



A Joint Multitask Learning Model for Cross-sectional and Longitudinal Predictions of Visual Field Using OCT

Ryo Asaoka, MD, PhD,^{1,2,3,4,5} Linchuan Xu, PhD,^{6,7} Hiroshi Murata, MD,⁵ Taichi Kiwaki, PhD,⁶ Masato Matsuura, CO, PhD,^{5,8} Yuri Fujino, CO,^{5,8,9} Masaki Tanito, MD, PhD,⁹ Kazuhiko Mori, MD, PhD,¹⁰ Yoko Ikeda, MD, PhD,^{10,11} Takashi Kanamoto, MD, PhD,^{12,13} Kenji Inoue, MD, PhD,¹⁴ Jukichi Yamagami, MD, PhD,¹⁵ Kenji Yamanishi, PhD⁶

Purpose: We constructed a multitask learning model (latent space linear regression and deep learning [LSLR-DL]) in which the 2 tasks of cross-sectional predictions (using OCT) of visual field (VF; central 10°) and longitudinal progression predictions of VF (30°) were performed jointly via sharing the deep learning (DL) component such that information from both tasks was used in an auxiliary manner (The Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining [SIGKDD] 2021). The purpose of the current study was to investigate the prediction accuracy preparing an independent validation dataset.

Design: Cohort study.

Participants: Cross-sectional training and testing data sets included the VF (Humphrey Field Analyzer [HFA] 10-2 test) and an OCT measurement (obtained within 6 months) from 591 eyes of 351 healthy people or patients with open-angle glaucoma (OAG) and from 155 eyes of 131 patients with OAG, respectively. Longitudinal training and testing data sets included 7984 VF results (HFA 24-2 test) from 998 eyes of 592 patients with OAG and 1184 VF results (HFA 24-2 test) from 148 eyes of 84 patients with OAG, respectively. Each eye had 8 VF test results (HFA 24-2 test). The OCT sequences within the observation period were used.

Methods: Root mean square error (RMSE) was used to evaluate the accuracy of LSLR-DL for the cross-sectional prediction of VF (HFA 10-2 test). For the longitudinal prediction, the final (eighth) VF test (HFA 24-2 test) was predicted using a shorter VF series and relevant OCT images, and the RMSE was calculated. For comparison, RMSE values were calculated by applying the DL component (cross-sectional prediction) and the ordinary pointwise linear regression (longitudinal prediction).

Main Outcome Measures: Root mean square error in the cross-sectional and longitudinal predictions.

Results: Using LSLR-DL, the mean RMSE in the cross-sectional prediction was 6.4 dB and was between 4.4 dB (VF tests 1 and 2) and 3.7 dB (VF tests 1–7) in the longitudinal prediction, indicating that LSLR-DL significantly outperformed other methods.

Conclusions: The results of this study indicate that LSLR-DL is useful for both the cross-sectional prediction of VF (HFA 10-2 test) and the longitudinal progression prediction of VF (HFA 24-2 test). *Ophthalmology Science* 2021;1:100055 © 2021 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Glaucoma causes irreversible damage to the visual field (VF)¹ and currently is the leading cause of irreversible blindness worldwide.² The VF test is a principal measure for both diagnosing glaucoma and monitoring its progression, and the Humphrey Field Analyzer (HFA) 24-2 test (Carl Zeiss Meditec) is one of the most frequently used VF measurements clinically. In the clinical setting, assessing the progression of VF defects often relies on applying a simple ordinary least squares linear regression to VF measurements, as used in the software Progressor (Medisoft Ltd).³ However, the VF threshold fluctuates in both the short term⁴ and the long term,⁵ and measurements of VF are associated with considerable noise, even with good reliability indices,^{6,7} which hampers

the accurate estimation of the VF progression speed.⁸ Thus, numerous attempts have been made to improve the accuracy of the analysis of VF progression. Additionally, previous studies have suggested that it is also important to measure the central 10° of the VF in patients with glaucoma using a test such as the HFA 10-2 test.^{5–7} In the clinic, however, it is difficult to perform HFA 24-2 tests at a sufficient frequency,^{9,10} which implies that it is not realistic to conduct the 10-2 VF test beyond a central 24° VF test with a frequency of adequate density. Because of the structure–function relationship, the sensitivity of VF can be predicted from the retinal thickness, including the ganglion cell layer (GCL), in glaucoma.^{11,12} This may be particularly feasible with the HFA 10-2 test because the

spectral-domain OCT macular scanning area overlaps primarily with the retina within the central 10°. ¹³ Therefore, the use of spectral-domain OCT to measure retinal thickness could be beneficial for the prediction of VF sensitivity in the central 10°. Furthermore, the high reproducibility of OCT, ^{14–17} in contrast to VF measurements, could highlight the merit of this approach further.

