City Research Online

City St George's, University of London

Citation: Kihara, Y., Montesano, G., Chen, A., Amerasinghe, N., Dimitriou, C., Jacob, A., Chabi, A., Crabb, D. P. & Lee, A. Y. (2022). Policy-Driven, Multimodal Deep Learning for Predicting Visual Fields from the Optic Disc and OCT Imaging. Ophthalmology, 129(7), pp. 781-791. doi: 10.1016/j.ophtha.2022.02.017

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/32191/

Link to published version: https://doi.org/10.1016/j.ophtha.2022.02.017

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Ophthalmology. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

Ophthalmology. 2022 July ; 129(7): 781–791. doi:10.1016/j.ophtha.2022.02.017.

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

Yuka Kihara[1,*],

Giovanni Montesano[2,*],

Andrew Chen[1],

Nishani Amerasinghe[3],

Chrysostomos Dimitriou[4],

Aby Jacob[3],

Almira Chabi, MD[5],

David P. Crabb[2],

Aaron Y. Lee, MD[1]

[1]University of Washington, Department of Ophthalmology, Seattle, WA, USA

[2]City, University of London, Optometry and Visual Sciences, London, UK

[3]University Hospital Southampton NHS Foundation Trust, Southampton, UK

[4]Colchester Hospital, East Suffolk and North Essex NHS Foundation Trust, Colchester, UK

[5]Santen, Emeryville, CA, USA

Abstract

Corresponding author: Aaron Y. Lee, MD, Assistant Professor, Department of Ophthalmology, University of Washington, 325 Ninth Avenue, Box 359608, Seattle, WA 98104, USA, T: (206) 543-7250, F: (206) 685-7055, aaronylee@gmail.com.

*Authors contributed equally and should be considered as co-first authors

Author Contributions:

Research design: Kihara, Montesano, Chabi, Crabb, Lee

Data acquisition and/or research execution: Amerasinghe, Dimitriou, Jacob, Chabi

Data analysis and/or interpretation: Kihara, Montesano, Crabb, Lee

Manuscript preparation: Kihara, Montesano, Chen, Chabi, Crabb, Lee

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Meeting Presentations:

This work has not been presented elsewhere.

HUMAN SUBJECTS:

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 non-human subjects research as only de-identified clinical and imaging data were used.

No animal studies were included in this study

This article contains additional online-only material. The following should appear online-only: Supplementary Clips 1 and 2; Supplementary Appendix.

Supplementary Material available at http://www.aaojournal.org

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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

Design: Retrospective, cross-sectional database study

Participants: 6437 patients undergoing routine care for glaucoma in three clinical sites in the UK.

Methods: OCT and infrared reflectance (IR) optic disc imaging was paired with the closest VF within 7 days. Efficient-Net B2 was used to train two 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 two model predictions.

Main Outcome Measures: Pointwise Mean Absolute Error (PMAE)

Results: A total of 5078 imaging 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] dB compared to the IR image of the disc alone and 0.060 [0.047, 0.073] dB compared 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.

Abstract

We used a large, real-world dataset to train a multimodal interpretable deep learning method to predict visual field sensitivity values in patients with glaucoma.

Keywords

Diagnostic testing; Glaucoma; Perimetry; Posterior segment optical coherence tomography; Deep learning; Artificial intelligence; Structure-function; Perimetry; Glaucoma; Visual field; Optical coherence tomography

INTRODUCTION

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. VF 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 Optical Coherence Tomography (SD-OCT)[6]. The latter has the advantage of providing cross-sections of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

retina and of the ONH that allow precise evaluation of their layers. Structural damage from glaucoma often manifests itself as thinning of the retinal nerve fibre 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 two 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 non-linear spatial mapping of SAP locations to the ONH. Although empirical maps exist [7–10] to describe average trajectories, these are often difficult to customize for individual patients [11–14] and become unreliable with increasing distance from the ONH[15]. 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[16]. Previous work has shown the potential of machine learning techniques to model such a complex problem[17]. More recently, advancements in the field of Artificial Intelligence (AI) and deep learning (DL), especially for image analysis with the introduction of Convolutional Neural Networks (CNN), have reignited the interest around structure-function predictions in glaucoma [18–20]. Recent attempts focused on predicting SAP sensitivity from segmented Cp-RNFL thickness profiles[21–23]. 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 CNN architectures.

Our work tested the hypothesis that a hybrid DL method combining information from paired ONH and Cp-OCT images can not only improve the prediction of SAP sensitivity directly from structural data without segmentation but also lead to a fully AI-derived structure function mapping. We used a large clinical database acquired from routine glaucoma clinics to develop two separate DL sub-models, each able to provide an estimate of VF sensitivity from either OCT or infrared SLO images. A third policy DL model then combines the two predictions by evaluating feature vectors generated by the two sub-models. Compared to previous work[21–23], 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 non-human subjects research as only de-identified clinical and imaging data were used. A large database of anonymized clinical data was curated from three regionally different National Health Service glaucoma clinics in the United Kingdom. All data were recorded into an Electronic Medical Record (EMR) platform (Medisoft, v6, Leeds, UK). All VFs were done with a Humphrey Field Analyzer (HFA, Zeiss Meditec, Dublin, CA, USA)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

was used to generate VFs. We included all Humphrey visual fields (HVF) performed with a Swedish interactive threshold algorithm (SITA) Standard/Fast or Full-threshold strategy using a 24-2 VF pattern with a false positive rate ≤ 15%, and that could be matched with a Cp-OCT scan (see later). We decided not to take false negative rates into account because of their strong relationship with the level of glaucoma damage[24] and because, together with fixation losses, they have been shown to be poor predictors of test reliability[25]. 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 attributed 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 two eyes, as these are typical features of VF defects from post-chiasmatic lesions (details provided in Supplemental Figure 1; available at http://www.aaojournal.org). Additionally, we carefully excluded patients that could have retinal diseases, by identifying patients who received intravitreal injections or were labeled in the EMR as having any other retinal or ocular diseases (Figure 1).

