© Modeling and Artificial Intelligence in Ophthalmology 2021; 1:1-10
Review article
https://doi.org/10.35119/maio.v4i1.125

A review of deep learning in structure and function in glaucoma

Eduardo B. Mariottoni[1], Felipe Medeiros[2,3], Vital P. Costa[4]

[1]Department of Ophthalmology, Federal University of São Paulo, São Paulo, Brazil; [2]Vision, Imaging, and Performance Laboratory, Duke Eye Center and Department of Ophthalmology, Duke University, Durham, NC, USA; [3]Department of Electrical and Computer Engineering, Pratt School of Engineering, Duke University, Durham, NC, USA; [4]Department of Ophthalmology, State University of Campinas, Campinas, Brazil

Abstract

The relationship between structural damage and functional loss in glaucoma is of great importance for its diagnosis and management. The functional status is usually assessed through visual field examination, a subjective test that is burdensome and time-consuming. Moreover, it depends on patients' answers and there is a learning curve until accurate and reliable measurements are possible. Structural assessment, on the other hand, has remarkably improved since the development of optical coherence tomography, a fast test that allows for objective and quantitative analysis of retinal layers. The relationship between the two tests, however, is complex and nonlinear, and is influenced by interindividual variability. Thus, qualitative evaluation or the use of conventional statistics might not be appropriate. In recent years, we have seen a remarkable evolution of artificial intelligence algorithms and deep learning models. These techniques have proved adequate to model such complicated relationships. In this review, we summarize studies that investigate the structure and function relationship in glaucoma making use of artificial intelligence and deep learning, the challenges associated with predicting visual field information from structural measurements, and the strategies adopted to improve their accuracy.

Keywords: artificial intelligence, deep learning, function, glaucoma, optical coherence tomography, structure, visual field

Correspondence: Eduardo B. Mariottoni, MD, Department of Ophthalmology, Federal University of São Paulo, São Paulo, Brazil.
E-mail: eduardomariottoni@gmail.com

1. Introduction

Glaucoma is a progressive optic neuropathy in which the death of retinal ganglion cells, and corresponding axons in the retinal nerve fiber layer (RNFL), leads to characteristic visual field (VF) defects.[1] The relationship between structural damage and glaucomatous VF loss is essential to diagnose glaucoma and to differentiate it from other diseases that may affect the retina and the visual pathways. The evaluation of the VF in clinical practice is performed through standard automated perimetry (SAP), which is a burdensome and time-consuming test with high test-retest variability that is inherent to testing strategies.[2,3] Yet, there is still no better way to investigate the functional loss in glaucoma.

With the advent of optical coherence tomography (OCT), it became possible to perform a quantitative analysis of the thickness of retina layers, such as the ganglion cell layer and the RNFL. OCT identifies glaucomatous damage with more accuracy than the qualitative assessment of fundus photos and even before the development of detectable VF defects.[4] It is a faster test than SAP, it does not rely on patient collaboration, and has lower test-retest variability. The relationship between the two tests, however, is complex and nonlinear, and they are generally used in a complementary fashion.

Recent advances on artificial intelligence (AI), with more complex algorithms and techniques such as deep learning (DL), have allowed predictions and classifications from images and other types of data with human-level accuracy, in some cases even more accurate. In ophthalmology, it has been applied to diagnose retinal diseases and other conditions. In glaucoma, one of the proposed uses of AI and DL is to investigate the structure-function relationship, in particular to estimate VF data from OCT measurements. The purpose of this review is to summarize studies that describe algorithms and DL models that predict SAP summary metrics or individual sensitivity threshold values from OCT measurements of retinal layers.

2. Prediction of visual field summary metrics

SAP summary metrics are useful to gauge the severity of the disease, as well as to monitor the progression through trend analysis. A few studies have shown that it is possible to estimate the values of SAP summary metrics, such as mean deviation (MD), pattern standard deviation (PSD) and VF index (VFI), using DL models based on retinal thickness data, measured with OCT.

