Journal Pre-proof

Deep Learning Estimation of 10-2 Visual Field Map Based on Macular Optical Coherence Tomography Angiography Measurements

Golnoush Mahmoudinezhad MD, MPH , Sasan Moghimi MD , Jiacheng Cheng MS , Liyang Ru BS , Dongchen Yang MS , Kushagra Agrawal MS , Rajeev Dixit MS , Siavash Beheshtaein PhD , Kelvin H. Du BS , Kareem Latif BS , Gopikasree Gunasegaran MD , Eleonora Micheletti MD , Takashi Nishida MD, PhD , Alireza Kamalipour MD, MPH , Evan Walker MS , Mark Christopher PhD , Linda Zangwill PhD , Nuno Vasconcelos PhD , Robert N. Weinreb MD

PII: S0002-9394(23)00379-3
DOI: https://doi.org/10.1016/j.ajo.2023.09.014
Reference: AJOPHT 12679

To appear in: *American Journal of Ophthalmology*

Received date: March 29, 2023
Revised date: September 7, 2023
Accepted date: September 13, 2023

Please cite this article as: Golnoush Mahmoudinezhad MD, MPH , Sasan Moghimi MD , Jiacheng Cheng MS , Liyang Ru BS , Dongchen Yang MS , Kushagra Agrawal MS , Rajeev Dixit MS , Siavash Beheshtaein PhD , Kelvin H. Du BS , Kareem Latif BS , Gopikasree Gunasegaran MD , Eleonora Micheletti MD , Takashi Nishida MD, PhD , Alireza Kamalipour MD, MPH , Evan Walker MS , Mark Christopher PhD , Linda Zangwill PhD , Nuno Vasconcelos PhD , Robert N. Weinreb MD , Deep Learning Estimation of 10-2 Visual Field Map Based on Macular Optical Coherence Tomography Angiography Measurements, *American Journal of Ophthalmology* (2023), doi: https://doi.org/10.1016/j.ajo.2023.09.014

Highlights

- DL models enable the estimation of VF loss from OCTA images with high accuracy.
- Applying DL to the OCTA images may enhance clinical decision making.
- It may improve risk stratification of patients at risk for central VF damage.

Deep Learning Estimation of 10-2 Visual Field Map Based on Macular Optical Coherence Tomography Angiography Measurements

Golnoush Mahmoudinezhad, MD, MPH[1*]; Sasan Moghimi, MD[1*]; Jiacheng Cheng, MS[2]; Liyang Ru, BS[2]; Dongchen Yang, MS[3]; Kushagra Agrawal, MS[2]; Rajeev Dixit, MS[2]; Siavash Beheshtaein, PhD[4]; Kelvin H. Du, BS[1]; Kareem Latif, BS[1]; Gopikasree Gunasegaran, MD[1]; Eleonora Micheletti, MD[1]; Takashi Nishida, MD, PhD[1]; Alireza Kamalipour, MD, MPH[1]; Evan Walker, MS[1]; Mark Christopher, PhD[1]; Linda Zangwill, PhD[1]; Nuno Vasconcelos, PhD[2]; Robert N. Weinreb, MD[1]

[1]Hamilton Glaucoma Center, Shiley Eye Institute, Viterbi Family Department of Ophthalmology, UC San Diego, La Jolla, CA, United States.

[2]Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, United States.

[3]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, United States

[4]L3Harris Technologies, Torrance, CA, United States.

*These authors had equal contributions as co-first authors.

Running Title: Estimating 10-2 Visual Field Map from Macula OCTA Vessel Density

Correspondence: Robert N. Weinreb, MD, Shiley Eye Institute, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0946. E-mail: rweinreb@ucsd.edu

Meeting presentations: ARVO 2023

Abstract

Purpose: To develop deep learning (DL) models estimating the central visual field (VF) from optical coherence tomography angiography (OCTA) vessel density (VD) measurements.

Design: Development and validation of a Deep Learning model

Methods:A total of 105110-2 VF-OCTA pairs from healthy, glaucoma suspects, and glaucoma eyes were included. DL models were trained on enface macula VD images from OCTA to estimate 10-2 mean deviation (MD), pattern standard deviation (PSD), 68 total deviation (TD) and pattern deviation (PD) values and compared with a linear regression (LR) model with the same input. Accuracy of the models was evaluated by calculating the average mean absolute error (MAE) and the R^2 (Squared Pearson correlation coefficients) of the estimated and actual VF values.

Results: DL models predicting 10-2 MD achieved R^2 of 0.85 (95% confidence interval [CI], 74–0.92) for 10-2 MD and MAEs of 1.76 dB (95% CI, 1.39–2.17 dB) for MD. This was significantly better than mean linear estimates for 10-2 MD. The DL model outperformed the LR model for the estimation of pointwise TD values with an

average MAE of 2.48 dB (95% CI: 1.99, 3.02) and R^2 of 0.69 (95% CI: 0.57, 0.76) over all test points. The DL model outperformed the LR model for the estimation of all sectors.

Conclusions:DL models enable the estimation of VF loss from OCTA images with high accuracy. Applying DL to the OCTA images may enhance clinical decision making. It also may improve individualized patient care and risk stratification of patients who are at risk for central VF damage.

Keywords: Glaucoma, Optical Coherence Tomography Angiography, 10-2, Visual Field, Artificial Intelligence, Deep Learning.

Introduction

Glaucoma is a progressive disease associated with characteristic structural changes to the eye and loss of visual function.[1] Early diagnosis and monitoring of disease progression are crucial to prevent significant loss of vision and blindness.[1,2]

Visual field (VF) testing is among the most important tool for the diagnosis and monitoring of glaucoma. VF results are crucial for estimating the current level and future risk of functional impairment for individuals affected by this disease.[2,3] Although VF damage in glaucoma predominantly affects the peripheral field in most patients, glaucomatous damage to the central field has been found to occur more often than expected, even in the early stages of the disease. More than

50% of eyes with mild to moderate glaucoma had defects within the central ±3°.[3,4] Moreover, the 24-2 VF test misses central macular damage confirmed by the 10-2 VF test.[5] In addition, the central VF is an important predictor of vision-related quality of life.[6,7] Therefore, being able to detect glaucomatous central VF damage effectively is important in the clinical management of glaucoma patients.

A recognized limitation with VF testing is the high test variability, which can make it difficult to detect true change over time.[8,9] A high variability can result in missed or late identification of glaucomatous progression, and this then delays interventions with consequent worse visual outcomes.[9-11] Moreover, it is time-consuming and costly to carry out a 10-2 VF test in addition to a central 24-degree VF test with sufficient frequency.[11]

Optical Coherence Tomography Angiography (OCTA) is a noninvasive optical imaging technology that provides information about retinal vasculature in the form of vessel density (VD) measurements.[12] Prior studies have shown that OCTA VD is lower in glaucoma and these measurements are associated with VF mean deviation (MD).[13,14] When performed with OCT, OCTA enhances early diagnosis, detection of progression, and risk assessment of glaucoma.[15,16] Moreover, it is notable that OCTA measurements continue to provide valuable information even when OCT measurements have reached their lower limit.[17] The longitudinal correlation between macular VD with 10–2 MD has been shown to be higher than that of macular ganglion cell complex and 10-2 MD.[18] Macular OCTA has good reproducibility and repeatability over time.[19-21] It could therefore be beneficial to estimate VF sensitivity in the central 10° from OCTA. While DL techniques have shown promise in improving the correlation between ganglion cell measurements obtained from OCT and central visual field damage,[22] there is still limited information regarding the

relationship between superficial macula microvasculature damage and visual function.

Unlike conventional statistical tools, deep learning (DL) does not make assumptions to adequately model relevant structure-function relationships. Prior studies have focused on Artificial Intelligence (AI) algorithms for evaluating 24-2 or 10-2 VF metrics using SD-OCT measurements.[22-29] The aim of the present study was to develop and validate a DL model to estimate the 10-2 global metrics and pointwise (total deviation (TD) and pattern deviation (PD)) values from macular OCTA scans.

Methods

The cohort included healthy, glaucoma suspect, and glaucoma participants. These participants were followed up with semiannual visits, which included OCTA imaging and VF testing. Study participants were selected from participants in the Diagnostic Innovations in Glaucoma Study (DIGS) at the University of California, San Diego (UCSD) and from the medical records of patients examined in the Glaucoma Clinic (GC) at the UCSD Shiley Eye Institute. The DIGS is an ongoing prospective, longitudinal study conducted at the Hamilton Glaucoma Center, UCSD. Details of the DIGS protocol have been described elsewhere.[30] All methods adhered to the tenets of the Declaration of Helsinki and the Health Insurance Portability and Accountability Act and were approved by the UCSD Institutional Review Board. Written informed consent was obtained from all subjects in DIGS and GC patients.

To be included in the study, participants had to meet the following criteria at study entry: 20/40 or better best-corrected visual acuity, at least 2 consecutive reliable standard automated perimetry VF tests, and intraocular pressure of less than 22 mmHg for healthy participants.[22] 24-2 VF tests with reliable results (false Positive ≤15%; false negative ≤33%; and fixation loss ≤33%) were included. For this analysis, inclusion in the glaucoma group required repeatable, abnormal 24-2 VF results, defined as a glaucoma hemifield test (GHT) outside normal limits or a pattern standard deviation (PSD) with a p-value <0.05. Inclusion in the glaucoma suspect group required the presence of visible glaucomatous optic neuropathy on fundus photographs without repeated, abnormal VF results. The healthy group consisted of participants without detectable glaucomatous optic neuropathy on fundus photography or optic nerve head examination, or repeatable abnormal 24-2 VF results.

OCTA Images

The commercially available Avanti Angiovue (Optovue Inc. Fremont, California) combines OCTA and OCT imaging in a single system (software version 2018.1.1.63). The Avanti system for measuring VD and tissue thickness has been described previously.[31]

For macular OCTA images, VD at the superficial macular slab was obtained from 6- × 6-mm^2 OCTA images comprised of 304 × 304 A-scans centered on the fovea. OCTA VD was calculated from the same scan slab, as follows. The Avanti AngioVue OCTA system uses the split-spectrum amplitude-decorrelation angiography method to capture the dynamic motion of the red blood cells and provide a high-resolution 3-dimensional visualization of perfused retinal vasculature at various user-defined layers of the retina at the capillary level.[31] VD was calculated

as the percent area occupied by flowing blood vessels in the selected region defined as pixels having decorrelation values above a set threshold level. The retinal layers of each scan were segmented automatically by the AngioVue software to visualize the superficial retinal capillary plexuses in a slab from the internal limiting membrane to 10μm above the inner plexiform layer. For this study, whole en face image VD (wiVD) was derived from the entire 6×6-mm^2 scan, and perifoveal VD was measured in an annular region centered on the fovea with an inner diameter of 1 mm and outer diameter of 6 mm.

