

Estimating the Severity of Visual Field Damage From Retinal Nerve Fiber Layer Thickness Measurements With Artificial Intelligence

Xiaoqin Huang^{1,*}, Jian Sun^{1,2,*}, Juleke Majoor³, Koenraad Arndt Vermeer³, Hans Lemij³, Tobias Elze⁴, Mengyu Wang⁴, Michael Vincent Boland⁵, Louis Robert Pasquale⁶, Vahid Mohammadzadeh⁷, Kouros Nouri-Mahdavi⁷, Chris Johnson⁸, and Siamak Yousefi^{1,9}

¹ Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, TN, USA

² German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany

³ Rotterdam Ophthalmic Institute, Rotterdam Eye Hospital, Rotterdam, The Netherlands

⁴ Eye Research Institute of Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA

⁵ Department of Ophthalmology, Massachusetts Eye and Ear, Boston, MA, USA

⁶ Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁷ Jules Stein Eye Institute, University of California Los Angeles, Los Angeles, CA, USA

⁸ Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, IA, USA

⁹ Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA

Correspondence: Siamak Yousefi, 930 Madison Ave., Suite 726, Memphis, TN 38163, USA. e-mail: siamak.yousefi@uthsc.edu

Received: March 31, 2021

Accepted: July 16, 2021

Published: August 16, 2021

Keywords: artificial intelligence (AI); visual fields; optical coherence tomography (OCT); estimating visual field

Citation: Huang X, Sun J, Majoor J, Vermeer KA, Lemij H, Elze T, Wang M, Boland MV, Pasquale LR, Mohammadzadeh V, Nouri-Mahdavi K, Johnson C, Yousefi S. Estimating the severity of visual field damage from retinal nerve fiber layer thickness measurements with artificial intelligence. *Transl Vis Sci Technol.* 2021;10(9):16, <https://doi.org/10.1167/tvst.10.9.16>

Purpose: The purpose of this study was to assess the accuracy of artificial neural networks (ANN) in estimating the severity of mean deviation (MD) from peripapillary retinal nerve fiber layer (RNFL) thickness measurements derived from optical coherence tomography (OCT).

Methods: Models were trained using 1796 pairs of visual field and OCT measurements from 1796 eyes to estimate visual field MD from RNFL data. Multivariable linear regression, random forest regressor, support vector regressor, and 1D convolutional neural network (CNN) models with sectoral RNFL thickness measurements were examined. Three independent subsets consisting of 698, 256, and 691 pairs of visual field and OCT measurements were used to validate the models. Estimation errors were visualized to assess model performance subjectively. Mean absolute error (MAE), root mean square error (RMSE), median absolute error, Pearson correlation, and R-squared metrics were used to assess model performance objectively.

Results: The MAE and RMSE of the ANN model based on the testing dataset were 4.0 dB (95% confidence interval = 3.8–4.2) and 5.2 dB (95% confidence interval = 5.1–5.4), respectively. The ranges of MAE and RMSE of the ANN model on independent datasets were 3.3–5.9 dB and 4.4–8.4 dB, respectively.

Conclusions: The proposed ANN model estimated MD from RNFL measurements better than multivariable linear regression model, random forest, support vector regressor, and 1-D CNN models. The model was generalizable to independent data from different centers and varying races.

Translational Relevance: Successful development of ANN models may assist clinicians in assessing visual function in glaucoma based on objective OCT measures with less dependence on subjective visual field tests.

Introduction

Primary open angle glaucoma (POAG) is a leading cause of irreversible blindness worldwide.^{1,2} Glaucoma causes the slow degeneration and eventual death of retinal ganglion cells (RGCs) and their attendant axons³ accompanied by characteristic structural changes and patterns of visual field loss.⁴ Major risk factors for POAG include elevated intra-ocular pressure (IOP), African ancestry, family history, and older age.^{5,6} Most of the affected individuals at the early stages of disease are unaware they have glaucoma, which in turn leads to delayed care and irreversible vision loss.⁷ Visual deterioration accelerates at comparatively late stages of glaucoma,^{8,9} concurrent with sharp increases in the costs of treatment.^{10–12} Therefore, early detection and proactive management of glaucoma would positively impact clinical and public health.

Currently, glaucoma is mainly diagnosed by visual field testing and optic nerve assessment through fundus photography or optical coherence tomography (OCT) imaging.¹³ Visual field testing through standard automated perimetry (SAP) provides a subjective psychophysical test that typically takes a few minutes to complete.¹⁴ Deterioration rates based on visual field tests have been used as functional surrogate end points in most recent glaucoma clinical trials.^{6,15,16} Although widely used, visual field testing generates surprisingly inconsistent and variable results, especially as the visual field deteriorates, and particularly in patients with suspected glaucoma.^{17,18}

These weaknesses underlie a general concern that visual field tests may be too insensitive or imprecise, or both, to adequately measure treatment efficacy in clinical trials over a short duration. In an attempt to address this deficiency, methods that rely on longitudinal visual field analysis and thus require either more frequent visual field tests over time or acquiring visual field tests over a longer period of time, are being used with the goal of generating a statistically reliable outcome with which to monitor glaucoma development. Both of these approaches incur more cost.^{19–22} Given the limitations of current visual field testing, there has been growing interest in OCT as a more reliable glaucoma assessment. Spectral domain OCT, a relatively newer generation ophthalmic imaging technique based on the principle of optical interferometry, noninvasively provides 2- and 3-dimensional high-resolution images of optic nerve head and surrounding peripapillary retinal layers as well as quantified measurements, all generated within a few seconds.²³

Artificial intelligence (AI) has made significant advancement in ophthalmology and glaucoma over the past few years.^{24–31} This early success has led to critical questions regarding whether OCT measurements are predictive of glaucoma status and, if so, whether AI might be utilized to provide objective, OCT-based monitoring of visual functional loss and glaucoma status. A successful solution might augment or replace subjective visual field testing with objective OCT imaging for glaucoma assessment.