In a work by Christopher *et al.,*[5] a DL algorithm was developed to predict MD, PSD, and pattern deviation (PD) averages for sectors derived from the Garway-Heath structure and function map.[6] They investigated the performance of DL models using three different inputs: RNFL thickness maps, RNFL *enface*, and

confocal scanning laser ophthalmoscope (CSLO) images. To predict SAP MD, the best performance was achieved using RNFL *enface* images as input. The mean absolute error (MAE) was 2.5 dB and the R^2 with measured values was 70%, while for RNFL thickness map the values were 2.8 dB and 63%, and for CSLO images the values were 3.1 dB and 48%. All DL models were superior to the predictions using macular RNFL thickness (MAE = 3.8 dB; R^2 = 40%) and circumpapillary RNFL thickness (MAE = 3.7 dB; R^2 = 45%). The RNFL *enface* images model was also the best one to predict PSD and the sectoral PD averages, except for the central region, in which the CSLO model achieved the best performance.

Yu *et al.* developed a three-dimensional convolutional neural network (3D-CNN) model to predict SAP VFI and MD from OCT volumes centered at the macula, the ONH or both.[7] The median estimate's errors for SAP VFI were 3.11% and 3.53% for the models that used the volumes centered at the macula and the optic nerve head (ONH), respectively. The model that used both volumes as input had a median error of 2.70%. To estimate the SAP MD, the model that used both volumes also achieved the best performance, with a median error of 1.57 dB, while the model with macula or ONH volume as input had median errors of 1.63 dB and 1.86 dB, respectively. Interestingly, although combining the macula and ONH centered volumes improved the overall performance of the DL model, the improvement was particularly better for more advanced disease, where the estimations had higher errors, both for VFI and MD. In a work from the same group of authors, George *et al.* showed a similar performance to estimate SAP VFI from ONH centered volumes, although the main goal of their work was to detect glaucomatous VF defect.[8]

While the studies presented above showed complex DL models, using different types of inputs, the work done by Huang *et al.* showed a model that predicted SAP MD from the RNFL thickness averaged into 64 sectors.[9] Although they use a simpler model, its performance in their internal test set was comparable to those of the previous studies, with a MAE of 4.0, a root mean squared error (RMSE) of 5.2, and a median absolute error of 3.1 dB. They also showed the results of external validation in three different datasets, with similar results in two of them. The third external dataset had data extracted from a different OCT machine than the one used to train the model. Not surprisingly, the predictions had larger errors and lower correlation with the actual SAP MD. This finding underscores the lack of generalizability regarding the source of information*, i.e.*, models trained with data from one OCT cannot be used to analyze data from a different OCT.

In summary, the DL models currently available in the literature show that it is possible to accurately predict SAP summary metrics from structural data derived from OCT measurements. Different strategies have been employed to improve the predictions. Although they present good overall performance, there is a trend for higher errors in more advanced disease, which can be a result of higher variability or lower availability of data, both expected to impact a model's performance.

Nonetheless, they have the potential to provide functional information for patients that are incapable of VF testing, or even increase the information available without the burden of additional testing.

3. Prediction of 24-2 visual field sensitivity threshold values

Although summary metrics offer useful information about the VF, they have limited value to define the location and the pattern of the glaucomatous defect, which can be assessed by the determining threshold values in all points tested by the 24-2 SAP. Predicting the whole 24-2 SAP, however, is a harder task than predicting summary metrics, because they are much more influenced by test-retest variability and individual variations. Nevertheless, many previous studies have attempted to predict the 24-2 SAP sensitivity threshold values from structural information.

The first attempts to use AI algorithms for this task were reported by Zhu *et al.*[10,11] In their works, they present the development of a machine learning algorithm, namely a Bayesian radial basis function, to predict the 24-2 SAP using RNFL thickness information from scanning laser polarimetry. Their algorithm was able to accurately predict the sensitivity threshold values with a MAE of 2.9 dB, a far superior performance than that of a linear regression model (MAE = 4.9 dB), developed for comparison purposes. There was a trend towards larger prediction errors in lower sensitivity values, although the predictions were within test-retest limits,[2] and clearly more accurate than the ones from the linear model. The superiority of the machine learning algorithm could also be noted in the gray scale images presented in the study.

With the emergence and increasing access to OCT technology, new studies attempted to estimate the VF from structural glaucomatous damage. Guo *et al.* developed a machine learning algorithm to predict the 24-2 SAP sensitivity threshold values from RNFL and ganglion cell layer plus inner plexiform layer (GCL + IPL) thickness measured by OCT.[12] In their study, a wide field composite OCT was acquired and divided into 54 sectors to match the locations tested by 24-2 SAP. They used different strategies to select which sectors of each layer would result in better predictions. The best performance was achieved by the model that included the GCL + IPL thickness of the VF location to be predicted and the RNFL thickness of sectors along the axonal route to the optic disc, a strategy they called “retinal ganglion cell axonal complex (RGC-AC) optimized”. They reported a correlation of 0.74 between the estimated and the real values, and a RMSE of 5.42 dB. In a qualitative assessment, the predicted VFs presented patterns and degrees of defects that were similar, although not identical, to the actual VFs.