OCTA image quality review was completed according to University of California San Diego Imaging Data Evaluation and Analysis reading center standard protocol. Segmentation errors were manually corrected. Images with a quality index (QI) of <4, poor clarity, residual motion artifacts visible as irregular vessel patterns, or disc boundaries on the enface angiogram, image cropping, or local weak signal due to vitreous opacity, or segmentation errors that could not be corrected were excluded.

VF testing was performed at each visit for all participants using 10-2 testing patterns and the Swedish interactive thresholding algorithm standard protocol (Humphrey Field Analyzer II; Carl Zeiss Meditec, Inc). 10-2 VF test results that showed more than 33% fixation losses, 33% false-negative errors, or 15% false-positive errors were excluded. VF results were processed and evaluated for quality according to standard protocols by the University of California, San Diego, Visual Field Assessment Center. Quantitative global VF metrics including mean deviation (MD) and pattern standard deviation (PSD) along with total deviation (TD) and pattern deviation (PD) values at individual 68 test locations were included in the analysis.

For each eye, the OCTA images were paired with the 10-2 test results acquired closest in time and within 180 days. This resulted in a set of 1051 OCTA and 10-2 VF pairs. These datasets were used to train and evaluate all models described herein.

Training the Deep Learning Models

The DL architecture used in this study was ResNet50.[32] A transfer learning approach was adopted by initializing model weights by training on a large, general image recognition dataset (ImageNet).[33] In order to match image dimensions expected by the pretrained network, macula VD image were resized to a size of 224 × 224 pixels, and pixel values were replicated in each red-green-blue color channel. Model weights were then fine-tuned on training dataset of OCTA VF pairs. DL regression neural networks were constructed to estimate quantitative VF metrics (MD, PSD, TD, and PD) for 10-2 results based on VDs. Notably, models were provided with only OCTA enface image, and no labels indicating disease status or severity. The linear regression (LR) model was fitted using the training dataset of OCTA VF pairs. Both the DL and the LR models were evaluated on testing data.

To construct independent datasets for training, validation, and testing, the dataset was divided randomly by participant in an 80%–10%–10% split. The training, validation, and testing datasets were separated by participant instead of image in order to avoid patient scans being included in both the training and validation or test set. Augmented data were generated by applying random geometric transformations to the VD, including random horizontal flipping as well as random rotations, random cropping, and scaling. The models were trained for 100 epochs with a batch size of 32 (except for pointwise prediction that we used a batch size of 64) and a learning

rate of 0.001 using optimized by the Adam optimizer.[34] The Rectified Linear Unit (ReLU) was used as activation function for ResNets. To preserve the spatial information encoded in the feature maps, the final global average pooling layer was replaced by a fully-connected layer in the sector-wise and point-wise models. The final model was selected after periodic evaluation of the validation set. Notably, no clinical assumptions were explicitly included in the model (e.g., superior hemiretina is associated with inferior hemiretina). Therefore, no previous clinical knowledge was used to derive the topographic relationship. DL training and testing were performed using Pytorch libraries.[35]

Evaluating Structure–Function Estimates

Testing data were used to evaluate the training models using the mean absolute error (MAE) and R^2 (squared Pearson correlation coefficient) metrics. LR models were also constructed to estimate VF metrics using OCTA images as a basis for comparison. LR is a straightforward and widely used method as a baseline for comparing with more complex models (i.e., DL) and numerous previous studies have used it for evaluation of the relationship between structure and function.[22,36,37] It also allows us to determine if the additional complexity of DL provides a significant improvement in predictive accuracy. The LR model was implemented using the Scikit-learn package.[38] Each 2D image was transformed into a 1D vector, and all 224*224 pixels were used as inputs for both global and pointwise predictions. The LR models were trained with the same data split as the DL models.

Separate analyses were performed, and principal component analysis (PCA)[39] was employed to reduce the dimensionality of the image, address the correlation between pixels, and extract features from en face OCTA images. For TD

and MD prediction, threshold sensitivity values in VF were converted from decibels (dB) to 1/Lambert values with the following formula: $1/Lambert=10^{dB/10}$. LR was applied to the extracted features and predicted VF metrics. Then, the results were converted to (dB) scale.

To summarize the sectoral performance of the models, we averaged the measured TD and PD values for the 10-2 clusters proposed by Hood et al., as demonstrated in Figure 1.[40] We then predicted sectoral VF sensitivity based on DL and LR.

Performance of different models was evaluated at 3 different levels: a) Global MAE/R^2 for MD and PSD, b) Pointwise and average MAE/R^2 for 68 VF points (TD and PD), and c) For each sector of the VF map (as demonstrated in Figure 1), the sectorial mean MAE/R^2 is averaged over the VF points within each sector.

Performance of different models at the global level was also evaluated in 4 different categories (Healthy, glaucoma suspect, mild glaucoma (24-2 MD≥-6), and moderate to advanced glaucoma (24-2 MD<-6))

Visualizing the Structure–Function Relationship

In order to identify which regions of the OCTA images contributed most to model decisions, we computed class activation maps (CAMs) using the trained DL models.[41] We used a version of Grad-CAM that extended original techniques to regression models.[42] These maps quantify the contribution of each pixel in the input images (layer VD maps) to the model prediction. The resulting heat map is superimposed on the OCTA image to visualize the retinal regions utilized by the model to perform its prediction. CAM images were generated for each of the OCTAs in the testing set.

Statistical Analysis

A bootstrap resampling procedure was used to estimate 95% confidence intervals (CIs) for MAE and squared Pearson correlation coefficient .[43] R^2 between predictions and measured values were compared with a test of equality of correlations.[44] The cluster of data for the study subject was used as the unit of resampling since multiple measurements of the same subject are likely to be correlated. In order to compare the MAEs of the estimated and the actual measurements between the DL and the LR models, a random-effects mixed model was used to account for multiple tests from the same participant.[45] *P*-values of less than 0.05 were considered statistically significant.

Results

A summary of the study population used to train and test models to estimate 10-2 VF results is provided (Table 1).

The dataset comprised 1051 pairs of 10-2 VF tests and OCTA scans of 384 separate individuals. A total of 616 individual eyes were included and the majority were diagnosed with glaucoma (n = 353, 57.3%). Glaucoma suspect eyes (n = 131, 21.3%), and healthy eyes (n =132, 21.4%) eyes were also included. The mean age (95% CI) in healthy, glaucoma suspects, and glaucoma groups was 59.5 (95% CI: 53.2, 65.8), 62.9 (95% CI: 59.7, 66.1), and 67.5 (95% CI: 0.55, 1.05) years old, respectively. The mean image quality in healthy, glaucoma suspects, and glaucoma groups was 7.5 (95% CI: 7.1, 7.9), 7.2 (95% CI: 6.8, 7.6), and 6.6 (95% CI: 6.3, 6.9), respectively.

Table 2 compares the performance of the DL and LR models for 10-2 MD and PSD estimation. The DL approach significantly (P < 0.001) outperformed the LR model, for both R^2 and MAE.

The mean (95% CI) MAE and R^2 between the DL predicted and ground truth MD were 1.76 (95% CI: 1.39, 2.17) dB and 0.85 (95% CI: 0.74, 0.92) dB which were significantly better than those of LR (MAE =3.36 (95% CI: 2.58, 4.22) dB, R^2=0.34 (95% CI: 0.18, 0.50), P<0.001). Similarly, for 10-2 PSD predictions, the MAE and R^2 were 0.79 dB (95% CI: 0.55, 1.05), and 0.89 (95% CI: 0.82, 0.95) for the DL model, and 2.57 dB (95% CI: 2.05, 3.10) dB, and 0.29 (95% CI: 0.14, 0.47) for the LR model, respectively, (P<0.001). Our separate analysis using PCA and non-dB values of VFs as ground truths showed similar results. DL outperformed LR in all the analyses in the prediction of MD and PSD (P<0.001), Supplementary Table 1.

The performance of DL and LR in each severity group (healthy, suspect, early, moderate to advanced) eyes are provided in Table 2. Our results showed that DL outperformed the prediction based on LR in all healthy, suspect, early, and moderate to advanced eyes (All P<0.04 for MAE and R^2) except for, R^2 in the prediction of MD (P =1.0), and PSD (P=0.60) in healthy eyes, MAE in the prediction of MD for glaucoma suspects eyes (P=0.16), and R^2 in the prediction of MD (P=0.28) and PSD (P=0.48) in glaucoma suspect eyes.

Similar results were found for point-wise comparisons (Figure 2 and supplementary Figure1). The DL model outperformed the LR model in pointwise prediction of TD measurements. Pointwise comparison of the predicted 68 TD values revealed an MAE of 2.48(95% CI: 1.99, 3.02) dB and the R^2 of 0.69(95% CI: 0.57, 0.76). Pointwise prediction based on DL were significantly better than those based

LR model, which achieved pointwise MAE =4.28 (95% CI: 3.46, 5.15) dB, and R^2=0.19 (95% CI: 0.09, 0.30), P < 0.001), (Figure1). Similarly, DL model achieved an MAE dB 2.08 (95% CI: 1.61, 2.63) and R^2 of 0.67 (95% CI: 0.54, 0.76) in estimating PD points. This was significantly better than pointwise PD estimation based on LR model (3.73 dB (95% CI: 2.96, 4.57) dB, and 0.12 (95% CI: 0.05, 0.21), P < 0.001), (Supplementary Figure1). Our separate analysis using PCA and non-dB values of VFs as ground truths showed similar results. DL outperformed LR in all the analyses in the pointwise prediction of TDs and PDs measurements (P<0.001), Supplementary Table 1.

Figure 3 displays a matrix of correlations between pairs of predicted TD values at different 10-2 VF test locations. Each subplot shows the correlation between the TD prediction at one location and those at all other locations. This figure illustrates that the predicted TD values display are highly correlated for adjacent locations with the correlation decreasing as the distance between the test locations of interest increases.