OCT scan patterns of the optic nerve from Heidelberg Spectralis (either 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 manually selected central point of the optic disc during acquisition by the photographer was used to crop a 260x260 pixel image at a native resolution as an eight-bit grayscale image. The circular OCT B-scan was resized to a 1:1 aspect ratio and resized to 260x260 as an eight-bit grayscale image. Each of the paired HVFs was collapsed to a one-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 dimensional vector at the time of evaluation. The VF sensitivity values were clipped between 0 and 40 dB and normalized to be between 0 and 1. The upper bound was chosen as it is commonly considered the highest plausible dB value in visually healthy subjects in standard testing conditions[26]. The imaging data was 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 [27], 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 260x260 pixels. EfficientNet has been shown to achieve better performance compared to other well-known CNN architectures such as ResNet and DenseNet with a smaller amount of parameters[27]. For comparison, we report the prediction performance of our model with different architectures in Supplemental Table 2 and 3 (available at http://www.aaojournal.org). The batch size was set to 32, NAdam was used as the optimizer[28], and the initial learning-rate was set to $5x10^{-3}$. The last layer of the model had 54 nodes

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

corresponding to the 54 VF sensitivity points on the 24-2. Our dataset was imbalanced; the majority of VF sensitivity values were around 30 dB, which caused the model to fail to learn patterns of advanced VF loss. To overcome this limitation, we employed the sensitivity-weighted mean square error for the loss function

$$mean((1/(1+\alpha \times yTrue)) \times (yTrue - yPred)^2)$$

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 α = 2 for our training. The MAE reported for the test dataset was calculated without applying any sensitivity-weights.

A policy deep learning model was then constructed to take the feature maps from both models and combine using a 1x1 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 OCT prediction being closer to true (observed) HVF. The output of the policy network was a 54 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. Since this value was generated by a sigmoid 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 two 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. el. A network diagram of the complete end-to-end multimodal DLdeep learning model is shown in Supplemental Figure 2 (available at http://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 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 (Supplemental Clips 1 and 2; available at http://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 (STD) 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 a total of 24,248 patients. After applying inclusion and exclusion criteria, the total number of eligible study patients was 6,437 (11,025 eyes). A flow chart of the patient selection process is shown in Figure

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1. The training, validation, and held-out test set contained 11,417 HVF-OCT pairs for 3,725 patients, 4,845 pairs for 1,323 patients, and 5,078 pairs for 1,389 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 three mutually exclusive sets of patients are shown in Table 1.

The training curves for the three DL models are provided in Supplemental Figure 3 (available at http://www.aaojournal.org). The final PMAE of the DL model using the IR SLO image of the disc alone was 3.62 dB (95% CI: 3.54 to 3.68 dB). The PMAE of the single modality DL model using the OCT alone was slightly better at 3.19 dB (95% CI: 3.13 to 3.25 dB). The best performing model was the policy fusion DL model that combined the predictions from the OCT and disc sub-models with a PMAE of 3.13 dB (95% CI: 3.07 to 3.19 dB). 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 sub-model and the policy fusion DL model was 0.060 dB (95% CI: 0.047 to 0.073 dB). The paired difference in PMAE between the disc sub-model and the policy fusion DL model was 0.485 [0.438, 0.533] dB. 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 Supplemental Figure 4 (available at http://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 (Figure 3). In order 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. Video clips of the perturbation experiments are shown in Supplemental Clips 1 and 2 (available at http://www.aaojournal.org) for the disc and OCT respectively for one 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 (**for the Policy**). The structure-function mapping for the whole held-out test set is reported in Figure 5 for the Cp-RNFL. The map was built by plotting the distribution of the circumpapillary angles at which the highest peak of the heat-map was located, in polar coordinates. The map was also compared with previously reported structure-function maps based on anatomical data [9,10,14,15,29]. 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 EMR. As expected, the performance was generally worse, but confirmed the improvement brought by the policy network (Supplementary material).

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 on what basis the policy network makes such a choice. The observation that the preference of the policy network appeared to be influenced by the quality of each image raises an interesting question: is the policy choice linked with the actual superiority on the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

prediction performance of the two models? To answer this question, we added several levels of noise to either 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 Supplemental Figure 5 (available at http://www.aaojournal.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 multi-modal deep learning 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 whole process simultaneously provides three distinct outputs (prediction from OCT, prediction from SLO of the disc, and policy-based fusion of the two results), with each producing directly interpretable results (HVF sensitivity maps). Importantly, the policy model is able to selectively choose between predictions from the two sub-models, thereby improving, for example, the spatial localization of deep defects (Figure 2B). The preference of the policy network appeared to be influenced by the quality of each image. Therefore, the policy model seems adept at working with real-world 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 Material (available at http://www.aaojournal.org) by simulating different amounts of noise for either the OCT or the disc image.

