric for screening.

In the current study, we develop and validate a novel deep learning algorithm that has been trained using the BMO-MRW from SDOCT to detect glaucomatous optic neuropathy on fundus photographs and to predict the quantitative amount of neuroretinal damage.

---

## METHODS

THE DATA USED IN THIS STUDY WERE DRAWN FROM THE Duke Glaucoma Repository, a database of electronic research and medical records developed by the Duke University Vision, Imaging and Performance (VIP) Laboratory. The Duke Institutional Review Board approved this study with a waiver of informed consent owing to the retrospective nature of this research. The study protocol adhered to the tenets of the Declaration of Helsinki and was conducted in accordance with the Health Insurance Portability and Accountability Act.

All included subjects were adults at least 18 years of age. Records were reviewed for diagnosis, medical history, and results from comprehensive ophthalmic examination including visual acuity, intraocular pressure, slit-lamp biomicroscopy, gonioscopy, and dilated fundus examination. In addition, stereoscopic optic disc photographs (Nidek 3DX, Nidek, Japan) and Spectralis SDOCT (software version 5.4.7.0; Heidelberg Engineering, GmbH, Dossenheim, Germany) images and associated data were collected. Standard automated perimetry (SAP) acquired with a size III stimulus using the 24-2 test pattern and Swedish interactive threshold algorithm

(Carl Zeiss Meditec, Inc, Dublin, California, USA) was included if the test met reliability parameters such as fewer than 33% fixation losses and less than 15% false-positive errors. Patients with other ocular or systemic diseases that could affect the optic nerve or visual field were excluded.

Eyes were categorized with glaucoma if they had evidence of glaucomatous optic neuropathy documented on dilated fundus examination (ie, cupping, diffuse or focal rim thinning, optic disc hemorrhage, or RNFL defects) and a reproducible visual field defect on 2 consecutive SAP with pattern standard deviation <5% or glaucoma hemifield test outside normal limits. Glaucoma suspects had ocular hypertension, positive family history, or a suspicious-appearing optic nerve on clinical examination. Normal subjects had to have a normal dilated examination with no evidence of ocular hypertension or abnormality on SAP.

• **SPECTRAL-DOMAIN OPTICAL COHERENCE TOMOGRAPHY:** Images of the optic nerve head (ONH) were acquired using Spectralis SDOCT (software version 5.4.7.0; Heidelberg Engineering, GmbH). The device has a dual-beam SDOCT as well as a confocal laser-scanning ophthalmoscope that uses a superluminescent diode light with a center wavelength of 870 nm as well as an infrared scan to provide simultaneous images of ocular microstructures. In order to decrease geometric errors, scans were acquired with reference to the fovea-BMO axes of the patient's eye,<sup>27</sup> and data were sectorized based on these axes. A live B-scan was used to detect the fovea and the center of the BMO. A radial scan pattern with 24 equidistant radial B-scans that each subtended 15 degrees was centered on the BMO and used to compute the neuroretinal rim parameters. Each B-scan was averaged from 1536 A-scans per B-scan acquired at a scanning speed of 40,000 A-scans per second. Automated software (Glaucoma Module Premium Edition, version 6.0; Heidelberg Engineering) was used to identify the ILM and BMO. The BMO-MRW was calculated as the minimum distance from the ILM to the BMO averaged over the 24 radial B-scans.<sup>28</sup> This was computed automatically for both the global value and the sectors by reference to Garway-Heath structure-function maps.<sup>30</sup>

During image acquisition, the device's image registration and eye-tracking software was used to reduce the effect of eye movements and to ensure that the same location was scanned over time. Images were manually reviewed to ensure scan centration, signal strength of at least 15 dB, and high image quality without image inversion or clipping. SDOCT images with coexistent retinal pathology were excluded. In order to improve the heterogeneity of the dataset for deep learning, we used all available optic disc photographs acquired over time and matched them to the closest SDOCT ONH scan within 6 months of the photograph date.

• **DEVELOPMENT OF THE DEEP LEARNING ALGORITHM:** We trained a deep learning algorithm to predict the SDOCT BMO-MRW from assessment of optic disc

**TABLE 1.** Demographic and Clinical Characteristics of the Eyes and Patients Included in the Training and Test Samples

	Normal	Suspect	Glaucoma
<b>Training sample<sup>a</sup></b>			
Number of eyes	124	178	441
Number of images	1046	1924	4570
Age (years)	65.6 ± 10.9	66.9 ± 11.2	71.3 ± 10.6
Sex, % female	67.8	58.2	52.0
Race (%)			
White	58.1	70.8	69.6
African American	41.9	29.2	30.4
SAP MD (dB)	-0.88 ± 1.68	-0.26 ± 1.63	-7.12 ± 7.24
SAP PSD (dB)	2.18 ± 0.91	1.59 ± 0.24	6.37 ± 4.02
Global BMO-MRW (μm)	301.0 ± 52.7	250.1 ± 58.3	205.3 ± 64.7
<b>Test sample<sup>b</sup></b>			
Number of eyes	42	34	108
Number of images	340	432	970
Age (years)	68.3 ± 9.7	72.6 ± 9.1	72.0 ± 11.0
Sex, % female	66.7	68.6	51.7
Race (%)			
White	61.9	81.4	80.8
African American	38.1	18.6	19.2
SAP MD (dB)	-0.39 ± 1.35	-0.04 ± 1.35	-5.62 ± 5.9
SAP PSD (dB)	2.14 ± 1.07	1.62 ± 0.20	5.11 ± 3.33
Global BMO-MRW (μm)	318.0 ± 51.2	230.1 ± 49.8	192.0 ± 60.2

