tvst

Article

Visual Field Inference From Optical Coherence Tomography Using Deep Learning Algorithms: A Comparison Between Devices

Jonghoon Shin[1,2], Sungjoon Kim[3], Jinmi Kim[4], and Keunheung Park[5]

[1] Department of Ophthalmology, College of Medicine, Pusan National University Yangsan Hospital, Yangsan, South Korea
[2] Department of Ophthalmology, Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, Yangsan, South Korea
[3] Department of Ophthalmology, Pusan National University College of Medicine, Busan, Korea
[4] Department of Biostatistics, Clinical Trial Center, Biomedical Research Institute, Pusan National University Hospital, Busan, Korea
[5] Department of Ophthalmology, Busan Medical Center, Busan, Korea

Correspondence: Keunheung Park, Department of Ophthalmology, Busan Medical Center, 359, World cup-daero, Yeonje-gu, Busan 47527, South Korea.
e-mail: climyth@naver.com.

Received: June 29, 2020
Accepted: April 28, 2021
Published: June 4, 2021

Keywords: deep learning; optical coherence tomography; visual field

Citation: 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

Purpose: To develop a deep learning model to estimate the visual field (VF) from spectral-domain optical coherence tomography (SD-OCT) and swept-source OCT (SS-OCT) and to compare the performance between them.

Methods: Two deep learning models based on Inception-ResNet-v2 were trained to estimate 24-2 VF from SS-OCT and SD-OCT images. The estimation performance of the two models was evaluated by using the root mean square error between the actual and estimated VF. The performance was also compared among different glaucoma severities, Garway-Heath sectorizations, and central/peripheral regions.

Results: The training dataset comprised images of 4391 eyes from 2350 subjects, and the test dataset was obtained from another 243 subjects (243 eyes). In all subjects, the global estimation errors were 5.29 $\pm$ 2.68 dB (SD-OCT) and 4.51 $\pm$ 2.54 dB (SS-OCT), and the estimation error of SS-OCT was significantly lower than that of SD-OCT ($P < 0.001$). In the analysis of sectors, SS-OCT showed better performance in all sectors except for the inferonasal sector in normal vision and early glaucoma. In advanced glaucoma, the estimation error of the central region was worsened in both OCTs, but SS-OCT was still significantly better in the peripheral region.

Conclusions: Our deep learning model estimated the VF 24-2 better with a wide field image of SS-OCT than did with retinal nerve fiber layer and ganglion cell–inner plexiform layer images of SD-OCT.

Translational Relevance: This deep learning method can help clinicians to determine the VF from OCT images. OCT manufacturers can equip this system to provide additional VF data.

Introduction

To detect glaucomatous damage in the optic nerve head (ONH), red-free fundus photography and optical coherence tomography (OCT) are usually performed.[1,2] The visual field (VF) test is also an essential examination to identify the corresponding functional damage.[3] However, VF examination includes many random errors and fluctuations[4–6] and could lead to an inaccurate diagnosis. In contrast with the VF, OCT is easy to perform, and the result is very objective and highly reproducible.[7–9] If we can estimate the VF from SS-OCT, the benefit will be substantial.

The diagnostic performance to differentiate normal vision from glaucoma has been proven in many previous studies.[10–12] Spectral-domain OCT (SD-OCT) is an improved version of conventional time-domain OCT, and it offers higher scanning rates and better resolution. The intravisit and intervisit reproducibility of SD-OCT is superior to that of time-domain OCT,[13] indicating the potential usefulness of SD-OCT

Copyright 2021 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

in monitoring glaucoma progression. The diagnostic ability of SD-OCT is also better than that of time-domain OCT.[14] More recently, swept-source OCT (SS-OCT), a new generation of OCT, has been developed. This technique allows clinicians to obtain a high-quality, wide-angle image that includes coverage of the whole optic disc and macula and has a rapid scan speed. Previous studies have revealed that the diagnostic ability of peripapillary retinal nerve fiber layer (RNFL) and macular ganglion cell–inner plexiform layer thickness using SS-OCT is comparable with that of SD-OCT and have suggested that a single wide scan in SS-OCT could replace separate peripapillary and macular scans in patients with glaucoma.[15]

With the recent tremendous advances in deep learning algorithms, there have been several trials for estimating the VF based on the structural damage displayed in OCT.[16,17] Because structural changes measured by OCT are closely related to the functional changes in VF examination,[18–20] OCT images are considered to provide useful data to estimate VF damage. Although these studies have found that deep learning systems using OCT have a high accuracy for estimating VF defects in glaucoma, there is no report that directly compares the accuracy of various OCTs used with estimate VF defects. In terms of the deep learning approach, it is necessary to compare which data have shown greater estimation accuracy, because the process of selecting data is an important factor to increase accuracy. Therefore, we determined whether the thickness map in SD-OCT or SS-OCT influences the accuracy of estimating functional damage in glaucoma.

The purpose of this study was to develop a deep learning architecture to estimate the VF from different OCT images and to compare the performance between SD-OCT and SS-OCT images. We built a model with a state-of-the-art deep learning architecture and compared its performance globally and regionally in different glaucoma severities.

Methods

This retrospective study was performed in accordance with the tenets of the Declaration of Helsinki. The study was approved by the Institutional Review Board of Yangsan Pusan National University Hospital, South Korea. The patients' consent was waived by the institutional review board owing to the retrospective nature of the investigation.

All training and test data were collected from subjects who visited the glaucoma clinic at Yangsan Pusan National University Hospital from 2015 to 2019. In Table 1, the demographic characteristics of the training dataset are summarized. The total number of images in the training dataset was 4391 eyes from 2350 subjects. Of the 4391 training images, 2171 images were from SD-OCT and 2220 were SS-OCT. The mean age was 61.4 ± 17.0 years (SD-OCT) and 61.4 ± 13.7 years (SS-OCT). Unlike the test dataset, which contains only patients with glaucoma, the training dataset was not labeled by diagnosis. Therefore, it contains not only normal subjects but also patients with glaucoma and other optic neuropathies. However, eyes with severe retinal disease or media opacity (corneal opacity, cataract, etc.) were excluded.

Apart from the training dataset, a completely different test dataset was obtained from another 243 subjects (243 eyes) (Table 2). To evaluate performance more objectively, we obtained only one eye dataset from each patient. All subjects in the test group had both SD-OCT and SS-OCT performed within 6 months from the date of the VF. We retrospectively reviewed all test patients' medical records to obtain detailed results of the ophthalmic examination, which included age, gender, Goldmann applanation tonometry, spherical equivalence of refractive error (ARK-510A; NIDEK, Hiroshi, Japan), central corneal thickness (Pachmate; DGH Technology, Exton, PA), and axial length (IOLMaster, Carl Zeiss Meditec, Dublin, CA). Glaucomatous optic neuropathy was defined if one or more of the following criteria were met: vertical cup-to-disc ratio asymmetry of 0.2 or more, focal or diffuse neuroretinal rim thinning, localized notching, and the presence of RNFL defects that corresponded with VF defects.[21] Normal subjects were defined as those with no history of ocular disease, an intraocular pressure of less than 21 mm Hg, an absence of a glaucomatous optic disc appearance, and a normal VF. To ensure representation of the full range of disease, normal subjects also included those who were clinically suspected of having glaucoma (based on optic disc or RNFL appearance or elevated intraocular pressure) but had a normal VF. Patients with a corneal or ocular media opacity, refractive error of ±6.0 diopters or more, optic neuropathies other than glaucoma, or recent ocular surgery or trauma were excluded.