Table 3 illustrates the predictive performance of the DL and LR models over the different sectors of the 10-2 VF map defined in Figure 1 based on mean TD values. DL has significantly lower TD MAE and higher R^2 than LR for all sector predictions (All P<0.001, except for superior temporal band R^2: P=0.009). The DL TD MAE ranged from 1.45 (1.16, 1.79) dB (R^2= 0.82 (0.64, 0.91) in the inferior temporal sector to MAE 2.99 (2.21, 3.78)dB (R^2=0.75(0.58, 0.88)) in the superior nasal sector of 10-2 VF. Supplementary Table 2 illustrates the predictive performance of the DL and LR models over different sectors of the 10-2 VF map for mean PD values. The DL PD MAE ranged from MAE of 1.35 (0.94, 1.83)dB (R^2=0.65 (0.40, 0.87)) in the inferior temporal sector to 2.91 (2.31, 3.59)dB (R^2=0.81 (0.67, 0.90) in the superior

temporal sector of 10-2 VF. Similarly, DL has significantly lower PD MAE and higher R^2 than LR for prediction of 10-2 sectors for all sectors (All P<0.001 except for superior temporal band R^2: P=0.02). Our separate analysis using PCA and non-dB values of VFs as ground truths showed similar results. DL outperformed LR in all the analyses in the prediction of sectors based on TDs and PDs measurements (P<0.02 and P<0.04, respectively), Supplementary Table 3 and 4.

The scatter plots of Figure 4 illustrate the relationship between the estimated and ground truth VF values (MD and PSD) for both the DL and LR models. The DL estimates are in much closer agreement with the groundtruth than those of LR.

In addition, Supplementary Figure 2 shows the predictive performance of the DL model across the continuum of actual 10-2 TD measurements."

To help understand how the DL models estimate VF metrics, we applied visualization techniques to highlight informative image regions. The resulting images are shown along with corresponding VD and VF groundtruth in Figure 5. The activation maps shows that the model relies on the anatomically correct regions to make its prediction. Moreover, supplementary Figure 3 shows the two cases where DL did not perform well in the prediction of MD and PSD.

Discussion

Given the impact of 10-2 on assessment of quality of life[46] and central damage in early glaucoma patients[47], accurate methods for estimating central VF are important. In the present study, DL approaches estimated VF MD and PSD using macula-centered OCTA scans with a high accuracy and outperformed estimates based on LR mean VD. DL approaches also outperformed LR for prediction of individual test locations of 10-2 VF map. Moreover, in the sector-wise analysis, the

MAEs ranged from 1.45 (1.16, 1.79)dB to 2.99 (2.21, 3.78)dB depending on the sectors and outperformed LR prediction.

Our DL model's accurate prediction of central VF from macular OCTA scans can assist clinicians in providing better individualized patient care for glaucoma. This is because estimation through DL is less prone to the natural variability of VF, which can be a challenge in assessing functional loss from structure.[48,49] Additionally, DL estimation can be performed with less burden on clinical settings in terms of time and cost, compared to obtaining a reliable central VF through 10-2 testing. Given extra amount of time required for performing a reliable central VF and also considering its variability, the proposed DL approach may decrease the number of VF exams needed for monitoring central damage in glaucoma. The impact of COVID-19 on our clinical practice,[50] further emphasizes the potential value of the DL approach in reducing the number of patient visits for disease monitoring. Therefore, using macular OCTA scans has the potential to reduce the need for central 10° VF testing.

In previous studies, a priori information on spatial information, and the form of the relationship between structural and functional measurements (e.g., linear or logarithmic) was used. For example, the displacement of the retinal ganglion cells (RGC) from the fovea was taken into account.[51-55] Raza et al. found improvements in structure function correlations after considering this displacement.[55] The variation in fovea-optic disc distance has also been proposed as a potentially important factor.[56] Notably, these studies reported the relationship between structure (i.e., OCT thickness) and function in glaucoma, with the strength of the association ranging from almost fair relationship to a strong structure–function correlation.[14,18,57-60] Nevertheless, the structure-function relationship in glaucoma is considered to be

nonlinear.[61-64] We hypothesized that the DL model could learn the influence of anatomical variations from the OCTA scans. Therefore, this information was not separately included as prior information during training. CAM visualization supports that the model was well trained to predict VF parameters based on corresponding regions of OCTA scans. DL models can also learn anatomical features of the macula from OCTA scans. More importantly, they can comprehend the correspondence of central VF locations with regions of the macula without assumptions about the nature of the structure function relationship.

Structural and functional damage caused by glaucoma typically follows distinct patterns, with correlations existing between specific locations of loss.[65-67] In Figure 3, we explore the correlation between the TD values at each test location of the central VF with all the other 67 test locations. The locations immediately below or above the temporal horizontal meridian appear to be independent, which is consistent with the fact that glaucomatous damage respects the temporal horizontal raphe.[68,69]

Several groups also have applied DL strategies to estimate 24-2 and 10-2 VF outcomes. Some previous studies used VF results as input to predict subsequent VF.[70,71] Other estimated VF parameters through various OCT inputs.[23,24,72-74] Christopher et al, also predicted 10-2 MD (MAE=1.9 dB, $R^{2=}$ 0.82) and 10-2 PSD (MAE=1.5 dB, $R^{2=}$ 0.69) from combination of all layers of macula thickness map scans.[23] Hashimoto et al., used a convolutional neural network (CNN) model to predict 10-2 VF from the thickness of the three macular layers and achieved an MAE of 2.84±2.98 dB ($R^{2=}$ 0.74) .[75] Kamalipour et al, also predicted 10-2 map with MAE of 2.88 dB(R^2= 0.55) for MD and MAE of 2.31 dB(R^2= 0.35) for PSD from circumpapillary spectral-domain optical coherence tomography (SD-OCT) RNFL

thickness measurements.[36] The reported test–retest variabilities in VF tests are between 1 and 2 dB (central area) and 4–6 dB (at 27°).[76,77] The R^2 and MAE values in the current study are better than previous published studies for MD (MAE:1.76 (1.39, 2.17), R^2 : 0.85 (0.74, 0.92)) and PSD (MAE: 0.79 (0.55, 1.05), R^2 : 0.89 (0.82, 0.95)). However, it is challenging to compare results across studies due to differences in study populations and disease severity. In each of these studies, the more severe disease the poorer the prediction.[23,24,36,72-74] Direct comparison between OCTA and OCT DL predictions of 10-2 VF using the same study population is needed to determine whether OCTA provides better prediction than OCT.

Previous studies have reported that OCTA may be particularly useful in eyes with more severe glaucoma as it continues to provide valuable information even when OCT measurements have reached their lower limit.[17] The benefits of DL over LR in Table 2 are particularly pronounced in Moderate/Advanced glaucoma eyes, with the MAE being reduced by more than half. This is despite the fact that functional measures are less reliable in severely damaged areas. Indeed, it appears that the DL-estimated measures are better for eyes with advanced glaucoma in Figure 4 than the measured 10-2. This suggests that the potential benefits of the DL technique with OCTA are highest when glaucoma is most severe.

There are few studies comparing the performance of combined OCT and OCTA metrics to individual OCT- and OCTA-based models,[16,59,78,79] with promising findings for diagnostic accuracy for glaucoma detection and the evaluation of structure-function relationship.[16,59,78,79] Also, Shoji and colleagues demonstrated that serial macular OCTA measurements were able to detect microvascular loss in glaucoma eyes without apparent evidence of ganglion cell complex (GCC) thickness alteration over an average follow-up duration of <14 months.[79] In addition, Hou et al

reported that in the POAG group, more than two-thirds of the eyes showed faster macular vessel density decrease than GCC thinning; faster macular vessel density decrease rate was associated significantly with worse glaucoma severity over 2.6 years of follow-up.[80] Moreover, it has recently been shown that longitudinal OCTA measurements may complement OCT-derived structural metrics for the evaluation of functional VF loss in patients with glaucoma.[16] Therefore, it seems that OCTA may be able to complement OCT metrics for central VF prediction.

The point-wise prediction is especially crucial when considering the future application of DL models in real-world clinical settings. A few studies estimated 68 points of VF based on OCT.[36,72,75,81,82] Asano et al., used a CNN model to estimate 10-2 test points from the combined OCT-measured macular retinal layers' thicknesses with MAE of 9.5 ± 9.4 and 9.4 ± 9.3 dB Using VGG and ResNet models.[81] Similarly, Kamalipour et al, also predicted 10-2 maps with an MAE of 4.04 dB from circumpapillary SD-OCT RNFL thickness measurements.[36] In our current study, we estimated 68 points of VF with an MAE of 2.48 (R^2=0.69) based on macula OCTA scans. Considering the inherent variability of individual VF test locations and the severity of VF damage at these locations, which can vary up to 4.4 dB in different 24-2 VF test locations,[25] our DL model performed well and better than published studies in point-wise estimation of 10-2 VF.

Sector-wise and point-wise analyses revealed that prediction accuracy in terms of MAE was better in the infero-temporal area compared to other areas. This was consistent with previous predictions of inferotemporal regions of 10-2 VF.[36,75,81] Weber et al reported that there is a preserved 'centro-coecal isle' corresponding to the maculopapillary bundle (the least vulnerable region to glaucomatous damage) in infero-temporal area which is 'still functioning and maintaining the rest of the VF in

patients with advanced glaucoma,[40,83] The smaller absolute prediction error values in this area may be due to the relatively preserved visual sensitivity in this region and therefore low values for TD and PD. Smaller TD and PD values will result in lower MAE. Moreover, the smaller variation of visual sensitivity in this area may be another possible reason.

This study has some limitations. First, we did not include an external independent test set to estimate the generalizability of our results to other populations. As the performance of the model is generally lower in external test sets,[84] future studies utilizing other external large datasets are necessary to validate our current estimations. Furthermore, we excluded OCT-A scans with poor quality, which might have led to better model performance than if real-world scans were included.[85] In addition, the image quality varies among different groups, including healthy, suspect, and glaucoma eyes. Lower image quality can affect DL prediction since blurry images can decrease the variance between adjacent pixels. We have reported previously that around 11% of OCTA images have segmentation errors that need correction, and that the segmentation error in 3.4% of images cannot be corrected.[85] Therefore, our model may work better on images without segmentation errors. In addition, DL did not perform similarly across all severities. VF values were overpredicted in less severe MDs (healthy or glaucoma suspect eyes) and underestimated in severe cases, although there were some cases that had good predictions. While prediction using DL algorithms from images is promising, this study was not longitudinal and did not aim to assess glaucoma progression. Future studies are needed to evaluate the role of DL in the prediction of VF progression using OCTA.