Structure-Function Mapping

Our structure-function mapping was agnostically learned by the two sub-models from the data alone. Remarkably, the OCT-based mapping yielded results in agreement with other mapping schemes based on anatomical features of the RNFL bundles[9,10,14,15,29]. In Figure 5, we represent the structure-function mapping for the Cp-OCT scans in the test set by plotting the angle at which the peak of each heatmap (see Figure 4 C) 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 Cp-RNFL scans. In many

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of the locations, this matched the inter-individual variability observed by Jansonius et al.[29] 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 [30,31]. One notable difference is the mapping of the macular locations, which in our model are shifted more towards the superior and inferior poles of the ONH (Figure 5) compared with Jansonius et al.[29] This is another important line of evidence to support the novelty of our technique because Jansonius et al.[29] did not account for the retinal ganglion cell displacement in the macular region [32,33] A similar difference was recently reported by Turpin and McKendrick[14], whose model is also reported in Figure 5 as a comparison.

From the example 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. Similarly, the heatmap produced for the Cp-OCT scans mainly focused on the RNFL (example in Figure 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[34]. 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 [35,36] but few have attempted pointwise predictions of the VF. Park et al.[19] and Hashimoto et al.[18] used fully segmented OCT thickness maps. Mariottoni et al. also predicted pointwise VF sensitivities agnostically from peripapillary OCT RNFL thickness profiles[21]. 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[34,37]. Fewer attempts have instead been made to utilize optic disc images to predict VF data.[38] Interestingly, our data indicate the existence of additional information gained by combining the two imaging modalities. The error profile of our prediction stratified by sensitivity was similar to previous publications,[21] 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 achievable predictive performance, and by the previously mentioned floor effect in the structural data[16]. Such a feature is clearly demonstrated by the positive bias at lower sensitivities (Figure 3). Notably, non-linear methods have proven more successful than linear models at overcoming this issue [17,18,21,39], but the problem persists in all these attempts.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Practical Implications and Future Directions

VF 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 progression[40] or into the VF test itself[41–43] 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 and/or practice patterns. We plan to prospectively evaluate and validate this DLdeep learning policy model in a clinical trial setting in the future to determine its generalizability. Moreover, our evaluation of the performance was based on MAE 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 one condition (i.e. glaucoma), one VF pattern (24-2), and one imaging modality (SLO/Cp-OCT obtained with a Spectralis SD-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 two 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 focussed 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 like 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 derived metric would require a normative dataset of imaging-predicted HVFs, 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[44] It is important to note that these are only illustrative examples. Moreover, TD and PD maps derived a *posteriori* from predicted sensitivities

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

might not necessarily be consistent with the original and could differ from the output of an algorithm trained to predict the same maps directly.

In conclusion, 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 two results. The DL models generate structure-function maps compatible with established anatomical 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to 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 would also like to thank Linda McInerney from Hinchingbrooke Hospital, North West Anglia NHS Foundation Trust, UK. Support with manuscript editing and formatting was provided by Katie Alexander, PhD, Helios Medical Communications, Alderley Park, Cheshire, UK, which was funded by Santen Inc., Emeryville, CA, USA.

Financial Disclosure:

This analysis was sponsored by Santen Inc., Emeryville, CA, USA, who participated in the data analysis, interpretation of the data, preparation, review, and approval of the manuscript.

Abbreviations and Acronyms:

AI	artificial intelligence
CI	confidence interval
CNN	convolutional neural networks
Cp-OCT	circumpapillary optical coherence tomography
Cp-RNFL	circumpapillary retinal nerve fiber layer
DL	deep learning
F	female
FP	false positive
FT	full threshold
HFA	Humphrey Field Analyzer
HVF	Humphrey visual field
I	inferior
IR	infrared reflectance

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

IVT	intravitreal
M	male
MAE	mean absolute error
MD	mean deviation
N	nasal
OCT	optical coherence tomography
ONH	optic nerve head
PMAE	pointwise mean absolute error
PSD	pattern standard deviation
RNFL	retinal nerve fiber layer
S	superior
SAP	standard automated perimetry
SD-OCT	spectral domain optical coherence tomography
SITA	Swedish interactive threshold algorithm
SLO	scanning laser ophthalmoscopy
STD	standard deviation
T	temporal
VF	visual field

REFERENCES

1. Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in early glaucoma. Br J Ophthalmol 1998;82:352–361. [PubMed: 9640180]
2. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. JAMA 2014;311:1901–1911. [PubMed: 24825645]
3. Hart WM Jr, Becker B. The onset and evolution of glaucomatous visual field defects. Ophthalmology 1982;89:268–279. [PubMed: 7088510]
4. 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. [PubMed: 12147599]
5. 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. Investigative Opthalmology & Visual Science 2003;44:4787. Available at: 10.1167/iovs.03-0023.
6. Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. British Journal of Ophthalmology 2014;98:ii15–ii19. Available at: 10.1136/bjophthalmol-2013-304326. [PubMed: 24357497]
7. 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. [PubMed: 23099334]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

