UCLA

Title

Prediction of Central Visual Field Measures From Macular OCT Volume Scans With Deep Learning.

Permalink

https://escholarship.org/uc/item/7wp8w95b

Journal

Translational Vision Science and Technology, 12(11)

Authors

Mohammadzadeh, Vahid

Vepa, Arvind

Li, Chuanlong

et al.

Publication Date

2023-11-01

DOI

10.1167/tvst.12.11.5

Peer reviewed

tvst

Prediction of Central Visual Field Measures From Macular OCT Volume Scans With Deep Learning

Vahid Mohammadzadeh[1,*], Arvind Vepa[2,*], Chuanlong Li[3], Sean Wu[4], Leila Chew[1], Golnoush Mahmoudinezhad[1], Evan Maltz[5], Serhat Sahin[2], Apoorva Mylavarapu[1], Kiumars Edalati[1], Jack Martinyan[1], Dariush Yalzadeh[1], Fabien Scalzo[2], Joseph Caprioli[1], and Kouros Nouri-Mahdavi[1]

[1] Glaucoma Division, Stein Eye Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
[2] Department of Computer Science, University of California Los Angeles, Los Angeles, CA, USA
[3] Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
[4] Department of Computer Science, Pepperdine University, Malibu, CA, USA
[5] Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA, USA

Purpose: Predict central 10° global and local visual field (VF) measurements from macular optical coherence tomography (OCT) volume scans with deep learning (DL).

Methods: This study included 1121 OCT volume scans and 10-2 VFs from 289 eyes (257 patients). Macular scans were used to estimate 10-2 VF mean deviation (MD), threshold sensitivities (TS), and total deviation (TD) values at 68 locations. A three-dimensional (3D) convolutional neural network based on the 3D DenseNet121 architecture was used for prediction. We compared DL predictions to those from baseline linear models. We carried out 10-fold stratified cross-validation to optimize generalizability. The performance of the DL and baseline models was compared based on correlations between ground truth and predicted VF measures and mean absolute error (MAE; ground truth − predicted values).

Results: Average (SD) MD was −9.3 (7.7) dB. Average (SD) correlations between predicted and ground truth MD and MD MAE were 0.74 (0.09) and 3.5 (0.4) dB, respectively. Estimation accuracy deteriorated with worsening MD. Average (SD) Pearson correlations between predicted and ground truth TS and MAEs for DL and baseline model were 0.71 (0.05) and 0.52 (0.05) ($P < 0.001$) and 6.5 (0.6) and 7.5 (0.5) dB ($P < 0.001$), respectively. For TD, correlation (SD) and MAE (SD) for DL and baseline models were 0.69 (0.02) and 0.48 (0.05) ($P < 0.001$) and 6.1 (0.5) and 7.8 (0.5) dB ($P < 0.001$), respectively.

Conclusions: Macular OCT volume scans can be used to predict global central VF parameters with clinically relevant accuracy.

Translational Relevance: Macular OCT imaging may be used to confirm and supplement central VF findings using deep learning.

Introduction

Glaucoma is a major cause of visual disability and diminished quality of life and is the second leading cause of irreversible blindness worldwide.[1] The hallmark of glaucoma is progressive loss of retinal ganglion cells (RGCs) and their axons and supporting cells that project visual information to the central nervous system, resulting in progressive loss of visual function.[2–4] Up to 50% of RGC complement in humans is located in the macula.[5,6] Macular imaging with optical coherence tomography (OCT) has become the standard modality to

assess macular ganglion cell health in glaucoma patients.[7–9]

Moderately strong cross-sectional relationships between central macular thickness measurements and central functional measurements have been found in glaucoma eyes.[10–19] We have previously reported a weak to fair longitudinal structure–function (SF) relationship between macular OCT thickness changes and changes in central 10° visual field (VF) measurements.[12,13,16] One reason for this might be high variability of perimetric measurements in eyes with moderate to advanced glaucoma.[20–22] On the other hand, macular OCT thickness measurements have lower variability and high reproducibility.[19,23] Central 10-2 VF testing is a more demanding test compared to standard 24-2 VFs, as more locations (68 vs. 54) are examined; hence, patient performance may be suboptimal, especially in the elderly.[24,25]

Artificial intelligence is being increasingly explored in the field of ophthalmology, most frequently for detection of diabetic retinopathy and in glaucoma diagnostics.[26–31] Convolutional neural networks (CNNs) are frequently utilized as an efficient image analysis approach capable of analyzing big image databases.[32,33] Recent studies using CNNs have demonstrated potentially superior performance compared to traditional statistical methods for the detection of glaucoma.[28,30] Deep learning (DL) has also been implemented for the prediction of VF measures using structural modalities in glaucoma.[34–36] Prior SF models utilized topographical matching of structural and functional data and considered anatomical retinal ganglion cell displacement to further enhance the SF relationship.[16,18,37] However, given the ability of CNNs to learn patterns within the input data, providing this additional information to the DL network may not be necessary. One study used circular optic nerve head B-scans, without additional information, for predicting 24-2 VF threshold sensitivities with good predictive performance.[38]

The purpose of this study was to design and validate a DL model to predict central 10° VF mean deviation (MD), threshold sensitivity (TS), and total deviation (TD) measurements at individual locations based on macular OCT volume scans in a cohort of eyes with a wide range of glaucoma severity.