BMO-MRW = minimum rim width relative to Bruch membrane opening; MD = mean deviation; PSD = pattern standard deviation; SAP = standard automated perimetry; SDOCT = spectral-domain optical coherence tomography.

<sup>a</sup>Training sample consists of 7540 pairs of disc photographs and SDOCT optic nerve head scans from 743 eyes of 392 subjects.

<sup>b</sup>Test sample consists of 1742 pairs of disc photographs and SDOCT optic nerve head scans from 184 eyes of 98 subjects.

photographs. The target value, or variable we wanted to predict from analysis of optic disc photographs, was the SDOCT BMO-MRW measurement. Separate models were used for the BMO-MRW global value and for each of the following sector values: superior temporal, superior nasal, temporal, nasal, inferior temporal, and inferior nasal. A pair of train-target for training the neural network consisted of the optic disc photograph and the SDOCT BMO-MRW values. The sample of pairs of photo-OCT was split into a training plus validation (80%) and test sample (20%). Random sampling was performed at the patient level so that no data of any patient were present in both the training and test samples. This helped to prevent leakage and biased estimates of test performance.

All optic disc stereophotographs were first preprocessed to derive data for the deep learning algorithm. Each stereoscopic photograph was split to create a pair of photographs from the stereo views. The images were subsequently down-sampled to 256 × 256 pixels and pixel values were scaled to range from 0 to 1. The heterogeneity of the photographs was improved with data augmentation, which helped to mitigate the risk of overfitting as well as allow the algorithm to learn the most relevant features of the image.

Data augmentation included random lighting, consisting of subtle changes in image balance and contrast; random rotation, consisting of rotations of up to 10 degrees in the image; and random horizontal image flips. Vertical flips were not performed to preserve the orientation of superior and inferior sectors.

We used the Residual deep neural Network (ResNet34) architecture that had been previously trained on the ImageNet dataset.<sup>31</sup> The ResNet is a deep residual network that allows relatively rapid training of very deep convolutional neural networks.<sup>32</sup> These networks use identity shortcut connections that skip one or more layers and decrease the vanishing gradient problem when training deep networks. Since the recognition task for the current investigation differs from that of ImageNet, we performed additional training by unfreezing the last two layers. Next, all layers were unfrozen and training was performed using differential learning rates. The network was trained with minibatch gradient descent of size 64. The best learning rate was determined using the cyclical learning method with stochastic gradient descents with restarts.

In order to assess the areas of the optic disc photographs that were most important in explaining the deep learning

**TABLE 2.** Mean Global and Sectoral Minimum Rim Width Relative to Bruch Membrane Opening Values Obtained From Spectral-Domain Optical Coherence Tomography and Corresponding Mean Values for the Predictions From the Deep Learning Algorithm in the Test Sample<sup>a</sup>

	Deep Learning Prediction From Fundus Photographs, Mean ± SD, μm	SDOCT BMO-MRW, Mean ± SD, μm	<i>P</i> for Difference in Means	<i>r</i> (R <sup>2</sup> )	<i>P</i> for Correlation	MAE, μm
Global	228.8 ± 63.1	226.0 ± 73.8	.415	0.88 (77%)	<.001	27.8
Temporal inferior	221.8 ± 78.9	228.4 ± 91.3	.141	0.81 (66%)	<.001	41.7
Temporal superior	206.8 ± 60.9	208.6 ± 75.7	.667	0.78 (62%)	<.001	36.8
Temporal	165.9 ± 34.8	163.1 ± 54.7	.487	0.68 (46%)	<.001	32.2
Nasal superior	253.3 ± 72.9	254.6 ± 90.4	.784	0.82 (67%)	<.001	38.8
Nasal inferior	273.6 ± 77.8	277.6 ± 104.3	.522	0.78 (60%)	<.001	51.9
Nasal	257.7 ± 71.2	253.9 ± 90.1	.459	0.82 (68%)	<.001	40.4