The development of deep learning (DL) methods represents a revolutionary advance in imaging recognition. ¹⁸ We recently reported the benefits of applying DL to OCT measurements regarding various aspects of VF, such as the diagnosis of glaucoma ¹⁹ and the cross-sectional prediction of the HFA 10-2 test. ²⁰ Moreover, we recently proposed the use of deeply regularized latent space linear regression (DLLR), in which the DL component contributes to improving the prediction accuracy of VF (HFA 24-2 test) progression. ²¹ Because both of these tasks (i.e., the cross-sectional prediction of the HFA 10-2 test results and the progression prediction of the HFA 24-2 test results) use DL to obtain knowledge of the transforming information from the OCT domain to the VF domain, it may be advantageous to apply the knowledge learned in one of the tasks to another task. Additionally, this application is clinically relevant because patients typically undergo measurement in the clinical setting using both the HFA 24-2 test longitudinally and OCT measurements, but the HFA 10-2 test usually is optional. Multitask learning ^{22,23} is particularly useful when the 2 tasks are performed simultaneously, where individual tasks have relatedness and can lend strengths to each other. A typical example is the joint prediction of VF progression (HFA 10-2 test) and the prediction of VF progression (HFA 24-2 test). This implies that the cross-sectional prediction of the HFA 10-2 test results can be improved by constructing such a multitask model in which the progression of the HFA 24-2 test is predicted simultaneously. This has clinical merit when we consider that in the clinical setting, the use of the HFA 24-2 test and OCT measurement usually results in a large amount of clinically available longitudinal VF data, but the HFA 10-2 test does not. Therefore, we constructed a multitask DL model (latent space linear regression and DL [LSLR-DL]) to predict the HFA 10-2 test results in a cross-sectional manner, along with the simultaneous longitudinal prediction of VF findings (HFA 24-2 test) via sharing the DL component, such that both sets of information were used in an auxiliary manner for both tasks; however, the prediction accuracy was validated merely using an internal cross-validation. ²⁴ The purpose of the current study was to investigate the prediction accuracy preparing an independent validation dataset.

Methods

This study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at the University of Tokyo, Osaka University, Kyoto Prefectural University of Medicine, Shimane University, and Hiroshima Memorial Hospital. The study complied with the tenets of the Declaration of Helsinki. Patients provided written consent for their information to be stored in the hospital database and used for research. In other cases in which the study protocols did not require that each patient

provide written informed consent (based on the Ethical Guidelines for Medical and Health Research Involving Human Subjects issued by the Japanese government), the protocol was instead posted at the outpatient clinic to notify study participants.

Data Collection

We obtained all data in the present study from Tokyo University Hospital, Osaka University Hospital, Hospital of Kyoto Prefectural University of Medicine, Oike-Ikeda Eye Clinic, Shimane University Hospital, Inouye Eye Hospital, and Hiroshima Memorial Hospital. Inclusion criteria were (1) patients in whom glaucoma was the only disease causing VF damage and (2) patients with at least 8 VF measurements obtained via the 10-2 HFA (Carl Zeiss Meditec). All studied patients had primary open-angle glaucoma (OAG), which was defined as (1) the presence of typical glaucomatous changes in the optic nerve head, such as a rim notch with a rim width of 0.1 disc diameter or less or a vertical cup-to-disc ratio of more than 0.7, a retinal nerve fiber layer defect with its edge at the optic nerve head margin greater than a major retinal vessel, diverging in an arcuate or wedge shape, or both and (2) gonioscopically wide-open angles of grade 3 or 4 based on the Shaffer classification. Exclusion criteria were (1) age younger than 20 years and (2) possible secondary ocular hypertension in either eye. All patients had prior experience of VF measurement. We applied these criteria to both the training and testing data sets.

Visual Field Measurement

In the cross-sectional prediction, we obtained VF measurements using the HFA with the 10-2 program (Swedish Interactive Threshold Algorithm Standard), whereas in the longitudinal prediction, we obtained VF measurements using the HFA with the 24-2 program. Only reliable VFs were included, defined as a fixation loss rate of less than 33%, a false-positive rate of less than 33%, and a false-negative rate of less than 33%.