Methods

Eyes from the Stein Eye Institute's clinical and research databases meeting the inclusion criteria were enrolled. The current study was carried out in accor-
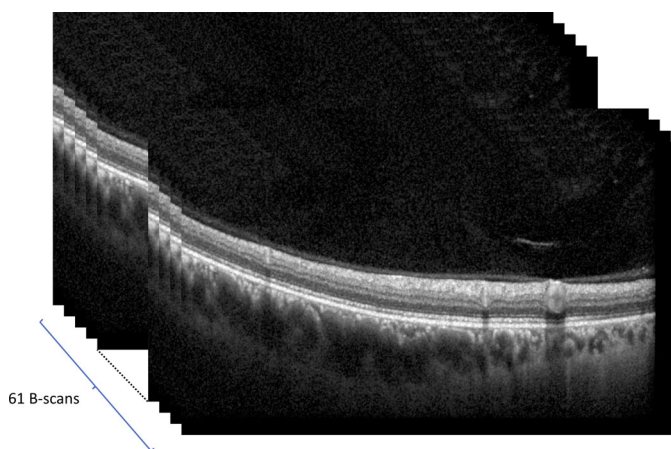

Figure 1. The macular volume scan from the SPECTRALIS OCT consisting of 61 raw macular B-scans was used as the input for the convolutional neural network. The 61 B-scans were concatenated into a 3D image volume. The 3D image volumes were resized and interpolated to 192 × 224 × 224 pixels (corresponding to depth, height, and weight, respectively).

dance with the tenets of the Declaration of Helsinki and the Health Insurance Portability and Accountability Act, and the protocols were approved by the Human Research Protection Program at the University of California Los Angeles (NCT01742819) and the Institutional Review Board (19-000953). The findings were reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement checklist. Patients had a diagnosis of glaucoma and were required to have had at least one visit with good-quality macular OCT and reliable central 10° VF performed within 6 months.

OCT Imaging

SPECTRALIS OCT (Heidelberg Engineering, Heidelberg, Germany) acquires a 30° × 25° macular volume scan aligned to the fovea–Bruch's membrane opening. The volume scans were required to have a quality factor > 15 and no major artifacts or confounding macular pathologies. The 61 B-scans from each volume scan were used as input for the CNN (Fig. 1). The macular volume scans were then assembled by concatenating along the B-scan axis (Fig. 1). Left eyes were flipped along vertical axis to match the right eye format, removing directional variability. Three-dimensional (3D) image volumes were resized and interpolated to 192 × 224 × 224 pixels (corresponding to number of scans, scan depth, and scan width, respectively). We used the area method for interpolation for both downsampling and upsampling during the resizing process. We found on initial exploratory analyses that there was an increase in validation

performance when we interpolated the first dimension (corresponding to number of scans) to 192.

Visual Field Measurements

The 10-2 testing pattern of the Humphrey Field Analyzer (Carl Zeiss Meditec, Dublin, CA) evaluates the central 10° of the VF. Threshold sensitivity is measured at 68 locations within a 10° radius from the fixation point. There is 2° spacing between the 68 locations in vertical and horizontal directions. The TD represents the difference between measured sensitivity at a given test location in an individual eye and the sensitivity in the age-matched normative database. The standard Swedish interactive thresholding algorithm (SITA) was used for all of the VF tests. Tests with false-positive rates > 15% were excluded. Main outcomes of interest were the predicted VF MD and the TS and TD values at 68 locations. We also considered flooring all TD values below −10 dB to −10 dB and carried out a repeat round of TD predictions and generated metrics based on the floored data.[12]

Network Architecture

We implemented a novel CNN that utilizes 3D OCT volume scans as data input to predict the VF measurements. We processed input data with a 3D CNN based on a 3D DenseNet121 with an encoder–decoder architecture. In the encoder, salient 3D visual features were generated from volume scans using 3D convolutional blocks with dimensions of $1024 \times 6 \times 7 \times 7$. In the decoder, VF measurements were generated from the visual features with fully connected neural networks. For the decoder, unlike the standard 3D DenseNet121, rather than utilizing dimensionality reduction of convolutional features based on channel dimension using adaptive pooling to generate $1024 \times 1 \times 1 \times 1$ volumes, we performed dimensionality reduction using 1×1 convolutional layers to generate $5 \times 6 \times 7 \times 6$ volumes; this was done to increase spatial information from the downsampled input volume and capture spatial relations in the VF measurements. A fully connected layer was then applied to the features to generate predicted VF measures (Supplementary Fig. S1). We trained separate independent models for estimating MD, TS, and unfloored and floored TD.

Model Training and Evaluation

The model was trained utilizing the mean squared error (MSE) between predicted and ground truth mean and total deviations as the main metric. Adam Optimizer was used with a learning rate of 1e-4, batch

size of 4, and a drop-out rate of 0.25 on the final fully connected layer. While training the CNN, random 90° rotations were applied with a 20% probability with an axis of rotation passing through the origin with direction (1, 0, 0) or, equivalently, a vector running lengthwise to the macular surface. Models were trained for 70 epochs, and the best checkpoint on the validation set was saved. Five percent of the training set was used for validation during training.

Model performance was assessed based on the mean absolute error (MAE) and Pearson correlation coefficient (r). These metrics were estimated between the ground truth, measured MD, TS, and TD, and the predicted values. We performed 10-fold stratified cross-validation to ensure that each fold had a similar distribution of ground truth VF values.[39] The training and validation splits did not share patients.