8. 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. [PubMed: 19539641]
9. 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. Available at: https://pubmed.ncbi.nlm.nih.gov/11013178/ [Accessed May 9, 2021].
10. 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. Available at: [Accessed May 9, 2021]. [PubMed: 19324839]
11. Lamparter J, Russell RA, Zhu H, et al. The influence of intersubject 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. [PubMed: 23882689]
12. 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. [PubMed: 22969077]
13. 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. Available at: [Accessed May 9, 2021]. [PubMed: 26436884]
14. Turpin A, McKendrick AM. Improving Personalized Structure to Function Mapping From Optic Nerve Head to Visual Field. Transl Vis Sci Technol 2021;10:19–19. Available at: [Accessed May 9, 2021].
15. 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. [PubMed: 19539641]
16. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. Prog Retin Eye Res 2007;26. Available at: https://pubmed.ncbi.nlm.nih.gov/17889587/ [Accessed May 9, 2021].
17. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. Invest Ophthalmol Vis Sci 2010;51. Available at: https://pubmed.ncbi.nlm.nih.gov/20505207/ [Accessed May 9, 2021].
18. 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. [PubMed: 32593978]
19. Park K, Kim J, Lee J. A deep learning approach to predict visual field using optical coherence tomography. PLoS One 2020;15:e0234902. [PubMed: 32628672]
20. 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. [PubMed: 32672618]
21. Mariottoni EB, Datta S, Dov D, et al. Artificial Intelligence Mapping of Structure to Function in Glaucoma. Transl Vis Sci Technol 2020;9:19–19. Available at: [Accessed May 9, 2021].
22. 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;10:4–4. Available at: [Accessed November 24, 2021].
23. Datta S, Mariottoni 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. Available at: [Accessed November 24, 2021]. [PubMed: 33414495]
24. 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. Available at: [Accessed November 24, 2021]. [PubMed: 10892863]
25. Yohannan J, Wang J, Brown J, et al. Evidence-based Criteria for Assessment of Visual Field Reliability. Ophthalmology 2017;124. Available at: https://pubmed.ncbi.nlm.nih.gov/28676280/ [Accessed November 24, 2021].
26. Heijl A, Patella VM, Bengtsoon B. The Field Analyzer Primer: Effective Perimetry. 2012.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

27. Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: International Conference on Machine Learning. PMLR; 2019:6105–6114. Available at: [Accessed June 17, 2021].
28. Dozat T Incorporating Nesterov Momentum into Adam. 2016. Available at: https://openreview.net/pdf?id=OM0jvwB8jlp57ZJjtNEZ [Accessed June 17, 2021].
29. 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. [PubMed: 23099334]
30. 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. [PubMed: 1577288]
31. Wall M, Lee EJ, Wanzek 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. [PubMed: 32108691]
32. 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. [PubMed: 17320143]
33. 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.
34. 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. [PubMed: 22427540]
35. 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. [PubMed: 31718841]
36. 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. [PubMed: 32826205]
37. 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. Available at: [Accessed June 11, 2021]. [PubMed: 26426403]
38. 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. [PubMed: 33273643]
39. 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. [PubMed: 28796875]
40. 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. Available at: [Accessed June 11, 2021]. [PubMed: 22467579]
41. 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. Available at: https://pubmed.ncbi.nlm.nih.gov/24049720/ [Accessed June 11, 2021].
42. Montesano G, Rossetti LM, Allegrini D, et al. Improving Visual Field Examination of the Macula Using Structural Information. Transl Vis Sci Technol 2018;7. Available at: https://pubmed.ncbi.nlm.nih.gov/30619656/ [Accessed June 11, 2021].
43. Ganeshrao SB, McKendrick AM, Denniss J, Turpin A. A perimetric test procedure that uses structural information. Optom Vis Sci 2015;92:70–82. Available at: [Accessed June 11, 2021]. [PubMed: 25415282]
44. Marín-Franch I, Swanson WH. The visualFields package: a tool for analysis and visualization of visual fields. J Vis 2013; 13. Available at: 10.1167/13.4.10.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