Several teams have previously attempted to estimate visual field parameters from OCT measurements using conventional statistical and conventional machine learning approaches.^{32–34} Recently, deep learning models have been offered to estimate global and local visual field damage from raw OCT scans and quantified thickness measurements.^{35–37} Zhu et al. developed linear and nonlinear regression models to estimate visual fields from RNFL thickness measurements obtained from scanning laser polarimetry (SLP).³² The mean absolute error (MAE) between the observed and estimated visual field threshold sensitivity values was approximately 3.9 dB. However, both linear and nonlinear models significantly underestimated the true sensitivity values of eyes in the early stages and overestimated the sensitivity values of eyes in the moderate and advanced stages of glaucoma.

Bogunovic and colleagues developed several learning models, including support vector regressor machines (SVM), to estimate visual field from quantified OCT (Spectralis) measurements from 122 subjects and achieved a root mean square error (RMSE) of approximately 3.7 dB averaged across all visual field test locations.³³ The same team used quantified measurements of RNFL and ganglion cell and inner plexiform layers captured by Spectralis OCT for estimating visual field sensitivity values a few years later.³⁴ Using another small sample size with fewer than 100 subjects, the model significantly underestimates or overestimates true visual field values at both ends of the glaucoma spectrum. Moreover, small sample sizes make it challenging to generalize the findings.

Sugiura et al. developed a complex deep learning-based model to estimate sensitivity at visual field test locations using several OCT-derived retinal layer thickness profiles and achieved an RMSE of about 6.1 dB.³⁵ Christopher et al. proposed a deep learning model to both diagnose glaucoma and to estimate global visual field parameters from OCT-derived RNFL thickness maps, OCT en face images, and confocal laser scanning ophthalmoscopy (CSLO) images.³⁶ In detecting glaucomatous visual field damage, these deep learning models estimated mean deviation (MD) with MAE

of about 2.9 dB. However, in the absence of any visualization of the error distributions, it is challenging to understand whether this level of MAE is mainly due to existence of a greater number of normal eyes and eyes at the early stages of glaucoma, which typically have smaller estimation errors compared to eyes at the later stages of glaucoma. In the absence of visualization, it is also not obvious to assess whether the proposed model is biased toward the two ends of the glaucoma spectrums.

Yu et al. combined OCT images from macula and optic disc to estimate visual field global parameters using a 3-D deep learning model.³⁷ The best MD estimation accuracy of RMSE approximately 2.4 dB and MAE approximately 2.3 dB for MD achieved when OCT data of both macula and optic disc were combined. Despite relatively low mean error rates, the distribution of errors was heavily skewed and reflected the tendency toward overestimating visual field indices for eyes at the moderate to advanced stages of glaucoma. Furthermore, most of the proposed conventional and deep learning models, including Yu et al., were not validated using independent datasets to assess generalizability.

In this paper, we describe an artificial neural network (ANN) model to estimate MD based on a large dataset of RNFL thickness measurements from OCT circle scans. We provide validation of the model using three independent datasets from different races, different instruments, and different scanning types. We show that our model is significantly simpler than most of the recently proposed deep learning models, yet (1) achieves a competing degree of accuracy, (2) performs well at estimating visual field parameters of eyes at the early stages of glaucoma, while also providing reasonable accuracy in later stages of glaucoma, and (3) is generalizable to unseen cohorts. Moreover, our results also help establish the degree to which the accuracy visual fields can be estimated from OCT parameters and demonstrate how AI models might avoid overestimation of visual field parameters from OCT images of eyes at the later stages of glaucoma.

Methods

Subjects and Datasets

Four independent cohorts with different ethnicities were used in this study to develop and validate the AI model. Participants gave written informed consent, and institutional review boards (IRBs) were approved at the respective sites. Methods adhered to the tenets of the Declaration of Helsinki. All visual field tests

were collected from the Humphrey Field Analyzer II (Carl Zeiss Meditec, Inc., Dublin, CA, USA) using standard 24-2 testing pattern with the Swedish interactive thresholding algorithm and global parameters, including MD and PSD, were exported. Tests with greater than 33% fixation losses, 20% false-negative or false-positive error rates were excluded. For all eyes, the corresponding OCT images were required to be within 180 days from the visual field testing date.

The OCT data of the training/testing and the first two validation subsets were collected from Spectralis instruments (Heidelberg Engineering, Heidelberg, Germany) using 3.46 mm circular scans centered on optic disc. The OCT data of the third validation subset was collected from Cirrus instruments (Carl Zeiss Meditec, Inc.) using Optic Disc Cube protocol (200 × 200) within a 6 × 6 mm area. Spectralis OCT scans with signal strength <15 and Cirrus OCT scans with signal strength <6 were excluded.

The training/testing dataset included 1796 visual field and OCT pairs from 1796 eyes (1796 patients), the dataset was split at a ratio of 0.7, 0.1, and 0.2 for training, validation, and testing, respectively. This first validation dataset included 698 visual field and OCT pairs from 698 eyes of patients with glaucoma of the Rotterdam Eye Hospital. The second validation dataset included 256 visual field and OCT pairs from 64 eyes of 64 patients who visited the Jules Stein Eye Institute, University of California Los Angeles (UCLA), and 691 visual field and OCT pairs from 691 eyes of 691 patients visiting the Massachusetts Eye and Ear (MEE) glaucoma service. For the MEE dataset, the circular scans were estimated from the cube scan in 256 sectors. More specifically, the A-scans closest to 256 sectors on the 3.46 mm circle around the optic disc were selected from the original 200 × 200 and appropriate smoothing and interpolation was applied. This procedure is inspired by the circle scan that Cirrus approximates from cube scans and provides on printout.