To generate model predictions, we performed test-time augmentation with 90° rotations randomly applied to the height and width dimensions of 3D input volume scans.[40] Additionally, we generated occlusion maps for each VF location. To create occlusion maps, we generated model predictions by sliding a mean-occlusion volume (with full input depth) along the height and width of the input volume and calculated differences from model predictions without occlusion to obtain prediction sensitivity at each location in the two-dimensional (2D) plane. We also explored the correlation of TS/TD at each location with the TS/TD at the remaining 68 locations for both ground truth and predicted TSs/TDs.

Two-Dimensional CNN Using Macular Thickness Maps

Finally, we investigated how 2D thickness maps would predict central VF measurements compared to macular B-scans. For this purpose, we designed a 2D CNN with a DenseNet backbone that used the ganglion cell layer (GCL) thickness maps as the input. We used the same 10-fold cross-validation with a similar train/test split as the 3D CNN model using the macular volume scans. This model was used to predict the central VF MD and the TS and TD values for all 68 VF locations. We compared Pearson correlation coefficients and MAEs from the DL model using the GCL thickness maps to the model using the macular volume scans with a paired *t*-test.

Baseline Model

We fit a linear model in log–log units (base model) in order to predict the MD and the TD and TS values for

each central VF location from ganglion cell/inner plexiform layer (GCIPL) thickness measurements. GCIPL thickness measurements from the central 24° × 24° region of macular volume scans (8 × 8 superpixel matrices) were exported and summed to calculate GCIPL thickness measurements in 64 superpixels. The GCIPL thickness measurements at superpixels were matched with VF locations after adjusting for perifoveal RGC displacements, as proposed by Drasdo et al.[37]

$$Y = a + b * \log(GCIPL) + error$$

where Y = TS or TD at each VF location, a = intercept, and b = rates of sensitivity change per 1 log10 unit decrease or increase in the GCIPL.

For estimating the correlation between the GCIPL and MD, we averaged the GCIPL thickness measurements to a global value and applied the same formula for estimating the correlation between GCIPL global measurements and the central VF MD. A paired *t*-test was used to compare the Pearson correlation coefficient and MAE between the base model and the DL model.

Results

The dataset included 1121 pairs of macular OCT and central 10° VF from 289 eyes (257 patients). Table 1 provides the clinical and demographic characteristics of study eyes. The mean (SD) age of the study patients was 68.3 (12.1) years. The average (SD) MD for the 10-2 and 24-2 VFs were −9.3 (7.7) and −9.5 (8.4) dB, respectively. The median (interquartile range [IQR]) number of testing sessions was 1 (1–8). The proportions of eyes with mild (24-2 MD > –6 dB), moderate (MD between −12 and −6 dB) and severe (MD < –12dB) glaucoma were 44%, 17%, and 39%, respectively.

Table 2 compares performance of the final DL and the base models (BMs). The mean (SD) correlations

Table 1. Demographic and Clinical Characteristics of the Study Eyes (257 Patients, 289 Eyes)

Characteristic	
Age (y), mean ± SD	68.3 ± 12.1
Gender, female/male, *n* (%)	173/116 (60/40)
Race, *n* (%)	
Caucasian	144 (50)
African American	47 (16)
Asian	42 (15)
Hispanic	14 (5)
Unknown	47 (14)
Glaucoma severity, *n* (%)	
Mild	127 (44)
Moderate	49 (17)
Severe	113 (39)
Number of visits, median (IQR)	1 (1–8)
10-2 VF MD	
Mean ± SD (dB)	−9.3 ± 7.7
Range	−31.4 to 1.2
24-2 VF MD	
Mean ± SD (dB)	−9.5 ± 8.4
Range	−31.6 to 2.15
Baseline GCIPL thickness (µm), mean ± SD	47.6 ± 6.7

and MAEs between the predicted and ground truth MD for the DL and base models were 0.74 (0.09) and 0.42 (0.17) and 3.5 (0.4) and 4.6 (0.6) dB, respectively. The mean (SD) correlations and MAEs between the predicted and ground truth TS were 0.71 (0.05) and 6.5 (0.6) dB for the DL algorithm, respectively, and were significantly better than those of the BM (r = 0.52 [0.05], P < 0.001; MAE = 7.5 [0.5] dB, P < 0.001). The mean (SD) correlations and MAEs between the predicted and ground truth unfloored TD values for the DL and BS models were 0.69 (0.02) and 0.48 (0.05) (P < 0.001) and 6.1 (0.5) and 7.8 (0.5) dB

Table 2. Performance of the DL and Baseline Models for Prediction of Central VF MD and Unfloored and Floored TD and Threshold Sensitivity Measurements From Macular OCT Volume Scans

	Correlation Coefficient (SD)			Mean Absolute Error (SD), dB		
	Deep Learning Model	Baseline Model	*P*	Deep Learning Model	Baseline Model	*P*
Mean deviation	0.74 (0.09)	0.42 (0.17)	<0.001	3.59 (0.41)	4.68 (0.69)	<0.001
Threshold sensitivities[a]	0.71 (0.05)	0.52 (0.05)	<0.001	6.53 (0.62)	7.57 (0.50)	<0.001
Unfloored TD	0.69 (0.02)	0.48 (0.05)	<0.001	6.18 (0.52)	7.82 (0.51)	<0.001
Floored TD	0.68 (0.03)	0.54 (0.06)	<0.001	2.29 (0.18)	2.77 (0.19)	0.001