OCT Measurement

We obtained OCT data using the RS 3000 (Nidek Co. Ltd) and axial length measurements using the OA-2000 (TOMEY). All spectral-domain OCT measurements were obtained after pupil dilation with 1% tropicamide, and we performed OCT imaging using the laser scan protocol. We carefully excluded data with apparent eye movement or involuntary blinking or saccade from the measurement. Following the manufacturer's recommendation, we also excluded imaging data with a quality factor of less than 7. Similar to our previous report that analyzed OCT data, ²⁵ the fovea was identified automatically as the pixel with the thinnest retinal thickness close to the fixation point and a square imaging area (30° × 30°) was centered on the fovea to exclude the area of the optic disc and parapapillary atrophy. We corrected the magnification effect on the basis of the formula provided by the manufacturer, which was based on Littman's equation, ^{26,27} using the measured axial length value. Using software supplied by the manufacturer, we calculated the thicknesses of (1) the macular retinal nerve fiber layer (RNFL), (2) macular GCL and inner plexiform layer (IPL), and (3) outer segment (OS) and retinal pigment epithelium (RPE). Unlike our previous study in which the mean thicknesses in the entire field were analyzed, these were exported as images of 512 × 128 pixels and resized to 224 × 224 pixels in the DL models using the pixel area relationship, which is the preferred method for image decimation. Besides the macular RNFL and GCL+IPL layers, we included the thickness of the OS+RPE because the structure–function relationship becomes stronger by including this layer, ²⁸ probably because the interindividual variation of the retinal layer

thicknesses can be considered. These thicknesses of macular RNFL, GCL+IPL, and OS+RPE were the 3 channels in DL.

Training and Testing Data Sets

Cross-sectional training and testing data sets included VF (HFA 10-2 test) and OCT measurement from 86 eyes of 43 healthy participants and 505 eyes of 304 patients with OAG and 155 eyes of 131 patients with OAG, respectively. Visual field (HFA 10-2 test) and OCT measurements were conducted within a period of 6 months. Longitudinal training and testing data sets included 7984 VF results (HFA 24-2 test) from 998 eyes of 592 patients with OAG and 1184 VF results (HFA 24-2 test) from 148 eyes of 84 patients with OAG, respectively. For each eye, a time series of 8 VF results (HFA 24-2 test) were available. We obtained OCT sequences for each eye within the same observation period. No overlap occurred between the training and testing data sets, both in the cross-sectional and longitudinal data.

Prediction of Visual Field

Latent Space Linear Regression and Deep Learning. Briefly, the proposed LSLR-DL model worked as follows.

1. In the cross-sectional prediction, the LSLR-DL model was trained to predict the sensitivity of VF at each test point (68 points in the HFA 10-2 test) using the thicknesses of GCC, macular RNFL, and OS+RPE (224×224 pixels with 3 channels). In the longitudinal prediction, the LSLR-DL model was trained first to predict the VF threshold at each test point (52 points in HFA 24-2 test mode) in the eighth VF test (HFA 24-2 test) using the first and second VF results (HFA 24-2 test) and also the thicknesses of GCC, macular RNFL, and OS+RPE (224×224 pixels with 3 channels) obtained in the same period. Subsequently, similar predictions were performed using different VF sequences (between the first third and first to seventh VFs by the HFA 24-2 test) and corresponding thicknesses of GCC, macular RNFL, and OS+RPE obtained in the same period. Predictions with different lengths of training VF sequences were performed separately because each of them corresponded to a unique scenario. The cross-sectional prediction and longitudinal prediction shared the DL component, such that they were performed simultaneously. We trained the model so that the summed loss with both the cross-sectional and longitudinal training data sets was minimized.
2. In the longitudinal prediction, all HFA 24-2 test points of a single eye were projected to a latent space to consider the correlations among HFA 24-2 test points. They shared parameters of linear regression in the latent space.
3. In the longitudinal prediction, the OCT sequences were transformed to VF (HFA 24-2 test) sequences via DL, such that the progression of VF (HFA 24-2 test) threshold was regularized by the progression of OCT. The transformation of the OCT sequences (224×224 pixels with 3 channels) into VF (HFA 24-2 test) sequences was realized by VGG16,²⁹ a convolutional neural network (CNN), considering the spatial relationships among the retina voxels and the nonlinear relationship between the thicknesses of the OCT-measured retinal layers and VF (HFA 24-2 test) threshold. The CNN currently is the state-of-the-art model for capturing spatially related information and nonlinear transformation from one data space to another. VGG16 won second place in the image classification task of ImageNet Large Scale Visual Recognition Challenge 2014 and has become a popular CNN model for many image classification and regression tasks. The

pretraining of VGG16 was conducted using images of the ImageNet database (<http://www.image-net.org/>). Figure 1 illustrates the architecture of VGG16.