Development of the Artificial Neural Network Model

Circular RNFL thickness values were averaged to generate 64 sectors for all OCT data (Fig. 1). An ANN model was then developed to estimate visual field MD values from 64 RNFL sectors using the training, validation, and testing dataset (Fig. 2). We developed several ANN models with different numbers of layers and neurons and eventually selected the simplest model with only one hidden layer and 256 neurons. Stochastic gradient descent (SGD) was used as the optimizer, root mean squared error (RMSE) was used as the loss

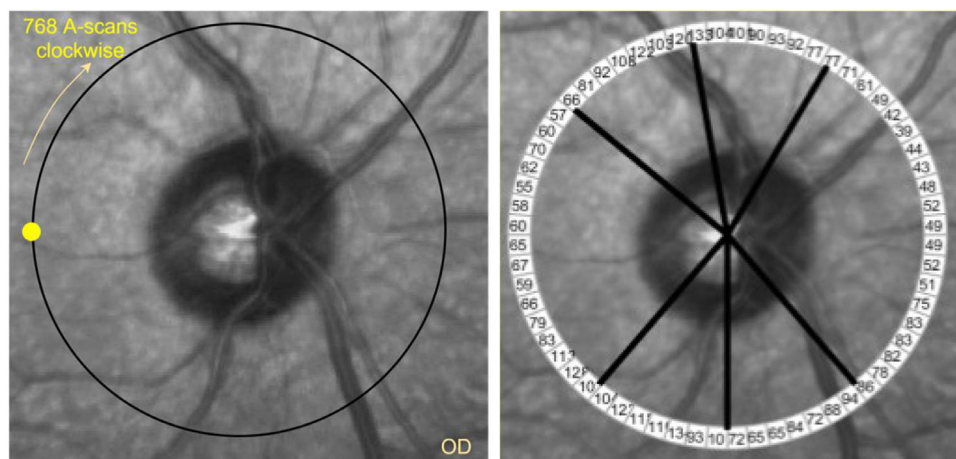


Figure 1. Circle scan around the optic disc of a sample right eye (OD). **Left:** A total of 768 A-scans are captured starting from the yellow circle clockwise. **Right:** Every 12 A-scans were averaged to generate 64 sectors around the optic disc.

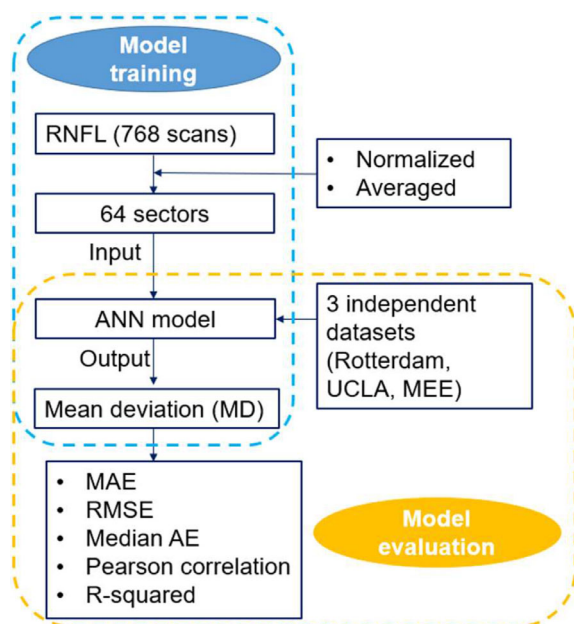


Figure 2. Diagram of the Artificial Neural Network (ANN) model for estimating visual fields from circumpapillary RNFL thickness measurements.

function for backpropagation, and the learning rate was set to 0.001. The model was trained for up to 1000 epochs.

Development of Multivariable Linear Regression, Random Forest, Support Vector Regressor, and 1-D CNN Models

Multivariable linear regression (LR) model was implemented by inputting seven RNFL global and sectoral parameters as the inputs to the model and

visual field MD as the output of the model. Random forest (RF), support vector regressor (SVR), and 1-D CNN models were implemented with 64 sectors as inputs and evaluated by MAE, RMSE and R-squared.

The number of trees was optimized for the RF model. A total of 100 estimators and default values of other parameters generated the least RMSE.

For SVR model, different kernels and regularization parameters were examined and optimized by a grid search and the model with radial basis function (RBF) kernel and C parameter (regularization) of 100 generated the least RMSE. The number of layers, neurons, optimizers, and learning rate were optimized by a grid search for the 1-D CNN model. The best performance of the 1-D CNN model was achieved with one convolutional layer with 256 neurons, kernel size of 3, one dense layer with 512 neurons, dropout of 0.25, and using an SGD optimizer at a learning rate of 0.001.

Evaluating Models

The accuracy of the ANN model in estimating visual field MD was assessed using MAE, RMSE, R-squared, and Pearson correlation. In addition to objective metrics, the distributions of errors were visualized using scatter plots to assess bias along the glaucoma spectrum.

We also performed ablation test on ANN model to uncover which sectors were more important in estimating MD from RNFL data. In each experiment, we excluded several sectors of the estimated visual field MD based on the remaining RNFL sectors. We then compared the accuracy of the model in terms of R-squared to see which group of excluded sectors impacted the accuracy significantly. We also performed

a similar experiment based on the RF model. More specifically, we identified and ranked more important sectors (features) in the RF model for estimating visual field MD from input RNFL sectors.

Results

The average age of the subjects in the testing and Rotterdam, and MEE independent datasets were 65.8, 66.0, and 60.8 years, respectively. About 56% of participants in the Rotterdam dataset were women. Table 1 shows the glaucoma severity level of eyes in all datasets. Figure 3 shows the distribution of MD for eyes in the testing and independent datasets. Whereas patients in the MEE dataset were mostly normal or in the early stages of glaucoma, patients in the UCLA dataset were at the later stages of glaucoma.