For floored measurements, TD values below −10 dB were converted to −10 dB.
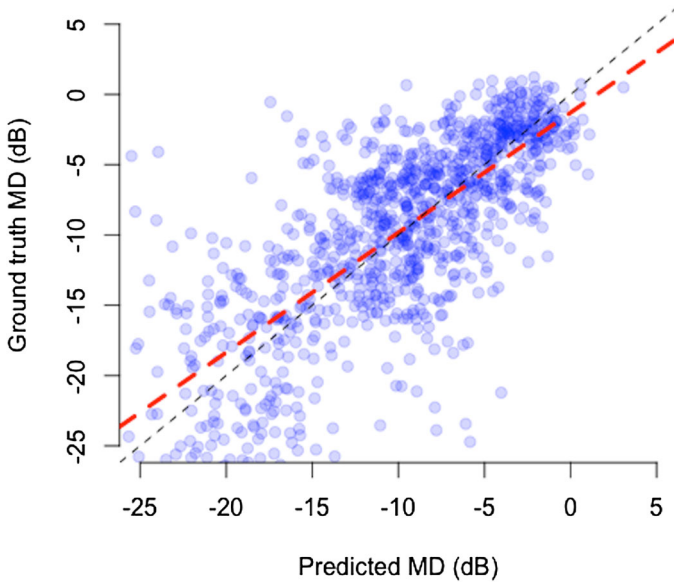[a]Only unfloored values.

Figure 2. Correlation between the predicted and observed central VF MD based on the deep learning algorithm developed in the study.

($P < 0.001$), respectively. Figure 2 displays the scatterplot of predicted against ground truth MD. The preci-

sion of the prediction varied as a function of glaucoma severity with less accurate predictions as the MD value diminished. Figure 3 provides boxplots for the ground truth and predicted TD and TS values at individual test locations for the DL and baseline models. For the DL model, the median ground truth and predicted values for the TD and TS values followed the same trend for the range of TD values. For very low (negative) actual TD values and low TS values, the IQRs for the predicted TD and TS tended to be low. In contrast, for the baseline model, the predictions were flat (floor values) for lower values of TD and TS.

Although there was no significant change in correlation coefficients when TD values were floored, the MAEs of VF predictions improved compared to unfloored results (Table 2). The correlations (SD) between predicted and ground truth TDs for the DL and base models were 0.68 (0.03) and 0.54 (0.06), respectively ($P < 0.001$), and the MAEs for the DL and base models were 2.2 (0.1) dB and 2.7 (0.1) dB, respectively ($P = 0.001$).

Figure 4 provides the occlusion maps highlighting regions on the macular OCT map where the model is applying the most attention for prediction of TS. The
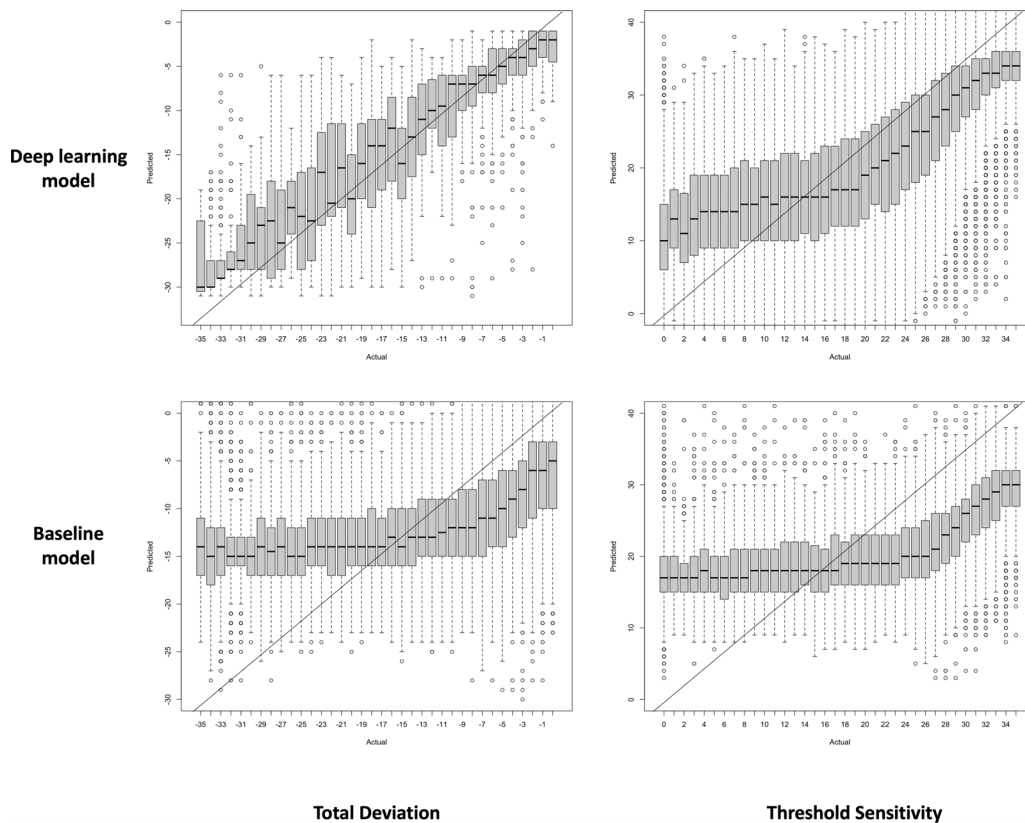