Subsequent studies have relied on DL models for such task. Mariottoni *et al* developed a one-dimensional CNN to predict the 24-2 SAP results based on circumpapillary RNFL thickness, measured by spectral domain (SD)-OCT.[13] The predictions

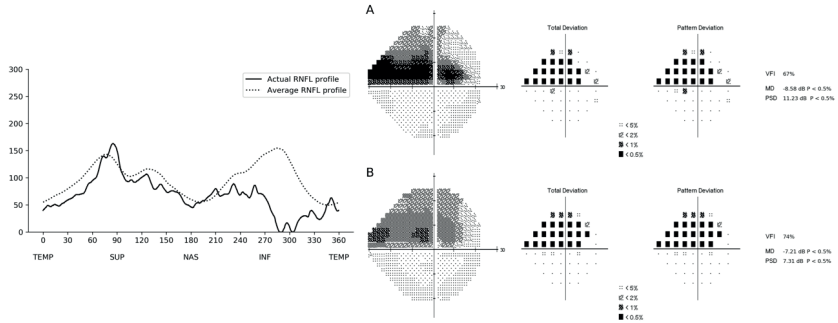

Fig. 1. Example of a case where the convolutional neural network (CNN) was able to predict the VF accurately using the retinal nerve fiber layer (RNFL) measurements. In the case illustrated, there is a large inferior temporal defect on the RNFL (left), that manifested on the VF as a superior arcuate defect (A, right). The CNN predicted a VF with a defect of similar shape and depth (B, right).

of their model had a correlation of 0.60 with measured values and a MAE of 4.25 dB, which was better than a linear model developed for comparison purposes (correlation of 0.52 and MAE of 4.96 dB). Figure 1 shows an example of a predicted VF from RNFL thickness data compared to the actual VF. Their main goal, however, was to develop a structure-function map based on the information captured by their model. To achieve that, artificial defects were simulated in a normal RNFL thickness profile on different locations and varying depths. The CNN was used to predict the VF based on the simulated RNFL profile and, as a result, the predictions showed the VF defect expected for each artificial RNFL thinning (Fig. 2). With a similar approach, Datta *et al.* developed a recurrent neural network (RNN) that analyzed the circumpapillary RNFL thickness, but used anatomic knowledge to improve predictions.[14] Their results were similar to those described by Mariottoni and colleagues.

While Mariottoni and Datta relied on circumpapillary RNFL thickness, other authors used OCT images as inputs to their models. Park *et al.* combined the macular GCL + IPL and peripapillary RNFL thickness maps as one image and developed a DL model to predict the 52 sensitivity values from 24-2 SAP.[15] The mean RMSE of their predictions was 4.79 dB and they were higher for eyes with glaucoma than healthy eyes (5.27 *vs* 3.27 dB). The prediction errors were also correlated with severity markers, namely SAP MD, macular GCL + IPL and peripapillary RNFL thickness, suggesting that predicting lower sensitivity values is a harder task. Subsequent studies from the same group compared different architectures for the DL model and different OCT technologies (SD and swept-source [SS] OCT).[16,17] Among the architectures tested, Inception-ResNet-v2 was superior to Inception-v3 and Inception-v4. Although the prediction errors were significantly lower, they were probably not clinically relevant. Furthermore, their findings may be specific