In conclusion, DL-based estimates of 10-2 VFs from OCTA scan showed high accuracy in estimating global and point-wise functional loss. These results suggest that OCTA information can be exploited for improved management of glaucoma. By applying DL to OCTA images, clinical decision-making, personalized patient care, and risk stratification may be improved for individuals with central VF damage.

References

1. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *Jama*. May 14 2014;311(18):1901-1911. doi:10.1001/jama.2014.3192
2. Ramulu P. Glaucoma and disability: which tasks are affected, and at what stage of disease? *Curr Opin Ophthalmol*. 2009;20(2):92.
3. Saunders LJ, Medeiros FA, Weinreb RN, Zangwill LM. What rates of glaucoma progression are clinically significant? *Expert Rev Ophthalmol*. 2016;11(3):227-234.
4. Schiefer U, Papageorgiou E, Sample PA, et al. Spatial pattern of glaucomatous visual field loss obtained with regionally condensed stimulus arrangements. *Invest Ophthalmol Vis Sci*. Nov 2010;51(11):5685-5689. doi:10.1167/iovs.09-5067
5. Grillo LM, Wang DL, Ramachandran R, et al. The 24-2 Visual Field Test Misses Central Macular Damage Confirmed by the 10-2 Visual Field Test and Optical Coherence Tomography. *Transl Vis Sci Technol*. Apr 2016;5(2):15. doi:10.1167/tvst.5.2.15
6. Abe RY, Diniz-Filho A, Costa VP, Gracitelli CP, Baig S, Medeiros FA. The impact of location of progressive visual field loss on longitudinal changes in quality of life of patients with glaucoma. *Ophthalmology*. 2016;123(3):552-557.
7. Sun Y, Lin C, Waisbourd M, et al. The Impact of Visual Field Clusters on Performance-based Measures and Vision-Related Quality of Life in Patients With Glaucoma. *Am J Ophthalmol*. Mar 2016;163:45-52. doi:10.1016/j.ajo.2015.12.006
8. Turpin A, Jankovic D, McKendrick AM. Retesting visual fields: utilizing prior information to decrease test-retest variability in glaucoma. *Invest Ophthalmol Vis Sci*. Apr 2007;48(4):1627-1634. doi:10.1167/iovs.06-1074
9. Russell RA, Crabb DP, Malik R, Garway-Heath DF. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Invest Ophthalmol Vis Sci*. Sep 6 2012;53(10):5985-5990. doi:10.1167/iovs.12-10428
10. 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*. Dec 2018;196:10-17. doi:10.1016/j.ajo.2018.08.010
11. Crabb DP, Russell RA, Malik R, et al. Frequency of visual field testing when monitoring patients newly diagnosed with glaucoma: mixed methods and modelling. *Health Serv Deliv Res*. 2014;2:1-102.
12. Yarmohammadi A, Zangwill LM, Diniz-Filho A, et al. Peripapillary and Macular Vessel Density in Patients with Glaucoma and Single-Hemifield Visual Field Defect. *Ophthalmology*. May 2017;124(5):709-719. doi:10.1016/j.ophtha.2017.01.004

13. Yarmohammadi A, Zangwill LM, Diniz-Filho A, et al. Relationship between Optical Coherence Tomography Angiography Vessel Density and Severity of Visual Field Loss in Glaucoma. *Ophthalmology*. Dec 2016;123(12):2498-2508. doi:10.1016/j.ophtha.2016.08.041
14. Penteado RC, Zangwill LM, Daga FB, et al. Optical Coherence Tomography Angiography Macular Vascular Density Measurements and the Central 10-2 Visual Field in Glaucoma. *J Glaucoma*. Jun 2018;27(6):481-489. doi:10.1097/IJG.0000000000000964
15. Kamalipour A, Moghimi S, Jacoba CM, et al. Measurements of OCT Angiography Complement OCT for Diagnosing Early Primary Open-Angle Glaucoma. *Ophthalmol Glaucoma*. May-Jun 2022;5(3):262-274. doi:10.1016/j.ogla.2021.09.012
16. Kamalipour A, Moghimi S, Khosravi P, et al. Combining Optical Coherence Tomography and Optical Coherence Tomography Angiography Longitudinal Data for the Detection of Visual Field Progression in Glaucoma. *Am J Ophthalmol*. Feb 2023;246:141-154. doi:10.1016/j.ajo.2022.10.016
17. Moghimi S, Bowd C, Zangwill LM, et al. Measurement Floors and Dynamic Ranges of OCT and OCT Angiography in Glaucoma. *Ophthalmology*. Jul 2019;126(7):980-988. doi:10.1016/j.ophtha.2019.03.003
18. Mohammadzadeh V, Moghimi S, Nishida T, et al. Longitudinal Structure-Function Relationship between Macular Vessel Density and Thickness and Central Visual Field in Early Glaucoma. *Ophthalmol Glaucoma*. Nov-Dec 2022;5(6):648-657. doi:10.1016/j.ogla.2022.06.004
19. Venugopal JP, Rao HL, Weinreb RN, et al. Repeatability of vessel density measurements of optical coherence tomography angiography in normal and glaucoma eyes. *Br J Ophthalmol*. 2018;102(3):352-357.
20. Moghimi S, Hou H, Rao H, Weinreb RN. Optical Coherence Tomography Angiography and Glaucoma: A Brief Review. *Asia Pac J Ophthalmol (Phila)*. Apr 4 2019;8(2):115-125. doi:10.22608/APO.201914
21. Manalastas PIC, Zangwill LM, Saunders LJ, et al. Reproducibility of Optical Coherence Tomography Angiography Macular and Optic Nerve Head Vascular Density in Glaucoma and Healthy Eyes. *J Glaucoma*. Oct 2017;26(10):851-859. doi:10.1097/IJG.0000000000000768
22. Christopher M, Bowd C, Proudfoot JA, et al. Deep Learning Estimation of 10-2 and 24-2 Visual Field Metrics Based on Thickness Maps from Macula OCT. *Ophthalmology*. Nov 2021;128(11):1534-1548. doi:10.1016/j.ophtha.2021.04.022
23. 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.
24. Mariottoni EB, Datta S, Dov D, et al. Artificial Intelligence Mapping of Structure to Function in Glaucoma. *Transl Vis Sci Technol*. Mar 2020;9(2):19. doi:10.1167/tvst.9.2.19
25. 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*. Aug 2 2021;10(9):16. doi:10.1167/tvst.10.9.16
26. 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*. Aug 1 2017;58(10):3975-3985. doi:10.1167/iovs.17-21832
27. Bogunovic H, Kwon YH, Rashid A, et al. Relationships of retinal structure and humphrey 24-2 visual field thresholds in patients with glaucoma. *Invest Ophthalmol Vis Sci*. Dec 9 2014;56(1):259-271. doi:10.1167/iovs.14-15885
28. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. Nov 2010;51(11):5657-5666. doi:10.1167/iovs.10-5239
29. 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.

30. Zangwill LM, Chan K, Bowd C, et al. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci*. Sep 2004;45(9):3144-3151. doi:10.1167/iovs.04-0202
31. Liu L, Jia Y, Takusagawa HL, et al. Optical Coherence Tomography Angiography of the Peripapillary Retina in Glaucoma. *Jama Ophthalmol*. Sep 2015;133(9):1045-1052. doi:10.1001/jamaophthalmol.2015.2225
32. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016:770-778.
33. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. Ieee; 2009:248-255.
34. Kingma DP, Ba J. Adam: A method for stochastic optimization. *In Proc 3rd International Conference on Learning Representations (ICLR) (ICLR, 2015),arXiv preprint arXiv:14126980*.
35. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Adv Neur In*. 2019;32
36. Kamalipour A, Moghimi S, Khosravi P, et al. Deep Learning Estimation of 10-2 Visual Field Map Based on Circumpapillary Retinal Nerve Fiber Layer Thickness Measurements. *Am J Ophthalmol*. Feb 2023;246:163-173. doi:10.1016/j.ajo.2022.10.013
37. Malik R, Swanson WH, Garway-Heath DF. 'Structure-function relationship' in glaucoma: past thinking and current concepts. *Clin Exp Ophthalmol*. May-Jun 2012;40(4):369-80. doi:10.1111/j.1442-9071.2012.02770.x
38. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011;12:2825-2830.
39. Jolliffe I. Principal Component Analysis,(Series: Springer Series in Statistics, Springer, NY). 2002: 338-372.
40. Hood DC, Raza AS, de Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res*. Jan 2013;32:1-21. doi:10.1016/j.preteyeres.2012.08.003
41. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. 2016:2921-2929.
42. Wang Z, Yang J. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *arXiv preprint arXiv:170310757*. 2017;
43. Mooney CZ, Mooney CF, Mooney CL, Duval RD, Duvall R. *Bootstrapping: A nonparametric approach to statistical inference*. sage; 1993.
44. Cohen P, West SG, Aiken LS. *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology press; 2014.
45. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. Dec 1982;38(4):963-974.
46. Blumberg DM, De Moraes CG, Prager AJ, et al. Association Between Undetected 10-2 Visual Field Damage and Vision-Related Quality of Life in Patients With Glaucoma. *Jama Ophthalmol*. Jul 2017;135(7):742-747. doi:10.1001/jamaophthalmol.2017.1396
47. De Moraes CG, Hood DC, Thenappan A, et al. 24-2 Visual Fields Miss Central Defects Shown on 10-2 Tests in Glaucoma Suspects, Ocular Hypertensives, and Early Glaucoma. *Ophthalmology*. 2017/10/01/ 2017;124(10):1449-1456. doi:https://doi.org/10.1016/j.ophtha.2017.04.021
48. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of Automated Perimetry: A Comparison between Standard Automated Perimetry with Stimulus Size III and V, Matrix, and Motion Perimetry. *Invest Ophthalmol Vis Sci*. Feb 2009;50(2):974-979. doi:10.1167/iovs.08-1789
49. Phu J, Khuu SK, Yapp M, Assaad N, Hennessy MP, Kalloniatis M. The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clin Exp Optom*. Jul 2017;100(4):313-332. doi:10.1111/cxo.12551
50. Mahmoudinezhad G, Moghimi S, Weinreb RN. COVID-19 Pandemic: Are We Back to Normal? *J Glaucoma*. Aug 2020;29(8):611-612. doi:10.1097/IJG.0000000000001590