Total Deviation   **Threshold Sensitivity**

Figure 3. Boxplots demonstrate the distribution of predicted total deviation (TD) and threshold sensitivity (TS) values against ground truth for the deep learning and the baseline models. When the deep learning model was used, for most of the range of the ground truth values, the variability of predicted TS was relatively constant. On the other hand, the predicted TD variance (the IQR) tended to be lower on both ends of the TD range. For the baseline model, the predictions tend to have a floor distribution for lower TD and TS values.
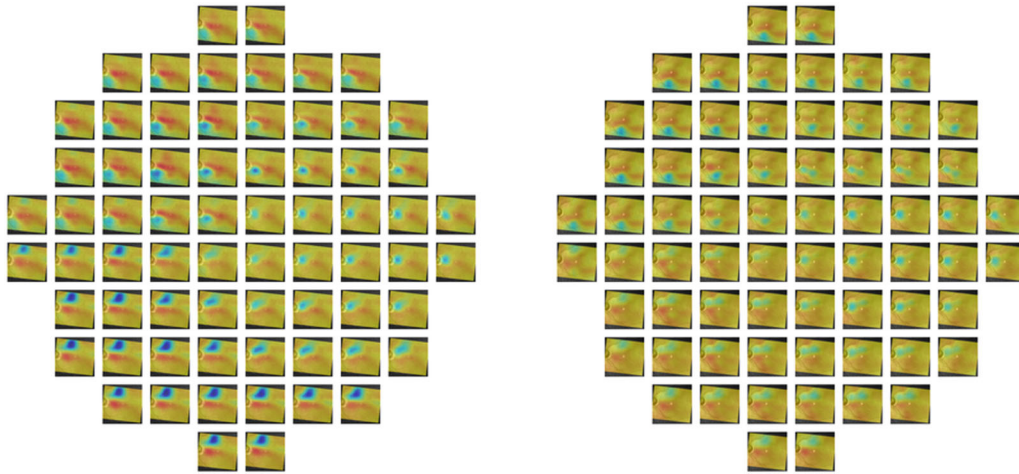
Figure 4. The occlusion maps generated for individual VF test locations. To create the occlusion maps, we generated model predictions by sliding an occlusion volume (with full input depth) along the height and width of the input volume and calculated the differences from the model predictions without occlusion. We then overlaid this on the infrared image of the OCT volume scan area. On the heatmap, the *green color* corresponds to zero contribution to the prediction, the *blue color* corresponds to a negative contribution to the prediction, and the *red color* corresponds to a positive contribution. Both *blue* and *red regions* are important for the prediction, but the *green regions* are not relevant. The regions with high prediction contribution respected the horizontal meridian and demonstrate correct SF mapping. The VF maps are in right-eye format.

regions with high prediction performance respected the horizontal meridian and demonstrated correct SF mapping in the two representative cases provided.

Figure 5 displays a compound matrix for the correlation of each of the 68 10-2 VF locations with the other 67 locations for both ground truth TD and TS (Figs. 5A, 5C) and predicted TD and TS (Figs. 5B, 5D) from the DL model. Two findings are notable: First, the correlation of each test location with the remaining locations is a function of the distance between the location of interest and other locations, and, second, between-location correlations respect the temporal horizontal meridian.

Supplementary Table S1 provides the prediction results of central VF measurements from 2D GCL thickness maps. For the MD, TS values, and unfloored and floored TD values, Pearson correlation coefficients were higher and the MAEs were lower for the DL model using the macular volume scans as the input compared to the DL model using the GCL thickness maps.

Discussion

We designed a CNN with 10-fold cross-validation to investigate the ability of macular OCT volume scans to predict global and pointwise central VF measures in a cohort of eyes with a wide range of glaucoma severity. We found strong correlations between the ground truth

and predicted VF parameters, ranging from 0.74 for MD to 0.71 and 0.69 on average for TS and TD values, respectively. Although estimated MAEs between actual and predicted pointwise TS and TD values were fairly high (6.5 and 6.1 dB, respectively, on average), those for the MD (3.5 dB) and floored TD (2.2 dB) were lower and more clinically relevant. Performance of the DL model for prediction of TS and TD values from macular volume scans was significantly better than a linear log–log model utilizing thickness measurements, a frequently used model for linking structural and functional measurements.