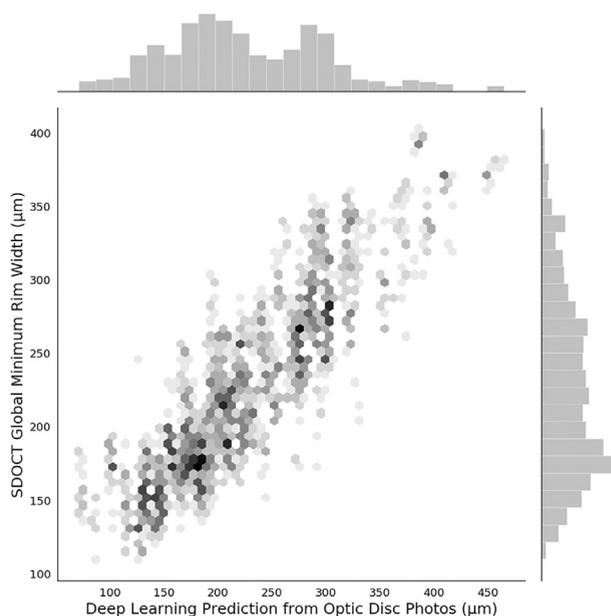
BMO-MRW = minimum rim width relative to Bruch membrane opening; MAE = mean absolute error; SDOCT = spectral-domain optical coherence tomography.

<sup>a</sup>The table also shows the Pearson correlation coefficient and mean absolute error between predictions and observations from the test sample.

algorithm predictions, we built heatmaps corresponding to the Gradient-weighted class activation maps over the input images.<sup>33,34</sup> For this analysis, we categorized SDOCT BMO-MRW into normal and abnormal categories based on the 95% cutoff from healthy eyes. These heatmaps indicate how important each location of the image is with respect to the class under consideration. This technique allows one to visualize the parts of the image that are most important in the deep neural network classification.

• **STATISTICAL ANALYSES:** We evaluated the performance of the deep learning algorithm for quantifying glaucomatous damage in optic disc photographs in the test sample by comparing the predictions of the algorithm with the actual SDOCT BMO-MRW global and sectoral values. Generalized estimating equations (GEE) were used to account for multiple measures within each patient. We calculated the mean absolute error of the predictions as well as the Pearson correlation coefficient.

We also evaluated the relationship between the predicted and observed values of sectoral BMO-MRW and the corresponding sectoral visual field sensitivities according to the Garway-Heath structure-function map.<sup>30</sup> Receiver operating characteristic curves were also used to compare the ability of the deep learning algorithm on optic disc photographs versus actual SDOCT BMO-MRW values to discriminate eyes with glaucomatous visual field loss from healthy eyes. Receiver operating characteristic (ROC) curves were plotted to demonstrate the tradeoff between the sensitivity and 1 – specificity. The area under the ROC curve (AUC) was used to assess the diagnostic accuracy of each parameter, with 1.0 representing perfect discrimination and 0.5 representing chance discrimination. Sensitivity at a fixed specificity of 95% was calculated. A bootstrap resampling procedure was used to derive 95%



**FIGURE 1.** Scatterplot illustrating the relationship between predictions obtained by the deep learning algorithm evaluating optic disc photographs and actual global minimum rim width relative to Bruch membrane opening (BMO-MRW) thickness measurements from spectral-domain optical coherence tomography (SDOCT). Data are from the independent test set.

confidence intervals and *P* values, where the cluster of data for the participant was considered as the unit of resampling. This procedure is commonly employed to adjust for the presence of multiple correlated measurements within the same subject,<sup>35</sup> and was used to adjust standard errors in this study since multiple images of both eyes in the same subject had been used.

**TABLE 3.** Correlations With 95% Confidence Interval Between Visual Field Sensitivity by Sectors and the Corresponding Minimum Rim Width Relative to Bruch Membrane Opening Deep Learning Predictions and Actual Spectral-Domain Optical Coherence Tomography Values

	Deep Learning BMO-MRW Prediction (95% CI)	SDOCT BMO-MRW (95% CI)
Temporal inferior	0.52 (0.42–0.63)	0.59 (0.49–0.70)
Temporal superior	0.43 (0.32–0.54)	0.50 (0.39–0.61)
Temporal	0.38 (0.22–0.54)	0.41 (0.29–0.52)
Nasal superior	0.40 (0.30–0.50)	0.42 (0.31–0.53)
Nasal inferior	0.39 (0.28–0.50)	0.48 (0.37–0.60)
Nasal	0.36 (0.23–0.50)	0.38 (0.27–0.50)

BMO-MRW = minimum rim width relative to Bruch membrane opening; CI = confidence interval; SDOCT = spectral-domain optical coherence tomography.

**TABLE 4.** Areas Under the Receiver Operating Characteristic Curves to Discriminate Eyes With Glaucoma From Healthy Eyes in the Test Sample for Spectral-Domain Optical Coherence Tomography Minimum Rim Width Relative to Bruch Membrane Opening and the Corresponding Predictions From the Deep Learning Algorithm Assessing Fundus Photographs

	Deep Learning BMO-MRW Prediction	SDOCT BMO-MRW	P
Global	0.945 (0.874–0.980)	0.933 (0.856–0.975)	.587
Temporal inferior	0.939 (0.885–0.973)	0.927 (0.843–0.969)	.632
Temporal superior	0.935 (0.879–0.970)	0.904 (0.822–0.969)	.334
Temporal	0.915 (0.829–0.962)	0.843 (0.734–0.927)	.180
Nasal superior	0.949 (0.886–0.980)	0.927 (0.848–0.974)	.365
Nasal inferior	0.942 (0.884–0.979)	0.952 (0.894–0.986)	.689
Nasal	0.919 (0.837–0.973)	0.898 (0.794–0.958)	.447

BMO-MRW = minimum rim width relative to Bruch membrane opening; ROC = receiver operating characteristic; SDOCT = spectral-domain optical coherence tomography.