Table 2 presents the overall accuracy of all models in estimating visual field MD from RNFL data based on the testing subset. The ANN model estimated MD from RNFL data with R-squared of 0.64, MAE of 4.0 dB, and RMSE of 5.2 dB. Table 3 illustrates the performance of the ANN model based on the testing subset and the independent validation subsets. The R-squared, MAE, and RMSE of the ANN model using three independent subsets were in the range of 0.3–0.67 dB, 3.3–5.9 dB, and 4.4–8.4 dB, respectively.

Table 1. Average Value of RNFL and MD in Training and Three Independent Datasets

Dataset	RNFL (SD); μm	MD (SD); dB
Training		
Rotterdam	71.3 (20.3)	−8.5 (8.4)
UCLA	69.8 (20.0)	−6.7 (7.9)
MEE	61.0 (13.2)	−9.1 (6.3)
MEE	83.8 (14.5)	−3.7 (5.1)

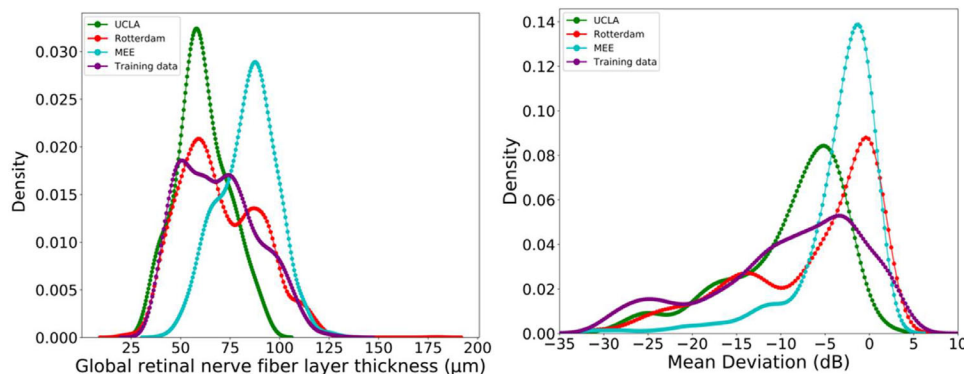


Figure 3. Distribution of eyes in the training and three independent datasets across glaucoma spectrum. **Left:** Distribution of eyes based on the global retinal nerve fiber layer thickness. **Right:** Distribution of eyes based on visual field mean deviation.

Figure 4 left shows the scatter plot of the true versus estimated visual field MD of the ANN model based on the testing subset. Figure 4 right shows the scatter plot of the true versus estimated visual field MD of the linear regression model based on the testing subset. Figure 5 demonstrates the scatter plot of the true versus estimated visual field MD of the ANN model based on three independent validation subsets.

Estimating visual field MD using a linear regression based on seven RNFL summary parameters, including global RNFL thickness and average RNFL thickness in temporal, temporal-superior, temporal-inferior, nasal, nasal-superior, and nasal-inferior, resulted in R-squared of 0.51 dB, MAE of 5.4 dB, and RMSE of 6.5 dB, in which all were significantly ($P < 0.01$) higher than the error rates of the ANN model (see Table 2).

Figure 5 demonstrates the scatter plot of the true versus estimated visual field MD of the ANN model based on the Rotterdam, UCLA, and MEE subsets. Table 3 illustrates the accuracy of the ANN model in estimating MD for eyes in the early and moderate to advanced stages of the glaucoma. R-squared of the ANN model based on the MEE dataset was negative. As we calculated the R-squared based on the sum of squares of the residual and total error, for MEE, the model provided a fit worse than a straight line.

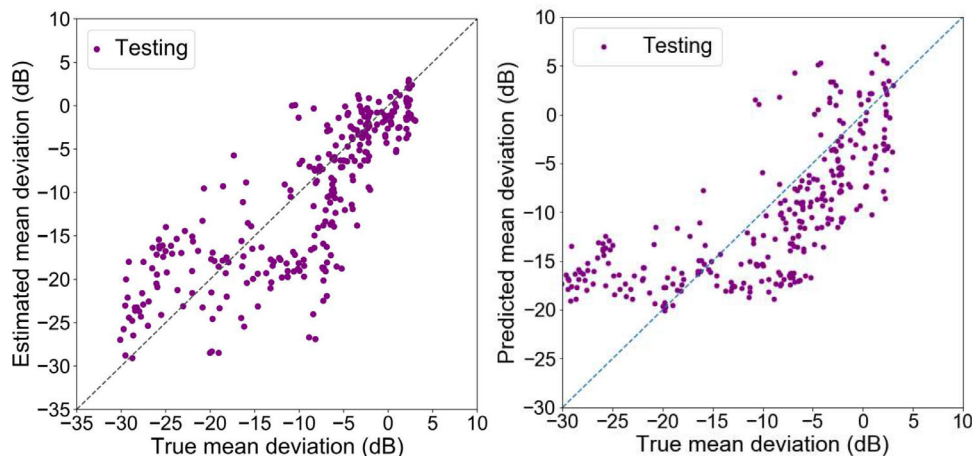
Table 4 presents the accuracy of the ANN model in estimating MD for eyes at different severity levels of glaucoma. Using testing subset, the ANN model estimated MD from RNFL data of eyes in the early stages of glaucoma ($\text{MD} \geq -6$ dB) with MAE of 3.6 dB and RMSE of 4.8 dB. The model's MAE and RMSE for eyes at the later stages of glaucoma ($\text{MD} < -6$ dB) were 4.6 dB and 5.9 dB, respectively. The MAE and RMSE of the model for eyes at the early stages of glaucoma ($\text{MD} \geq -6$ dB) using three independent subsets were in the range of 2.4–5.0 dB