4. In the longitudinal prediction, eyes with a similar HFA 24-2 test had similar regression parameters to avoid overfitting, based only on their own data.
5. We used the cross-sectional and longitudinal testing data sets to evaluate the accuracies of the cross-sectional and longitudinal predictions, respectively.

More technical descriptions of the LSLR-DL model are introduced as follows as well as in Figure 2. In the cross-sectional prediction, the HFA 10-2 test of an eye was transformed from an OCT measurement using the DL component as follows:

$$f_c(C_i; \theta_c) = V_i + \varepsilon, \quad (1)$$

where $C_i \in R^{3 \times H \times W}$ is a 3-dimensional array representing the thickness of the OCT-measured retinal layers (the layers of macular RNFL, GCL+IPL, and OS+RPE) of the i th eye; H and W are the height and width of the measurement, respectively; $V_i \in R^{68}$ is the VF in the 10-2 test mode; $f_c(\cdot; \theta_c)$ is a function parameterized by VGG16; θ_c contains the parameters to be estimated; and ε contains the errors of the transformation.

For the longitudinal prediction, the projection of VF (HFA 24-2 test) threshold into a latent space and the LSLR were realized by matrix factorization denoted as follows:

$$F_i = G_i W_i P_i^T + \varepsilon, \quad (2)$$

where $F_i \in R^{D \times T_i}$ is a matrix consisting of the sequence of VF results (HFA 24-2 test) for the i th eye, D is the number of HFA 24-2 test points, T_i is the number of time stamps at which the HFA 24-2 tests were conducted, $G_i \in R^{D \times S}$ and $W_i \in R^{S \times 2}$ are 2-factor matrices to be estimated, S is a predefined hyperparameter, and P_i^T is the transpose of $P_i \in R^{T_i \times 2}$, which was defined as $P_i = [t_i, 1_i]$. Because $t_i \in R^{T_i \times 1}$ was a column vector of time stamps and $1_i \in R^{T_i \times 1}$ was a vector of ones, $W_i P_i^T$ realized the linear regression of VF (HFA 24-2 test) threshold on time in the latent space. The first column and the second column of W_i correspond to the coefficient and intercept, respectively. G_i then can be interpreted as a projection matrix that realizes the transformation of VF (HFA 24-2 test) threshold, and ε contains the errors of the projection. In equation (2), the regression parameters in W_i were shared by all VF (HFA 24-2 test) points to force them to share the same progression pattern in the latent space. The differences among the VF (HFA 24-2 test) points in the raw data space were preserved in the projection matrix G_i .

The transformation of the OCT sequences into VF (HFA 24-2 test) threshold sequences also was realized by VGG16. Particularly, each OCT measurement was transformed into a VF (HFA 24-2 test) threshold interpolated by LSLR at the time at which no VF measurement was available. The transformation was realized by the following equation:

$$f_l(C_l^t; \theta_l) = G_l W_l [t_l^t, 1]^T + \varepsilon, \quad (3)$$

where $C_l^t \in R^{3 \times H \times W}$ is a 3-dimensional array representing the l th thicknesses of the OCT-measured retinal layer of the i th eye and t_l^t is the time at which the OCT-measured thickness was obtained; $f_l(\cdot; \theta_l)$ is a function parameterized by VGG16, and θ_l contains the

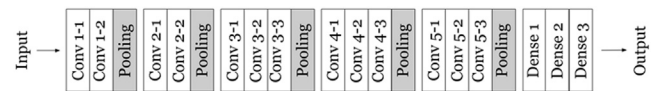


Figure 1. Diagram showing the architecture of VGG16. Transformation of the sequences of OCT (224×224 pixels with 3 channels) into visual field sequences was realized by VGG16.