## RESULTS

THE STUDY INCLUDED 9282 PAIRS OF OPTIC DISC PHOTOGRAPHS and SDOCT optic nerve head scans from 927 eyes of 490 subjects, divided into training and validation (80%) and test sample (20%). The test sample consisted of 1742 pairs of disc photographs and SDOCTs from 184 eyes of 98 subjects. [Table 1](#) shows demographic and clinical characteristics of the subjects and eyes in the training and test samples.

[Table 2](#) shows mean BMO-MRW predictions in the test set from the deep learning algorithm for global and sectoral areas, as well as corresponding actual SDOCT BMO-MRW measurements. For global BMO-MRW, the mean prediction from the deep learning algorithm was  $228.8 \pm 63.1 \mu\text{m}$ , whereas the mean actual SDOCT BMO-MRW was  $226.0 \pm 73.8 \mu\text{m}$  ( $P = .415$ ; GEE). A strong correlation was seen between predictions and observed values, with  $r = 0.88$  ( $R^2 = 77\%$ ;  $P < .001$ ; GEE). [Figure 1](#) shows a scatterplot and corresponding histograms illustrating the relationship between deep learning predictions and actual SDOCT measurements. The mean absolute deviation was  $27.8 \mu\text{m}$ , or approximately 12%. Pearson correlation coefficients ranged

from 0.68 to 0.82 for the different sectors around the optic disc (all correlations with  $P < .001$ ; GEE).

Deep learning global BMO-MRW predictions were significantly associated with SAP MD ( $r = 0.46$ ;  $P < .001$ ; GEE). SDOCT global BMO-MRW also showed a similar association with SAP MD ( $r = 0.49$ ;  $P < .001$ ; GEE). [Table 3](#) shows the associations between sectoral deep learning BMO-MRW predictions and the visual field sensitivities of the corresponding sectors according to the structure-function map. The associations had similar strength to those observed for the actual SDOCT BMO-MRW measurements. ROC curve areas to discriminate glaucomatous from healthy eyes are shown in [Table 4](#) and they were also similar for deep learning predictions and actual SDOCT measurements. For global BMO-MRW, deep learning predictions had an ROC curve area of 0.945 (95% confidence interval [CI]: 0.874–0.980) versus 0.933 (95% CI: 0.856–0.975) for SDOCT measurements.

[Figure 2](#) illustrates examples of optic disc photographs and corresponding activation maps (heatmaps) of the deep learning network for a normal (Top) and a glaucoma (Bottom) eye. The heatmaps show that the activations were most strongly found in the area of the optic nerve, indicating

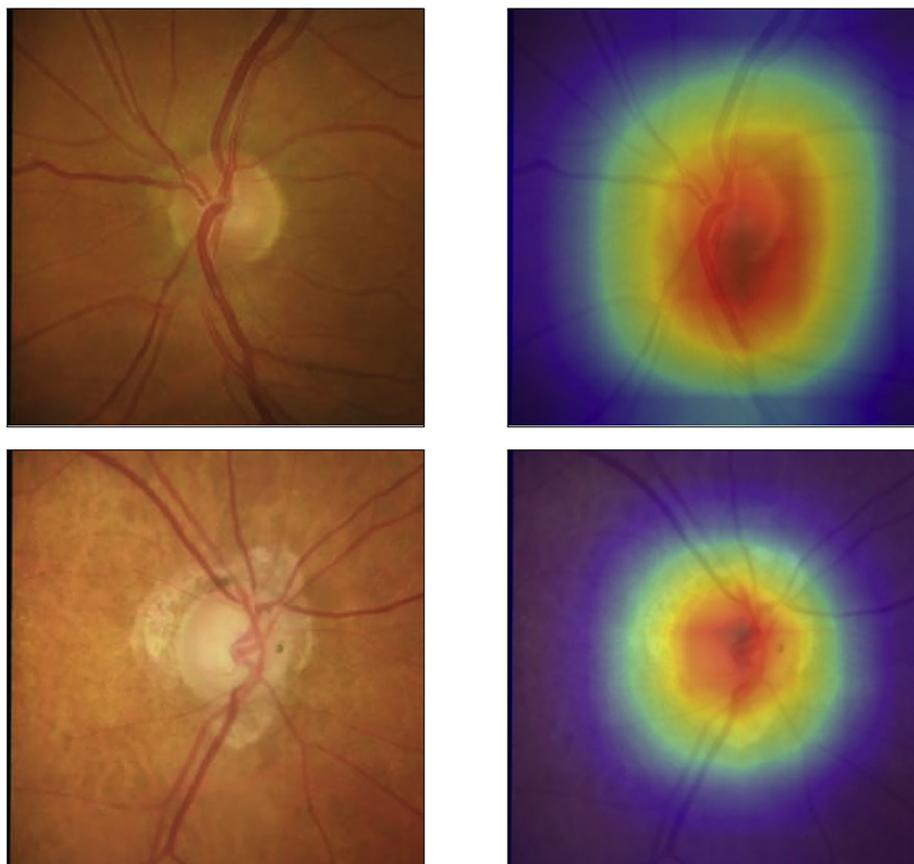


FIGURE 2. Activation heatmaps showing the areas of the optic disc photograph that were most important for the deep learning algorithm predictions in an example of a healthy (Top row) and glaucomatous (Bottom row) eye.

