



Policy-Driven, Multimodal Deep Learning for Predicting Visual Fields from the Optic Disc and OCT Imaging

Yuka Kihara, MS,¹ Giovanni Montesano, MD,^{2,3} Andrew Chen, MD,¹ Nishani Amerasinghe, MD,⁴ Chrysostomos Dimitriou, MD,⁵ Aby Jacob, MD,⁴ Almira Chabi, MD,⁶ David P. Crabb, PhD,² Aaron Y. Lee, MD, MSCI¹

Purpose: To develop and validate a deep learning (DL) system for predicting each point on visual fields (VFs) from disc and OCT imaging and derive a structure–function mapping.

Design: Retrospective, cross-sectional database study.

Participants: A total of 6437 patients undergoing routine care for glaucoma in 3 clinical sites in the United Kingdom.

Methods: OCT and infrared reflectance (IR) optic disc imaging were paired with the closest VF within 7 days. EfficientNet B2 was used to train 2 single-modality DL models to predict each of the 52 sensitivity points on the 24-2 VF pattern. A policy DL model was designed and trained to fuse the 2 model predictions.

Main Outcome Measures: Pointwise mean absolute error (PMAE).

Results: A total of 5078 imaging scans to VF pairs were used as a held-out test set to measure the final performance. The improvement in PMAE with the policy model was 0.485 (0.438, 0.533) decibels (dB) compared with the IR image of the disc alone and 0.060 (0.047, 0.073) dB with to the OCT alone. The improvement with the policy fusion model was statistically significant ($P < 0.0001$). Occlusion masking shows that the DL models learned the correct structure–function mapping in a data-driven, feature agnostic fashion.

Conclusions: The multimodal, policy DL model performed the best; it provided explainable maps of its confidence in fusing data from single modalities and provides a pathway for probing the structure–function relationship in glaucoma. *Ophthalmology* 2022;129:781–791 © 2022 by the American Academy of Ophthalmology



Supplemental material available at www.aojournal.org.

Glaucoma is a disease of the optic nerve head (ONH) characterized by loss of retinal ganglion cells and their axons.^{1,2} This in turn can lead to progressive damage to the visual field (VF) with specific patterns determined by the arcuate trajectories followed by the axons as they enter the ONH.³ In clinical practice, glaucomatous damage can be detected and monitored through functional and structural evaluations. Visual field tests are known for requiring strong cooperation from the patient and, as currently implemented, are affected by considerable test–retest variability, especially in people with advanced VF damage.^{4,5} Structural evaluations of the macular region and the ONH are often done with imaging devices, most commonly fundus cameras and spectral domain OCT.⁶ The latter has the advantage of providing cross-sections of the retina and the ONH that allow precise evaluation of their layers. Structural damage from glaucoma often manifests itself as thinning of the retinal nerve fiber layer (RNFL, where the ganglion cell axons reside). Circumpapillary OCT (Cp-OCT) scans are a quick and common OCT scanning pattern that capture a circular OCT section

around the ONH. In some devices, this is coupled with a 2-dimensional scanning laser ophthalmoscopy (SLO) image of the ONH, often used to track eye movements during the acquisition.

Matching VF testing and structural data in glaucoma has proven problematic for several reasons. The first challenge is posed by the complex nonlinear spatial mapping of standard automated perimetry locations to the ONH. Although empirical maps exist^{7–9,10} to describe average trajectories, they are often difficult to customize for individual patients^{11–14} and become unreliable with increasing distance from the ONH.¹⁰ Moreover, structural measurements are often affected by a strong measurement floor effect, whereby little change in structure is observed for large variations in function, especially for more advanced damage.¹⁵ Previous work has shown the potential of machine learning techniques to model such a complex problem.¹⁶ More recently, advancements in the field of artificial intelligence and deep learning (DL), especially for image analysis with the introduction of convolutional neural networks, have reignited the interest around

structure–function predictions in glaucoma.^{17–19} Recent attempts focused on predicting standard automated perimetry sensitivity from segmented circum papillary RNFL thickness profiles.^{20–22} Although simple, such an approach relies on segmentations being available and correct. Moreover, it discards valuable information from OCT reflectivity. Finally, little effort has been put into predicting VF data from ONH planar pictures. Despite providing only limited direct quantitative data, these images are rich with contextual information, for example, on the general shape of the ONH and the position of the blood vessels, that can be efficiently exploited by convolutional neural network architectures.

Our work tested the hypothesis that a hybrid DL method combining information from paired ONH and Cp-OCT images cannot only improve the prediction of standard automated perimetry sensitivity directly from structural data without segmentation but also lead to a fully artificial intelligence–derived structure–function mapping. We used a large clinical database acquired from routine glaucoma clinics to develop 2 separate DL submodels, each able to provide an estimate of VF sensitivity from OCT or infrared SLO images. A third policy DL model then combines the 2 predictions by evaluating feature vectors generated by the 2 submodels. Compared with previous work,^{20–22} our work focused on developing a model to predict VF sensitivity values from a combination of different imaging modalities in an agnostic fashion, particularly without any use of segmented OCT data.

Methods

This study was conducted in accordance with the Declaration of Helsinki. This retrospective database study was approved by the Western Institutional Review Board and deemed to be nonhuman subjects research because only deidentified clinical and imaging data were used. A large database of anonymized clinical data was curated from 3 regionally different National Health Service glaucoma clinics in the United Kingdom. All data were recorded into an electronic medical record platform (Medisoft, version 6). A Humphrey field analyzer (HFA, Zeiss Meditec) was used to generate VFs. We included all Humphrey VFs (HVF) performed with a Swedish interactive threshold algorithm standard/fast or full-threshold strategy using a 24-2 VF pattern with a false-positive rate $\leq 15\%$ and that could be matched with a Cp-OCT scan. We decided not to take false-negative rates into account because of their strong relationship with the level of glaucoma damage²³ and because, together with fixation losses, they have been shown to be poor predictors of test reliability.²⁴ A flowchart of the selection steps is shown in Figure 1. Particular care was taken to exclude eyes with VF defects that could likely be due to neurological damage. To this aim, we applied a detection method to identify vertical asymmetries in VF defects of each eye and the correspondence of the defect between overlapping regions of the VF from the 2 eyes because these are typical features of VF defects from postchiasmatic lesions (details provided in Fig S1, available at www.aaojournal.org). Additionally, we carefully excluded patients who could have retinal diseases by identifying patients who received intravitreal injections or were labeled in the electronic medical record (EMR) as having any other retinal or ocular diseases (Fig 1).

OCT scan patterns of the optic nerve from Heidelberg Spectralis (OCT-1 or OCT-2) were identified in the database. Each HVF

was linked to OCT scans performed within 7 days. For the training set, we retained all possible pairs because any variation in the data within such a short time interval could be attributed to test–retest variability; this effectively served as a form of data augmentation. For the validation and test sets, we only retained pairs with the OCT scan closest to the HVF within the selected time frame. The infrared reflectance (IR) image of the optic disc and the circular B-scan set at 3.5 mm were then preprocessed. For the disc IR image, the central point of the optic disc manually selected by the photographer during acquisition was used to crop a 260×260 pixel image at a native resolution as an 8-bit grayscale image. The circular OCT B-scan was resized to a 1:1 aspect ratio and resized to 260×260 as an 8-bit grayscale image. Each of the paired HVFs was collapsed to a 1-dimensional, 54 element vector representing each sensitivity point on the 24-2 VF. Two blind spot locations ($X = 15$; $Y = +/- 3$ degrees for a right eye) were discarded from the 54 element, 1-dimensional vector at the time of evaluation. The VF sensitivity values were clipped between 0 and 40 decibels (dB) and normalized to be between 0 and 1. The upper bound was chosen because it is commonly considered the highest plausible decibel value in visually healthy subjects in standard testing conditions.²⁵ The imaging data were normalized to be between 0 and 1. The data were partitioned into 60%, 20%, and 20% for training, validation, and held-out test sets at the patient level, respectively.