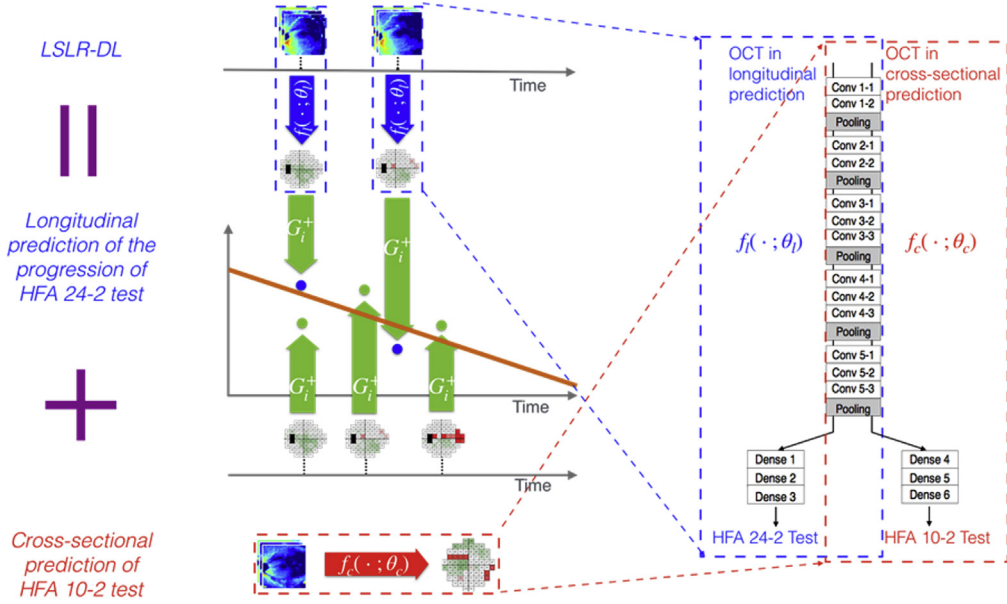


Figure 2. Diagram showing the architecture of the latent space linear regression and deep learning (LSLR-DL) model. The LSLR-DL model is a DL-based model that simultaneously predicts the Humphrey Field Analyzer (HFA) 10-2 test in a cross-sectional manner and longitudinally predicts the progression of the HFA 24-2 test, via sharing the DL component, such that both sets of information are used in an auxiliary manner for both tasks. G_i^+ is the pseudoinverse of G_i in equation (2).

parameters to be estimated. We propose performing the OCT transformation because the underlying model can be shared by the cross-sectional prediction. In this way, both the longitudinal prediction and the cross-sectional prediction can use additional information for better performance. In particular, $f_l(\cdot; \theta_l)$ shares parameters in the convolutional layers with $f_c(\cdot; \theta_c)$, which was for the cross-sectional prediction in equation (1). $f_l(\cdot; \theta_l)$ is not exactly the same as $f_c(\cdot; \theta_c)$, because the VF test points are not shared between the longitudinal prediction (HFA 24-2 test) and the cross-sectional prediction (HFA 10-2 test). Figure 2 illustrates the shared convolutional layers and the difference between $f_l(\cdot; \theta_l)$ and $f_c(\cdot; \theta_c)$. The OCT in both the cross-sectional prediction and the longitudinal prediction was processed by the same convolutional layers. The processed OCT in the cross-sectional prediction then was transformed into VF (HFA 10-2 test) using fully connected layers, whereas the processed OCT in the longitudinal prediction was transformed into VF (HFA 24-2 test) using different fully connected layers.

Linear regression of VF (HFA 24-2 test) in the latent space is regularized by the OCT information through equation (3), in which the VF interpolated at the time stamp when an OCT was measured was regularized by the VF results predicted from the OCT, as illustrated in Figure 2. Figure 2 shows that 3 VFs (HFA 24-2 test) are projected into the latent space by multiplying the pseudoinverse of G_i . At the same time, 2 VF results predicted from 2 respective OCTs are transformed into the same latent space. The linear regression of VF results (HFA 24-2 test) over time is fitted on the 5 data points, including both the VF (HFA 24-2 test) data points and the OCT data points, instead of only the VF (HFA 24-2 test) data points.