51. Mohammadzadeh V, Rabiolo A, Fu Q, et al. Longitudinal Macular Structure-Function Relationships in Glaucoma. *Ophthalmology*. Jan 2020 Jul 1;127(7):888-900. doi:10.1016/j.ophtha.2020.01.023
52. Nouri-Mahdavi K, Fatehi N, Caprioli J. Longitudinal Macular Structure-Function Relationships in Glaucoma and Their Sources of Variability. *Am J Ophthalmol*. 11 2019;207:18-36. doi:10.1016/j.ajo.2019.04.034
53. Hirasawa K, Matsuura M, Fujino Y, et al. Comparing Structure-Function Relationships Based on Drasdo's and Sjostrand's Retinal Ganglion Cell Displacement Models. *Invest Ophthalmol Vis Sci*. Apr 9 2020;61(4):10. doi:10.1167/iovs.61.4.10
54. 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*. Oct 2007;47(22):2901-2911. doi:10.1016/j.visres.2007.01.007
55. Raza AS, Cho J, de Moraes CG, et al. Retinal ganglion cell layer thickness and local visual field sensitivity in glaucoma. *Arch Ophthalmol*. Dec 2011;129(12):1529-1536. doi:10.1001/archophthalmol.2011.352
56. 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*. Jun 2019;103(6):821-825. doi:10.1136/bjophthalmol-2018-312162
57. Yohannan J, Boland MV. The evolving role of the relationship between optic nerve structure and function in glaucoma. *Ophthalmology*. 2017;124(12):S66-S70.
58. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res*. Nov 2007;26(6):688-710. doi:10.1016/j.preteyeres.2007.08.001
59. Wong D, Chua J, Tan B, et al. Combining OCT and OCTA for Focal Structure-Function Modeling in Early Primary Open-Angle Glaucoma. *Invest Ophthalmol Vis Sci*. Dec 1 2021;62(15):8. doi:10.1167/iovs.62.15.8
60. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Invest Ophthalmol Vis Sci*. Jul 29 2011;52(8):5794-5803. doi:10.1167/iovs.10-7111
61. 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*. Jan 2012;21(1):49-54. doi:10.1097/IJG.0b013e31822af27a
62. Garway-Heath DF, Caprioli J, Fitzke FW, Hitchings RA. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci*. Jun 2000;41(7):1774-1782.
63. Rao HL, Zangwill LM, Weinreb RN, Leite MT, Sample PA, Medeiros FA. Structure-function relationship in glaucoma using spectral-domain optical coherence tomography. *Arch Ophthalmol*. Jul 2011;129(7):864-871. doi:10.1001/archophthalmol.2011.145
64. Altangerel U, Spaeth GL, Rhee DJ. Visual function, disability, and psychological impact of glaucoma. *Curr Opin Ophthalmol*. Apr 2003;14(2):100-105. doi:10.1097/00055735-200304000-00009
65. Hood DC, Tsamis E, Bommakanti NK, et al. Structure-Function Agreement Is Better Than Commonly Thought in Eyes With Early Glaucoma. *Invest Ophthalmol Vis Sci*. Oct 2019;60(13):4241-4248. doi:10.1167/iovs.19-27920
66. Hood DC. Improving our understanding, and detection, of glaucomatous damage: An approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res*. Mar 2017;57:46-75. doi:10.1016/j.preteyeres.2016.12.002
67. Kim S, Lee JY, Kim SO, Kook MS. Macular structure-function relationship at various spatial locations in glaucoma. *Br J Ophthalmol*. Oct 2015;99(10):1412-1418. doi:10.1136/bjophthalmol-2014-306538

68. 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*. Aug 1 2017;58(10):4211-4217. doi:10.1167/iovs.17-21995
69. 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*. Aug 2017;58(10):4310-4317. doi:10.1167/iovs.17-21961
70. Wen JC, Lee CS, Keane PA, et al. Forecasting future Humphrey visual fields using deep learning. *PloS one*. 2019;14(4):e0214875.
71. Park K, Kim J, Lee J. Visual field prediction using recurrent neural network. *Sci Rep*. 2019;9(1):1-12.
72. 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*. Oct 2020;218:304-313. doi:10.1016/j.ajo.2020.04.037
73. Masumoto H, Tabuchi H, Nakakura S, Ishitobi N, Miki M, Enno H. Deep-learning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity. *J Glaucoma*. Jul 2018;27(7):647-652. doi:10.1097/IJG.0000000000000988
74. 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 preprint arXiv:190801428*. 2019:1-12.
75. Hashimoto Y, Asaoka R, Kiwaki T, et al. Deep learning model to predict visual field in central 10 degrees from optical coherence tomography measurement in glaucoma. *Br J Ophthalmol*. Apr 2021;105(4):507-513. doi:10.1136/bjophthalmol-2019-315600
76. Parrish RK, 2nd, Schiffman J, Anderson DR. Static and kinetic visual field testing. Reproducibility in normal volunteers. *Arch Ophthalmol*. Oct 1984;102(10):1497-1502. doi:10.1001/archopht.1984.01040031217021
77. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. Aug 15 1989;108(2):130-135. doi:10.1016/0002-9394(89)90006-8
78. Bowd C, Belghith A, Proudfoot JA, et al. Gradient-Boosting Classifiers Combining Vessel Density and Tissue Thickness Measurements for Classifying Early to Moderate Glaucoma. *Am J Ophthalmol*. Sep 2020;217:131-139. doi:10.1016/j.ajo.2020.03.024
79. Shoji T, Zangwill LM, Akagi T, et al. Progressive Macula Vessel Density Loss in Primary Open-Angle Glaucoma: A Longitudinal Study. *Am J Ophthalmol*. Oct 2017;182:107-117. doi:10.1016/j.ajo.2017.07.011
80. Hou H, Moghimi S, Proudfoot JA, et al. Ganglion Cell Complex Thickness and Macular Vessel Density Loss in Primary Open-Angle Glaucoma. *Ophthalmology*. Aug 2020;127(8):1043-1052. doi:10.1016/j.ophtha.2019.12.030
81. 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*. Jan 26 2021;11(1):2214. doi:10.1038/s41598-020-79494-6
82. 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*. Dec 2021;1(4):100055. doi:10.1016/j.xops.2021.100055
83. Weber J, Schultze T, Ulrich H. The visual field in advanced glaucoma. *Int Ophthalmol*. Jan 1989;13(1-2):47-50. doi:10.1007/BF02028637
84. Fan R, Alipour K, Bowd C, et al. Detecting Glaucoma from Fundus Photographs Using Deep Learning without Convolutions: Transformer for Improved Generalization. *Ophthalmol Sci*. Mar 2023;3(1):100233. doi:10.1016/j.xops.2022.100233
85. Kamalipour A, Moghimi S, Hou H, et al. OCT Angiography Artifacts in Glaucoma. *Ophthalmology*. Oct 2021;128(10):1426-1437. doi:10.1016/j.ophtha.2021.03.036

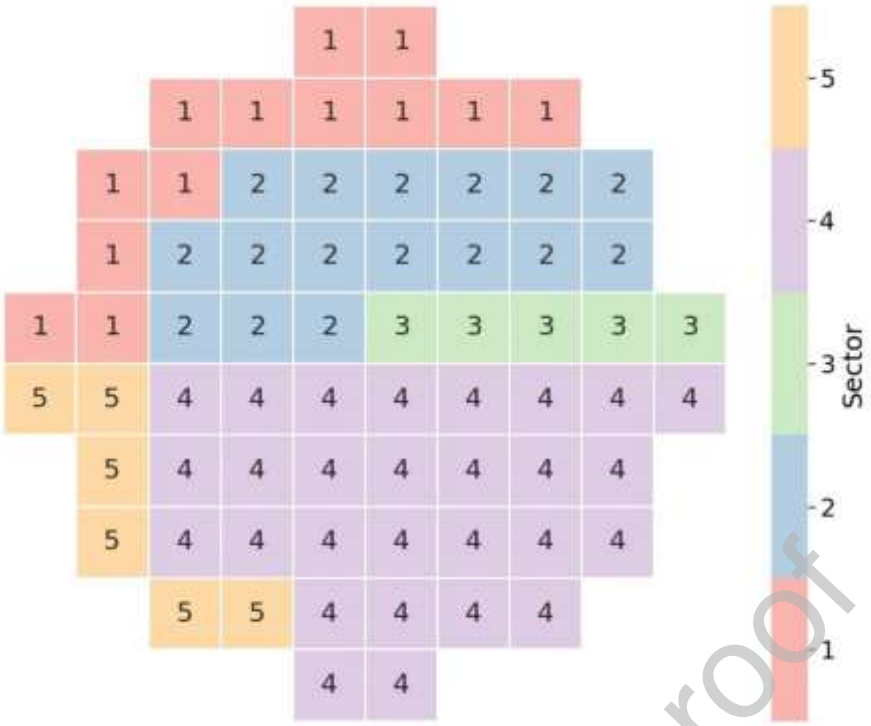


Figure1. Sectoral 10-2 visual field (VF) map (right eye view). The 68 individual 10-2 VF test points were clustered as suggested by Hood et al.[40] Note that this map assumes 5 distinct VF zones based on their vulnerability to damage in the macula. Zone 1 = superior nasal (SN) zone; zone 2 = superior temporal (ST) zone; zone 3 = superior temporal band (STB) zone; zone 4 = inferior temporal (IT) zone; and zone 5 = inferior nasal (IN) zone.