Two separate DL models were then trained using an EfficientNet B2 architecture,²⁶ one for the disc and one for the OCT. EfficientNet is designed for a specific input resolution and optimized by a scaling method that uniformly scales each dimension of depth, width, and resolution with a fixed set of coefficients. The input resolution for EfficientNet B2 is 260×260 pixels. EfficientNet has been shown to achieve better performance than other well-known convolutional neural network architectures such as ResNet and DenseNet with a smaller amount of parameters.²⁶ For comparison, we report the prediction performance of our model with different architectures in Tables S2 and S3 (available at www.aaojournal.org). The batch size was set to 32, NAdam was used as the optimizer,²⁷ and the initial learning rate was set to 5×10^{-3} . The last layer of the model had 54 nodes corresponding to the 54 VF sensitivity points on the 24-2. Our dataset was imbalanced; the majority of VF sensitivity values were approximately 30 dB, which caused the model to fail to learn patterns of advanced VF loss. To overcome this limitation, we used the sensitivity-weighted mean square error for the loss function

$$\text{mean} \left(\frac{1}{1 + \alpha \times yTrue} \times (yTrue - yPred)^2 \right),$$

where $yTrue$ is 1×54 true sensitivities and $yPred$ is 1×54 predicted sensitivities and the mean was calculated over samples in the same training batch. In this case, α is a hyperparameter that controls how strongly the model weighs lower sensitivities. We used $\alpha = 2$ for our training. The mean absolute error reported for the test dataset was calculated without applying any sensitivity weights.

A policy DL model was then constructed to take the feature maps from both models and combine using a 1×1 convolutional layer after concatenation of the feature maps. For each training sample, we generated our target of a binary 54 element vector, where each location was labeled as 0 for the disc prediction being closer to the true (observed) HVF and 1 for the OCT prediction being closer to the true (observed) HVF. The output of the policy network was a 54 element, 1-dimensional vector, where each element was a floating point value from 0 to 1, where 0 represents an absolute preference for the disc model prediction and 1 represents an absolute preference for the OCT model prediction at each VF location. Because this value was generated by a sigmoid

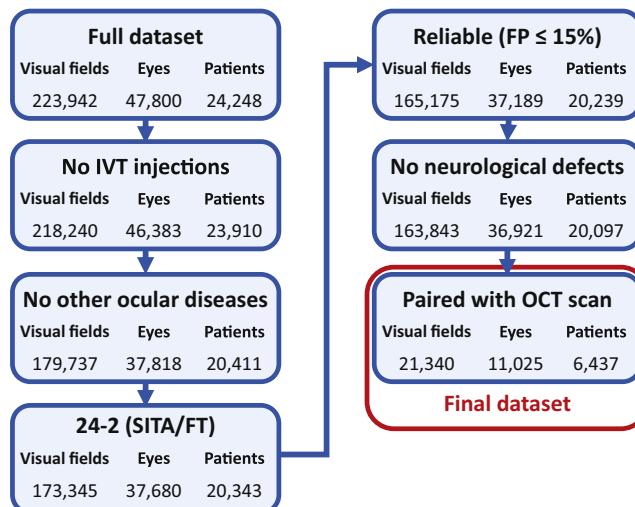


Figure 1. Flowchart of the selection process from the full dataset of visual field tests. FP = false-positive; FT = full-threshold; IVT = intravitreal; SITA = Swedish Interactive Testing Algorithm.

function, we chose 0.5 as the threshold to decide between adopting the prediction from the disc or OCT model. Because the policy network used a final fully connected layer, it was able to learn the relevant spatial patterns and correlations between the feature maps of the 2 base models. We used the same training settings (learning rate, optimizer, and batch size) as for the base models, with a weighted binary cross-entropy loss function. A network diagram of the complete end-to-end multimodal DL model is shown in Figure S2 (available at www.aaojournal.org), along with a detailed description of the loss function for the policy network.

To visualize the trained model predictions, a mean-occlusion mask was iteratively applied to each pixel position, and the change in the predicted VF sensitivities was quantified. Videos of the occlusion affecting the HVF were generated by iterating the occlusion over each position as a separate frame. The heatmap describing the magnitude of each difference was projected back onto the 24-2 to determine the occlusion locations that led to the greatest change in the predicted HVF (Videos 1 and 2, available at www.aaojournal.org). The final performance of the disc, OCT, and multimodal policy DL models was assessed in the final held-out test set at the end of the study. Pointwise mean absolute error (PMAE) and standard deviation were assessed by taking the mean of the absolute difference for each sensitivity point between the predicted VF and the true clinical VF.

Results

Clinical data, imaging data, and HVF data were extracted for 24,248 patients. After applying inclusion and exclusion criteria, the total number of eligible study patients was 6437 (11,025 eyes). A flowchart of the patient selection process is shown in Figure 1. The training, validation, and held-out test set contained 11,417 HVF-OCT pairs for 3725 patients, 4845 pairs for 1323 patients, and 5078 pairs for 1389 patients, respectively. Note that each VF could be paired with multiple OCTs in the training set but only with the closest OCT for the validation and test sets. Descriptive statistics for the 3 mutually exclusive sets of patients are shown in Table 1.

The training curves for the 3 DL models are provided in Figure S3 (available at www.aaojournal.org). The final PMAE of the DL model using the IR SLO image of the disc alone was 3.62 dB (95% confidence interval [CI], 3.54–3.68 dB). The PMAE of the single modality DL model using the OCT alone was better at 3.19

dB (95% CI, 3.13–3.25). The best performing model was the policy fusion DL model that combined the predictions from the OCT and disc submodels with a PMAE of 3.13 dB (95% CI, 3.07–3.19). The details are shown in Table S4. When tested with a paired *t* test, all pairwise comparisons of PMAE were highly statistically significant (all $P < 0.000001$). The paired difference in PMAE between the OCT submodel and the policy fusion DL model was 0.060 dB (95% CI, 0.047–0.073). The paired difference in PMAE between the disc submodel and the policy fusion DL model was 0.485 dB (95% CI, 0.438–0.533). Clinical examples of the policy DL model are shown in Figure 2 with interpretability plots of the policy model displaying the confidence in choosing the model output from disc or OCT. Failure examples where the model predictions deviated from the clinical HVF are shown in Figure S4 (available at www.aaojournal.org).