Zeiss SD-OCT

The Cirrus SD-OCT instrument (Carl Zeiss Meditec) was used to acquire macular ganglion cell–inner plexiform layer and peripapillary RNFL (pRNFL) thickness maps. Two consecutive OCT exams, namely, the 6 mm × 6 mm macular cube scan 200 × 200 protocol and the 6 mm × 6 mm optic disc cube 200 × 200 scan, were performed at the same time to obtain both macular ganglion cell–inner plexiform

Table 1. Demographic Characteristics of the Training Dataset

	Zeiss SD-OCT	Topcon SS-OCT
Total number of eyes	2171	2220
Total number of patients	1230	1120
Age (mean ± standard deviation)	61.4 ± 17.0	61.4 ± 13.7
Number of eyes binned by VF MD		
MD ≥ −6 dB	1344 (61.9%)	1196 (53.9%)
−6 dB > MD ≥ −12 dB	380 (17.5%)	399 (18.0%)
−12 dB > MD	447 (20.6%)	625 (28.2%)

Table 2. Demographic Characteristics of the Test Dataset

		Glaucoma ($n = 157$)		
	Normal ($n = 86$)	Early ($n = 87$)	Advanced ($n = 70$)	*P* Value
Age (years)	62.0 ± 13.3	60.3 ± 14.7	66.6 ± 11.2	0.011[a]
Gender (male/female)	36/50	37/50	30/40	0.017[b]
Intraocular pressure (mm Hg)	15.97 ± 3.92	15.46 ± 4.46	15.00 ± 4.38	0.369[a]
Refractive error (diopters)	−0.91 ± 2.57	−1.37 ± 2.89	−1.43 ± 2.46	0.381[a]
Central corneal thickness (µm)	550.3 ± 68.6	546.5 ± 44.7	542.3 ± 36.6	0.654[a]
Axial length (cm)	23.91 ± 1.47	24.19 ± 1.69	24.13 ± 1.29	0.496[a]
VF examination				
MD (dB)	−1.23 ± 1.24	−3.16 ± 1.79	−13.35 ± 7.40	<0.001[c]
PSD (dB)	1.71 ± 0.54	3.03 ± 1.69	8.93 ± 3.27	<0.001[a]
VF index (%)	98.7 ± 1.2	95.2 ± 3.7	63.2 ± 25.0	<0.001[a]
Zeiss SD-OCT macular ganglion cell analysis				
Signal strength	7.63 ± 1.11	7.36 ± 8.37	7.20 ± 1.12	0.048[a]
Average GCIPL thickness (µm)	80.2 ± 5.3	73.1 ± 8.4	63.6 ± 8.1	<0.001[c]
Zeiss SD-OCT ONH and RNFL analysis				
Signal strength	7.72 ± 0.95	7.45 ± 1.03	7.21 ± 1.06	0.008[a]
Superior RNFL thickness (µm)	109.5 ± 14.9	94.9 ± 15.9	78.6 ± 19.5	<0.001[c]
Temporal RNFL thickness (µm)	69.4 ± 12.5	65.1 ± 14.1	55.5 ± 13.7	<0.001[a]
Nasal RNFL thickness (µm)	65.3 ± 9.4	64.1 ± 10.1	59.3 ± 10.1	<0.001[c]
Inferior RNFL thickness (µm)	111.8 ± 18.1	87.7 ± 18.3	67.1 ± 15.2	<0.001[c]
Topcon SS-OCT				
Image-quality score	59.5 ± 6.1	59.1 ± 5.9	57.8 ± 6.1	0.209[a]
Superior RNFL thickness (µm)	119.3 ± 18.4	97.8 ± 22.9	74.2 ± 27.5	<0.001[c]
Temporal RNFL thickness (µm)	81.1 ± 13.1	72.6 ± 15.5	57.9 ± 16.1	<0.001[c]
Nasal RNFL thickness (µm)	69.3 ± 14.5	64.7 ± 13.3	55.6 ± 16.9	<0.001[c]
Inferior RNFL thickness (µm)	122.8 ± 21.0	90.1 ± 24.2	58.1 ± 19.2	<0.001[c]

GCIPL, ganglion cell to inner plexiform layer.
[a]Analysis of variance test.
[b]χ^2 test.
[c]Kruskal-Wallis test.

layer and pRNFL thickness maps. For quality control, only good quality scans defined as having a minimum signal strength of 6, having no involuntary eye movements or blinking artifacts, and being without misalignment or segmentation failures were used for analysis.

Topcon SS-OCT

Wide-angle scanning using the SS-OCT (DRI-OCT-1 Atlantis; Topcon, Tokyo, Japan) was performed on each subject within 6 months of the VF. Wide-angle scanning uses a wide-angle 12 × 9 mm lens, with

the scan centered on the fovea, for 256 B-scans, each comprising 512 A-scans, for a total of 131,072 axial scans per volume. A scan time of 1.3 seconds per 12 × 9 mm^2 scan, which was previously shown to be sufficient for acquiring total images, was used here. Poor quality images (image quality scores of <40, poorly focused, or decentered during fovea scanning) or those acquired after segmentation failures or with artifacts owing to eye movements or blinking were excluded. Built-in DRI-OCT-1 software (version 9.12) automatically identified the outer boundary of the RNFL, from the internal limiting area to the retinal ganglion cells and the outer boundary of the inner plexiform layer.

VF Examination

A VF test was performed on all training and test subjects using a Humphrey Field Analyzer 750i instrument (Carl Zeiss Meditec) with the Swedish interactive threshold algorithm 24-2 or 30-2. Of the 54 test points of the 24-2 test pattern, 2 points of physiologic scotoma were excluded, and the remaining 52 test points of the total threshold value were used as the ground truth VF of the training and test sets. Adequate reliability was defined as having a false-positive rate of less than 33%, false-negative rate of less than 33%, and fixation loss of less than 20%. Normal VFs were defined as those with a glaucoma hemifield test within the normal limits and with a mean deviation (MD) and pattern standard deviation (PSD) within 95% of the Humphrey Field Analyzer's normative database. Glaucoma VFs were defined as those who met at least one of the following criteria: glaucoma hemifield test outside the normal limits and/or PSD probability outside of 95% of the Humphrey Field Analyzer's normative database. Glaucoma severity was determined by the MD of the VF: early greater than –6 dB and moderate to severe (advanced) –6 dB or less.

To evaluate performance sectorally, we mapped 52 VF test points according to two different methods. First, Garway-Heath sectorization[22] was used (Supplementary Fig. S1A). This method assigns VF test points to the corresponding ONH sectors and is known to be useful to analyze structure-function relationships. The second method was central and peripheral mapping (Supplementary Fig. S1B). The central region contains test points that overlap with the VF 10-2 points,[23] and the peripheral region contains the rest of the test points outside the central region.