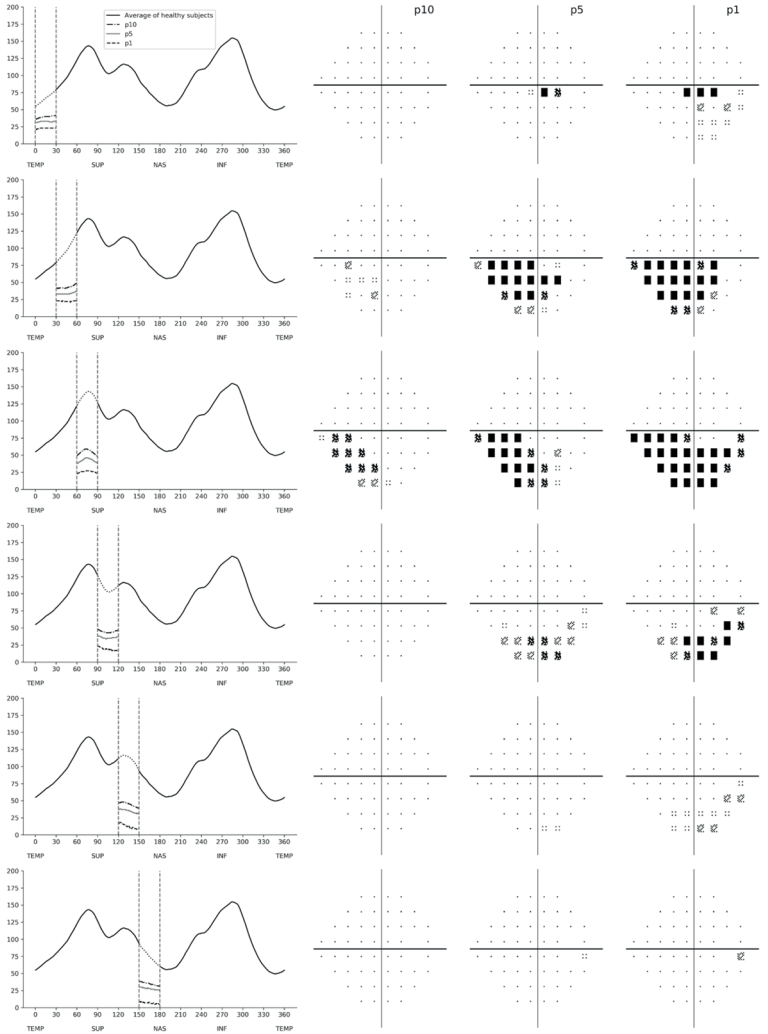

Fig. 2. Patterns of visual field loss predicted from the convolutional neural network when simulating retinal nerve fiber layer (RNFL) defects in the superior hemiretina. The RNFL profile is shown on the left, with dashed vertical lines showing the location of each simulated RNFL defect. For each simulated defect in a particular location, there were three simulated depths representing the 10th (p10), 5th (p5), and 1st (p1) percentiles. The corresponding predicted standard automated perimetry pattern deviation plots are shown on the right.

to the dataset available for development and evaluation of the DL models. On the other hand, the DL model developed to analyze SS-OCT images showed a better performance than the SD-OCT model, which was justified by the larger area scanned by SS-OCT in comparison to SD-OCT. There was also an imbalance of training data available for each device, which could also have influenced their performance.

The studies presented above confirm that predicting the sensitivity threshold values for each location tested by the 24-2 SAP is a harder task than predicting summary metrics such as MD and VFI. The higher test-retest and interindividual variabilities negatively influence the performance of DL models, especially in lower ranges of measured values. However, the performance shown by these algorithms indicate that it is possible to acquire information about the functional status based on structural information, which could be useful when actual VF testing is not possible.

4. Prediction of 10-2 visual field sensitivity threshold values

Although the mainstay of functional testing in glaucoma is the 24-2 SAP, testing the central 10 degrees of the VF is beneficial in cases of advanced glaucoma, where most of the peripheral vision is lost, or in early cases, in which the 6-degree spacing of the 24-2 SAP may miss a small paracentral glaucomatous defect. However, to test both 24-2 and 10-2 SAP is costly and burdensome. An alternative would be to predict the 10-2 SAP from OCT measurements assisted by DL models. The studies presented bellow describe the development of such DL models, all done by the same group of authors. They focused on the macular scan, rather than peripapillary RNFL, given the overlap between the regions tested and the good correlation between retinal layer thicknesses and the sensitivity threshold values of the central 10 degrees of the VF.[18,19]

In their first attempt to predict 10-2 SAP from OCT, Sugiura and colleagues developed a DL model to predict the 68 sensitivity values from 10-2 SAP from the thickness of three retinal layers in the macula: the RNFL, the GCL + IPL, and the rod and cones layer.[20] In order to improve the predictions of the DL model, they used unpaired VF data and unsupervised learning to create patterns of glaucomatous VF defects. Those patterns were used as regularization of the predictions. The result of this pattern-based regularization was an improvement of the RMSE from 6.76 to 6.16 dB. This model was later validated with an external dataset in a work presented by Hashimoto and colleagues.[21] The prediction's errors were slightly higher in the external validation (MAE = 5.47 dB) in comparison to the internal validation (MAE = 4.71 dB), but the DL model still outperformed other machine learning models presented in the study.