Table 2. Estimation Error of Different Models Based on the Testing Subset

Model	MAE (dB) (95% Confidence Interval)	RMSE (dB) (95% Confidence Interval)	R-Squared (95% Confidence Interval)
ANN	4.0 (3.8–4.2)	5.2 (5.1–5.4)	0.64 (0.59, 0.68)
RF	4.0 (3.8–4.2)	5.4 (5.2–5.4)	0.47 (0.43–0.51)
SVR	4.2 (4.0–4.4)	5.7 (5.5–5.9)	0.41 (0.36–0.47)
1-D CNN	4.1 (3.9–4.3)	5.5 (5.3–5.7)	0.45 (0.35–0.54)
LR (7 summary parameters)	5.4 (5.2–5.6)	6.5 (6.3–6.7)	0.51 (0.45–0.56)
LR (64 sectors)	5.2 (5.0–5.4)	6.7 (6.5–6.9)	0.17 (0.14–0.23)

Table 3. Accuracy of the Artificial Neural Network (ANN) Model in Estimating Visual Field Mean Deviation (MD) From Retinal Nerve Fiber Layer (RNFL) Thickness Measurements

Dataset	Testing	Rotterdam	UCLA	MEE
Mean absolute error (MAE); dB; 95% CI	4.0 (3.8, 4.2)	3.3 (2.77, 3.83)	3.9 (3.58, 4.36)	5.9 (5.3, 6.6)
Root mean square error (RMSE); dB; 95% CI	5.2 (5.1, 5.4)	4.4 (3.72, 5.08)	5.3 (4.88, 5.83)	8.4 (7.4, 10.4)
Median absolute error; dB; 95% CI	3.1 (2.7, 3.7)	2.6 (2.23, 3.16)	2.9 (2.58, 3.37)	3.5 (3.1, 4.9)
Pearson correlation; 95% CI	0.81 (0.80, 0.83)	0.84 (0.81, 0.86)	0.61 (0.52, 0.68)	0.62 (0.57, 0.66)
R-squared; 95% CI	0.64 (0.59, 0.68)	0.67 (0.47, 0.86)	0.30 (0.11, 0.43)	−1.74 (−2.86, −0.77)

**Figure 4.** Scatter plots of the true versus estimated mean deviations (MD) of the testing dataset. **Left:** Outcome of the Artificial Neural Network model based on 64 RNFL sectors. **Right:** Outcome of the linear regression based on seven RNFL summary parameters.

and 3.1–7.4 dB, respectively. The MAE and RMSE of the model for eyes at the later stages of glaucoma ($MD < -6$ dB) using three independent subsets were in the range of 4.2–10.0 dB, and 5.2–11.9 dB, respectively.

Table 5 shows the outcome of the ablation test on the ANN model based on the testing subset. We observed that excluding sectors 41–64 impacted the accuracy of the model more than the other sectors.

Figure 6 shows the feature importance of the 64 sectors superimposed on the fundus photograph for easier interpretation.

Discussion

The ability of AI models to accurately estimate visual field damage from OCT images has several advantages. It may complement and subsequently reduce the burden of subjective visual field testing in patients with glaucoma. It could support less frequent visual field testing and individualization of testing requirements to individual patients. Eventually, it may even fulfil the long-term hope of replacing subjective, time-consuming, and inconsistent visual field testing with more rapid, objective, and more reproducible

Table 4. Accuracy of the Artificial Neural Network Model in Estimating Mean Deviation (MD) From Retinal Nerve Fiber Layer (RNFL) Thickness Measurements for Eyes in the Early ($MD \geq -6$) and Moderately Severe to Advanced ($MD < -6$) Stages of Glaucoma

Dataset	Testing		Rotterdam		UCLA		MEE	
MD Interval (dB)	$MD \geq -6$	$MD < -6$	$MD \geq -6$	$MD < -6$	$MD \geq -6$	$MD < -6$	$MD \geq -6$	$MD < -6$
Mean absolute error (MAE); dB	3.6 (3.3, 4.3)	4.6 (4.5, 4.8)	2.8 (2.4, 3.3)	4.2 (3.5, 4.9)	2.4 (1.9, 3.1)	4.9 (4.6, 5.2)	5.0 (4.5, 5.5)	10.0 (8.6, 11.6)
Root mean square error (RMSE); dB	4.8 (4.3, 5.5)	5.9 (5.7, 6.1)	3.9 (3.4, 4.5)	5.2 (4.3, 6.1)	3.1 (2.5, 3.9)	6.4 (5.9, 6.8)	7.4 (6.7, 8.1)	11.9 (10.2, 13.8)
Median absolute error; dB	3.0 (2.8, 4.0)	3.7 (3.6, 3.8)	2.1 (1.8, 3.2)	3.7 (3.1, 4.1)	2.0 (1.3, 2.6)	4.0 (3.5, 4.4)	2.8 (2.5, 3.1)	8.9 (7.1, 11.1)

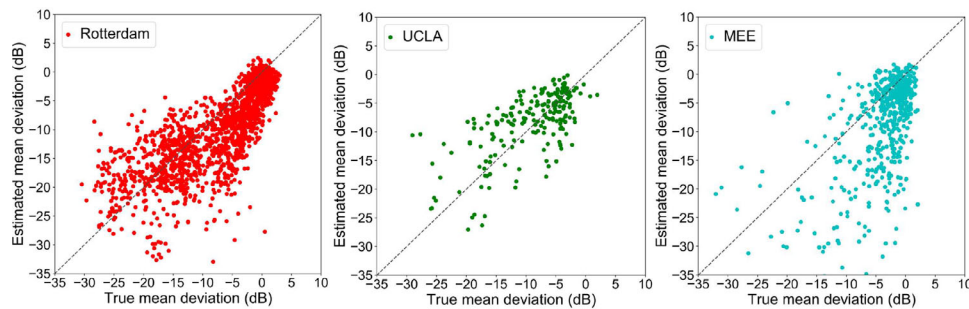


Figure 5. Scatter plots of the true versus estimated mean deviations (MD) of the ANN model based on independent subsets. **Left:** A subset with 691 visual fields and OCT pairs from Rotterdam eye hospital. **Middle:** A subset with 256 visual fields and OCT pairs from UCLA. **Right:** A subset with 691 visual fields and OCT pairs from MEE. The MEE subset included Cirrus cube scans while other subsets included Spectralis circle scans.