Input Image Generation

We developed custom Windows software using Microsoft Visual Studio 2015 and C# language with a dot net library to generate combined input images. Our custom software used two report images exported from Zeiss SD-OCT: (1) ganglion cell analysis of both eyes: macular cube 200 × 200, and (2) ONH and RNFL analysis of both eyes: optic disc cube 200 × 200 protocol. The custom software automatically detected the location of these blue-toned thickness maps by searching for the rectangular boundary of the blue image, starting from a predefined location, and cropped and combined them (Fig. 1 left). Another custom software for SS-OCT cropped the ganglion cell layer thickness map and RNFL thickness map in the SS-OCT report image and combined it together (Fig. 1, right). All left eye images were flipped horizontally to match the right eye format.

Deep Learning Architectures and Training

The open source deep learning library, Keras,[24] running on top of the TensorFlow backend (Google, Mountain View, CA) python (version 3.5) API r1.10, was used. The CUDA toolkit 9.0 and cuDNN 7.0 library were installed to use the GPU computation power. The hardware environment used for training and test runs was an Intel i5-8400 CPU, 32 GB RAM, and a GeForce Titan XP (NVIDIA, Santa Clara, CA).

The final deep neural network architecture to estimate the total threshold values is shown in Fig. 1. Google's state-of-the-art CNN architecture, Inception-Resnet V2, was used as the backbone structure at the beginning of architecture to extract global features. A bottleneck layer of the InceptionResnet V2 backbone was modified by one global average pooling layer followed by four consecutive fully connected layers (dense layers 1–4 in Fig. 1). The ReLu (rectified linear unit) was used as the activation function in all 4 dense layers. A combined OCT image, which had a size of 322 × 161 (width × height) for SD-OCT or 480 × 200 (width × height) for SS-OCT, was fed into the input layer of the CNN architecture. The output of the InceptionResnet V2 backbone architecture produced a shape of 3 × 8 × 2048 (height × width × depth) global features. The global average pooling layer flattened the output of the inception backbone and averaged 2048 features. Four dense layers gradually condensed these features into 52 final output values, which corresponded with 52 total threshold values (two points of physiologic scotoma were excluded from estimation).

Before training began, InceptionResnet pretrained on the ImageNet dataset was downloaded and applied. No layer was frozen during training and all layers were fine tuned. Training data records were randomly split into training and validation datasets in a 9:1 ratio and

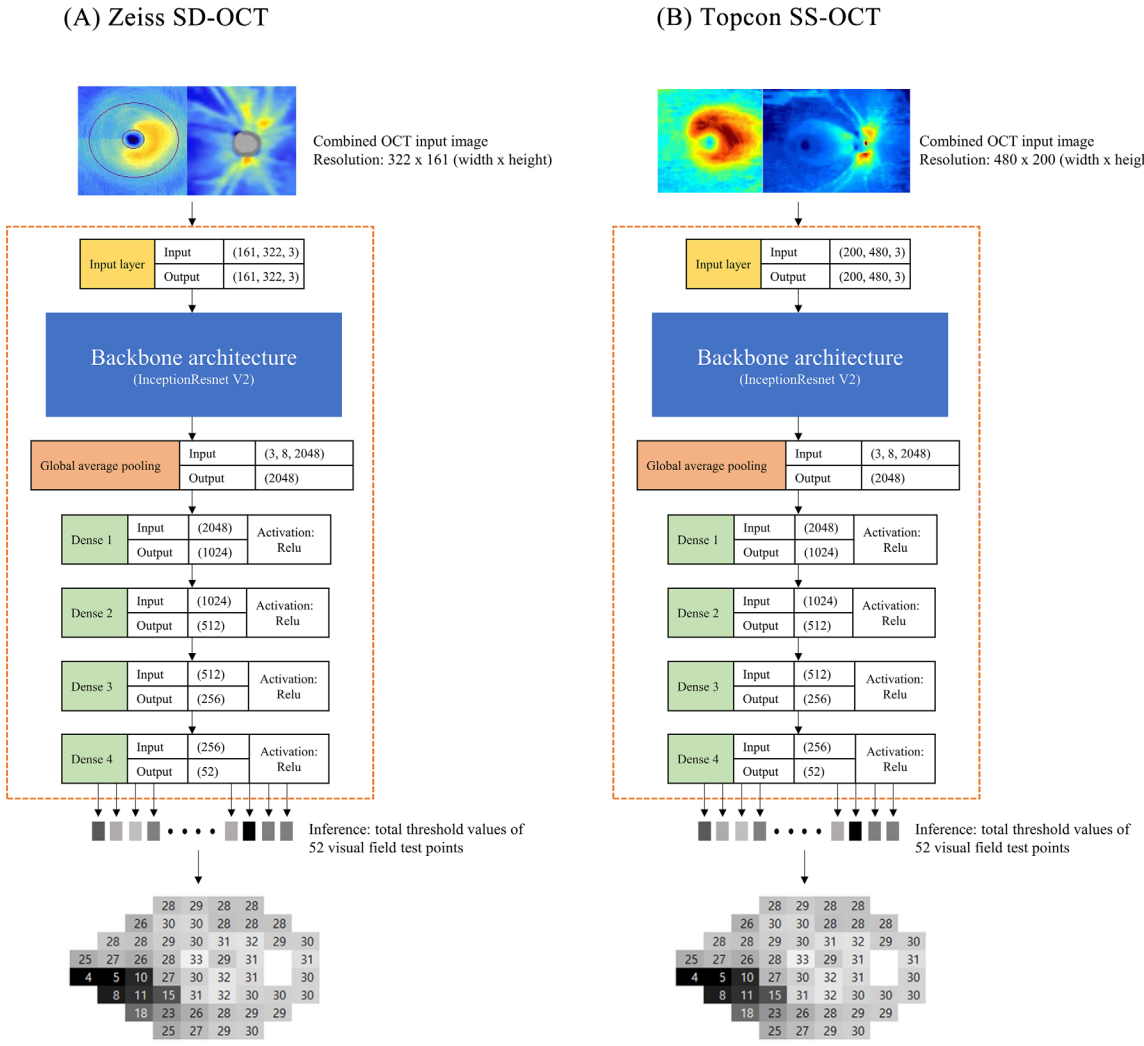

Figure 1. Diagram of the deep learning architecture.

batches of 64 were supplied to the neural network. The optimizer was 'rmsprop' and the loss function was 'mean squared error.' Training was monitored by reference to the loss trends of both the training and validation sets. When no further performance gain was observed over 100 epochs, training finished. To prevent overfitting, the repeated random subsampling cross-validation technique[25] was used. The training data were again randomly split in a 9:1 ratio, the last trained weight file was loaded, and training resumed until no further performance gain was evident over 100 epochs. This process was repeated five times.

Statistical Analysis

The Shapiro–Wilk test was performed to check the normality of the continuous data distribution. To compare values among all three groups (normal, early glaucoma, advanced glaucoma), we used the Kruskal–Wallis test, Friedman's test, or analysis of variance depending on the normality of the data. To perform post hoc analysis, we used the Mann–Whitney *U* test or Wilcoxon's signed rank test. The chi-square test was used for categorical variables. The VF estimation error was calculated as the root mean square error (RMSE) using the following formula:

$$RMSE = \sqrt{\sum_{n=1}^{N} \frac{(true\ THV_n - predicted\ THV_n)^2}{N}}$$

$n = n^{th}$ test point of the visual field exam, N = Total number of values, THV=visual field threshold value.