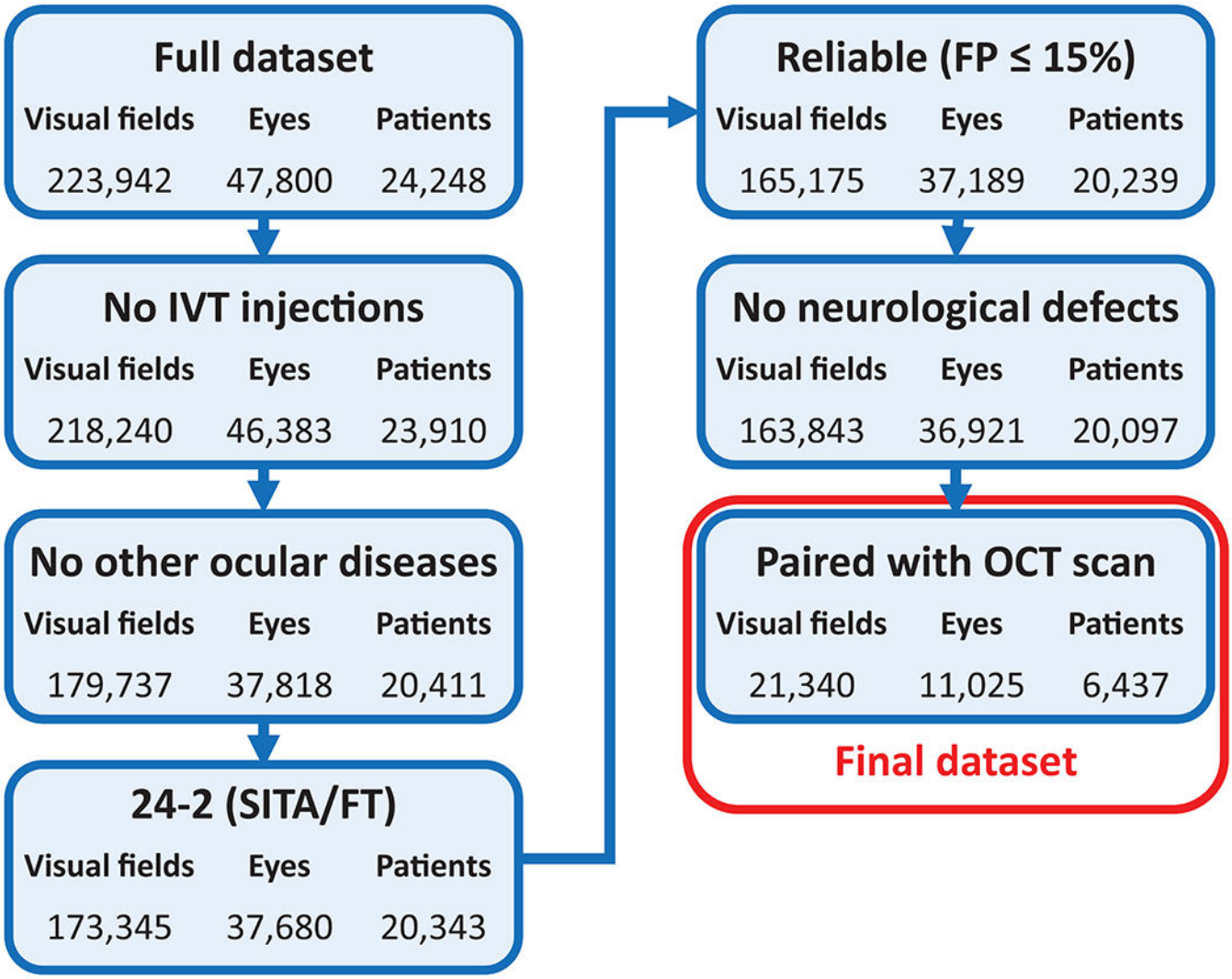

Figure 1.
Flowchart of the selection process from the full dataset of VF tests. FP = false positive; FT = full-threshold; IVT = intravitreal; SITA = Swedish interactive threshold algorithm; VF = visual field.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