To fairly assess the final model performance across the full range of sensitivities, the pointwise error was plotted for each sensitivity threshold from 0 to 36 dB in the held-out test set (Fig 3). To probe the structure–function mapping learned by the model, a mean-occlusion mask was iteratively applied to each pixel position on the disc and OCT, thereby measuring the effect on HVF predictions. Clips of the perturbation experiments are shown in Videos 1 and 2 (available at www.aaojournal.org) for the disc and OCT, respectively, for 1 example patient with apparently healthy disc/OCT and HVF. A remapping of the areas leading to the largest magnitude change for both models is shown as heatmaps in Figure 4B (for the disc SLO), Figure 4C (for the OCT), and Figure 4D, E (for the policy). The structure–function mapping for the entire held-out test set is reported in Figure 5 for the circum papillary RNFL. The map was built by plotting the distribution of the circum papillary angles at which the highest peak of the heatmap was located in polar coordinates. The map was also compared with previously reported structure–function maps based on anatomic data.^{7–10,14} Because high myopia can negatively affect imaging and, therefore, predictions, we identified 21 HVF-imaging pairs from 14 eyes labeled as high myopia in the electronic medical record. As expected, the performance was generally worse but confirmed the improvement brought by the policy network (Table S5 and Supplementary data, available at www.aaojournal.org).

We demonstrated that the policy network implicitly optimizes the multimodal architecture accuracy by explicitly learning to choose between single modality networks; however, it is not clear

Table 1. Descriptive Statistics of the Selected Sample

	Training Set	Test Set	Validation Set
No. of patients	4786	1585	1541
No. of eyes	8883	2902	2840
No. of visual fields	22 866	7370	7411
No. of OCT scans	21 963	7042	7110
Quality index (dB)	25 [22, 28]	25 [22, 27]	25 [22, 27.25]
MD (dB)	-1.71 [-4.39, -0.23]	-1.65 [-4.27, -0.18]	-1.63 [-4.25, -0.19]
PSD (dB)	2.07 [1.57, 4.44]	2.06 [1.57, 4.24]	2.1 [1.56, 4.34]
FP (%)	1 [0, 4]	1 [0, 4]	1 [0, 5]
Sex (F/M)	2492/2294	791/794	743/798
Baseline age (yrs)	71 [62, 78]	71 [62, 79]	70 [62, 78]
Average IOP (mmHg)	17.5 [15, 20.25]	17.6 [15, 20.2]	17.33 [14.83, 20]
Cataract surgery (yes/no)	989/7894	314/2588	324/2516

Continuous variables are reported as median [interquartile range]. dB = decibels; F = female; FP = false-positive; IOP = intraocular pressure; M = male; MD = mean deviation; PSD = pattern standard deviation.

on what basis the policy network makes such a choice. The observation that the preference of the policy network seemed to be influenced by the quality of each image raises an interesting question: Is the policy choice linked with the actual superiority on the prediction performance of the 2 models? To answer this question, we added several levels of noise to SLO or OCT to degrade its performance on one of these models and then observed how that affected the policy choice. As the level of noise on SLO increased and as the prediction performance from the SLO network degraded (PMAE: from 3.62 to 4.61), the policy network adopted the prediction from OCT model more (policy choice: from 63.6% to 67.7%). On the contrary, it trusted the prediction from the SLO model more (policy choice: 63.6% to 50.4%) as the level of noise on OCT increased and the prediction performance from OCT network degraded (PMAE: from 3.19 to 8.84). Throughout the experiments, our policy network kept a stable performance (PMAE: from 3.13 to 3.16 with noise on SLO, PMAE: from 3.16 to 3.23 with noise on OCT). The detailed results are shown in Figure S5 (available at www.aojournal.org).

Discussion

Our work demonstrates an end-to-end DL approach for accurately predicting pointwise 24-2 HVF threshold sensitivities from the combination of the IR SLO ONH image and the Cp-OCT B-scan image. In addition, we found that the structure–function relationship that was learned directly from the imaging data in a fully agnostic, data-driven manner falls within prior manually derived efforts. In the present study, we aimed to develop and validate a multimodal DL model, and we have demonstrated differences in model results with statistical significance. Demonstrating clinically meaningful improvements with this model would be the subject of a different study design.

Explainability

The entire process simultaneously provides 3 distinct outputs (prediction from OCT, prediction from SLO of the disc, and policy-based fusion of the 2 results), with each producing directly interpretable results (HVF sensitivity maps). Of note, the policy model is able to selectively choose between

predictions from the 2 submodels, thereby improving, for example, the spatial localization of deep defects (Fig 2B). The preference of the policy network seemed to be influenced by the quality of each image. Therefore, the policy model seems adept at working with factual data, which might include a preponderance of artifactual image quality in OCT/SLO scans. The policy network still has the potential to fail if both the SLO and the OCT image are affected by artifacts or poor quality. We explored these aspects more in detail in the *Supplementary data* (available at www.aojournal.org) by simulating different amounts of noise for the OCT or the disc image.

Structure–Function Mapping

Our structure–function mapping was agnostically learned by the 2 submodels from the data alone. Remarkably, the OCT-based mapping yielded results in agreement with other mapping schemes based on anatomic features of the RNFL bundles.^{7–10,14} In Figure 5, we show the structure–function mapping for the Cp-OCT scans in the test set by plotting the angle at which the peak of each heatmap (Fig 4C) was found around the ONH. Our technique indicated some spread in the distribution of the peaks, suggesting adaptive mapping based on individual features of the circum papillary RNFL scans. In many of the locations, this matched the interindividual variability observed by Jansonius et al.⁹ Inevitably, some of the variability is due to measurement noise, as evident by the larger spread in the distribution for the temporal VF locations, where perimetric defects are less commonly detected.^{28,29} One notable difference is the mapping of the macular locations, which, in our model, are shifted more toward the superior and inferior poles of the ONH (Fig 5) than Jansonius et al.⁹ This is another important line of evidence to support the novelty of our technique because Jansonius et al.⁹ did not account for the retinal ganglion cell displacement in the macular region.^{30,31} A similar difference was recently reported by Turpin and McKendrick,¹⁴ whose model is also shown in Figure 5 as a comparison.