When we calculated the global RMSE (i.e., including all 52 test points), N was 52, and for the regional estimation error, N was only a select number of VF test points inside the target region. These regions are defined in Figure 1. For conducting statistical analy-

ses, SPSS (version 21.0 for Windows; SPSS, Chicago, IL) was used, and a P value of less than 0.05 (single comparison) and a P value of less than 0.017 (multiple comparisons) were considered to indicate statistical significance.

Results

Data from a total of 243 eyes from 243 patients were used for the test set. The test set included 86 normal subjects, 87 patients with early glaucoma, and 70 patients with advanced (moderate-to-severe stage) glaucoma. The demographic characteristics are summarized in Table 2. Age and gender were significantly different ($P = 0.011$ and $P = 0.017$, respectively) between the subject groups, whereas intraocular pressure, refractive error, central corneal thickness, and axial length were not. All VF parameters, such as the MD, PSD, and VF index, were significantly different (all P values < 0.001) between the subject groups. The average MD was –1.23 dB/–3.16 dB/–13.35 dB (normal/early/advanced glaucoma, respectively). Supplementary Fig. S2 shows the distribution of MD between subject group. All SD-OCT parameters were significantly different between the subject groups. In the SD-OCT macular ganglion cell analysis, the average GCIPL thickness was 80.2 µm (normal), 73.1 µm (early glaucoma), 63.6 µm (advanced glaucoma) and was significantly different between the subject groups (all P values < 0.001). In SD-OCT ONH and RNFL analysis, the sectoral RNFL thicknesses were 109.5 µm, 69.4 µm, 65.3 µm, and 111.8 µm for normal subjects, 94.9 µm, 65.1 µm, 64.1 µm, and 87.7 µm for patients with early glaucoma, and 78.6 µm, 55.5 µm, 59.3 µm, and 67.1 µm for patients with advanced glaucoma (superior, temporal, nasal, and inferior sector, respectively), and these values were all significantly different between groups (all $P < 0.001$). All sectoral RNFL thicknesses measured by SS-OCT were also significantly different between subject groups (all $P < 0.001$) except for the image quality score ($P = 0.209$).

Global and regional VF estimation error between ground truth and estimated are summarized in Table 3, and representative examples are shown in

Table 3. RMSE Between Ground Truth and Estimated Values of the VF

			Glaucoma			Post Hoc Analysis		
	All Subjects	Normal	Early	Advanced	P_{al}[a]	P_{NE}[b]	P_{EA}[c]	P_{NA}[d]
Zeiss SD-OCT								
Global	5.29 ± 2.68	3.75 ± 1.26	4.73 ± 2.28	7.84 ± 2.67	<0.001	0.003	<0.001	<0.001
Temporal	4.73 ± 3.38	3.06 ± 1.29	4.08 ± 2.93	7.59 ± 3.94	<0.001	0.138	<0.001	<0.001
Superotemporal	4.78 ± 3.09	3.77 ± 1.57	4.13 ± 2.06	6.80 ± 4.45	<0.001	0.287	<0.001	<0.001
Inferotemporal	4.85 ± 3.49	3.41 ± 1.42	4.39 ± 3.06	7.17 ± 4.57	<0.001	0.073	<0.001	<0.001
Nasal	4.42 ± 3.04	3.59 ± 1.73	3.89 ± 2.26	6.09 ± 4.35	<0.001	0.709	0.001	<0.001
Superonasal	5.17 ± 2.94	4.00 ± 1.41	4.73 ± 1.98	7.16 ± 4.17	<0.001	0.013	<0.001	<0.001
Inferonasal	5.23 ± 3.80	3.90 ± 1.82	5.13 ± 4.08	6.97 ± 4.62	<0.001	0.185	0.001	<0.001
Central	5.14 ± 3.07	3.65 ± 1.49	4.28 ± 2.39	8.05 ± 3.37	<0.001	0.182	<0.001	<0.001
Peripheral	5.26 ± 2.69	3.76 ± 1.26	4.81 ± 2.35	7.65 ± 2.82	<0.001	0.003	<0.001	<0.001
Topcon SS-OCT								
Global	4.51 ± 2.54	2.88 ± 0.92	3.77 ± 1.45	7.43 ± 2.54	<0.001	<0.001	<0.001	<0.001
Temporal	3.89 ± 3.37	1.94 ± 0.92	2.96 ± 2.00	7.42 ± 3.98	<0.001	<0.001	<0.001	<0.001
Superotemporal	3.65 ± 3.25	2.38 ± 1.30	2.79 ± 2.05	6.26 ± 4.51	<0.001	0.385	<0.001	<0.001
Inferotemporal	4.30 ± 3.46	2.60 ± 0.91	3.72 ± 2.01	7.11 ± 4.94	<0.001	<0.001	<0.001	<0.001
Nasal	3.41 ± 2.48	2.78 ± 1.25	2.91 ± 1.58	4.81 ± 3.74	<0.001	0.803	0.001	0.001
Superonasal	4.22 ± 2.90	3.05 ± 1.38	3.60 ± 1.93	6.39 ± 4.00	<0.001	0.126	<0.001	<0.001
Inferonasal	4.54 ± 2.99	3.59 ± 1.61	4.26 ± 2.65	6.06 ± 4.05	<0.001	0.222	0.002	<0.001
Central	4.13 ± 3.33	2.22 ± 1.15	3.00 ± 1.82	7.88 ± 3.63	<0.001	0.007	<0.001	<0.001
Peripheral	4.51 ± 2.46	3.01 ± 0.97	3.89 ± 1.54	7.11 ± 2.67	<0.001	<0.001	<0.001	<0.001

[a] P value among all subject groups (Kruskal–Wallis test, significance level $P < 0.05$).
[b] P value between normal and early glaucoma (Mann–Whitney U test, significance level $P < 0.017$).
[c] P value between early and advanced glaucoma (Mann–Whitney U test, significance level $P < 0.017$).
[d] P value between normal and advanced glaucoma (Mann–Whitney U test, significance level $P < 0.017$).

Table 4. RMSE Between Ground Truth and Estimated Values of the VF in the Global Sector

	Zeiss SD-OCT	Topcon SS-OCT	*P* Value[a]
All subjects	5.29 ± 2.68	4.51 ± 2.54	<0.001
Normal	3.75 ± 1.26	2.88 ± 0.92	<0.001
Early glaucoma	4.73 ± 2.28	3.77 ± 1.45	<0.001
Advanced glaucoma	7.84 ± 2.67	7.43 ± 2.54	0.218

[a]Wilcoxon's signed rank test.