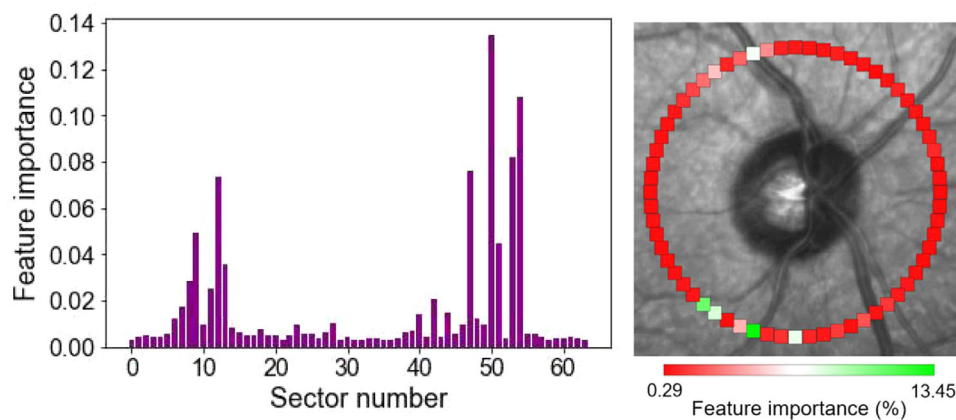


Figure 6. Feature (sector) ranking based on the random forest regressor (RF) model. **Left:** Sectors that were more important in estimating visual field mean deviation from 64 RNFL sectors. **Right:** Importance sectors were color coded and superimposed on fundus photograph to provide a user-friendly visualization. More important sectors are presented in greenish colors.

Table 5. Ablation Rest on Artificial Neural Network (ANN) Based on the Testing Subset

Sectors Excluded	R ² (95% Confidence Interval)
None	0.64 (0.59, 0.68)
1–10	0.62 (0.56–0.67)
1–20	0.59 (0.52–0.65)
1–30	0.57 (0.50–0.63)
1–40	0.51 (0.40–0.59)
41–50	0.55 (0.48–0.62)
51–60	0.60 (0.52–0.66)
41–60	0.46 (0.37–0.53)
41–64	0.42 (0.36–0.51)

OCT imaging. However, even with recent advancements in AI, we have yet to reach these ultimate goals.

We developed several linear and nonlinear models for estimating the visual field MD from RNFL thickness measurements. More complex models are often

believed to perform better; however, it is known that more complex models typically make more assumptions, leading to narrower application and less generalizability. Occam's razor theory suggests selecting simpler machine learning models may be desirable, particularly if the accuracy is not significantly compromised.³⁸ We showed that a simple ANN model can estimate global visual field MD without compromising the accuracy when compared with other linear or more complex 1-D CNN models.

We trained and tested the ANN model using a relatively large subset of OCT and visual field pairs, and validated the model using three different subsets. The error in estimating visual field MD in terms of MAE and RMSE was 4.0 dB and 5.2 dB, respectively. Our model's error is lower than the model reported by Sugiura et al.³⁵ that achieved an RMSE of about 6.1 dB, and comparable to the error level (MAE of 3.9 dB) of the model developed by Zhu et al.³² that used RNFL profiles quantified from SLP images. The RMSE about

3.7 dB in two studies^{33,34} that used several retinal layers to estimate MD is lower than ours, however, there are caveats that make generalization of their results challenging: (1) the number of subjects in the first study was approximately 120 and in the second study approximately 100, which makes generalization of findings challenging; (2) models significantly underestimate the true sensitivity values of eyes in the early stages and significantly overestimate the sensitivity values of eyes in the moderate and advanced stages of glaucoma. In contrast, the error distribution of our model is relatively symmetric (see Fig. 4 left).

A critical step in our training was to select the OCT and visual field pairs uniformly across all stages of glaucoma severity (see Fig. 3). However, this is not the case for almost all the previously published papers, which included significantly greater number of eyes in the early stages of glaucoma. For instance, two deep learning models proposed previously^{36,37} have reported significantly smaller MAEs in the range of 2.3–2.9 dB. The models have used OCT en face images, CSLO images, or OCT images from the macula and optic disc to estimate visual field parameters. Although the numbers of samples in both studies are large, there are several other concerns regarding both studies: (1) the number of eyes at the early stages of glaucoma is significantly larger than the eyes at the later stages of glaucoma, and (2) it is unknown whether the distribution of estimation error is relatively symmetric or biased at the ends of the glaucoma spectrum. The first concern is critical because models typically perform better for estimating visual field MD of normal eyes and eyes at the earlier stages of glaucoma compared to eyes at the moderate and end stages of glaucoma. As such, a model that uses relatively larger numbers of normal eyes and eyes at the early stages of glaucoma may misleadingly generate lower error rates compared to models that exploit eyes selected uniformly across the full glaucoma spectrum. The second concern is critical because in the absence of appropriate visualization of the error distributions using scatter or Bland-Altman plots, it is challenging to understand whether the model is biased toward one end or both ends of the glaucoma spectrum.