To avoid overfitting the data from each eye, the regression parameters of the eyes with similar VF (HFA 24-2 test) measurements were also forced to be similar. The relationship between one eye and the other eyes was defined as follows:

$$W_i = \sum_{j \neq i}^N z_{i,j} * W_j + \varepsilon \quad (4)$$

where $z_{i,j} \in R$ is the similarity constant between the i th eye and the j th eye. $z_{i,j}$ was quantified by the following equation:

$$z_{i,j} = \exp \left\{ - \frac{\left(\sum_{k=1}^{T_i} \sum_{d=1}^D (F_i^{(d,k)} - F_j^{*(d,k)}) \right)^2}{(T_i D \sigma_i)^2} \right\}, \quad (5)$$

where $F_i^{(d,k)}$ is the value at the d th row and the k th column of the VF (HFA 24-2 test) matrix of the i th eye, and F_j^* contains the VF (HFA 24-2 test) thresholds interpolated by pointwise linear regression (PLR), because the j th eye may have different time stamps on which VF (HFA 24-2 test) tests were conducted from those of the i th eye. σ_i was the median of values $\left(\sum_{k=1}^{T_i} \sum_{d=1}^D (F_i^{(d,k)} - F_j^{*(d,k)}) \right)^2 / (T_i D)^2$ computed from the combinations of the i th eye and all other eyes. The definition of similarity between 2 eyes, $z_{i,j}$, followed the Gaussian kernel, ³⁰ and worked well as demonstrated by our previous study.²¹

By combining the LSLR and DL, the objective function for a test eye indexed by 0 was quantified as follows:

$$L = \sum_{i=0}^N \|F_i - G_i W_i P_i^T\|_F^2 + \lambda_1 \sum_{j=1}^N z_{0,j} \|W_0 - W_j\|_F^2 + \lambda_2 \sum_{i=0}^N \sum_l \|f_l(C_i^l; \theta) - G_i W_i [t_i^l, 1]^T\|_F^2, \quad (6)$$

where the eyes of indices from $j = 1$ to N were used as training information for the learning of linear regression for the test eye, $\|\cdot\|_F^2$ is the square of the Frobenius norm, and $\lambda_1, \lambda_2 \in R$ are hyperparameters.

Because the cross-sectional prediction and longitudinal prediction shared the DL component, the 2 prediction tasks were performed jointly by having the following unified objective function:

$$L = \sum_{i=0}^N \|F_i - G_i W_i P_i^T\|_F^2 + \lambda_1 \sum_{j=1}^N z_{0j} \|W_0 - W_j\|_F^2 + \lambda_2 \left(\sum_{i=0}^N \sum_l \|f_i(C_i^l; \theta) - G_i W_i [t_i^l, 1]^T\|_F^2 + \sum_{i=0}^M \|f_c(C_i; \theta) - V_i\|_F^2 \right), \quad (7)$$

where M is the number of VFs (HFA 10-2 test) in the cross-sectional prediction task.

The optimization algorithms designed for DL (e.g., Adam³¹) can be used to solve the objective function. We implemented the LSLR-DL model in Pytorch and applied Adam as the learning algorithm. The values of hyperparameters λ_1 and λ_2 were determined by grid search, and the value of the hyperparameter S was chosen as 4 because our previous study²¹ suggested that a small value would work well.

Other Models (Cross-sectional Prediction of Visual Field from OCT). The following models were also constructed for the purpose of comparison.

1. Multiple linear regression (MLR): The VF (HFA 10-2 test) threshold at each test point was predicted simply by applying MLR to 150 528 ($224 \times 224 \times 3$) variables of GCC, macular RNFL, and OS+RPE thicknesses.
2. Support vector regression (SVR) constructs a hyperplane, which provides the largest separation margin between 2 classes. A soft margin allows some errors to occur between the separation hyperplane, and a kernel function maps the data into higher-dimensional space, which allows a linear separation in a nonlinear classification problem.³² At each test point, the VF (HFA 10-2 test) threshold was predicted by applying SVR to 150 528 ($224 \times 224 \times 3$) variables of GCC, RNFL, and OS+RPE.
3. Deep learning: At each test point, the VF (HFA 10-2 test) threshold was predicted simply by applying VGG16 to the

channels). Convolutional neural network and tensor regression is a recent method that combines a CNN with tensor regression.

Other Models (Longitudinal Prediction of Visual Field).