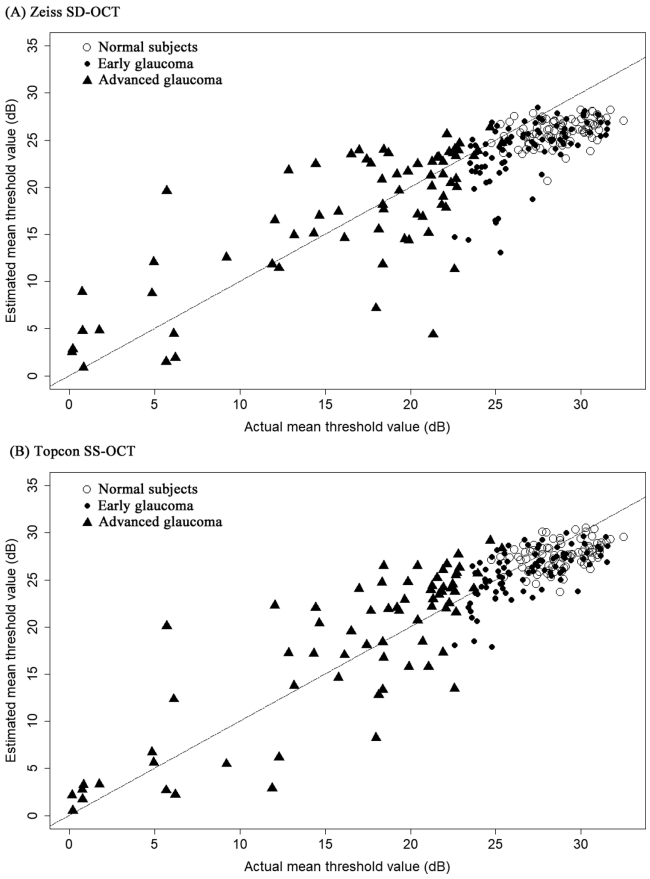

Figure 2. Scatter plot of estimated versus actual mean threshold value.

Supplementary Fig. S3. Globally, the estimation error became significantly worsened as glaucoma progressed in both OCTs (all $P \leq 0.003$). In SD-OCT, the global RMSE was 5.29 ± 2.68 dB for all patients, 3.75 ± 1.26 dB for normal subjects, 4.73 ± 2.28 dB for patients with early glaucoma, and 7.84 ± 2.67 dB for patients with advanced glaucoma. In SS-OCT, the global RMSE was 4.51 ± 2.54 dB for all patients, 2.88 ± 0.92 dB for normal subjects, 3.77 ± 1.45 dB for patients with early glaucoma, and 7.43 ± 2.54 dB for patients with advanced glaucoma. Supplementary Table S1 lists the intraclass variabilities of the VF total threshold values. Data from a second VF examination performed within 6 months after the original test set were collected and the mean absolute differences between the two consecutive examinations were calculated. The intra-lass variabilities of the VF mean threshold values were 4.28 ± 1.94 (all subjects), 3.27 ± 1.07 dB (normal subjects), 3.87 ± 1.48 dB (patients with early glaucoma), and 5.56 ± 2.27 dB (patients with advanced glaucoma).

In the Garway–Heath sectorization, the RMSEs of the temporal side sectors (temporal, superotemporal, and inferotemporal sectors), which are more important for glaucomatous damage, were all lower than those of the corresponding nasal side sectors (nasal, superonasal, and inferonasal sectors). The lowest RMSE sector among the temporal side was the temporal sector (4.73 dB) in SD-OCT and the superotemporal sector (3.65 dB) in SS-OCT. All sectoral estimation errors became significantly worsened as the glaucoma progressed. However, between normal vision and early glaucoma, not all sectors were significantly different. In SD-OCT, only the superonasal sector was significantly different between normal vision and early glaucoma, whereas in SS-OCT, the temporal and inferotemporal sectors were significantly different. In central and peripheral sectorization, the estimation error of the central region was generally lower than that of the peripheral region, but in advanced glaucoma, it was the opposite. The estimation error of SD-OCT was 5.14 dB and 5.26 dB (central and peripheral) and that of SS-OCT was 4.13 dB and 4.51 dB (central and peripheral) in all patients.

A global comparison of the estimation error (RMSE) between the two OCT devices is summarized in Table 4 and Figure 2. In general, the global estimation error of SS-OCT was significantly lower than that of SD-OCT (all $P < 0.001$), except for the advanced glaucoma group. In advanced glaucoma, the estimation error of SS-OCT (7.43 dB) was still better than that of SD-OCT (7.84 dB), but it was not significant ($P = 0.218$). The plot of estimation error (Fig. 3) showed that the Topcon SS-OCT data were skewed to the left compared with the Zeiss-SD-OCT data. This finding indicated that the overall estimation error was lower for the Topcon SS-OCT.

Bar plots in Figure 4 show the sectoral estimation error between SD-OCT and SS-OCT. In normal

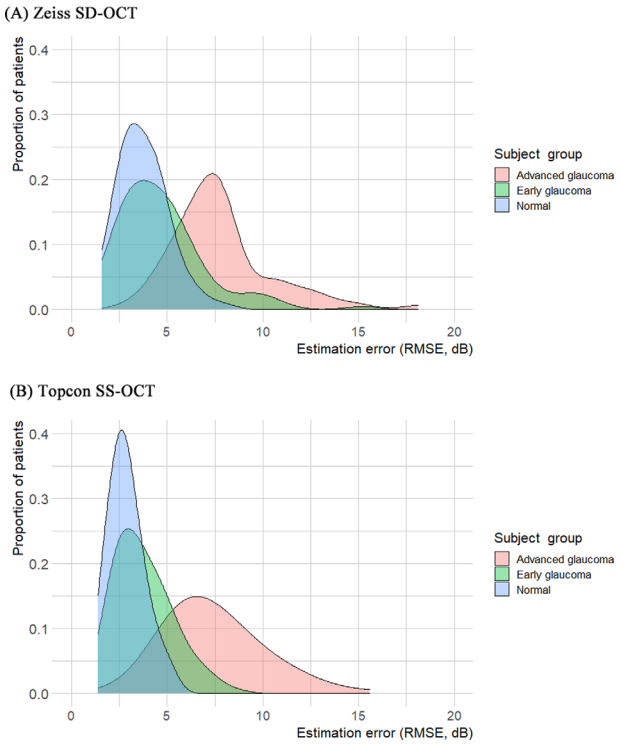

Figure 3. Distribution plot of estimation error.

subjects and the early glaucoma groups, the estimation error of SS-OCT was better than that of SD-OCT in all sectors, except for the inferonasal sector. However, in the advanced glaucoma group, all temporal side sectors (temporal, superotemporal, and inferotemporal) showed no significant performance difference between both OCTs, while in all nasal sectors (nasal, superonasal, and inferonasal), SS-OCT showed significantly better performance than SD-OCT. In the advanced glaucoma group, the peripheral estimation error of SS-OCT was significantly better than that of SD-OCT, but there was no significant difference in the central region. In the normal and early glaucoma groups, SS-OCT was significantly better than SD-OCT in both the central and peripheral regions.