A unique aspect of our study is the inclusion of three independent subsets from different centers, different instruments, and even different scan types in order to validate models. Using Rotterdam and UCLA subsets, we achieved MAE and RMSE lower than 5.3 dB, similar to the error rates using the testing subset. Whereas the training subset was selected from a local pool of data, the Rotterdam subset was from a different race reflecting a significant degree of generalizability

of the model to other races and data from other institutes (see Fig. 5 left). However, the degree of generalizability of the model using MEE subset was not similar to other two validation subsets (see Fig. 5 right). This may be due to two reasons: (1) the OCT data from MEE were collected from Cirrus instruments, whereas the OCT data from other subsets were collected from Spectralis instruments. (2) The OCT data from MEE were in cube scan format that were approximated to circular scan computationally for the sake of comparing models. This approximation included interpolation and smoothing as well that may deviate from the true values of circular A-scans. (3) The eyes in the MEE dataset had a significantly different distribution of global RNFL and MD compared to all other subsets (see Fig. 3).

A major problem facing most models that attempt to estimate visual field parameters from OCT data is the floor effect that the instrument is unable to detect further RNFL loss beyond a certain point. We combined all datasets and observed that OCT reaches a “floor” at global RNFL thickness of about 40 microns, below which no useful data is obtained. Therefore, no matter how severe the visual field defect, the OCT is unable to reflect structural damage beyond this floor. This fact may explain why most of the models in the literature significantly underestimate the severity of visual fields for eyes at the advanced stages of the glaucoma.

A critical question would be, what is the highest degree of accuracy feasible for models that estimate visual field parameters from OCT. It is well known that visual field test is variable, particularly in the late stage of the disease, where significant disease variability exists.^{17,18} We used sequences of visual field test results from a cohort consists of 133 eyes from 71 patients with POAG, in which visual fields were collected once a week for an average of 10 consecutive weeks, thus appropriate for assessing test-retest variability.³⁹ For each visual field test point, we subtracted the total deviation (TD) values from the subsequent test for each eye and repeated this process for all visual field test points and all the sequences in this subset. The test-retest variability of visual field test points was in the range of 1.6–4.4 dB for visual field test points. Based on this test-retest experiment, visual field tests have inherent variability close to 4.4 dB. Thus, it may be a realistic goal for AI models based on OCT to be able to estimate visual field parameters up to 4.4 dB error, which is inherent to visual fields.

To examine the clinical relevance of findings and to see which RNFL sectors were more important in estimating visual field MD from 64 RNFL sectors, we

performed two tests. First, the ablation test based on the ANN model by excluding some of the sectors and observing the accuracy of the model. The outcome of the ablation test revealed that sectors in the temporal-inferior region were more important in estimating visual field severity from RNFL data (see Table 5). A finer experiment was performed by the RF regressor in which we observed that NRFL sectors in the temporal-inferior and temporal-superior were more important in estimating MD from 64 RNFL sectors (see Fig. 6). Findings agreed with previous literature.^{40,41}

We used large datasets to train AI models, developed several models, including 1-D CNNs, selected the simplest model with the highest accuracy, and validated the results using three different subsets from different centers, instruments, and scan types; however, our study has limitations as well. Our models did not benefit from 2-D convolutional deep neural networks as we did not have access to raw OCT images. We also did not estimate each visual field test point and only estimated visual field MD. Follow-up studies would be desirable to incorporate raw images and estimate each visual field test locations as MD estimates do not provide information on the regional nature of the visual field loss.

In conclusion, we developed an ANN to estimate visual field MD from input RNFL data. We validated our algorithm with three independent subsets and demonstrated that the performance of the model is close to test-retest variability in visual fields. Our study suggests that successful development of AI models to estimate visual field parameters from OCT data could augment or even replace subjective and tedious visual field testing with objective and rapid OCT imaging.

Acknowledgments

Supported by NIH Grants EY030142 and EY031725 and a Challenge Grant from Research to Prevent Blindness (RPB), New York. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure: **X. Huang**, None; **J. Sun**, None; **J. Majoor**, None; **K.A. Vermeer**, None; **H. Lemij**, None; **T. Elze**, None; **M. Wang**, None; **M.V. Boland**, None; **L.R. Pasquale**, None; **V. Mohammadzadeh**, None; **K. Nouri-Mahdavi**, None; **C. Johnson**, None; **S. Yousefi**, None

* XH and JS contributed equally to this study.

References

1. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311(18):1901–1911.
2. Quigley HA. Glaucoma. *The Lancet*. 2011;377(9774):1367–1377.
3. Kwon YH, Fingert JH, Kuehn MH, Alward WL. Primary open-angle glaucoma. *N Engl J Med*. 2009;360(11):1113–1124.
4. Bathija R, Gupta N, Zangwill L, Weinreb RN. Changing definition of glaucoma. *J Glaucoma*. 1998;7(3):165–169.
5. Worley A, Grimmer-Somers K. Risk factors for glaucoma: what do they really mean? *Aust J Prim Health*. 2011;17(3):233–239.
6. Gordon MO, Beiser JA, Brandt JD, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120(6):714–720; discussion 729–730.
7. Sarfraz MH, Mehboob MA, ul Haq RI. Correlation between central corneal thickness and visual field defects, cup to disc ratio and retinal nerve fiber layer thickness in primary open angle glaucoma patients. *Pakistan Journal of Medical Sciences*. 2017;33(1):132.
8. Pathak M, Demirel S, Gardiner SK. Nonlinear Trend Analysis of Longitudinal Pointwise Visual Field Sensitivity in Suspected and Early Glaucoma. *Transl Vis Sci Technol*. 2015;4(1):8.
9. Chen A, Nouri-Mahdavi K, Otarola FJ, Yu F, Afifi AA, Caprioli J. Models of glaucomatous visual field loss. *Invest Ophthalmol Vis Sci*. 2014;55(12):7881–7887.
10. Lee PP, Walt JG, Doyle JJ, et al. A multicenter, retrospective pilot study of resource use and costs associated with severity of disease in glaucoma. *Arch Ophthalmol*. 2006;124(1):12–19.
11. Wilson R, Walker AM, Dueker DK, Crick R. Risk factors for rate of progression of glaucomatous visual field loss: A computer-based analysis. *Arch Ophthalmol*. 1982;100(5):737–741.
12. Traverso CE, Walt JG, Kelly SP, et al. Direct costs of glaucoma and severity of the disease: a multinational long term study of resource utilisation in Europe. *Br J Ophthalmol*. 2005;89(10):1245–1249.
13. National Institute for Health and Care Excellence (UK). In: *Glaucoma: diagnosis and management*. In: National Institute for Health and Care Excellence: Clinical Guidelines. London, UK: National Institute for Health and Care Excellence (UK); 2017.