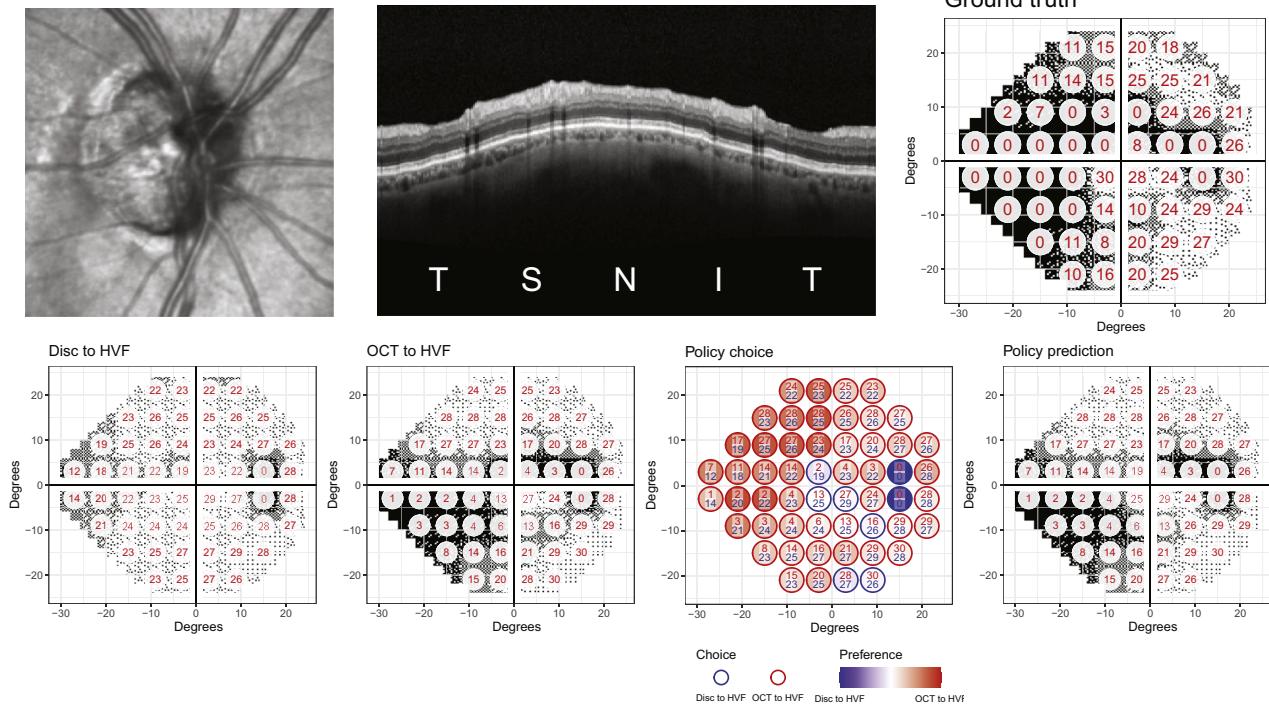
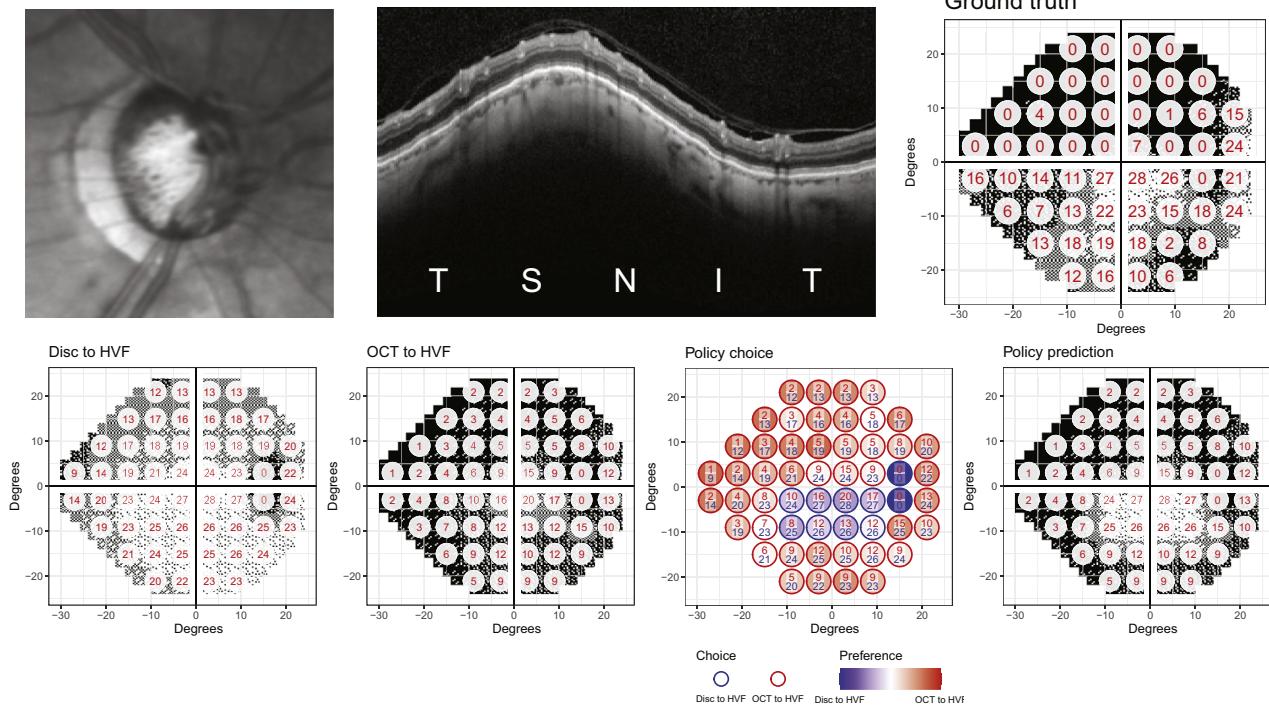
A**B**

Figure 2. Prediction examples from the test set. The top of each panel shows the disc and OCT images used for predictions alongside the ground-truth Humphrey visual field (HVF). The bottom of each panel shows, in order, the predictions from the 2 submodels, the choice score of the policy network for each location, and the final prediction of the policy network. **A**, Both submodels predict a similar location of the defect, but the OCT to HVF is more accurate in predicting the magnitude of glaucoma damage; the policy network correctly selects this prediction for the superior hemifield. **B**, The prediction from either individual submodel is wrong, showing either little damage or diffused advanced loss; however, the policy network correctly selects predictions for each location to obtain a result very close to the ground truth, better characterizing the spared paracentral visual field. I = inferior; N = nasal; S = superior; T = temporal.