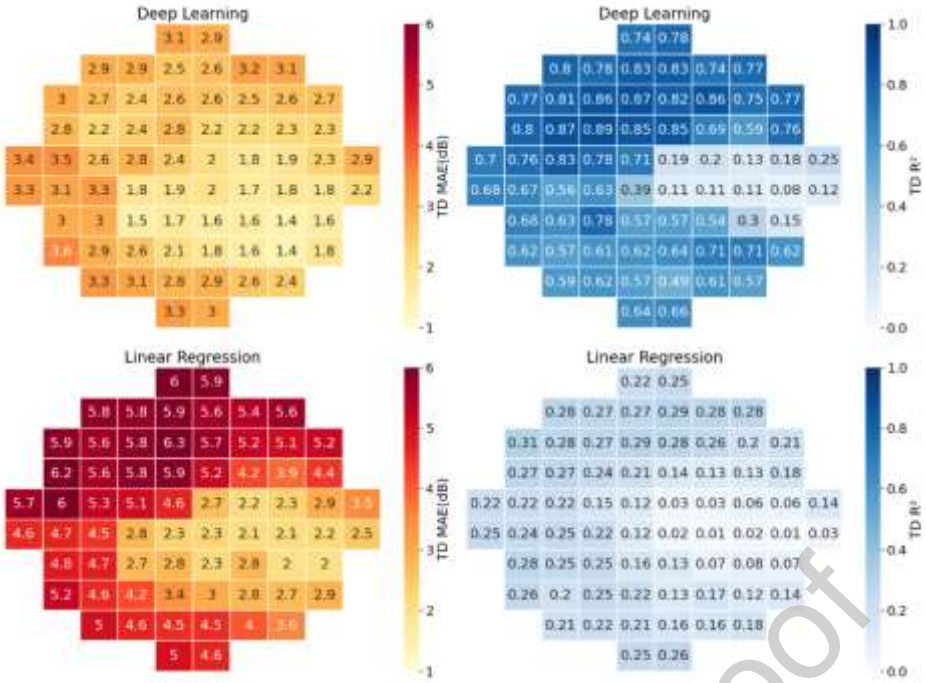


Figure 2. Models' performance for the total deviation estimation of individual 10-2 test points (Right eye view). The top row illustrates pointwise mean absolute errors (MAE) and the R^2 between estimated and the actual 10-2 total deviation (TD) values for the deep learning (DL) models. MAE is shown on the left and on R^2 the right. The bottom row illustrates the same quantities for the LR models. Darker color represents the test points with a higher MAE and a higher R^2.

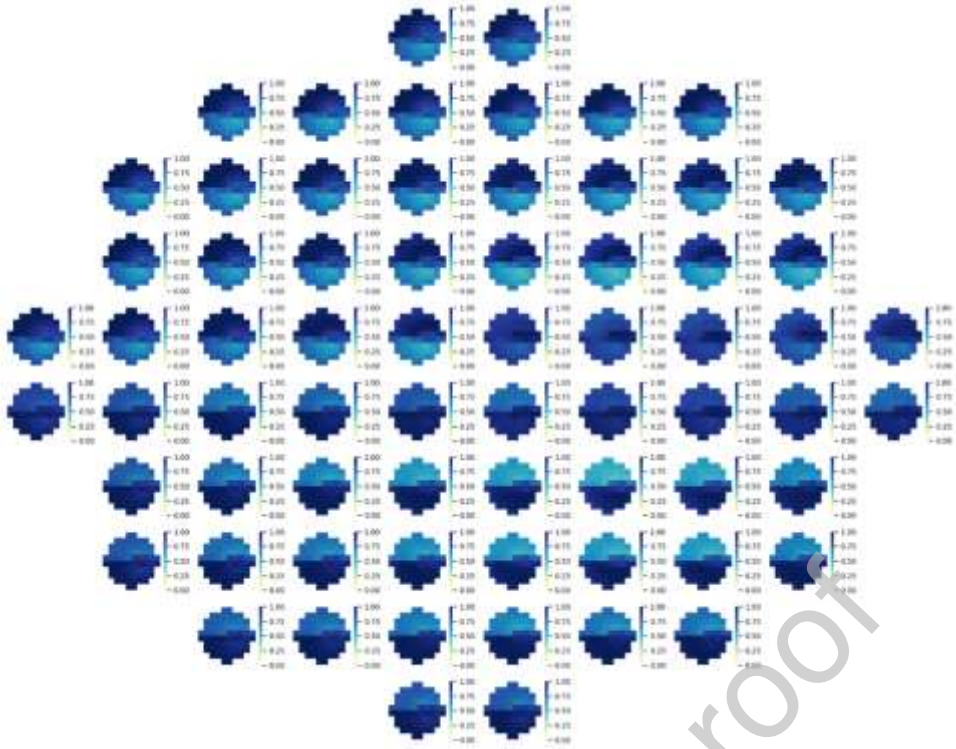


Figure 3. Correlation between total deviations (TD) predicted for each location and the 67 remaining test locations of the 10-2 VF. Darker colors indicate higher correlation. Correlation declines as distance increases between the location of interest and the other locations in the visual field. Correlations are very low for pairs of locations in opposite hemifields, except in the nasal region. This confirms that similar to glaucoma pattern damage in retinal ganglion cells, vessel density loss also respects to the temporal raphe.

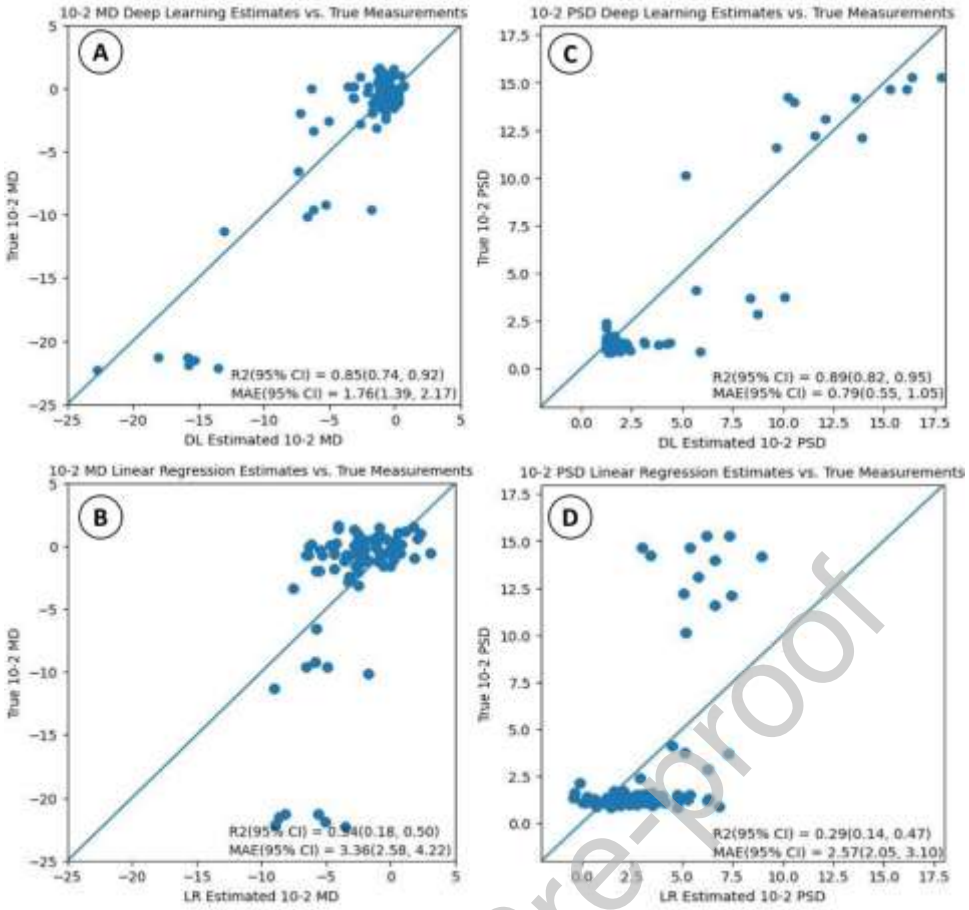


Figure 4. Scatterplots of the estimates of 10-2 mean deviation (MD) and pattern standard deviation (PSD) vs. the measured groundtruth values in decibels, for deep learning (DL) (A and C) and linear regression (LR) (B and D).