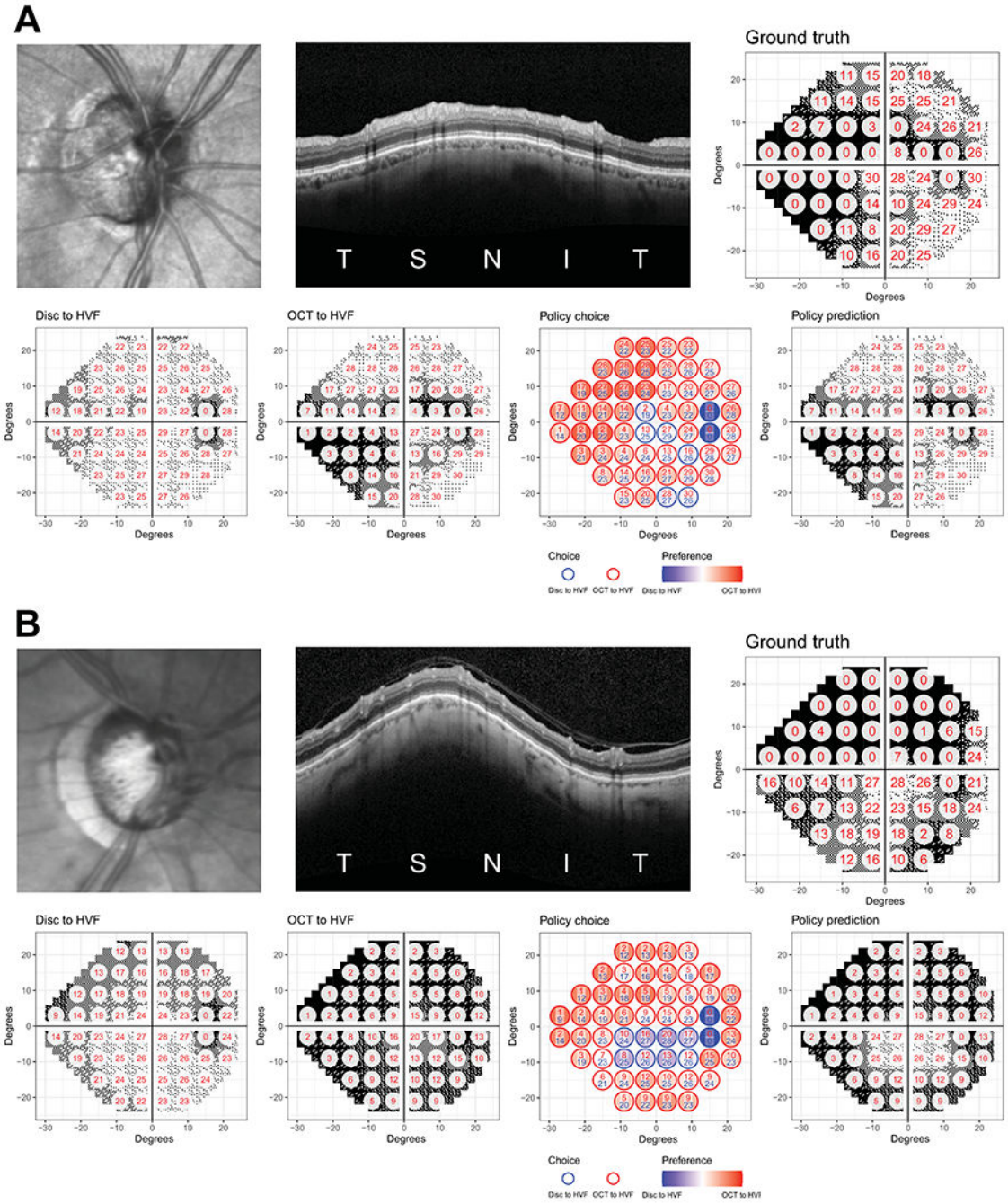

Figure 2.

Prediction examples from the test set. The top of each panel reports the disc and OCT images used for predictions alongside the ground- truth HVF. The bottom of each panel shows, in order, the predictions from the two sub-models, the choice score of the policy network for each location, and the final prediction of the policy network. **A**, Both sub-models 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 sub-model is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 VF. HVF = Humphrey visual field; I = inferior; N = nasal; OCT = optical coherence tomography;
S = superior; T = temporal.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

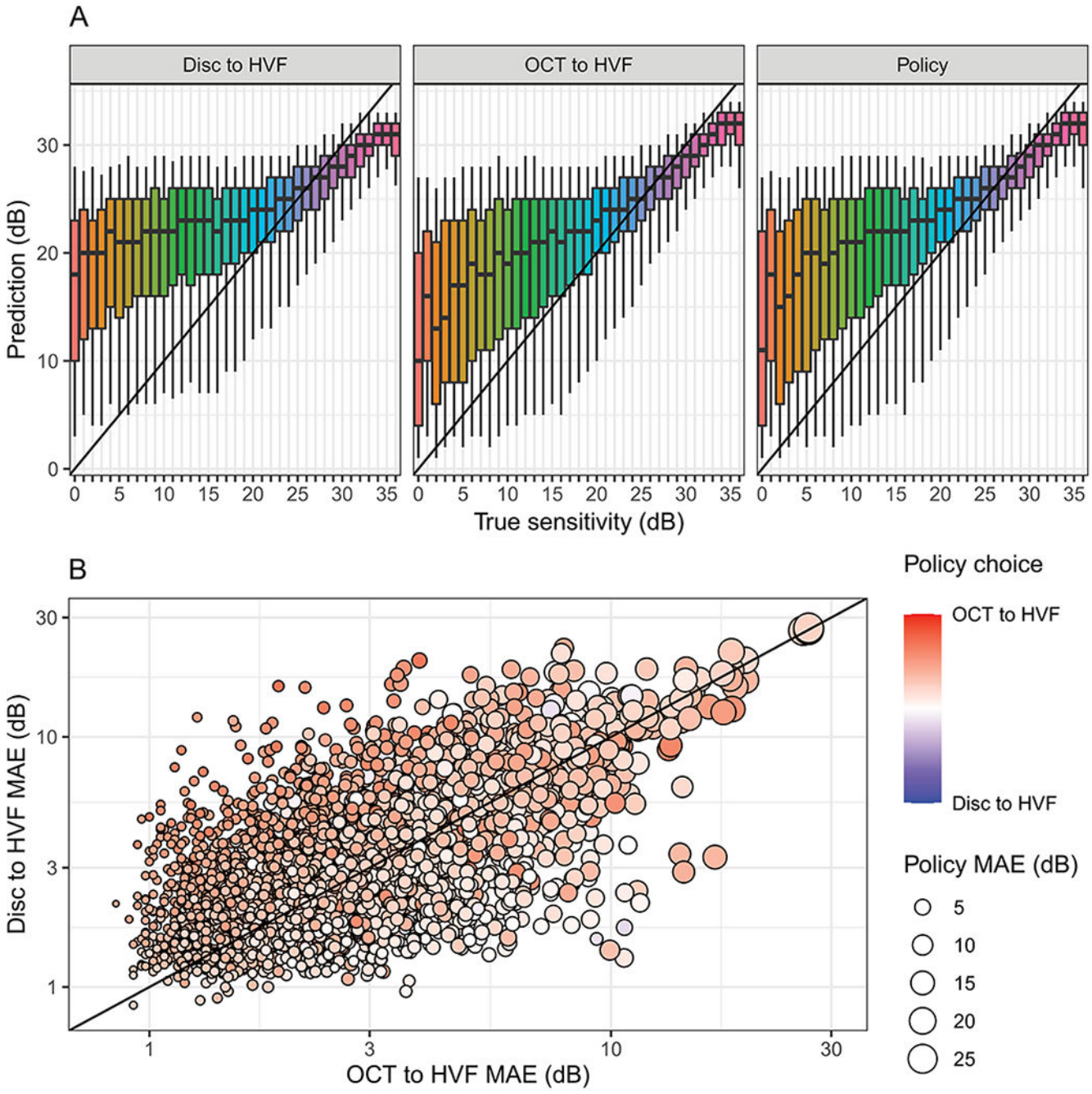

Figure 3.