Our study is distinct from two points of view compared to prior studies implementing DL for the prediction of central VF measures from macular structural measurements. The first point is that we performed 10-fold cross-validation on the entire dataset in contrast to the frequently used approach of applying the algorithm to the testing subset only once; this substantially increased the stability and validity of the results and provided a more realistic perspective on how this algorithm will perform on other datasets.[39] Of the eyes in this study, 56% had moderate to severe glaucoma. Because we trained the DL model on the entire dataset with 10-fold cross-validation, the information from all of the eyes with more advanced glaucoma was used to make predictions. Therefore, the performance of our model would potentially be generalizable regardless of glaucoma severity. The second point is that our approach does not require segmentation of macular volume scans. This task can be
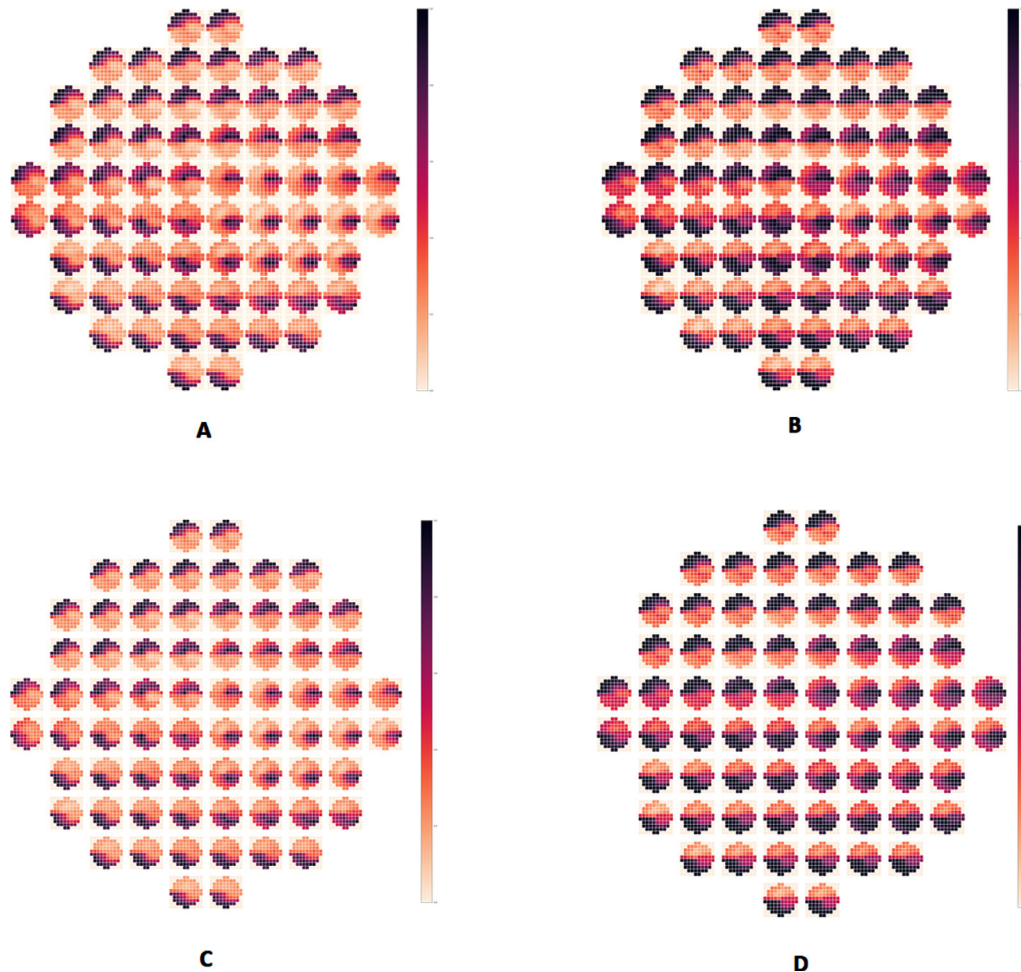
Figure 5. The correlation matrix visualizes the correlation of the ground truth (**A**) and predicted (**B**) total deviation (TD) values and ground truth (**C**) and predicted (**D**) threshold sensitivity at each 10-2 VF location with the other 67 test locations across the VFs. The correlation diminishes with increasing distance between the location of interest and the other VF locations across the VF. The correlations fall to be very low levels when the location of interest belong to the opposite hemifield except in the nasal region. This is consistent with the independence of the retinal ganglion cell damage across the temporal raphe.

challenging in eyes with suboptimal image quality or advanced glaucoma and often introduces significant measurement noise.

The implications of the current study are manyfold. Timely detection or confirmation of glaucoma progression is crucial to prevent further loss of vision; conversely, if disease stability is established based on structural measurements, functional testing may be needed less frequently, especially in eyes with moderate to advanced glaucoma. Structural measurements may also be weighted more in eyes with high long-term VF fluctuation.[21,25] Our algorithms will be helpful in establishing the functional significance of macular thickness changes. Similarly, clinicians may have to rely solely on structural tests for people unable to perform reliable VF testing. Because OCT imaging is faster, more efficient, and less costly than perimetry, reducing VF testing frequency

may lead to significant savings in clinic time and resources.

Our study utilized raw macular OCT volume scans to predict global and pointwise central VF measures; the performance of this model was significantly better than a base model consisting of an exponential fit between the GCIPL thickness measurements and VF measurements. In order to further demonstrate the effectiveness of the macular volume scans, we trained a separate CNN with GCL thickness maps and compared the results with those from the volume scans. For all of the central VF measures, the correlation coefficients were significantly higher and the MAEs were significantly lower for the DL model using the volume scans. We previously demonstrated that macular B-scans can predict the future course of structural damage in the macula using generative adversarial DL networks.[41] Prior studies have reported that

volumetric optic nerve head and macular OCT scans performed well for predicting global 24-2 VF parameters.[38,42,43] Kihara et al.[38] used infrared images and circular optic nerve B-scans for pointwise prediction of 24-2 VFs. They found that this multimodal approach had higher performance than using either modality individually. Yu and colleagues[44] investigated macular and optic nerve head volume scans for the prediction of 24-2 MDs; they reported a correlation coefficient of 0.87 between actual and predicted MD. Median MD was −3.4 dB in their study compared to −9.5 dB in our study. With advancing glaucoma, VF prediction accuracy diminishes; expansion of the data cloud around the line of best fit in eyes with lower MD in Figure 2 is consistent with worse MD predictions. The wider range of MD or TS and TD values to be predicted in our study likely made overall prediction more challenging.

Many studies evaluating SF relationships considered RGC displacement from the fovea as proposed by Drasdo et al.[16,18,19,37,45] Raza and coauthors[16] found improvements in SF correlations after applying this displacement correction. Other anatomic variations could affect SF relationships. Bedggood et al.[46,47] and others[48,49] investigated variations in horizontal raphe position. Variations in fovea–optic disc distance have been proposed as a potentially important factor.[50] We hypothesized that the DL model would be able to learn the influence of such anatomical variations from macular volume scans; therefore, this information was not separately included as prior information for training. The patterns observed on the occlusion maps demonstrated that the model was utilizing information from the expected regions of the macular volume scans to predict pointwise VF measures (Fig. 4). These findings suggest that a well-trained CNN has the ability to learn anatomical features of the macula from macular volume scans and, more importantly, can comprehend the correspondence of central VF locations with matching regions of the macula.