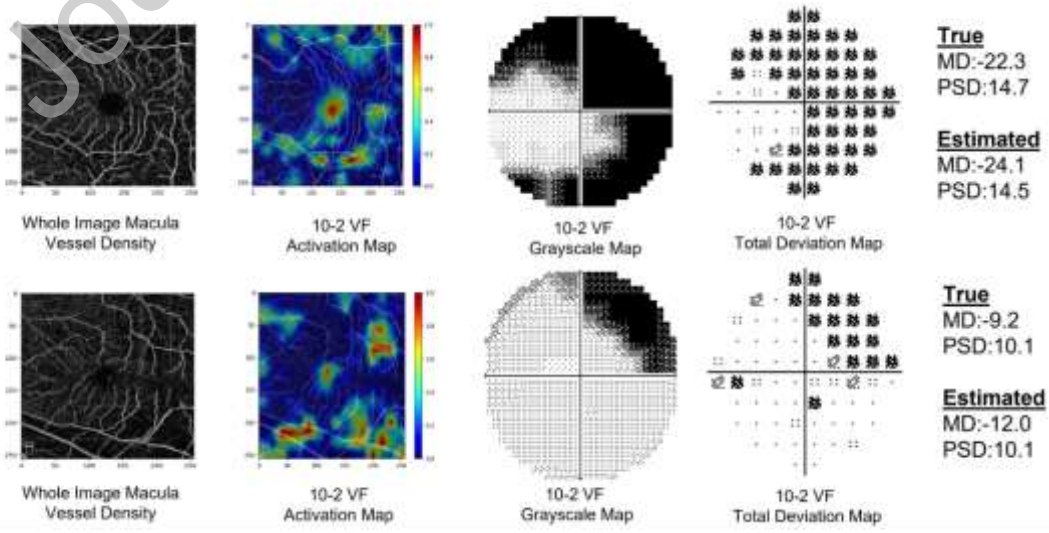
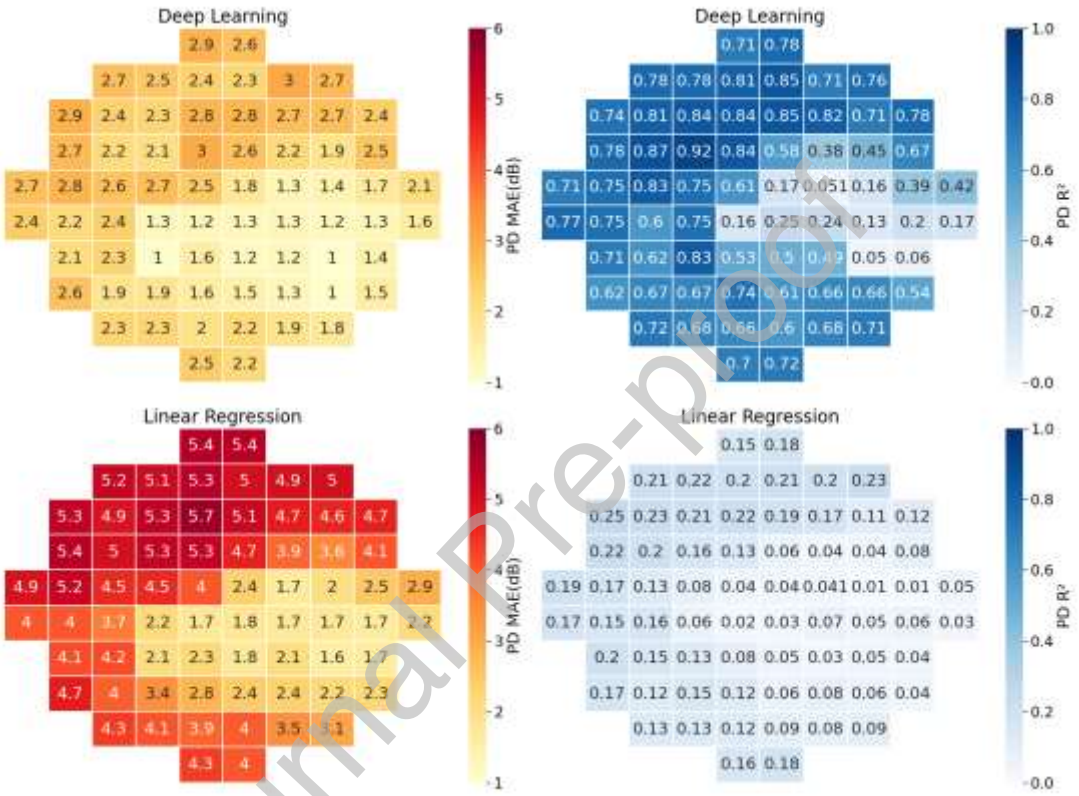
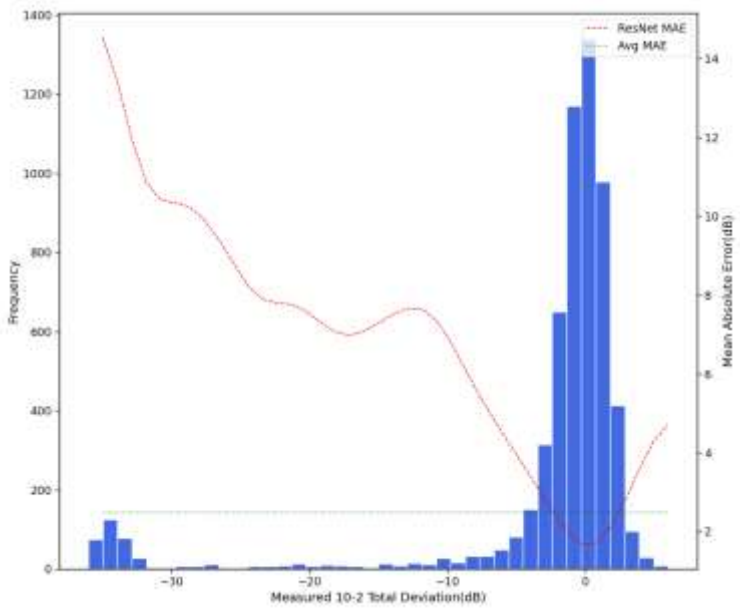


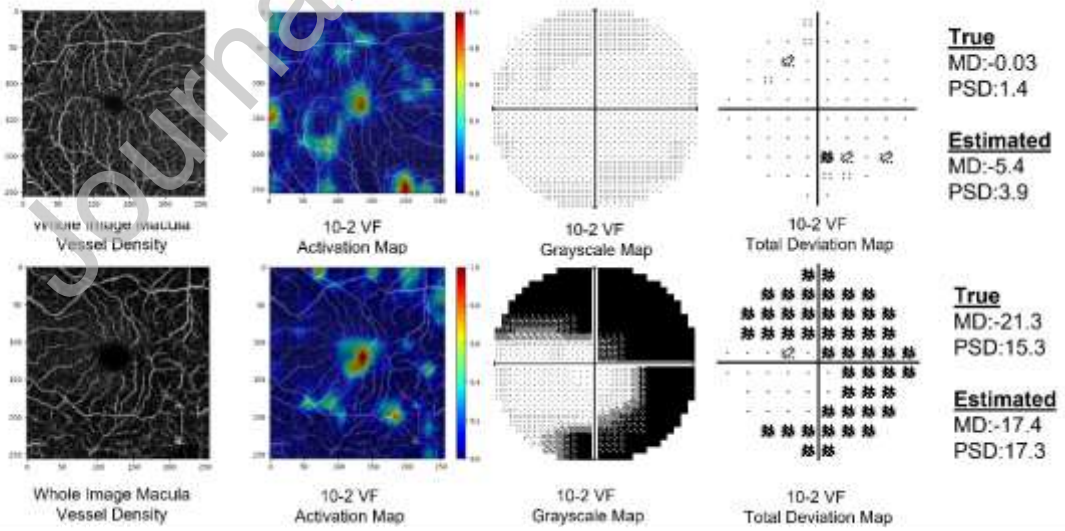
Figure 5. Visualizations of regions used by the deep learning models to estimate 10-2 visual field (VF) MD. Macular vessel densities are shown along with activation maps and VF groundtruth. Activation maps indicate the areas of the macula vessel density maps that are most important to the deep learning (DL) model predictions.(brighter regions as shown in the color bar). Both examples are left eyes.



Supplementary Figure 1. Models' performance for the pattern deviation estimation of individual 10-2 test points (Right eye view). The top row illustrates pointwise mean absolute errors (MAE) and the R^2 between estimated and the actual 10-2 pattern deviation (PD) values for the deep learning (DL) models. MAE is shown on the left and on R^2 the right. The bottom row illustrates the same quantities for the LR models. Darker color represents the test points with a higher MAE and a higher R^2.



Supplementary Figure 2. Histogram illustrating the distribution of the actual 10-2 total deviation values in the test set. The mean absolute errors (MAEs) between the ResNet50 model estimations using macula vessel density images and the actual 10-2 total deviation values are averaged according to the actual 10-2 values (dashed line). The dotted horizontal line represents the average MAE of the ResNet50 model.



Supplementary Figure 3. Visualizations of regions used by the deep learning models to estimate 10-2 visual field (VF) MD. Macular vessel densities are shown along with activation maps and VF groundtruth. Activation maps are shown for two

examples in which the deep learning (DL) model inaccurately estimated the 10-2 MD. Activation maps indicate the areas of the macula vessel density maps that are most important to the DL model predictions (brighter regions as shown in the color bar). Both examples are left eyes.

Table 1. Summary of the 10-2 Diagnostic Innovations in Glaucoma Study Datasets

Characteristic	Training and validation set	Test set	Total
Participants, n(%)	352(91.7)	32 (8.3)	384 (100)
DIGS	277(91.4)	26(8.6)	303(100)
VO	75(92.6)	6(7.4)	81(100)
Eyes, n(%)	567(92.0)	49 (8.0)	616 (100)
DIGS	452(91.5)	42(8.5)	494(100)
VO	115(94.3)	7(5.7)	122(100)
OCTA and VF pairs	965(91.8)	86 (8.2)	1051(100)
DIGS	812(91.2)	78(8.8)	890(100)
VO	153(95.0)	8(5.0)	161(100)
Age (yrs)	66.9 (65.3,68.4)	65.1 (60.6,69.6)	66.7 (65.2,68.2)
Sex (female), n (%)	191 (93.6)	13(6.4)	204(100)
Ethnicity, n (%)			
African American	60 (88.2)	8 (11.8)	68 (100)
Non-African American	292 (92.4)	24(7.6)	316 (100)
Diagnosis by eye, n (%)			
Healthy	121 (91.7)	11(8.3)	132(100)
Glaucoma suspect	112 (85.5)	19(14.5)	131 (100)
Glaucoma	334(94.6)	19(5.4)	353(100)
10-2 MD, dB	-3.4 (-3.8,-3.0)	-2.8 (-4.2,-1.5)	-3.3 (-3.7,-3.0)
10-2 PSD, dB	3.9 (3.6,4.2)	3.1 (2.2,4.0)	3.9 (3.6,4.1)
OCTA vessel density, %	42.9 (42.5,43.2)	44.4 (43.1,45.6)	43.0 (42.5,43.2)

VF: visual field, OCT: optical coherence angiography, MD: mean deviation, PSD: pattern standard deviation, n: number, VO: Viterbi Ophthalmology, DIGS: Diagnostic Innovations in Glaucoma Study. Data are represented as n (%) for categorical and mean ± standard deviation for continuous variables.

Table 2. Performance of the deep learning and Linear regression models for prediction of central visual field mean deviation (MD) and pattern standard deviation (PSD) measurements from macular optical coherence tomography Angiography (OCTA) scans.

Global 10-2 VF Estimations	DL		LR	
	MAE, dB (95% CI)	R^2 (95% CI)	MAE, dB (95% CI)	R^2 (95% CI)
Direct prediction				
10-2 MD	1.76 (1.39, 2.17)	0.85 (0.74, 0.92)	3.36 (2.58, 4. 22)	0.34 (0.18, 0.50)
10-2 PSD	0.79 (0.55, 1.05)	0.89 (0.82, 0.95)	2.57 (2.05, 3.10)	0.29 (0.14, 0.47)

Prediction in different categories				
Healthy				
10-2 MD	1.46 (0.90, 2.45)	0.10 (0.01, 0.39)	2.44 (1.65, 3.32)	0.10 (0.01, 0.46)
10-2 PSD	0.29 (0.19, 0.41)	0.07 (0.00, 0.30)	1.71 (1.19, 2.27)	0.14 (0.01, 0.59)
Glaucoma suspect				
10-2 MD	1.35 (1.00, 1.98)	0.02 (0.00, 0.16)	1.60 (1.17, 2.16)	0.15 (0.00, 0.47)
10-2 PSD	0.32 (0.23, 0.48)	0.11 (0.00, 0.34)	1.26 (0.83, 1.77)	0.06 (0.00, 0.39)
Early glaucoma				
10-2 MD	1.50 (1.04, 2.02)	0.61 (0.00, 0.82)	2.90 (2.05, 3.93)	0.12 (0.00, 0.61)
10-2 PSD	1.14 (0.48, 1.93)	0.87 (0.13, 0.99)	3.22 (2.16, 4.41)	0.13 (0.00,0.54)
Moderate to advanced				
10-2 MD	3.31 (1.78, 4.62)	0.87 (0.60, 0.97)	8.65 (3.93, 13.40)	0.18 (0.01,1.0)
10-2 PSD	2.63 (1.80, 3.39)	0.74 (0.38, 0.98)	5.81 (3.28, 8.35)	0.18 (0.01.1.0)

DL: Deep Learning, LR: Linear regression, TD: total deviation, MD: mean deviation, PSD: pattern standard deviation, VF: visual field, MAE: mean absolute error. In all comparisons, DL was significantly better than LR (All p<0.001) for direct MD and PSD prediction. DL outperformed the prediction based on LR in all healthy, suspect, early, and moderate to advanced eyes (All P<0.04 for MAE and R^2) except for R^2 in the prediction of MD (P =1.0), and PSD (P=0.60) in healthy eyes, MAE in prediction of MD for glaucoma suspects eyes (P=0.16), R^2 in the prediction of MD (P=0.28) and PSD (P=0.48) in glaucoma suspect eyes.