A,) Predictions from the two sub-models 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 two sub-models and the final average choice from the policy network. Higher MAE in the disc to HVF model is generally associated with a preference for the OCT to HVF model and vice-versa. The diagonal line indicates equivalent MAE. HVF = Humphrey visual field; MAE = mean absolute error; OCT = optical coherence tomography.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

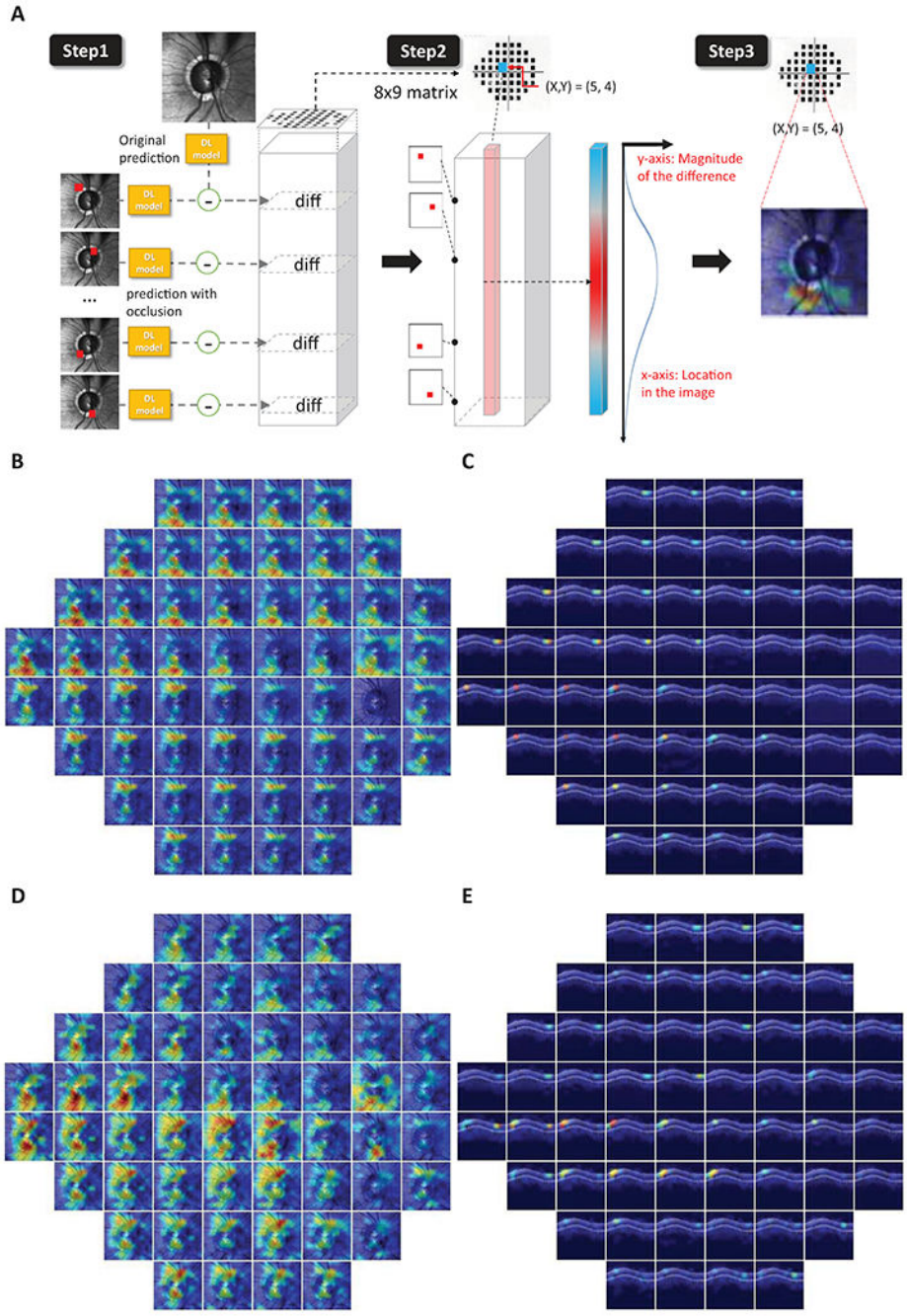

Figure 4.