Although prediction results for floored TDs seemed to indicate that the performance of the DL model improved when TD values were fixed at −10 dB, this might be a function of decreased dynamic range for prediction, as the correlation between the predicted and actual TD values did not increase. For MD values below −10 dB, the model had lower prediction accuracy (Fig. 2), likely because macular thickness measurements do not show significant additional thinning when the corresponding VF locations have reached TD values of −8 to −10 dB.[12,16,51–53] An interesting finding of our study was that, based on Figure 3, the IQR of the predicted TD was lower for more very negative (severe) ground truth TD values. This suggests

the potential utility of the DL model in predicting TD for locations demonstrating severe glaucoma damage.

To verify the prediction ability of the model, we explored the correlation between the ground truth and predicted TS and TDs at individual central VF locations with all other 67 test locations (Fig. 5). Figure 5 illustrates that actual and predicted TS and TD values display very high correlations with adjacent locations; correlations decreased as the distance between test locations of interest increased. The locations immediately below or above the temporal horizontal meridian tended to be independent, consistent with the fact that glaucomatous damage respects the temporal horizontal raphe.[46,54] This finding provides additional evidence that our DL model properly learned anatomical correspondence between VF locations and macular OCT volume scans.

External validation of our model on a different dataset of glaucoma eyes is needed before it can be implemented. Given the 10-fold cross-validation on the entire dataset, we expect our algorithm to perform well with other datasets. Also, as only raw macular volume scans were used as input for the DL algorithm, it will be easily adaptable to other OCT devices. As the broken-stick model is most useful for exploring localized SF relationships, we did not carry out this model for predicting the 10-2 MDs. SF models using global measures do not show a measurement floor at all or not until the very late stages of glaucoma. In the baseline model, TS and TD were the independent variables; the model was then inverted to provide predicted TS and TD.

In conclusion, our proposed DL model with 10-fold cross-validation of the entire dataset predicted the central VF MD from macular OCT volume scans with clinically relevant performance and potentially high generalizability. Accurate prediction of local TS and TD values is more challenging for a variety of reasons. When the findings are validated in external datasets, the resulting algorithms may be used for confirming or predicting functional damage or its progression. The proposed DL model provides global estimates of the central field of vision based on raw, unaltered macular OCT B-scans with clinically relevant accuracy. It can potentially be used to confirm or predict VF progression in patients with glaucoma when the algorithm has been validated.

Acknowledgments

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. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol.* 1989;107(5):453–464.
3. Quigley HA, Nickells RW, Kerrigan LA, Pease ME, Thibault DJ, Zack DJ. Retinal ganglion cell death in experimental glaucoma and after axotomy occurs by apoptosis. *Invest Ophthalmol Vis Sci.* 1995;36(5):774–786.
4. Medeiros FA, Gracitelli CP, Boer ER, Weinreb RN, Zangwill LM, Rosen PN. Longitudinal changes in quality of life and rates of progressive visual field loss in glaucoma patients. *Ophthalmology.* 2015;122(2):293–301.
5. Curcio CA, Allen KA. Topography of ganglion cells in human retina. *J Comp Neurol.* 1990;300(1): 5–25.
6. Zhang C, Tatham AJ, Weinreb RN, et al. Relationship between ganglion cell layer thickness and estimated retinal ganglion cell counts in the glaucomatous macula. *Ophthalmology.* 2014;121(12):2371–2379.
7. Mohammadzadeh V, Fatehi N, Yarmohammadi A, et al. Macular imaging with optical coherence tomography in glaucoma. *Surv Ophthalmol.* 2020;65(6):597–638.
8. Medeiros FA, Zangwill LM, Bowd C, Vessani RM, Susanna R, Weinreb RN. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. *Am J Ophthalmol.* 2005;139(1):44–55.
9. Leung CK, Chan WM, Yung WH, et al. Comparison of macular and peripapillary measurements for the detection of glaucoma: an optical coherence tomography study. *Ophthalmology.* 2005;112(3):391–400.
10. Pollet-Villard F, Chiquet C, Romanet JP, Noel C, Aptel F. 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.
11. Kim NR, Lee ES, Seong GJ, Kim JH, An HG, Kim CY. Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT in glaucoma. *Invest Ophthalmol Vis Sci.* 2010;51(9):4646–4651.
12. Miraftabi A, Amini N, Morales E, et al. Macular SD-OCT outcome measures: comparison of local structure-function relationships and dynamic range. *Invest Ophthalmol Vis Sci.* 2016;57(11): 4815–4823.
13. 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.

14. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res.* 2007;26(6): 688–710.

15. Hood DC, Raza AS, de Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res.* 2013;32:1–21.

16. 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.

17. Suda K, Hangai M, Akagi T, et al. Comparison of longitudinal changes in functional and structural measures for evaluating progression of glaucomatous optic neuropathy. *Invest Ophthalmol Vis Sci.* 2015;56(9):5477–5484.

18. Mohammadzadeh V, Rabiolo A, Fu Q, et al. Longitudinal macular structure–function relationships in glaucoma. *Ophthalmology.* 2020;127(7):888–900.

19. Nouri-Mahdavi K, Fatehi N, Caprioli J. Longitudinal macular structure-function relationships in glaucoma and their sources of variability. *Am J Ophthalmol.* 2019;207:18–36.