Table 3. Performance of the two models (Deep Learning and Linear Regression) for prediction of sectoral mean TD values in 10-2 VF Map.

Sectors	**DL**		**LR**	
	MAE, dB (95% CI)	**R^2 (95% CI)**	**MAE, dB (95% CI)**	**R^2 (95% CI)**
Superior nasal	2.99 (2.21, 3.78)	0.75(0.58, 0.88)	5.39 (4.20, 6.64)	0.30 (0.14, 0.47)
Superior temporal	2.75 (1.95, 3.66)	0.71 (0.54, 0.84)	4.85 (3.68, 6.15)	0.25 (0.10, 0.41)
Superior temporal band	2.14 (1.50, 3.01)	0.27(0.11, 0.50)	2.34 (1.65, 3.32)	0.08 (0.01, 0.23)
Inferior nasal	2.70 (1.98, 3.52)	0.79 (0.67, 0.90)	4.52 (3.25, 5.99)	0.27 (0.12, 0.43)
Inferior Temporal	1.45 (1.16, 1.79)	0.82 (0.64, 0.91)	2.58 (1.94, 3.30)	0.27 (0.12, 0.43)

DL: Deep Learning, LR: Linear regression, TD: total deviation, MD: mean deviation, VF: visual field, MAE: mean absolute error.* 10-2 VF map was clustered into 5 separate zones as proposed by Hood et al.,.[40] In all comparisons DL was significantly better than LR(All p<0.001).

Supplementary Table 1. Performance of the deep learning and Linear regression models for prediction of central visual field mean deviation (MD) and pattern standard deviation (PSD) measurements from macular optical coherence tomography Angiography (OCTA) scans.

Global and pointwise 10-2 VF Estimations	DL		LR[a]	
	MAE, dB (95% CI)	R^2 (95% CI)	MAE, dB (95% CI)	R^2 (95% CI)
Direct prediction				
10-2 MD[b]	1.76 (1.39, 2.17)	0.85 (0.74, 0.92)	3.23 (2.53, 4.00)	0.45 (0.29, 0.59)
10-2 PSD	0.79 (0.55, 1.05)	0.89 (0.82, 0.95)	2.59 (2.12, 3.07)	0.34 (0.19, 0.50)
Pointwise prediction				
10-2 TDs[b]	2.48(1.99, 3.02)	0.69 (0.57, 0.76)	3.67 (2.78, 4.66)	0.27 (0.16, 0.38)
10-2 PDs	2.08 (1.61, 2.63)	0.67 (0.54, 0.76)	3.52 (2.73, 4.37)	0.16 (0.08, 0.26)

DL: Deep Learning, LR: Linear regression, TD: total deviation, MD: mean deviation, PSD: pattern standard deviation, VF: visual field, MAE: mean absolute error. a: LR after applying PCA on OCTA images. b: MD and TD values were transformed to non-dB scale for prediction based on LR and the final results were converted to (dB) scale.

Supplementary Table 2. Performance of the two models (DL and LR) for prediction of sectoral mean PD values in 10-2 VF Map

Sectors	DL		LR	
	MAE, dB (95% CI)	R^2 (95% CI)	MAE, dB (95% CI)	R^2 (95% CI)
Superior nasal	2.85 (2.20, 3.57)	0.77(0.55, 0.92)	4.66 (3.50, 5.91)	0.24 (0.08, 0.42)
Superior temporal	2.91 (2.31, 3.59)	0.81 (0.67, 0.90)	4.22 (3.08, 5.53)	0.16 (0.03, 0.30)
Superior temporal band	2.36 (1.84, 3.19)	0.35(0.04, 0.66)	1.84 (1.19, 2.88)	0.01 (0.00, 0.05)
Inferior nasal	2.25 (1.50, 3.20)	0.87 (0.71, 0.95)	3.89 (2.70, 5.30)	0.18 (0.05, 0.34)
Inferior temporal	1.35 (0.94, 1.83)	0.65 (0.40, 0.87)	1.88(1.35, 2.50)	0.14 (0.02, 0.30)

DL: Deep Learning, LR: Linear regression, PD: pattern deviation, MD: mean deviation, VF: visual field, MAE: mean absolute error.* 10-2 VF map was clustered into 5 zones as proposed by Hood et al.,.[40] In all comparisons, DL was significantly better than LR(All p<0.001).

Supplementary Table 3. Performance of the two models (Deep Learning and Linear Regression) for prediction of sectoral mean TD values in 10-2 VF Map.

Sectors[b]	DL		LR[a]	
	MAE, dB (95% CI)	R^2 (95% CI)	MAE, dB (95% CI)	R^2 (95% CI)
Superior nasal	2.99 (2.21, 3.78)	0.75(0.58, 0.88)	4.38 (3.06, 5.89)	0.41 (0.21, 0.59)
Superior temporal	2.75 (1.95, 3.66)	0.71 (0.54, 0.84)	4.00 (2.69, 5.43)	0.36 (0.18, 0.52)
Superior temporal band	2.14 (1.50, 3.01)	0.27(0.11, 0.50)	2.09 (1.40, 3.12)	0.13 (0.04, 0.29)
Inferior nasal	2.70 (1.98, 3.52)	0.79 (0.67, 0.90)	4.14 (2.82, 5.72)	0.27 (0.13, 0.46)
Inferior Temporal	1.45 (1.16, 1.79)	0.82 (0.64, 0.91)	2.29 (1.69, 3.02)	0.36 (0.19, 0.52)

DL: Deep Learning, LR: Linear regression, TD: total deviation, MD: mean deviation, VF: visual field, MAE: mean absolute error.* 10-2 VF map was clustered into 5 separate zones as proposed by Hood et al.,.[40] In all comparisons, DL was significantly better than LR (All p<0.02, except for the comparison between MAEs (P=0.82) and R^2 (P=0.95) in the superior temporal band. a: LR after applying PCA on OCTA images. b: Average TD values were in the non-dB scale for prediction based on LR and the final results were converted to the (dB) scale.

Supplementary Table 4. Performance of the two models (DL and LR) for prediction of sectoral mean PD values in 10-2 VF Map

Sectors	**DL**		**LR[a]**	
	MAE, dB (95% CI)	***R*² (95% CI)**	**MAE, dB (95% CI)**	***R*² (95% CI)**
Superior nasal	2.85 (2.20, 3.57)	0.77(0.55, 0.92)	4.47 (3.39, 5.67)	0.34 (0.16, 0.51)
Superior temporal	2.91 (2.31, 3.59)	0.81 (0.67, 0.90)	4.03 (2.92, 5.22)	0.26 (0.11, 0.41)
Superior temporal band	2.36 (1.84, 3.19)	0.35(0.04, 0.66)	1.66 (1.05, 2.64)	0.01 (0.00, 0.05)
Inferior nasal	2.25 (1.50, 3.20)	0.87 (0.71, 0.95)	3.80 (2.61, 5.20)	0.22 (0.10, 0.36)
Inferior temporal	1.35 (0.94, 1.83)	0.65 (0.40, 0.87)	1.77 (1.26, 2.39)	0.19 (0.07, 0.34)

DL: Deep Learning, LR: Linear regression, PD: pattern deviation, MD: mean deviation, VF: visual field, MAE: mean absolute error.* 10-2 VF map was clustered into 5 zones as proposed by Hood et al..,[40] In all comparisons, DL was significantly better than LR (All p<0.04, except for the comparison between R^2 (P=0.72) in the superior temporal band. a: LR after applying PCA on OCTA images.

Golnoush Mahmoudinezhad, MD, MPH, is currently a research fellow at UCSD's Hamilton Glaucoma Center under the supervision of Dr. Robert Weinreb and Dr. Sasan Moghimi. She received her Medical Degree and a Master of Public Health degree from Shiraz University of Medical Sciences in Iran. Her current projects focus on glaucoma diagnosis and progression with various imaging approaches, the use of artificial intelligence in predicting visual field maps from OCTA, and the detection of progression. Her career goals include training in ophthalmology in an academic setting with a long-term plan of becoming a successful clinician-scientist.

Declaration of Interest statement

Commercial Relationships Disclosure:

GM None

SM None

JC None

LR None

DY None

KA None

RD None

SB None

KHD None

KL None

GG None

EM None

TN C: Topcon

AK Fight for sight

EW None

MC P: AISight Health

LMZ C: Abbvie Inc., Topcon; F: National Eye Institute, Carl Zeiss Meditec Inc., Heidelberg Engineering GmbH, Optovue Inc., Topcon Medical Systems Inc.; P: Zeiss Meditec, AISight Health

NV F: Amazon, Nautilus cluster

RNW C: Abbvie, Alcon, Allergan, Amydis, Equinox, Iantrek, Implandata, IOPtic, , Nicox, Santen, Topcon Medical; F: Topcon Medical, Heidelberg Engineering, Carl Zeiss Meditec, Optovue, Centervue, Zilia; P: Toromedes, Carl Zeiss Meditec.

Grant Support

National Eye Institute Grants R01EY029058, R01EY034148, R01EY023704, R01EY019869, U10EY014267, R01EY026574, P30EY022589, R01EY034146, K99EY030942; an Unrestricted Grant from Research to Prevent Blindness (New York, NY); Bright Focus Foundation grant #2017122; The Glaucoma Foundation grant (No grant number); Tobacco-Related Disease Research Program T31IP1511; and National Science Foundation IIS-1924937 IIS-2041009.

The sponsor or funding organization had no role in the design or conduct of this research.

Conflict of Interest: A blanket statement that “no conflicting relationship exists for any author.”