Figure 5 shows the pointwise errors (RMSE) for SD-OCT and SS-OCT. SS-OCT (middle column) generally showed more error with respect to brightness than SD-OCT (left column), especially in subjects with normal vision and patients with early glaucoma. However, as the glaucoma progressed, the images became darker, and the difference between the two devices disappeared. The number of points showing non-significant differences (right column, darker color) between the two types of OCT was 2 (3.84%) for normal subjects, 6 (11.53%) for patients with early glaucoma, and 27 (51.9%) for patients with advanced

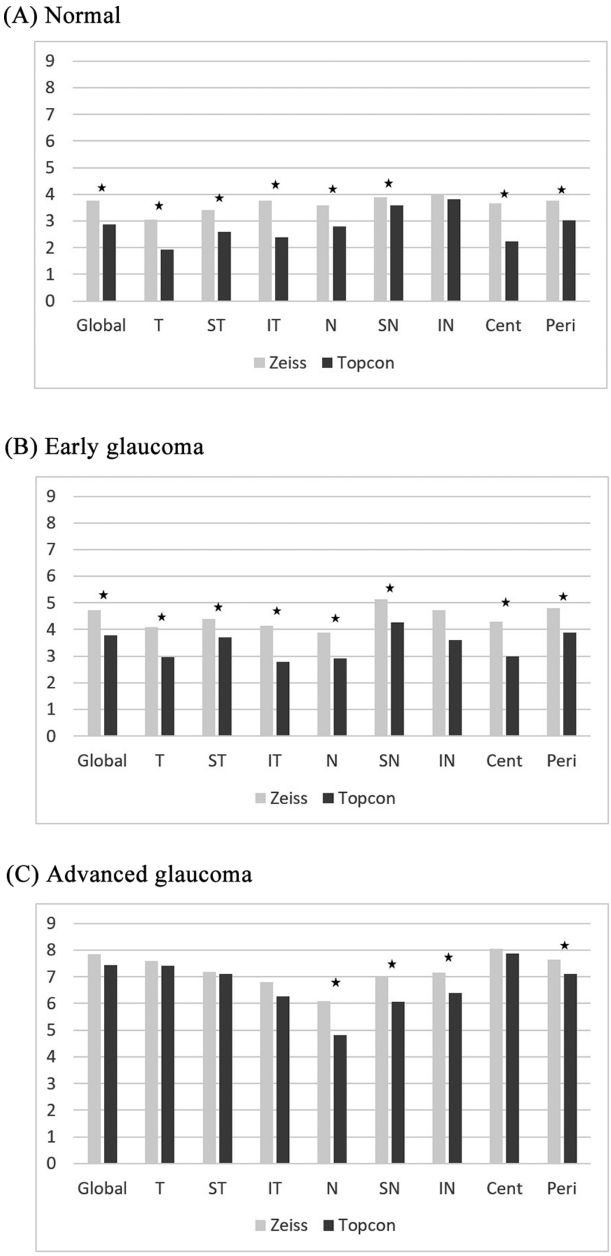

Figure 4. Regional estimation error. Cent, central; IN, inferonasal; IT, inferotemporal; N, nasal; SN, superonasal; Peri, peripheral; ST, superotemporal; T, temporal. An asterisk (*) denotes a significant difference between Zeiss and Topcon.

glaucoma. As glaucoma progressed, there were more of these points in the nasal periphery of the VF.

A multiple linear regression analysis was performed to identify factors associated with estimation error (Table 5). The outcome variable was estimation error (RMSE) for both the Zeiss SD-OCT and Topcon SS-OCT. Age, axial length, central corneal thickness, OCT signal strength, and VF MD were included as covariates. For both the SD-OCT and SS-OCT data, only VF MD was significantly associated with estimation error.

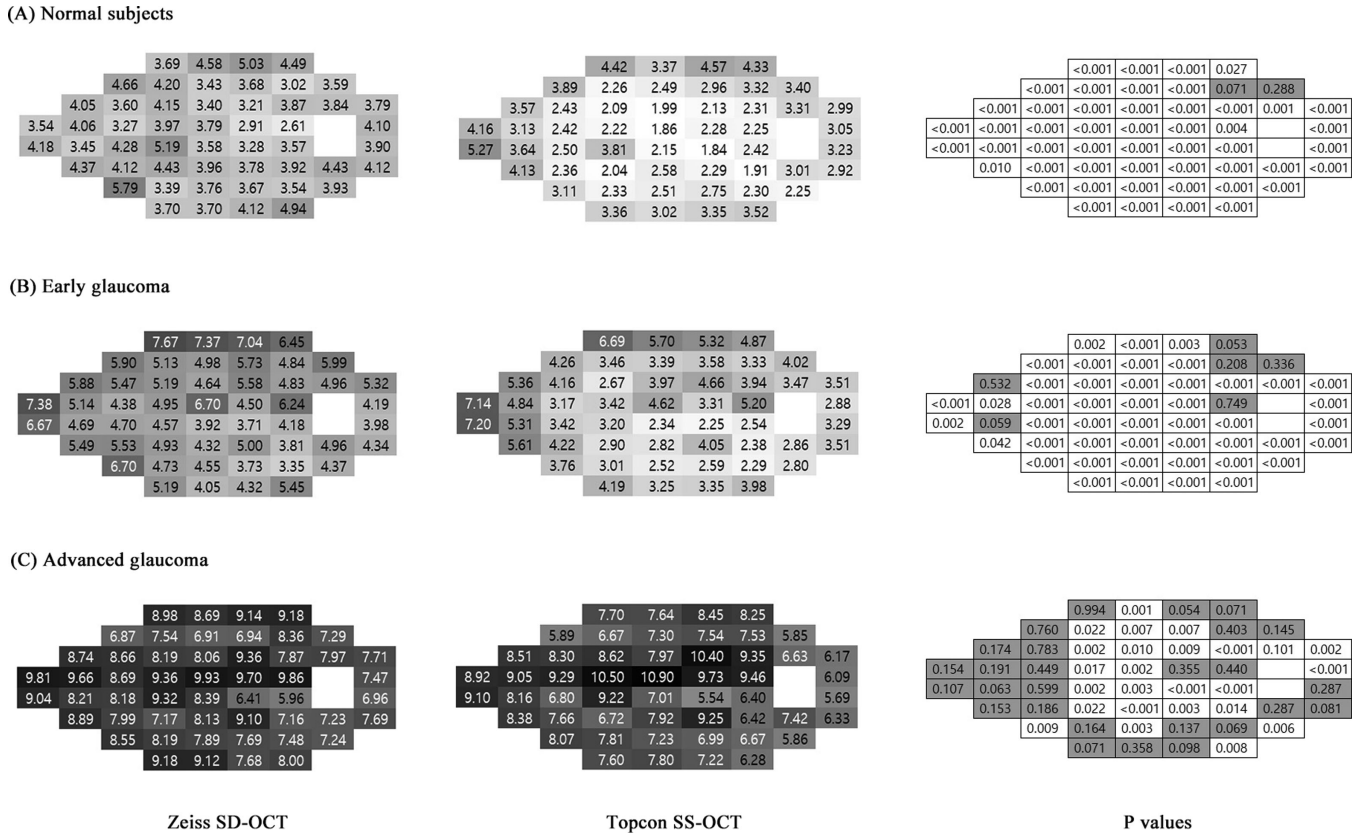

Figure 5. Pointwise RMSE between ground truth and the estimated value. In left and middle column, the darker the color is, the higher estimation error. Right column shows pointwise *P* values (the darker color means insignificant) comparing two OCTs (Wilcoxon's signed rank test). SS-OCT, swept source OCT.