20. Wall M, Doyle CK, Zamba KD, Artes P, Johnson CA. The repeatability of mean defect with size III and size V standard automated perimetry. *Invest Ophthalmol Vis Sci.* 2013;54(2):1345–1351.

21. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol.* 1989;108(2):130–135.

22. Wyatt HJ, Dul MW, Swanson WH. Variability of visual field measurements is correlated with the gradient of visual sensitivity. *Vision Res.* 2007;47(7):925–936.

23. Miraftabi A, Amini N, Gornbein J, et al. Local variability of macular thickness measurements with SD-OCT and influencing factors. *Transl Vis Sci Technol.* 2016;5(4):5.

24. Rabiolo A, Morales E, Kim JH, et al. Predictors of long-term visual field fluctuation in glaucoma patients. *Ophthalmology.* 2020;127(6):739–747.

25. Wu Z, Medeiros FA, Weinreb RN, Zangwill LM. Performance of the 10-2 and 24-2 visual field tests for detecting central visual field abnormalities in glaucoma. *Am J Ophthalmol.* 2018;196:10–17.

26. Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* 2016;57(13):5200–5206.

27. Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep.* 2018;8(1):14665.

28. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology.* 2018;125(8):1199–1206.

29. 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.

30. Asaoka R, Murata H, Hirasawa K, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol.* 2019;198:136–145.

31. Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA. Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans. *JAMA Ophthalmology.* 2020; 138(4):333–339.

32. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Barlett P, Pereira FCN, Burges JC, Bouttou L, Weinberger KQ, eds. *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Red Hook, NY: Curran Associates; 2012.

33. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data.* 2015;2(1):2–21.

34. 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.

35. 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. *Transl Vis Sci Technol.* 2021;10(13):28.

36. Asaoka R, Xu L, Murata H, et al. A joint multitask learning model for cross-sectional and longitudinal predictions of visual field using OCT. *Ophthalmol Sci.* 2021;1(4):100055.

37. Drasdo N, Millican CL, Katholi CR, Curcio CA. The length of Henle fibers in the human retina and a model of ganglion receptive field density in the visual field. *Vision Res.* 2007;47(22):2901–2911.

38. Kihara Y, Montesano G, Chen A, et al. Policy-driven, multimodal deep learning for predicting visual fields from the optic disc and OCT imaging. *Ophthalmology.* 2022;129(7):781–791.

39. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)* (pp. 1137–1145). San Francisco, CA: Morgan Kaufmann; 1995.

40. Moshkov N, Mathe B, Kertesz-Farkas A, Hollandi R, Horvath P. Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Sci Rep.* 2020;10(1):5068.

41. Hassan ON, Sahin S, Mohammadzadeh V, et al. Conditional GAN for prediction of glaucoma progression with macular optical coherence tomography. In: Bebis G, et al., eds. *Advances in Visual Computing, ISVC 2020. Lecture Notes in Computer Science* (Vol. 12510, pp. 761–772). Cham: Springer; 2020.

42. 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 J Biomed Health Inform.* 2020;24(12):3421–3430.

43. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R. Inference of visual field test performance from OCT volumes using deep learning. arXiv. 2019, arXiv:190801428.

44. Yu HH, 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.* 2021;4(1):102–112.

45. Hirasawa K, Matsuura M, Fujino Y, et al. Comparing structure-function relationships based on Drasdo's and Sjöstrand's retinal ganglion cell displacement models. *Invest Ophthalmol Vis Sci.* 2020;61(4):10.

46. Bedggood P, Nguyen B, Lakkis G, Turpin A, McKendrick AM. Orientation of the temporal nerve fiber raphe in healthy and in glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2017;58(10):4211–4217.

47. Bedggood P, Tanabe F, McKendrick AM, Turpin A. Automatic identification of the temporal retinal nerve fiber raphe from macular cube data. *Biomed Opt Express.* 2016;7(10):4043–4053.

48. Chauhan BC, Sharpe GP, Hutchison DM. Imaging of the temporal raphe with optical coherence tomography. *Ophthalmology.* 2014;121(11):2287–2288.

49. Ghassabi Z, Nguyen AH, Amini N, Henry S, Caprioli J, Nouri-Mahdavi K. The fovea-BMO axis angle and macular thickness vertical asymmetry across the temporal raphe. *J Glaucoma.* 2018;27(11):993–998.

50. Qiu K, Chen B, Yang J, et al. Effect of optic disc-fovea distance on the normative classifications of macular inner retinal layers as assessed with OCT in healthy subjects. *Br J Ophthalmol.* 2019;103(6):821–825.

51. Sung KR, Sun JH, Na JH, Lee JY, Lee Y. Progression detection capability of macular thickness in advanced glaucomatous eyes. *Ophthalmology.* 2012;119(2):308–313.

52. Bowd C, Zangwill LM, Weinreb RN, Medeiros FA, Belghith A. Estimating optical coherence tomography structural measurement floors to improve detection of progression in advanced glaucoma. *Am J Ophthalmol.* 2017;175:37–44.

53. Lavinsky F, Wu M, Schuman JS, et al. Can macula and optic nerve head parameters detect glaucoma progression in eyes with advanced circumpapillary retinal nerve fiber layer damage? *Ophthalmology.* 2018;125(12):1907–1912.

54. Sharifipour F, Morales E, Lee JW, et al. Vertical macular asymmetry measures derived from SD-OCT for detection of early glaucoma. *Invest Ophthalmol Vis Sci.* 2017;58(10):4310–4317.