Subsequently, the same group proposed methods to improve the accuracy of the DL models. In a work by Xu *et al.*, they applied a tensor regression on top of a

CNN and compared it to a regular CNN.[22] The proposed method was able to improve predictions, reducing the RMSE from 6.76 to 6.32 dB. In a subsequent work, Asano and colleagues used the sensitivity threshold values of the four most central points tested by 24-2 SAP, all within the central 10 degrees, to correct the 10-2 predictions.[23] The increase in accuracy translated as an improvement of the MAE from 9.4 to 5.3 dB. In a similar study, Hashimoto *et al.* combined the pattern-based regularization and the 24-2 SAP correction.[24] They found a MAE of 5.3 dB with pattern-based regularization alone *versus* 4.2 dB when combined with 24-2 SAP correction.

A common feature of all studies presented above was the decline in performance when predicting lower sensitivity values, which was also true in DL models to predict summary metrics and 24-2 SAP sensitivity threshold values. This is likely due to higher test-retest variability and smaller frequency of examples in the training data. In addition, there was a small number of eyes to evaluate the DL models, probably due to the lesser frequency in which the 10-2 SAP is tested, in comparison to 24-2 SAP. Consequently, their results must be generalized with caution, as it may not translate to other populations.

5. Conclusion

In this review we have summarized studies that attempted to estimate the VF loss from structural assessment, mostly by OCT. It was demonstrated that it is possible to predict summary metrics as well as individual sensitivity threshold values for both 24-2 and 10-2 SAP. The use of such predictions can be valuable for glaucoma management, either when VF is not possible or to increase the number of data points available to assess progression.

It is important to note that the evidence currently available does not warrant the substitution of SAP by DL predictions of the VF, but rather to complement it in the function assessment. This is due to the magnitude of prediction errors, even considering the advances achieved by the proposed methods. In particular, for advanced glaucoma, predictions tend to be higher than the measured values, underestimating the depth of the VF defect. It should also be noted that all AI algorithms could be influenced by the demographic characteristics of the population included in its development. For that reason, the performance of the AI algorithms should preferentially be demonstrated in external datasets, from a different geographical location, if possible, to demonstrate its generalizability.

Future work should focus on improving the performance of the AI algorithms, in particular for more advanced disease, and on how to employ the predicted VF in clinical practice.

References

1. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. JAMA. 2014;311(18):1901-1911. https://doi.org/10.1001/jama.2014.3192
2. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. Invest Ophthalmol Vis Sci. 2002;43(8):2654-2659.
3. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the Reliability of Standard Automated Perimetry in Regions of Glaucomatous Damage. Ophthalmology. 2014;121(7):1359-1369. https://doi.org/https://doi.org/10.1016/j.ophtha.2014.01.020
4. Kuang TM, Zhang C, Zangwill LM, Weinreb RN, Medeiros FA. Estimating Lead Time Gained by Optical Coherence Tomography in Detecting Glaucoma before Development of Visual Field Defects. Ophthalmology. 2015;122(10):2002-2009. https://doi.org/10.1016/j.ophtha.2015.06.015
5. Christopher M, Bowd C, Belghith A, et al. Deep Learning Approaches Predict Glaucomatous Visual Field Damage from OCT Optic Nerve Head En Face Images and Retinal Nerve Fiber Layer Thickness Maps. Ophthalmology. 2020;127(3):346-356. https://doi.org/10.1016/j.ophtha.2019.09.036
6. 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.
7. Yu H-H, Maetschke SR, Antony BJ, et al. Estimating Global Visual Field Indices in Glaucoma by Combining Macula and Optic Disc OCT Scans Using 3-Dimensional Convolutional Neural Networks. Ophthalmology Glaucoma. 2021;4(1):102-112. https://doi.org/https://doi.org/10.1016/j.ogla.2020.07.002
8. George Y, Antony BJ, Ishikawa H, Wollstein G, Schuman JS, Garnavi R. Attention-Guided 3D-CNN Framework for Glaucoma Detection and Structural-Functional Association Using Volumetric Images. IEEE Journal of Biomedical and Health Informatics. 2020;24(12):3421-3430. https://doi.org/10.1109/JBHI.2020.3001019
9. Huang X, Sun J, Majoor J, et al. Estimating the Severity of Visual Field Damage From Retinal Nerve Fiber Layer Thickness Measurements With Artificial Intelligence. Transl Vis Sci Technol. 2021;10(9):16. https://doi.org/10.1167/tvst.10.9.16
10. Zhu H, Crabb DP, Garway-Heath DF, editors. A Bayesian Radial Basis Function Model to Link Retinal Structure and Visual Function in Glaucoma. 3rd International Conference on Bioinformatics and Biomedical Engineering 2009 11-13 June 2009; Beijing, China.
11. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting Visual Function from the Measurements of Retinal Nerve Fiber Layer Structure. Investigative Ophthalmology & Visual Science. 2010;51(11):5657-5666. https://doi.org/10.1167/iovs.10-5239
12. Guo Z, Kwon YH, Lee K, et al. Optical Coherence Tomography Analysis Based Prediction of Humphrey 24-2 Visual Field Thresholds in Patients With Glaucoma. Invest Ophthalmol Vis Sci. 2017;58(10):3975-3985. https://doi.org/10.1167/iovs.17-21832
13. Mariottoni EB, Datta S, Dov D, et al. Artificial Intelligence Mapping of Structure to Function in Glaucoma. Transl Vis Sci Technol. 2020;9(2):19. https://doi.org/10.1167/tvst.9.2.19
14. Datta S, Mariottoni EB, Dov D, Jammal AA, Carin L, Medeiros FA. RetiNerveNet: using recursive deep learning to estimate pointwise 24-2 visual field data based on retinal structure. Scientific Reports. 2021;11(1):12562. https://doi.org/10.1038/s41598-021-91493-9
15. Park K, Kim J, Lee J. A deep learning approach to predict visual field using optical coherence tomography. PLoS One. 2020;15(7):e0234902. https://doi.org/10.1371/journal.pone.0234902