Table 5. Multiple Linear Regression Analysis for Association Between Estimation Error and Various Factors

	β	*P* Value
Zeiss SD-OCT		
Age	−0.106	0.205
Axial length	0.026	0.739
Central corneal thickness	−0.039	0.547
Macula OCT signal strength	−0.079	0.356
ONH OCT signal strength	0.029	0.722
VF MD	−0.493	<0.001
Topcon SS-OCT		
Age	−0.045	0.530
Axial length	−0.050	0.479
Central corneal thickness	−0.014	0.814
OCT signal strength	−0.063	0.292
VF MD	−0.612	<0.001

Table contains the results of two multiple linear regression models. The outcome variables are estimation error (root mean squared error) of Zeiss SD-OCT and estimation error of Topcon SS-OCT. Each model includes age, axial length, central corneal thickness, OCT signal strengths, VF MD factor as a covariate. The 'Enter' method was used.

Discussion

The present study found that SS-OCT showed a significantly higher accuracy than SD-OCT, except for the advanced glaucoma group. The estimation error became significantly worsened as glaucoma progressed in both OCTs. In the regional comparison, the estimation error of SS-OCT was significantly lower than that of SD-OCT in almost all sectors except for only the inferonasal sector in the normal subject and early glaucoma groups.

Several previous studies have estimated the VF from OCT images using non-deep learning methods. Tan et al.[26] developed a mathematical formula that converts RNFL thickness into "actual VF MD," which shows better reproducibility than "actual VF." However, actual VF MD is a global index. In another study, Zhang et al.[27] used multiple linear regression to estimate the 10-2 and 24-2 VFs simultaneously from a combined OCT image (similar to our study). In contrast with our study, however, their method predicted the probability of normality. The 10-2 and 24-2 VF estimates exhibited 82.2% agreement. In

another study based on a conventional artificial intelligence method, Guo et al.[28] used a "nine-field-per-eye" protocol, in which the patient fixated on a 3 × 3 grid of spots (total of nine OCT images). They constructed four topological structure–function maps and used a support vector machine algorithm to calculate the VF from those maps. Among the four predefined maps, the one showing the best performance had an RMSE of 5.42 dB. Thus, the performance was good for some of the structure–function maps. The highest prediction error among the maps was 7.24 dB.

Recently, researchers have begun to use deep learning algorithms to estimate the VF from OCT data. Christopher et al.[16] attempted to estimate the VF from SD-OCT images using a deep learning method. Among ONH en face images, peripapillary RNFL thickness maps, and confocal scanning laser ophthalmoscopic images, the best performance was achieved by the ONH en face image, with a prediction error of 2.5 dB. Although our method could estimate the entire VF, their deep learning model estimated the VF both globally and in terms of individual sectors, without a comparison among OCT sources. Sugiura et al.[29] described another deep learning method. As in our study, their deep learning architecture used both GCIPL and RNFL images as input, but the RNFL thickness maps were obtained in macular areas. Their method was designed for application to the entire Humphrey 10-2 VF. The RMSE value obtained was 6.16 dB. Additionally, Yu et al.[30] developed a deep learning architecture to estimate global indices of the VF (MD and PSD) from OCT images. As in our study, the architecture used by Yu et al. incorporated both macular cube images and ONH RNFL thickness maps, obtained by a Zeiss SD-OCT instrument. The median error in the VF MD was 1.57 dB.

In the present study, the estimation error of SS-OCT was significantly lower than that of SD-OCT. This outcome is probably because the scanning area of SS-OCT is wider than that of SD-OCT. SS-OCT covers an area of 12 × 9 mm (width × height), whereas SD-OCT covers 6 × 6 mm of peripapillary area and 6 × 6 mm of macular area. This wider area should contain much more information than the area in SD-OCT and could better reflect the structural damage corresponding to the functional loss. Hood et al.[31] reported that a single wide scan in SS-OCT contains the information needed to diagnose early glaucoma with excellent sensitivity and specificity. In another study,[32] structural changes shown on SS-OCT can detect or estimate VF damage even in preperimetric glaucoma.

However, in patients with advanced glaucoma, we found no significant performance difference between the two types of OCT. The training dataset for advanced glaucoma was smaller than that for early glaucoma, so more extensive training of the deep learning model was possible for early glaucoma. However, even with sufficient training data, the performance with respect to advanced glaucoma may not be greatly improved because the quality of data is also problematic: previous studies demonstrated a saturation effect in the structure–function relationship in advanced glaucoma, because of remnant glial cells and blood vessels providing a degree of thickness of the macula even after total loss of visual function.[33] This remnant thickness may hamper estimation of the VF from OCT-derived measurements of RNFL and ganglion cell layer thickness. As shown in Supplementary Table S1, the mean absolute VF total threshold value difference between two consecutive examinations (i.e., the intraclass variability) was relatively large considering the range of estimation error, and increased as glaucoma progressed. Thus, a large proportion of the estimation error was attributable to the intrinsic variability of the VF examination per se. This finding may explain why the two OCT devices performed similarly.

Glaucomatous changes typically begin in the peripheral field.[34] In this study, a pointwise comparison of estimation error showed that nearly one-half of the peripheral points exhibited higher error and no performance difference between the two types of OCT in patients with advanced glaucoma. Moreover, most retinal ganglion cell damage occurs in the early stage glaucoma. In advanced glaucoma, few retinal ganglion cells remain. Although the VF exhibits considerable alterations, changes in RGCs are small. This factor causes great difficulty in the detection of RNFL thickness changes using OCT.[35]

In the present study, the central region was significantly better estimated than the peripheral region in normal and early glaucoma, not advanced glaucoma. Hood et al.[36] reported that a 24-2 VF is based on a 6° grid, which can miss glaucomatous damage in the central macular region. In some previous studies, a 10-2 VF, which uses a narrower grid (2° grid), may help to better detect the preserved central island and glaucoma progression, especially in advanced glaucoma.[37,38] Thus, in advanced glaucoma, 24-2 VF have limited accuracy, especially in the central region, and mismatches between estimated and actual VFs may become larger. If we train another deep learning model to estimate the central VF area with a 10-2 VF, the estimation accuracy may be improved.

The limitation of this study is the training data imbalance between both OCTs. Despite our best efforts

to make it even, SD-OCT images contained a smaller number of moderate to severe glaucoma cases (38.1%) than SS-OCT images did (46.2%). The imbalance was not large, but could influence the relatively lower performance of SD-OCT. Another limitation of our model is that the estimation performance worsened as glaucoma progressed. We are preparing further studies to improve this problem by augmenting the information, such as angio-OCT or slab images, in SS-OCT reports. Finally, in our study, the OCT signal strength did not significantly affect the performance of the deep learning model, probably because only good-quality OCT images were included in the test data set. In future studies, the impact of OCT signal strength should be evaluated.