that these areas were the most important for the network predictions. Figure 3 shows some random examples of optic disc photographs from the test sample along with the corresponding global BMO-MRW deep learning predictions and SDOCT measurements. For illustrative purposes, the random examples were drawn from a sample where predictions closely agreed to observations (within 20  $\mu\text{m}$ ) and also from a sample where the predictions disagreed with the observations (greater than 50  $\mu\text{m}$  difference).

## DISCUSSION

IN THIS STUDY, WE DEVELOPED AND VALIDATED A NOVEL deep learning neural network capable of quantifying the amount of neuroretinal damage on an optic disc photograph. We accomplished this by training the algorithm to predict the SDOCT BMO-MRW when assessing photographs of glaucomatous optic nerves, glaucoma suspects, and normal nerves. We also demonstrated that these predicted structural values were significantly correlated with visual field loss on standard automated perimetry. To the best of our knowledge, this is the first investigation

to demonstrate that a deep learning network can accurately predict the BMO-MRW when assessing an optic disc photograph.

The task of grading disc photographs is notoriously laborious and time-intensive for human graders. Deep learning algorithms may offer an alternative mechanism for interpreting disc photographs that is both expedient and accurate. Such algorithms may eventually preclude the reliance on human graders. However, currently published machine learning algorithms that have been developed for the detection of glaucoma on optic disc photographs were trained using photographs that had been graded by human graders.<sup>17,18</sup> The reliance on subjective gradings for a reference standard presents several dilemmas. Human gradings on optic disc photographs have been shown to have poor reproducibility<sup>14,15</sup> and interrater reliability.<sup>8,15,16</sup> Moreover, misclassification can occur. Clinicians tend to underappreciate glaucomatous damage in eyes with small optic discs and myopic nerves but overdiagnose glaucoma in eyes with large physiologic cups.<sup>17</sup> Thus, if subjective human gradings are used as the reference standard to train a neural network, the resulting deep learning algorithm will yield similar diagnostic errors.