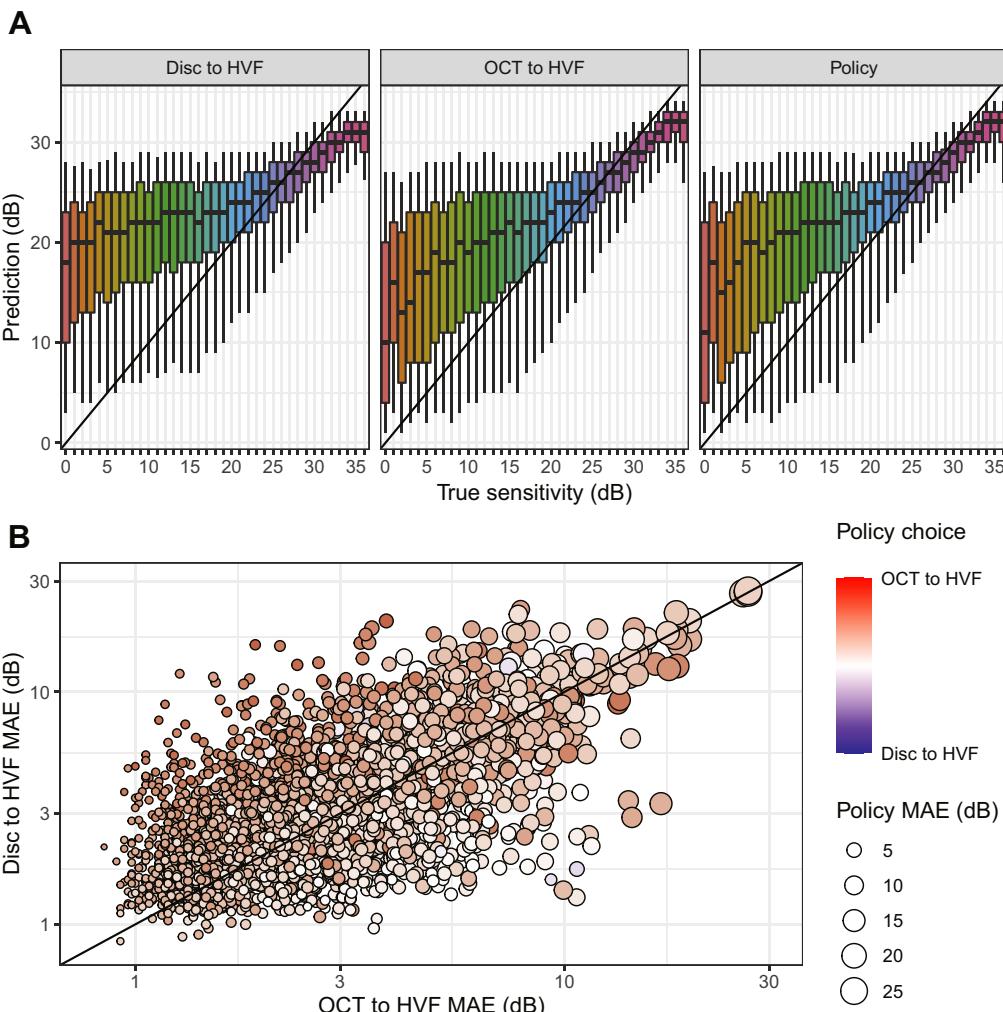


Figure 3. A, Predictions from the 2 submodels and the policy network for each level of sensitivity. The diagonal line indicates perfect agreement. B, Relationship between the prediction mean absolute error (MAE) from the 2 submodels and the final average choice from the policy network. Higher MAE in the disc to Humphrey visual field (HVF) model is generally associated with a preference for the OCT to HVF model and vice versa. The diagonal line indicates equivalent MAE. dB = decibels.

From the example shown in Figure 4B, the heatmap for the SLO predictions is mostly focused on the neuroretinal rim for the inferior hemifield; however, it appears more widespread for the location in the superior hemifield. Likewise, the heatmap produced for the Cp-OCT scans mainly focused on the RNFL (example in Fig 4C). This suggests that our technique learns the image features relevant for prediction without the need for segmentation, which is noteworthy. The disc heatmap commonly indicates some relevance of image features beyond the optic disc, suggesting a role of the reflectivity patterns induced by the loss of RNFL.³² In the heatmaps, the most frequent location of the blind spot ($X = 15$; $Y = -3$ degrees) was rendered inactive during the training, meaning that the network learned that the sensitivity at this point could not be predicted by any information present in the image. This is inconsequential for our results (both blind spot locations were excluded from our analyses) but provided us with the chance of confirming that the models were able to learn meaningful spatial features from the data.

Comparison with Previous Work

Others have used DL methods to predict global VF metrics,^{33,34} but few have attempted pointwise predictions of the VF. Park et al¹⁸ and Hashimoto et al¹⁷ used fully segmented OCT thickness maps. Mariottini et al²⁰ also predicted pointwise VF sensitivities agnostically from peripapillary OCT RNFL thickness profiles; however, such an approach not only requires an accurate segmentation to be available in the first place but also disregards other information contained in the OCT image that might inform prediction, such as the reflectivity of the RNFL.^{32,35} Fewer attempts have instead been made to use optic disc images to predict VF data.³⁶ Our data indicate the existence of additional information gained by combining the 2 imaging modalities. The error profile of our prediction stratified by sensitivity was similar to previous publications,²⁰ in that the error greatly increased below 15 dB. This could be linked to the high level of VF noise at low sensitivities,⁴ which imposes a lower bound on the

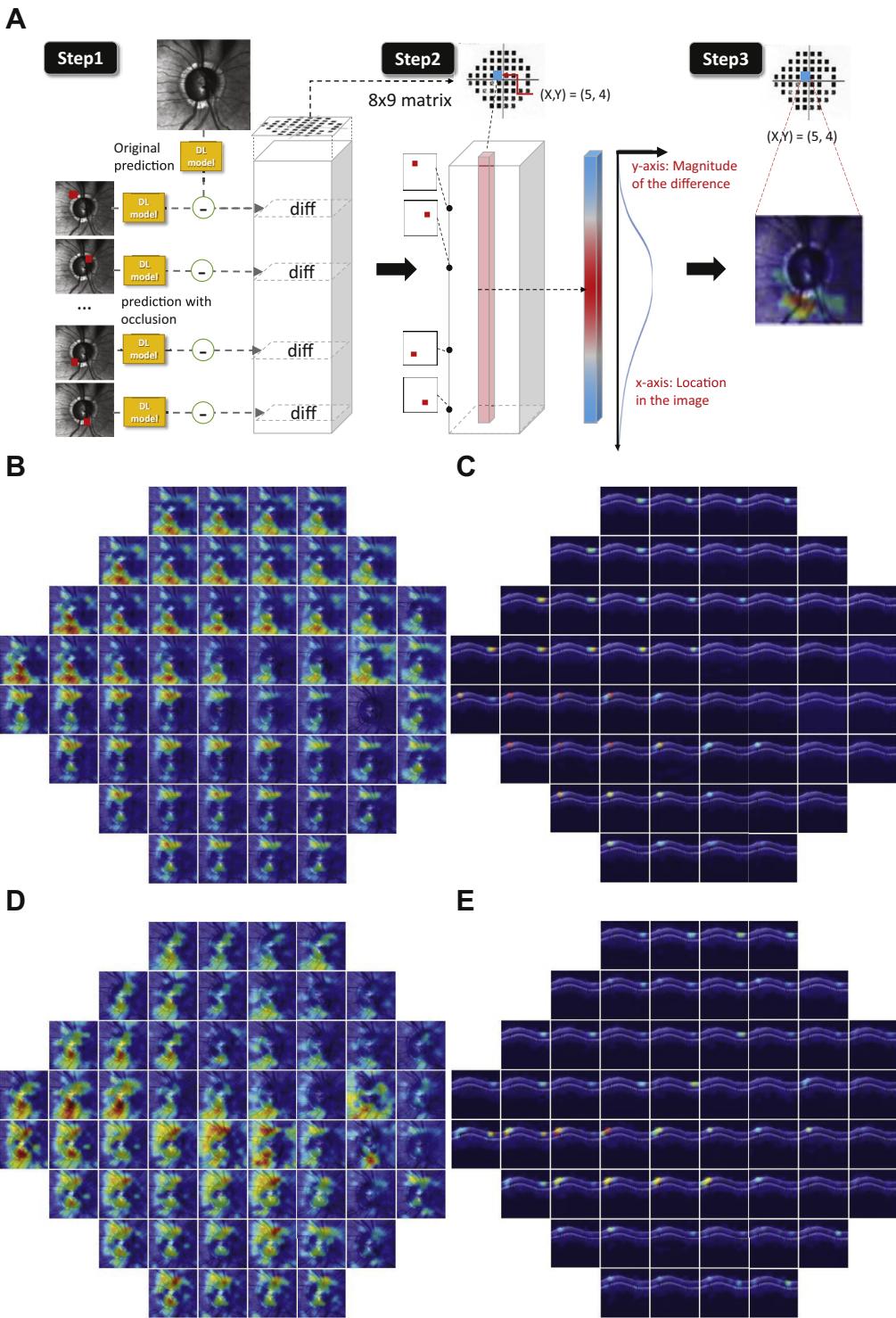


Figure 4. **A**, Process for generating heatmaps for the disc scanning laser ophthalmoscopy (SLO) and OCT models. We first took the difference in prediction between “with” and “without” occlusion at each occlusion location (Step 1), computed the distribution of magnitude of the difference over occlusion locations (Step 2), and projected the distribution back to the 2-dimensional image space to obtain a heatmap (Step 3). According to the location of the perturbation in the predicted visual field, the magnitude of changes reflects where the network focuses on during the predictions. **B**, Heatmap for 1 patient-generated example based on our disc SLO. **C**, Heatmap for 1 patient-generated example based on the OCT model. The perturbation processes of the corresponding example patient are also shown in Videos 1 and 2 (available at www.aaojournal.org). **D**, Heatmap for 1 patient-generated example from the policy network obtained by occluding the disc SLO. **E**, Heatmap for 1 patient-generated example from the policy network obtained by occluding the OCT image. DL = deep learning.