16. Park K, Kim J, Kim S, Shin J. Prediction of visual field from swept-source optical coherence tomography using deep learning algorithms. Graefes Arch Clin Exp Ophthalmol. 2020;258(11):2489-2499. https://doi.org/10.1007/s00417-020-04909-z
17. Shin J, Kim S, Kim J, Park K. Visual Field Inference From Optical Coherence Tomography Using Deep Learning Algorithms: A Comparison Between Devices. Transl Vis Sci Technol. 2021;10(7):4. https://doi.org/10.1167/tvst.10.7.4
18. Raza AS, Cho J, de Moraes CG, et al. Retinal ganglion cell layer thickness and local visual field sensitivity in glaucoma. Arch Ophthalmol. 2011;129(12):1529-1536. https://doi.org/10.1001/archophthalmol.2011.352
19. Lee JW, Morales E, Sharifipour F, et al. The relationship between central visual field sensitivity and macular ganglion cell/inner plexiform layer thickness in glaucoma. Br J Ophthalmol. 2017;101(8):1052-1058. https://doi.org/10.1136/bjophthalmol-2016-309208
20. Sugiura H, Kiwaki T, Yousefi S, Murata H, Asaoka R, Yamanishi K. Estimating Glaucomatous Visual Sensitivity from Retinal Thickness with Pattern-Based Regularization and Visualization. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; London, United Kingdom: Association for Computing Machinery; 2018. p. 783–792.
21. Hashimoto Y, Asaoka R, Kiwaki T, et al. Deep learning model to predict visual field in central 10° from optical coherence tomography measurement in glaucoma. Br J Ophthalmol. 2021;105(4):507-513. https://doi.org/10.1136/bjophthalmol-2019-315600
22. Xu L, Asaoka R, Kiwaki T, et al. Predicting the Glaucomatous Central 10-Degree Visual Field From Optical Coherence Tomography Using Deep Learning and Tensor Regression. Am J Ophthalmol. 2020;218:304-313. https://doi.org/10.1016/j.ajo.2020.04.037
23. Asano S, Asaoka R, Murata H, et al. Predicting the central 10 degrees visual field in glaucoma by applying a deep learning algorithm to optical coherence tomography images. Sci Rep. 2021;11(1):2214. https://doi.org/10.1038/s41598-020-79494-6
24. Hashimoto Y, Kiwaki T, Sugiura H, et al. Predicting 10-2 Visual Field From Optical Coherence Tomography in Glaucoma Using Deep Learning Corrected With 24-2/30-2 Visual Field. Translational Vision Science & Technology. 2021;10(13):28-28. https://doi.org/10.1167/tvst.10.13.28