1. Pointwise linear regression: Using ordinary least squares linear regression, VF (HFA 24-2 test) threshold was regressed against time at each VF test point.
2. Deeply regularized latent space linear regression: The DLLR

model uses OCT measurements to regularize the linear regression of VF measurements against time in a latent space. The OCT measurements lie in a completely different data space to VF, and the OCT measurements also may be obtained at different points in time to the VF measurements. To address the heterogeneity in data space, DLLR transforms both the extracted information of the OCT and VF measurements into the same latent space. Deeply regularized latent space linear regression focuses on the coefficient and the intercept of the latent space linear regression of the measurement sets to address the heterogeneity in time. This model outperformed PLR^{21,33} because glaucomatous VF damage results from the loss of retinal ganglion cells and VF threshold also fluctuates in both the short⁴ and long⁵ terms. Moreover, VF measurements are associated with considerable noise,^{6,7} which hampers the accurate estimation of the speed of VF progression,⁸ whereas OCT measurements are highly reproducible.^{14–17} The major difference between LSLR-DL and DLLR is the way in which OCTs are used to regularize the progression pattern of VFs. In particular, DLLR transforms all OCTs together into the latent space to regularize the coefficient directly and to intercept the linear regression while LSLR-DL transforms OCTs individually into VFs to regularize the VFs interpolated at time stamps when OCTs were measured.

Statistical Analysis

We evaluated cross-sectional prediction errors using RMSE, which is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{52} (\text{predicted VF threshold of the } i\text{th point} - \text{actual VF threshold of the } i\text{th point})^2}{52}}, \quad (8)$$

GCC, RNFL, and OS+RPE thicknesses (224×224 pixels with 3 channels).

4. Convolutional neural network and tensor regression (CNN-TR): At each test point, the VF (HFA 10-2 test) threshold was predicted by applying CNN-TR²⁰ to the GCC, RNFL, and OS+RPE thicknesses (224×224 pixels with 3

where i is the number of the predicted 52 test points.

Using a linear mixed model, we compared the RMSE values with LSLR-DL, MLR, SVR, DL, and CNN-TR models, whereby the prediction error values were nested within each patient and test point. The linear mixed model is equivalent to ordinary linear

Table 1. Demographic Information of the Cross-sectional Training and Testing Data Sets

Variable	Cross-sectional Training Data Set	Cross-sectional Testing Data Set
Eyes (left/right)	289/302	89/66
Sex (female/male)	191/156	70/61
Age (yrs)	55.1 ± 14.8	65.8 ± 12.2
Axial length (mm)	25.4 ± 2.7	24.6 ± 1.7
Threshold (HFA 10-2 test; dB)	24.1 ± 9.3	21.9 ± 7.8
MD (HFA 10-2 test; dB)	-8.8 ± 9.4	-10.4 ± 8.1
RNFL (μm)	30.5 ± 9.0	26.9 ± 8.4
GCL+IPL (μm)	39.7 ± 9.0	38.7 ± 7.5
OS+RPE (μm)	67.1 ± 3.8	65.6 ± 5.1

GCL = ganglion cell layer; HFA = Humphrey Field Analyzer; IPL = inner plexiform layer; MD = mean deviation; OS = outer segment; RNFL = retinal nerve fiber layer; RPE = retinal pigment epithelium.

Data are presented as no. or mean ± standard deviation.

regression in that the model describes the relationship between the predictor variables and a single outcome variable. However, standard linear regression analysis assumes that all observations are independent of each other. In the present study, the measurements were nested within patients and also test points, and hence are dependent on each other. Ignoring this measurement grouping results in the underestimation of standard errors of regression coefficients. The linear mixed model adjusts for the hierarchical structure of the data, modeling such a way that measurements are grouped within patients to reduce the possible bias derived from the nested structure of data.^{34,35} Using Holm's method, we adjusted for multiple comparisons.

We also calculated longitudinal prediction error as the RSME between the predicted and actual 52 threshold values of the eighth VF (HFA 24-2 test). Using the linear mixed model, we compared these values among PLR, DLLR, and LSLR-DL, whereby the values were nested within patients and test points and also the test points of the VF (HFA 24-2 test).

Results

Table 1 presents demographic information of the cross-sectional training and testing data sets. In the training data set, 191 patients were women and 156 patients were men. In the testing data set, 70 patients were women and 61 patients were men. The mean ± standard deviation age of the patients was 55.1 ± 14.8 years and 65.8 ± 12.2 years in the

training and testing data sets, respectively. In the training data set, the mean deviation of the HFA 10-2 test was -8.8 ± 9.4 dB, whereas this value was -10.4 ± 8.1 dB in the testing data set.