14. Anderson DR, Patella VM. *Automated Static Perimetry*. London, UK: Mosby; 1999.
15. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology*. 1999;106(11):2144–2153.
16. Sommer A. Collaborative normal-tension glaucoma study. *Am J Ophthalmol*. 1999;128(6):776–777.
17. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci*. 2000;41(2):417–421.
18. Mikelberg FS, Parfitt CM, Swindale NV, Graham SL, Drance SM, Gosine R. Ability of the Heidelberg Retina Tomograph to Detect Early Glaucomatous Visual Field Loss. *J Glaucoma*. 1995;4(4):242–247.
19. Bengtsson B, Heijl A. Inter-subject variability and normal limits of the SITA Standard, SITA Fast, and the Humphrey Full Threshold computerized perimetry strategies, SITA STATPAC. *Acta Ophthalmol Scand*. 1999;77(2):125–129.
20. Bengtsson B, Patella V, Heijl A. Prediction of glaucomatous visual field loss by extrapolation of linear trends. *Arch Ophthalmol*. 2009;127(12):1610–1615.
21. Viswanathan AC, Fitzke FW, Hitchings RA. Early detection of visual field progression in glaucoma: a comparison of PROGRESSOR and STATPAC 2. *Br J Ophthalmol*. 1997;81(12):1037–1042.
22. Heijl A, Lindgren G, Lindgren A, et al. Extended empirical statistical package for evaluation of single and multiple fields in glaucoma: Statpac 2. *Perimetry Update*. 1990;91:303–315.
23. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science (New York, NY)*. 1991;254(5035):1178–1181.
24. Yousefi S, Elze T, Pasquale LR, et al. Monitoring Glaucomatous Functional Loss Using an Artificial Intelligence-Enabled Dashboard. *Ophthalmology*. 2020;127:1170–1178.
25. Thakur A, Goldbaum M, Yousefi S. Predicting Glaucoma before Onset Using Deep Learning. *Ophthalmol Glaucoma*. 2020;3(4):262–268.
26. Norouzifard M, Nemati A, GholamHosseini H, Klette R, Nouri-Mahdavi K, Yousefi S. Automated glaucoma diagnosis using deep and transfer learning: Proposal of a system for clinical testing. Paper presented at: 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ), 2018. Available at: <https://ieeexplore.ieee.org/document/8634671>.
27. Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA. Assessment of a Segmentation-Free Deep Learning Algorithm for Diagnosing Glaucoma From Optical Coherence Tomography Scans. *JAMA Ophthalmol*. 2020;138(4):333–339.
28. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology*. 2018;125(8):1199–1206.
29. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158–164.
30. Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017;318(22):2211–2223.
31. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342–1350.
32. Zhu H, Crabb DP, Schlottmann PG, et al. Predicting visual function from the measurements of retinal nerve fiber layer structure. *Invest Ophthalmol Vis Sci*. 2010;51(11):5657–5666.
33. Bogunovic H, Kwon YH, Rashid A, et al. Relationships of retinal structure and humphrey 24-2 visual field thresholds in patients with glaucoma. *Invest Ophthalmol Vis Sci*. 2014;56(1):259–271.
34. Guo Z, Kwon YH, Lee K, et al. Optical Coherence Tomography Analysis Based Prediction of Humphrey 24-2 Visual Field Thresholds in Patients With Glaucoma. *Invest Ophthalmol Vis Sci*. 2017;58(10):3975–3985.
35. Sugiura H, Kiwaki T, Yousefi S, Murata H, Asaoka R, Yamanishi K. Estimating Glaucomatous Visual Sensitivity from Retinal Thickness with Pattern-Based Regularization and Visualization. *Kdd'18: Proceedings of the 24th Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*. 2018:783–792.
36. Christopher M, Bowd C, Belghith A, et al. Deep Learning Approaches Predict Glaucomatous Visual Field Damage from OCT Optic Nerve Head En Face Images and Retinal Nerve Fiber Layer Thickness Maps. *Ophthalmology*. 2020;127(3):346–356.
37. Yu HH, Maetschke SR, Antony BJ, et al. Estimating Global Visual Field Indices in Glaucoma

- by Combining Macula and Optic Disc OCT Scans Using 3-Dimensional Convolutional Neural Networks. *Ophthalmol Glaucoma*. 2021;4(1):102–112.
38. Schaffer J. What Not to Multiply Without Necessity. *Australasian Journal of Philosophy*. 2015;93(4):644–664.
 39. Garway-Heath DF, Zhu H, Cheng Q, et al. Combining optical coherence tomography with visual field data to rapidly detect disease progression in glaucoma: a diagnostic accuracy study. *Health Technol Assess*. 2018;22(4):1–106.
 40. Wang M, Shen LQ, Pasquale LR, et al. An Artificial Intelligence Approach to Assess Spatial Patterns of Retinal Nerve Fiber Layer Thickness Maps in Glaucoma. *Transl Vis Sci Technol*. 2020;9(9):41.
 41. Hood DC, Raza AS, de Moraes CG, Liebmann JM, Ritch R. Glaucomatous damage of the macula. *Prog Retin Eye Res*. 2013;32:1–21.