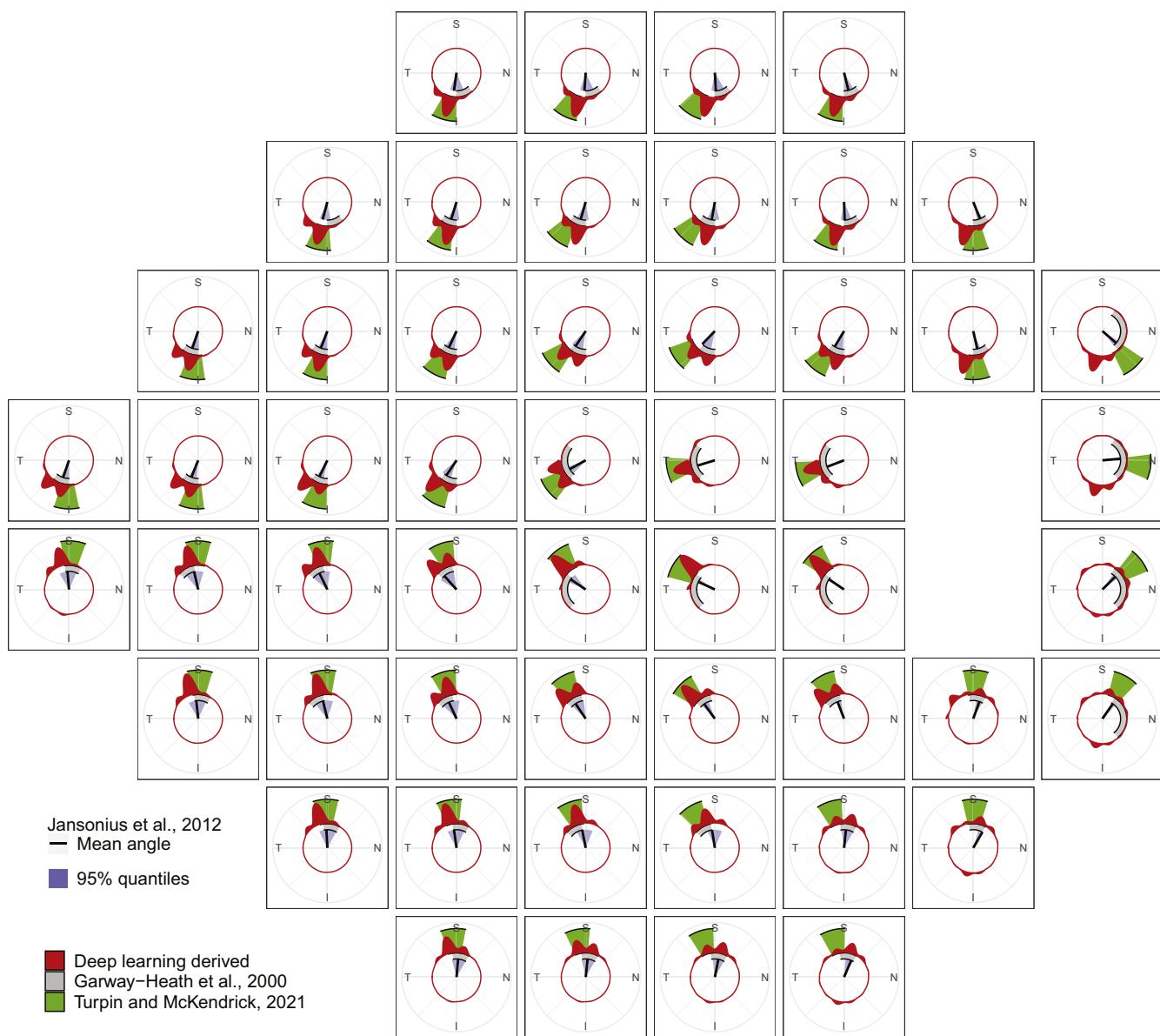


Figure 5. Structure–function mapping for the OCT to Humphrey visual field model generated from the heatmaps for the test set for each testing point in the 24-2 visual field pattern shown on polar coordinates. The map was built by plotting the distribution of the circumapillary angles at which the highest peak of the heatmap was located. Other mapping schemes are reported for comparison. I = inferior; N = nasal; S = superior; T = temporal.

achievable predictive performance, and by the previously mentioned floor effect in the structural data.¹⁵ Such a feature is clearly demonstrated by the positive bias at lower sensitivities (Fig 3). Notably, nonlinear methods have proven more successful than linear models at overcoming this issue,^{16,17,20,37} but the problem persists in all these attempts.

Practical Implications and Future Directions

Visual field estimation from imaging allows the conversion of the information contained in OCT/SLO structural data into a more clinically meaningful format. For example, it could allow for structural data to be seamlessly integrated into analyses of VF progression³⁸ or into the VF test itself.^{39–41} This could expedite perimetry or reduce

between-visit measurement variability when following a patient over time. The latter is appealing because this might improve the power to detect disease progression in a trial, and this is the subject of our future work.

The retrospective nature of the database introduces potential bias and confounders. In addition, the database was derived from a single region and may reflect regional differences in patient population or practice patterns. We plan to prospectively evaluate and validate this DL policy model in a clinical trial setting in the future to determine its generalizability. Moreover, our evaluation of the performance was based on the mean absolute error for comparison with other studies; however, this might not be the best metric to assess performance for clinical applications. Future developments will seek to build a normative dataset for our predictions so

that the performance in detecting VF defects (i.e., significant deviation from normality) could be assessed. By their nature, DL approaches behave as “black boxes,” and our method is no exception; however, we have made a conscious effort to make our results more interpretable by using a policy model and occlusion experiments. Moreover, as with many other ML and DL strategies, our method is specific to 1 condition (i.e., glaucoma), 1 VF pattern (24-2), and 1 imaging modality (SLO/Cp-OCT obtained with a Spectralis spectral domain OCT) and would not directly generalize to different contexts. One important aspect is that we did not perform any quality assessment of the images. This choice was partially due to the impossibility of establishing simple clear-cut parameters of image quality other than visual inspection. Moreover, one of the goals with our policy-based fusion network was to implement a strategy to maximally exploit the available information from 2 imaging resources, including cases when either one was noisy or of poor quality. Finally, this allowed us to test our algorithm on images that would be reflective of those acquired in standard clinical settings.