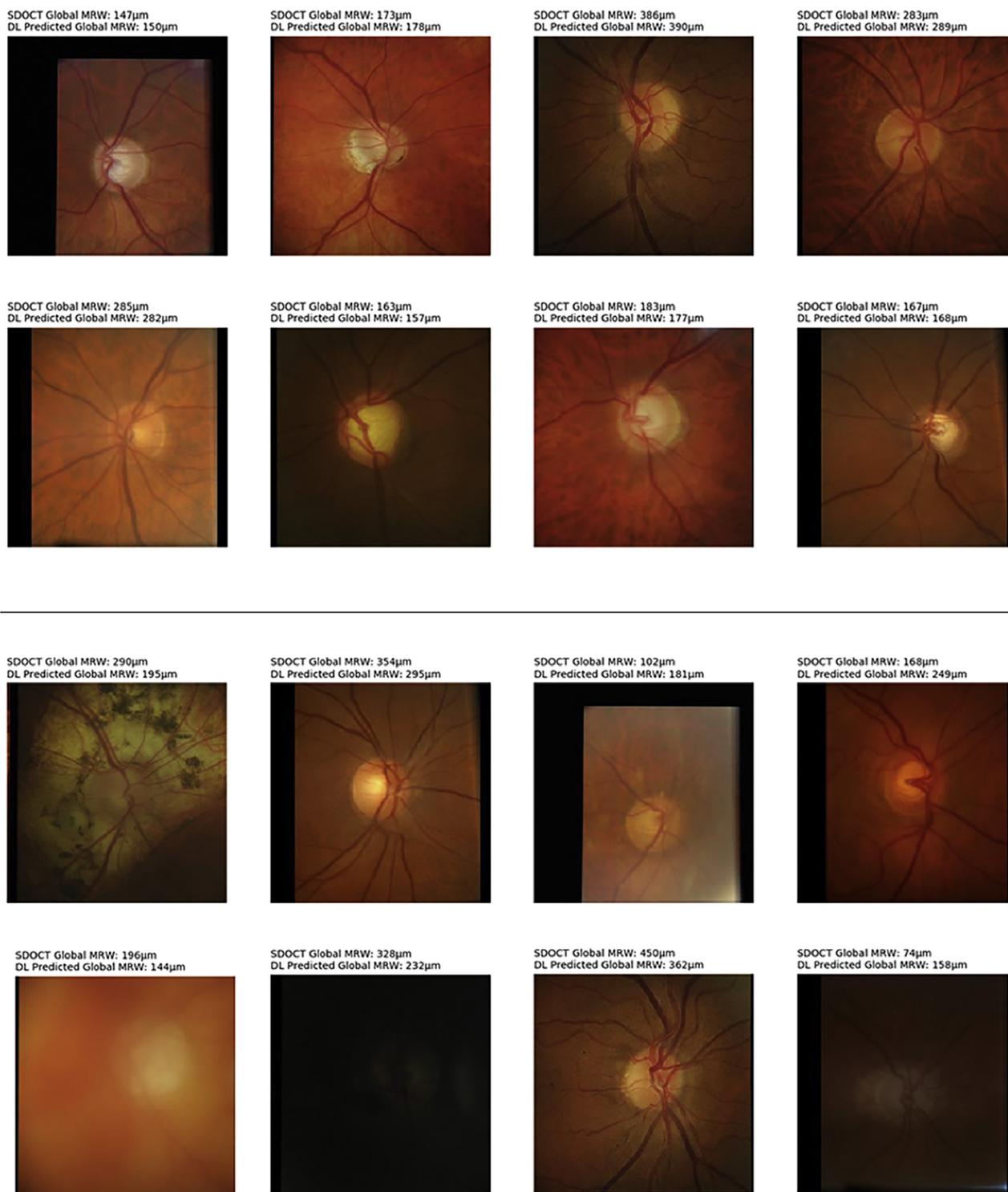


FIGURE 3. Random examples of optic disc photographs from the test sample along with the corresponding global minimum rim width relative to Bruch membrane opening (BMO-MRW) deep learning (DL) predictions and spectral-domain optical coherence tomography (SDOCT) measurements. (Top) Random examples where predictions closely agreed to observations (within 20  $\mu$ m). (Bottom) Random examples where the predictions disagreed with the observations (greater than 50  $\mu$ m difference).

Using an SDOCT parameter such as BMO-MRW as a reference standard for training the classifier offers several unique advantages. First, the data are objective rather than subjective, and bypass the need for human gradings. Second the parameter BMO-MRW is known to have high reproducibility,<sup>36</sup> unlike human gradings of photographs.<sup>14,15</sup> Third, the BMO-MRW parameter is highly accurate for glaucoma on SDOCT.<sup>25</sup> Previous groups have also suggested that certain sectors of the BMO-MRW, such as the superotemporal, superonasal, and nasal sectors, may perform better than regional RNFL thickness for the diagnosis of glaucoma.<sup>37</sup> We found that the superotemporal, superonasal, inferotemporal, and inferonasal sector values had high ROC curve areas for discriminating eyes with glaucomatous field loss from healthy eyes. Future studies should investigate whether neural network trained with SDOCT BMO-MRW perform better than those trained with other parameters such as RNFL.

BMO-MRW may also be advantageous in situations where the optic nerve is difficult to grade. For example, a previous study has shown that the use of BMO-MRW may improve the structure-function correlation in eyes with high myopia.<sup>25</sup> This is important since Li and associates' deep learning algorithm, which was trained with subjectively graded photographs, tended to underdiagnose glaucoma in high myopia, generating false-negatives.<sup>17</sup> Similarly, eyes with physiologically enlarged cups or pathologic myopia were frequently misclassified with glaucoma, leading to a false-positive result in their study.<sup>17</sup> Since our deep learning algorithm was trained using SDOCT BMO-MRW, we suspect that it may have better discriminatory ability in such challenging cases when compared to other algorithms trained by human gradings of photographs. However, additional work is needed to test this hypothesis. Finally, several groups have suggested that BMO-MRW may be better able to detect mild glaucoma<sup>22,29</sup> and glaucoma suspects<sup>24</sup> than other SDOCT parameters. Thus, neural networks trained using SDOCT BMO-MRW may be of particular value as a screening measure to detect preperimetric glaucoma before the onset of visual field loss.

Although SDOCT has become the *nonpareil* method for quantifying structural damage from glaucoma, the technology is currently too expensive and impractical to transport for it to be used for screening in a low-resource setting. On the other hand, fundus photographs can be acquired through a variety of affordable and portable mechanisms, from relatively inexpensive nonmydriatic fundus cameras<sup>38</sup> to cell phones.<sup>39</sup> However, numerous studies have found that human interpretations of monoscopic disc photographs have poor sensitivity as low as 50% for glaucoma.<sup>8-13</sup> Such reports have called into question the practical utility of relying on photographs for glaucoma screening. Although screening with SDOCT is not currently feasible, we have shown that a deep learning

algorithm trained to predict the BMO-MRW from fundus photographs shows very high correlation with the actual SDOCT values. Moreover, the correlation of the algorithm's predictions with visual field loss was similar to that found for the actual SDOCT measurements. This provides additional validation of the algorithm by demonstrating its ability to discriminate between glaucomatous field loss and normal fields. This is particularly important, since other recently published work that developed deep learning algorithms trained by subjectively graded optic disc photographs did not correlate their classifications with visual dysfunction or any other clinically relevant outcomes.<sup>11,12,17,18</sup>

Finally, since the algorithm we developed provides continuous predicted values similar to the BMO-MRW, it is possible that these values may be useful for detecting longitudinal change over time. This is a unique advantage over any classification algorithm that only provides a binary assessment of the optic disc as glaucomatous or not. Algorithms trained to produce quantitative data may make it possible to detect progression in patients followed over time with fundus photographs in low-resource settings where SDOCT is not available. Future longitudinal investigations will evaluate the plausibility of this approach.