In conclusion, our deep learning model estimated the VF 24-2 better with SS-OCT images than with SD-OCT images. As glaucoma progressed, the estimative performance worsened in both OCTs. In the regional analysis, SS-OCT showed significantly better performance in almost all sectors than SD-OCT in normal and early glaucoma. In advanced glaucoma, SS-OCT showed significantly better performance than SD-OCT in only the peripheral region.

Data Availability

All data are provided with supporting information file. Please refer 'datasheet.xlsx' spreadsheet. All python source code, trained model, and test images are available on Github web site: https://github.com/climyth/VFbyOCT-Comparison.

Acknowledgments

Disclosure: **J. Shin**, None; **S. Kim**, None; **J. Kim**, None; **K. Park**, None

References

1. Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol*. 2014;98(Suppl 2):ii15–ii19.
2. Sommer A, Miller NR, Pollack I, Maumenee AE, George T. The nerve fiber layer in the diagnosis of glaucoma. *Arch Ophthalmol*. 1977;95(12):2149–2156.
3. Hart WM, Becker B. The onset and evolution of glaucomatous visual field defects. *Ophthalmology*. 1982;89(3):268–279.
4. Fogagnolo P, Sangermani C, Oddone F, et al. Long-term perimetric fluctuation in patients with different stages of glaucoma. *Br J Ophthalmol*. 2011;95(2):189–193.
5. Langerhorst CT, Van den Berg T, Van Spronsen R, Greve EL. Results of a fluctuation analysis and defect volume program for automated static threshold perimetry with the scoperimeter. In: *Sixth International Visual Field Symposium*. New York: Springer; 1985:1–6.
6. Brenton RS, Argus WA. Fluctuations on the Humphrey and Octopus perimeters. *Invest Ophthalmol Vis Sci*. 1987;28(5):767–771.
7. Gürses-Özden R, Teng C, Vessani R, Zafar S, Liebmann JM, Ritch R. Macular and retinal nerve fiber layer thickness measurement reproducibility using optical coherence tomography (OCT-3). *J Glaucoma*. 2004;13(3):238.
8. Blumenthal EZ, Williams JM, Weinreb RN, Girkin CA, Berry CC, Zangwill LM. Reproducibility of nerve fiber layer thickness measurements by use of optical coherence tomography. *Ophthalmology*. 2000;107(12):2278–2282.
9. Carpineto P, Nubile M, Agnifili L, et al. Reproducibility and repeatability of Cirrus^{TM} HD-OCT peripapillary retinal nerve fibre layer thickness measurements in young normal subjects. *Ophthalmologica*. 2012;227(3):139–145.
10. Chen HY, Huang ML. Discrimination between normal and glaucomatous eyes using Stratus optical coherence tomography in Taiwan Chine se subjects. *Graefes Arch Clin Exp Ophthalmol.* 2005;243:894–902
11. DeLeón-Ortega JE, Arthur SN, McGwin G, Xie A, Monheit BE, Girkin CA. Discrimination between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective optic nerve head assessment. *Invest Ophthalmol Vis Sci*. 2006;47(8):3374–3380.
12. Huang M-L, Chen H-Y. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2005;46(11):4121–4129.
13. Schuman JS. Spectral domain optical coherence tomography for glaucoma (An AOS Thesis). *Trans Am Ophthalmol Soc*. 2008;106:426–458.
14. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology*. 2009;116(7):1257–1263.
15. Yang Z, Tatham AJ, Zangwill LM, Weinreb RN, Zhang C, Medeiros FA. Diagnostic ability of reti-

nal nerve fiber layer imaging by swept-source optical coherence tomography in glaucoma. *Am J Ophthalmol*. 2015;159(1):193–201.

16. 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.
17. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. 2010;51(11):5657–5666.
18. Leite MT, Zangwill LM, Weinreb RN, Rao HL, Alencar LM, Medeiros FA. Structure-function relationships using the Cirrus spectral domain optical coherence tomograph and standard automated perimetry. *J Glaucoma*. 2012;21(1):49–54.
19. Kim S, Lee JY, Kim SO, Kook MS. Macular structure-function relationship at various spatial locations in glaucoma. *Br J Ophthalmol*. 2015;99(10):1412–1418.
20. Na JH, Kook MS, Lee Y, Baek S. Structure-function relationship of the macular visual field sensitivity and the ganglion cell complex thickness in glaucoma. *Invest Ophthalmol Vis Sci*. 2012;53(8):5044–5051.
21. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol*. 2002;86(2):238–242.
22. 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.
23. Park SC, Kung Y, Su D, et al. Parafoveal scotoma progression in glaucoma: Humphrey 10-2 versus 24-2 visual field analysis. *Ophthalmology*. 2013;120(8):1546–1550.
24. Chollet F. *Keras*. 2015. Available at: https://github.com/fchollet/keras. Accessed June 24, 2020.
25. Dubitzky W, Granzow M, Berrar DP. *Fundamentals of Data Mining in Genomics and Proteomics*. New York: Springer Science & Business Media; 2007.
26. Tan O, Greenfield DS, Francis BA, Varma R, Schuman JS, Huang D. Estimating visual field mean deviation using optical coherence tomographic nerve fiber layer measurements in glaucoma patients. *Sci Rep*. 2019;9(1):18528.
27. Zhang X, Dastiridou A, Francis BA, et al. Comparison of glaucoma progression detection by optical coherence tomography and visual field. *Am J Ophthalmol*. 2017;184:63.
28. 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.
29. 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. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '18. London: Association for Computing Machinery; 2018:783–792, https://www.kdd.org/kdd2018/.
30. 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. *Ophthalmol Glaucoma*. Published online first: 11 July 2020.
31. Hood DC, De Cuir N, Blumberg DM, et al. A single wide-field OCT protocol can provide compelling information for the diagnosis of early glaucoma. *Transl Vis Sci Technol*. 2016;5(6):4.
32. Lee WJ, Kim YK, Jeoung JW, Park KH. Can probability maps of swept-source optical coherence tomography predict visual field changes in preperimetric glaucoma? *Invest Ophthalmol Vis Sci*. 2017;58(14):6257–6264.
33. Harwerth RS, Carter-Dawson L, Smith EL, Crawford ML. Scaling the structure– function relationship for clinical perimetry. *Acta Ophthalmol Scand*. 2005;83(4):448–455.
34. Hoffmann EM, Boden C, Zangwill LM, Bourne RR, Weinreb RN, Sample PA. Inter-eye comparison of patterns of visual field loss in patients with glaucomatous optic neuropathy. *Am J Ophthalmol*. 2006;141(4):703–708.
35. Medeiros FA, Zangwill LM, Bowd C, Mansouri K, Weinreb RN. The structure and function relationship in glaucoma: implications for detection of progression and measurement of rates of change. *Invest Ophthalmol Vis Sci*. 2012;53(11):6939–6946.
36. Hood DC, Raza AS, de Moraes CGV, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res*. 2013;32:1–21.
37. Rao HL, Begum VU, Khadka D, Mandal AK, Senthil S, Garudadri CS. Comparing glaucoma progression on 24-2 and 10-2 visual field examinations. *PLoS Ono*. 2015;10(5):e0127233.
38. Weber J, Schultze T, Ulrich H. The visual field in advanced glaucoma. *Int Ophthalmol*. 1989;13(1–2):47–50.