In our work, we focused on modeling HVF sensitivity and not any derived metric, such as pattern deviation (PD), total deviation (TD), or related probability maps. Our choice was partially influenced by the lack of TD and PD values in our dataset; however, modeling HVF sensitivity allowed us to use a metric that was more directly related to image features, better capturing the effect of factors such as eccentricity and aging. We believe this had an important impact on our structure–function mapping and on the performance of the algorithm. Moreover, predicting sensitivity values would allow us to calculate all derived metrics, but the opposite would not be possible. This makes our choice the most flexible for future

applications. Obviously, calculating a derived metric would require a normative dataset of imaging-predicted HVFs, which is not yet available to us; however, we show a proof of concept by calculating the TD and PD probability maps for a few examples using the functions available in the *visualFields* package for R.⁴² It is important to note that these are only illustrative examples. Moreover, TD and PD maps derived a posteriori from predicted sensitivities might not necessarily be consistent with the original and could differ from the output of an algorithm trained to predict the same maps directly.

We demonstrate the ability of DL models to predict VF pointwise sensitivities agnostically from Cp-OCT scans and IR SLO images of the disc, further improving predictions with a policy-based fusion of the 2 results. The DL models generate structure–function maps compatible with established anatomic features and are able to capture the functional consequence of relevant structural changes in glaucoma. Ubiquitously obtained IR/OCT data may contain information beyond conventional segmentation that could be used to better determine visual function and detect progression in patients with glaucoma.

Acknowledgments

The authors thank John Holdsworth, David Austin, Christian Martin, and Nick Kirby from Medisoft for their help in organizing the data and building the infrastructure to support this work. We also thank Linda McInerney from Hinchingbrooke Hospital (North West Anglia NHS Foundation Trust, United Kingdom). Support with manuscript editing and formatting was provided by Katie Alexander, PhD (Helios Medical Communications, Alderley Park, Cheshire, United Kingdom), which was funded by Santen Inc (Emeryville, California).

Footnotes and Disclosures

Originally received: August 2, 2021.

Final revision: January 28, 2022.

Accepted: February 15, 2022.

Available online: February 21, 2022. Manuscript no. OPHTHA-D-21-01540

¹ University of Washington, Department of Ophthalmology, Seattle, Washington.

² City, University of London, Optometry and Visual Sciences, London, United Kingdom.

³ NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust, UCL Institute of Ophthalmology, London, UK.

⁴ University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom.

⁵ Colchester Hospital, East Suffolk and North Essex NHS Foundation Trust, Colchester, United Kingdom.

⁶ Santen, Emeryville, California.

*Y.Y. and G.M. contributed equally and should be considered as co-first authors.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): A.Y.L.: Grants – Santen, during the conduct of the study; Grants – Carl Zeiss Meditec, Microsoft; Personal fees – Genentech, Gyroscope, Johnson & Johnson, Topcon, US Food & Drug Administration, Verana Health, outside the submitted work.

D.P.C.: Grants and nonfinancial support – Santen; Nonfinancial support – Medisoft, during the conduct of the study; Grants and personal fees –

Allergan, Apellis, Centervue; Personal fees – THEA, Roche, outside the submitted work.

A.J.: Nonfinancial support – Santen, Medisoft, during the conduct of the study; Financial support to attend the Moorfields International Glaucoma Symposium 2019, outside the submitted work.

N.A.: Nonfinancial support – Santen, Medisoft, during the conduct of the study; Personal fees – Santen (speaker honorarium), outside the submitted work.

A.C.: Financial support – Santen, during the conduct of the study.

C.D.: Nonfinancial support – Santen, during the conduct of the study; Personal fees – Thea; Grants – Glaukos Educational grant; Personal fees – AbbVie (Allergan), Ivantis, Rheon Medical, Iridex, Sight Sciences, outside the submitted work.

Y.K.: Grants and nonfinancial support – Santen; Nonfinancial support – Medisoft, during the conduct of the study.

G.M.: Grants and nonfinancial support – Santen; Nonfinancial support – Medisoft, during the conduct of the study; Personal fees – CenterVue, SpA, outside the submitted work.

Sponsored by Santen Inc., which participated in the data analysis, interpretation of the data, preparation, review, and approval of the manuscript.

HUMAN SUBJECTS: Human subjects were not included in this study. This retrospective database study was approved by the Western Institutional Review Board and deemed to be non-human subjects research as only de-identified IRC and imaging data were used. All research adhered to the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Kihara, Montesano, Chabi, Crabb, Lee

Data collection: Amerasinghe, Dimitriou, Jacob, Chabi.

Analysis and interpretation: Kihara, Montesano, Crabb, Lee

Obtained funding: N/A

Overall responsibility: Kihara, Montesano, Chen, Chabi, Crabb, Lee

Abbreviations and Acronyms:

CI = confidence interval; **Cp-OCT** = circum papillary OCT;

dB = decibels; **DL** = deep learning; **HFA** = Humphrey Field Analyzer;

HVF = Humphrey visual field; **IR** = infrared reflectance; **ONH** = optic

nerve head; **PD** = pattern deviation; **PMAE** = pointwise mean absolute error; **RNFL** = retinal nerve fiber layer; **SLO** = scanning laser ophthalmoscopy; **TD** = total deviation; **VF** = visual field.

Keywords:

Artificial intelligence, Deep learning, Glaucoma, OCT, Perimetry, Structure-function, Visual field.

Correspondence:

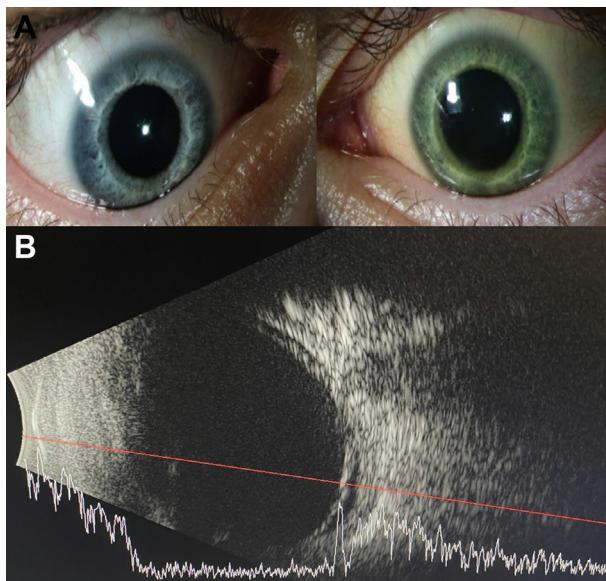
Aaron Y. Lee, MD, Department of Ophthalmology, University of Washington, 325 Ninth Avenue, Box 359608, Seattle, WA 98104. E-mail: aaronylee@gmail.com.