Our study has several limitations. We did not remove low-quality fundus photographs, which likely introduced some of the unexplained variability in the algorithm. On the other hand, photographs obtained through screening programs and in a clinical setting are likely to exhibit a range of quality. Algorithms trained only on optimal photographs may not be generalizable to real-world settings. We expect that our algorithm will perform well at quantifying neuroretinal damage on optic disc photographs acquired in a similar clinical setting. However, performance in a screening setting may differ, especially in racial and ethnic groups not represented in this study. External validation in different screening populations will be the subject of future studies. Variability of BMO-MRW on SDOCT may also contribute to some of the unexplained variability of the algorithm. Future studies will compare algorithms trained by different SDOCT parameters to algorithms trained by subjectively graded fundus photographs. Combining two or more parameters may also further improve the discriminatory ability of the algorithm.<sup>40</sup>

In summary, we used the BMO-MRW from SDOCT to train a deep learning algorithm to predict quantitative data about the degree of neuroretinal damage when assessing an optic disc photograph. The predicted quantitative data were correlated with the BMO-MRW value from SDOCT and showed high diagnostic discrimination for glaucomatous visual field loss. This approach overcomes substantial limitations of algorithms trained by human graders. These findings may have important implications for the diagnostic value of fundus photographs for glaucoma screening, especially in low-resource settings.

FUNDING/SUPPORT: SUPPORTED IN PART BY NATIONAL INSTITUTES OF HEALTH/NATIONAL EYE INSTITUTE GRANT EY025056 (F.A.M.), EY021818 (F.A.M.), and the Heed Foundation, San Francisco, CA (A.C.T.). The funding organizations had no role in the design or conduct of this research. Financial Disclosures: Alessandro A. Jammal: Conselho Nacional de Desenvolvimento Científico e Tecnológico (Financial Support [F]). Felipe A. Medeiros: Alcon Laboratories (Consultant [C], F, Recipient [R]), Allergan (C, F), Bausch & Lomb (F), Carl Zeiss Meditec (C, F, R), Heidelberg Engineering (F), Merck (F), nGoggle Inc (F), Sensimed (C), Topcon (C), Reichert (C, R). The following author has no financial disclosures: Atalie C. Thompson.