A,) Process for generating heatmaps for the disc SLO and OCT models. We first took the difference in prediction between 'with' and 'without' occlusion at each occlusion location (Step1), computed the distribution of magnitude of the difference over occlusion locations (Step2), and projected the distribution back to 2 dimensional image space to obtain a heatmap (Step3). According to the location of the perturbation in the predicted VF, the magnitude of changes reflects where the network focuses on during the predictions. **B**, Heatmap for one patient-generated example based on our disc SLO. **C**, Heatmap for one

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

patient-generated example based on the OCT model. The perturbation processes of the corresponding example patient are also shown in Supplemental Clips 1 and 2 (available at http://www.aaojournal.org). **D**, Heatmap for one patient-generated example from the policy network obtained by occluding the disc SLO. **E,** Heatmap for one patient-generated example from the policy network obtained by occluding the OCT image. OCT = optical coherence tomography;
SLO = scanning laser ophthalmoscopy; VF = visual field.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

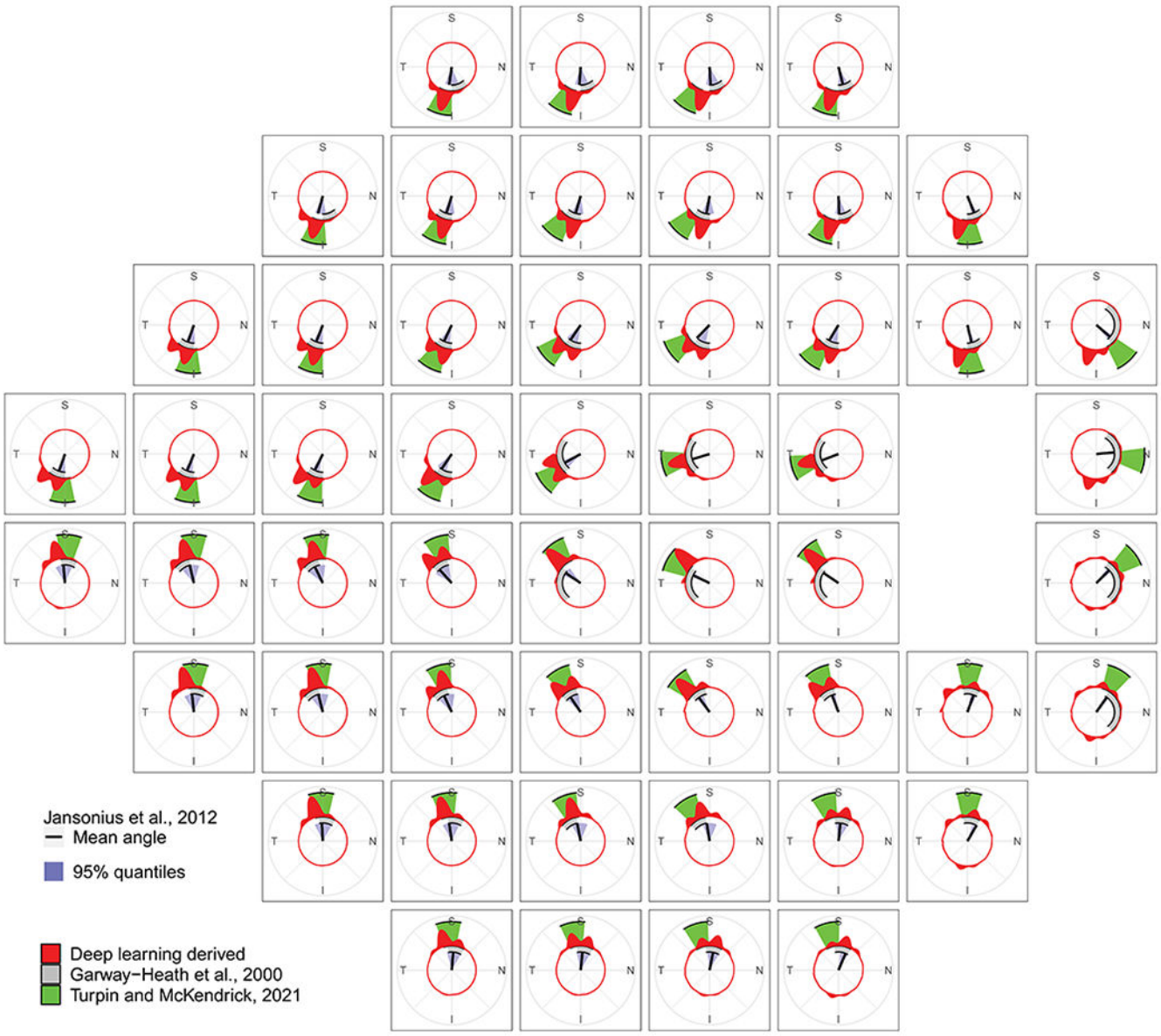

Figure 5.

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

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Descriptive Statistics of the Selected Sample. Continuous variables are reported as median [interquartile range].

	Training set	Test set	Validation set
# Patients	4,786	1,585	1,541
# Eyes	8,883	2,902	2,840
# Visual fields	22,866	7,370	7,411
# OCT scans	21,963	7,042	7,110
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)	2,492 / 2,294	791 / 794	743 / 798
Baseline age (years)	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 / 7,894	314 / 2,588	324 / 2,516

F = female; FP = false positive; IOP = intraocular pressure; M = male; MD = mean deviation; OCT = optical coherence tomography; PSD = pattern standard deviation.