References

- Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in early glaucoma. *Br J Ophthalmol*. 1998;82:352–361.
- Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311:1901–1911.
- Hart Jr WM, Becker B. The onset and evolution of glaucomatous visual field defects. *Ophthalmology*. 1982;89: 268–279.
- Artes PH, Iwase A, Ohno Y, et al. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43: 2654–2659.
- Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold estimates from full threshold, ZEST, and SITA-like strategies, as determined by computer simulation. *Invest Ophthalmol Vis Sci*. 2003;44:4787.
- Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol*. 2014;98:ii15–ii19.
- Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology*. 2000;107:1809–1815.
- Turpin A, Sampson GP, McKendrick AM. Combining ganglion cell topology and data of patients with glaucoma to determine a structure–function map. *Invest Ophthalmol Vis Sci*. 2009;50:3249–3256.
- Jansonius NM, Schiefer J, Nevalainen J, et al. A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: average course, variability, and influence of refraction, optic disc size and optic disc position. *Exp Eye Res*. 2012;105:70–78.
- Jansonius NM, Nevalainen J, Selig B, et al. A mathematical description of nerve fiber bundle trajectories and their variability in the human retina. *Vision Res*. 2009;49:2157–2163.
- Lamparter J, Russell RA, Zhu H, et al. The influence of inter-subject variability in ocular anatomical variables on the mapping of retinal locations to the retinal nerve fiber layer and optic nerve head. *Invest Ophthalmol Vis Sci*. 2013;54:6074–6082.
- Denniss J, McKendrick AM, Turpin A. An anatomically customizable computational model relating the visual field to the optic nerve head in individual eyes. *Invest Ophthalmol Vis Sci*. 2012;53:6981–6990.
- Qiu K, Schiefer J, Nevalainen J, et al. Influence of the retinal blood vessel topography on the variability of the retinal nerve fiber bundle trajectories in the human retina. *Invest Ophthalmol Vis Sci*. 2015;56:6320–6325.
- Turpin A, McKendrick AM. Improving personalized structure to function mapping from optic nerve head to visual field. *Transl Vis Sci Technol*. 2021;10:19.
- Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res*. 2007;26:688–710.
- Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. 2010;51:5657–5666.
- Hashimoto Y, Asaoka R, Kiwaki T, et al. Deep learning model to predict visual field in central 10° from optical coherence tomography measurement in glaucoma. *Br J Ophthalmol*. 2021;105:507–513.
- Park K, Kim J, Lee J. A deep learning approach to predict visual field using optical coherence tomography. *PLoS One*. 2020;15:e0234902.
- Asaoka R, Murata H, Matsuura M, et al. Improving the structure–function relationship in glaucomatous visual fields by using a deep learning-based noise reduction approach. *Ophthalmol Glaucoma*. 2020;3:210–217.
- Mariottini EB, Datta S, Dov D, et al. Artificial intelligence mapping of structure to function in glaucoma. *Transl Vis Sci Technol*. 2020;9, 19–19.
- Shin J, Kim S, Kim J, Park K. Visual field inference from optical coherence tomography using deep learning algorithms: a comparison between devices. *Transl Vis Sci Technol*. 2021;10:4.
- Datta S, Mariottini EB, Dov D, et al. RetiNerveNet: using recursive deep learning to estimate pointwise 24-2 visual field data based on retinal structure. *Sci Rep*. 2021;11:1–10.
- Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci*. 2000;41:2201–2204.
- Yohannan J, Wang J, Brown J, et al. Evidence-based criteria for assessment of visual field reliability. *Ophthalmology*. 2017;124:1612–1620.
- Heijl A, Patella VM, Bengtsson B. *The Field Analyzer Primer: Effective Perimetry*. Carl Zeiss Meditec; 2012.
- Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *International Conference on Machine Learning*. PMLR; 2019:6105–6114.
- Dozat T. Incorporating Nesterov Momentum into Adam. 2016. Available at: <https://openreview.net/pdf?id=OM0jvwB8jIp57ZJjtNEZ>. Accessed June 17, 2021.
- Pennebaker GE, Stewart WC. Temporal visual field in glaucoma: a re-evaluation in the automated perimetry era. *Graefes Arch Clin Exp Ophthalmol*. 1992;230:111–114.
- Wall M, Lee EJ, Wanzer RJ, et al. Temporal wedge defects in glaucoma: structure/function correlation with threshold automated perimetry of the full visual field. *J Glaucoma*. 2020;29:191–197.
- 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;47:2901–2911.

31. Montesano G, Ometto G, Hogg RE, et al. Revisiting the Drasdo model: implications for structure-function analysis of the macular region. *Transl Vis Sci Technol.* 2020;9:15.
32. van der Schoot J, Vermeer KA, de Boer JF, Lemij HG. The effect of glaucoma on the optical attenuation coefficient of the retinal nerve fiber layer in spectral domain optical coherence tomography images. *Invest Ophthalmol Vis Sci.* 2012;53:2424–2430.
33. 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;127:346–356.
34. Yu H-H, Maetschke SR, Antony BJ, et al. Estimating global visual field indices in glaucoma by combining macula and optic disc OCT scans using 3-dimensional convolutional neural networks. *Ophthalmol Glaucoma.* 2021;4:102–112.
35. Hood DC, Fortune B, Mavrommatis MA, et al. Details of glaucomatous damage are better seen on OCT en face images than on OCT retinal nerve fiber layer thickness maps. *Invest Ophthalmol Vis Sci.* 2015;56:6208.
36. Lee J, Kim YW, Ha A, et al. Estimating visual field loss from monoscopic optic disc photography using deep learning model. *Sci Rep.* 2020;10:21052.
37. Guo Z, Kwon YH, Lee K, et al. Optical coherence tomography analysis based prediction of Humphrey 24-2 visual field thresholds in patients with glaucoma. *Invest Ophthalmol Vis Sci.* 2017;58:3975–3985.
38. Russell RA, Malik R, Chauhan BC, et al. Improved estimates of visual field progression using Bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci.* 2012;53:2760.
39. Denniss J, McKendrick AM, Turpin A. Towards patient-tailored perimetry: automated perimetry can be improved by seeding procedures with patient-specific structural information. *Transl Vis Sci Technol.* 2013;2:3.
40. Montesano G, Rossetti LM, Allegrini D, et al. Improving visual field examination of the macula using structural information. *Transl Vis Sci Technol.* 2018;7:36.
41. Ganeshrao SB, McKendrick AM, Denniss J, Turpin A. A perimetric test procedure that uses structural information. *Optom Vis Sci.* 2015;92:70–82.
42. Marín-Franch I, Swanson WH. The visualFields package: a tool for analysis and visualization of visual fields. *J Vis.* 2013;13:10.

Pictures & Perspectives



Iris Heterochromia and Scleral Yellowing Secondary to Self-Injury

An 18-year-old man with a history of Potocki-Lupski syndrome, intellectual disability, and autism presented with 2 weeks of progressive scleral yellowing and change in blue iris color to green in the left eye (Fig A). The patient's mother reported a year-long history of frequent rubbing, hitting, and multiple attempts at self-enucleation. B-scan ultrasound of the left eye revealed a total retinal detachment (Fig B) without evidence of a foreign body. Clinical examination revealed an anterior chamber filled with yellow turbid fluid causing the patient's blue iris to appear green. Subsequent episcleral drainage resulted in the unilateral pseudo-icteric appearance (Magnified version of Fig A-B is available online at www.aoajournal.org).

GRANT A. JUSTIN, MD

JORDAN D. DEANER, MD

DILRAJ S. GREWAL, MD

Department of Ophthalmology, Duke University Medical Center, Durham, North Carolina