## REFERENCES

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006; 90(3):262–267.
2. Budenz DL, Barton K, Whiteside-de Vos J, et al. Prevalence of glaucoma in an urban West African population: the Tema Eye Survey. *JAMA Ophthalmol* 2013;131(5):651–658.
3. Hennis A, Wu SY, Nemesure B, et al. Awareness of incident open-angle glaucoma in a population study: the Barbados Eye Studies. *Ophthalmology* 2007;114(10):1816–1821.
4. Sathyamangalam RV, Paul PG, George R, et al. Determinants of glaucoma awareness and knowledge in urban Chennai. *Indian J Ophthalmol* 2009;57(5):355–360.
5. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA* 2014;311(18):1901–1911.
6. Owsley C, McGwin G Jr, Lee DJ, et al. Diabetes eye screening in urban settings serving minority populations: detection of diabetic retinopathy and other ocular findings using telemedicine. *JAMA Ophthalmol* 2015;133(2):174–181.
7. Maa AY, Patel S, Chasan JE, et al. Retrospective evaluation of a teleretinal screening program in detecting multiple nondiabetic eye diseases. *Telemed J E Health* 2017;23(1):41–48.
8. Chan HH, Ong DN, Kong YX, et al. Glaucomatous optic neuropathy evaluation (GONE) project: the effect of monoscopic versus stereoscopic viewing conditions on optic nerve evaluation. *Am J Ophthalmol* 2014;157(5):936–944.
9. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc* 1976;74:532–572.
10. Kumar S, Giubilato A, Morgan W, et al. Glaucoma screening: analysis of conventional and telemedicine-friendly devices. *Clin Exp Ophthalmol* 2007;35(3):237–243.
11. Maa AY, Evans C, DeLaune WR, et al. A novel tele-eye protocol for ocular disease detection and access to eye care services. *Telemed J E Health* 2014;20(4):318–323.
12. Marcus DM, Brooks SE, Ulrich LD, et al. Telemedicine diagnosis of eye disorders by direct ophthalmoscopy. A pilot study. *Ophthalmology* 1998;105(10):1907–1914.
13. Yogesan K, Constable IJ, Barry CJ, et al. Evaluation of a portable fundus camera for use in the teleophthalmologic diagnosis of glaucoma. *J Glaucoma* 1999;8(5):297–301.
14. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology* 1992; 99(2):215–221.
15. Abrams LS, Scott IU, Spaeth GL, et al. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology* 1994;101(10): 1662–1667.
16. Jampel HD, Friedman D, Quigley H, et al. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol* 2009;147(1):39–44 e1.
17. 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(8):1199–1206.
18. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318(22):2211–2223.
19. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316(22):2402–2410.
20. Raju M, Pagidimarri V, Barreto R, et al. Development of a deep learning algorithm for automatic diagnosis of diabetic retinopathy. *Stud Health Technol Inform* 2017;245:559–563.
21. Takahashi H, Tampo H, Arai Y, et al. Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic retinopathy. *PLoS One* 2017;12(6):e0179790.
22. Chauhan BC, O’Leary N, AlMobarak FA, et al. Enhanced detection of open-angle glaucoma with an anatomically accurate optical coherence tomography-derived neuroretinal rim parameter. *Ophthalmology* 2013;120(3):535–543.
23. Reis AS, O’Leary N, Yang H, et al. Influence of clinically invisible, but optical coherence tomography detected, optic disc margin anatomy on neuroretinal rim evaluation. *Invest Ophthalmol Vis Sci* 2012;53(4):1852–1860.
24. Pollet-Villard F, Chiquet C, Romanet JP, et al. Structure-function relationships with spectral-domain optical coherence tomography retinal nerve fiber layer and optic nerve head measurements. *Invest Ophthalmol Vis Sci* 2014;55(5): 2953–2962.
25. Reznicek L, Burzer S, Laubichler A, et al. Structure-function relationship comparison between retinal nerve fibre layer and Bruch’s membrane opening-minimum rim width in glaucoma. *Int J Ophthalmol* 2017;10(10):1534–1538.
26. Gmeiner JM, Schrems WA, Mardin CY, et al. Comparison of Bruch’s membrane opening minimum rim width and peripapillary retinal nerve fiber layer thickness in early glaucoma assessment. *Invest Ophthalmol Vis Sci* 2016;57(9): Oct575–Oct584.
27. Danthurebandara VM, Vianna JR, Sharpe GP, et al. Diagnostic accuracy of glaucoma with sector-based and a new total profile-based analysis of neuroretinal rim and retinal nerve fiber layer thickness. *Invest Ophthalmol Vis Sci* 2016;57(1): 181–187.
28. Gardiner SK, Ren R, Yang H, et al. A method to estimate the amount of neuroretinal rim tissue in glaucoma: comparison

- with current methods for measuring rim area. *Am J Ophthalmol* 2014;157(3):540–549.e1-2.
29. Gardiner SK, Boey PY, Yang H, et al. Structural measurements for monitoring change in glaucoma: comparing retinal nerve fiber layer thickness with minimum rim width and area. *Invest Ophthalmol Vis Sci* 2015;56(11):6886–6891.
  30. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000;107(10):1809–1815.
  31. Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009 June;248–255.
  32. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *ArXiv e-prints* 2015. Available at: <https://ui.adsabs.harvard.edu/#abs/2015arXiv151203385H>. Accessed December 1, 2015.
  33. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *ArXiv e-prints* 2016. Available at: <https://ui.adsabs.harvard.edu/#abs/2016arXiv161002391S>. Accessed October 1, 2016.
  34. Selvaraju RR, Das A, Vedantam R, et al. Grad-CAM: Why did you say that? *ArXiv e-prints* 2016. Available at: <https://ui.adsabs.harvard.edu/#abs/2016arXiv161107450S>. Accessed November 1, 2016.
  35. Medeiros FA, Sample PA, Zangwill LM, et al. A statistical approach to the evaluation of covariate effects on the receiver operating characteristic curves of diagnostic tests in glaucoma. *Invest Ophthalmol Vis Sci* 2006;47(6):2520–2527.
  36. Reis ASC, Zangalli CES, Abe RY, et al. Intra- and interobserver reproducibility of Bruch's membrane opening minimum rim width measurements with spectral domain optical coherence tomography. *Acta Ophthalmol* 2017;95(7):e548–e555.
  37. Kim SH, Park KH, Lee JW. Diagnostic accuracies of Bruch membrane opening-minimum rim width and retinal nerve fiber layer thickness in glaucoma. *J Korean Ophthalmol Soc* 2017;58(7):836–845.
  38. Miller SE, Thapa S, Robin AL, et al. Glaucoma screening in Nepal: cup-to-disc estimate with standard mydriatic fundus camera compared to portable nonmydriatic camera. *Am J Ophthalmol* 2017;182:99–106.
  39. Fleming C, Whitlock EP, Beil T, et al. Screening for primary open-angle glaucoma in the primary care setting: an update for the US preventive services task force. *Ann Fam Med* 2005;3(2):167–170.
  40. Park K, Kim J, Lee J. The relationship between Bruch's membrane opening-minimum rim width and retinal nerve fiber layer thickness and a new index using a neural network. *Transl Vis Sci Technol* 2018;7(4):14.