Table 2 shows the demographic information of the longitudinal training and testing data sets. The mean ± standard deviation age of the patients was 60.7 ± 13.5 years and 61.2 ± 10.4 years in the training and testing data sets, respectively. Visual field results (HFA 24-2 test) were obtained for 5.9 ± 1.9 years and 5.4 ± 1.1 years in the training and testing data sets, respectively. The mean number of OCT measurements was 5.6 ± 2.8, with a maximum of 12 and a minimum of 1 in the training data sets, and the number in the testing data sets was 5.2 ± 1.6, with a maximum of 9 and a minimum of 2. In the initial HFA 24-2 test, the mean total deviation (mTD) value was -6.2 ± 7.1 and -4.9 ± 4.6 dB in the training and testing data sets, respectively. The mTD progression rate with VFs from 1st to 10th (VF1-10) was -0.3 ± 0.8 dB/year and -0.3 ± 0.7 dB/year in the training and testing data sets, respectively.

Cross-sectional Prediction

Figure 3 shows the threshold of the HFA 10-2 test at each test point in the cross-sectional testing dataset. The mean value ranged from 14.6 to 29.3 dB. Figure 4 presents the

Table 2. Demographic Information of the Longitudinal Training and Testing Data Sets

Variable	Longitudinal Training Data Set	Longitudinal Testing Data Set
Eyes (right/left)	505/493	72/76
Sex (female/male)	296/296	46/38
Age (yrs)	60.7 ± 13.5	61.2 ± 10.4
Axial length (mm)	25.8 ± 1.9	24.9 ± 1.8
mTD of first VF (HFA 24-2 test; dB)	-6.2 ± 7.1	-4.9 ± 4.6
mTD progression rate with VF1-10 (HFA 24-2 test; dB/yr)	-0.3 ± 0.8	-0.3 ± 0.7
Sequences of OCT (no. of times)	5.6 ± 2.8	5.2 ± 1.6
Macular RNFL of first OCT (μm)	30.4 ± 9.0	30.4 ± 7.5
GCL+IPL of first OCT (μm)	42.7 ± 8.7	41.6 ± 8.1
OS+RPE of first OCT (μm)	65.4 ± 4.9	65.2 ± 5.1

GCL = ganglion cell layer; HFA = Humphrey Field Analyzer; IPL = inner plexiform layer; mTD = mean total deviation; OS = outer segment; RNFL = retinal nerve fiber layer; RPE = retinal pigment epithelium.

Data are presented as no. or mean ± standard deviation.

				14.6 (11.8)	15.7 (11.8)				
		16.1 (11.8)	16.4 (11.9)	16.0 (11.8)	16.1 (12.1)	16.5 (12.3)	16.2 (12.5)		
	17.6 (12.1)	19.5 (12.1)	19.6 (12.5)	18.5 (12.7)	17.9 (12.6)	17.6 (12.7)	16.7 (12.7)	16.3 (12.6)	
	21.4 (11.9)	23.6 (11.4)	23.5 (11.8)	23.5 (11.3)	21.5 (12.3)	19.0 (13.0)	18.4 (12.8)	17.4 (13.0)	
22.5 (10.8)	25.2 (10.0)	26.5 (9.6)	27.9 (8.7)	27.3 (8.5)	24.9 (10.6)	22.5 (11.8)	20.1 (12.6)	19.0 (12.4)	17.8 (12.3)
25.2 (9.1)	27.4 (7.9)	28.5 (8.0)	29.3 (7.6)	29.0 (6.9)	28.2 (7.9)	25.7 (10.7)	23.4 (11.6)	21.4 (11.9)	19.3 (12.7)
	27.1 (8.5)	28.2 (7.9)	28.5 (8.5)	28.4 (7.7)	26.5 (10.1)	25.0 (11.8)	23.1 (11.9)	20.2 (12.8)	
	25.2 (9.4)	26.2 (9.4)	25.7 (10.5)	25.5 (10.0)	23.9 (11.5)	23.4 (12.3)	21.5 (12.6)	19.0 (12.8)	
		22.4 (11.7)	22.5 (11.8)	21.8 (12.3)	21.7 (12.0)	20.2 (12.7)	19.1 (12.6)		
				18.8 (12.6)	18.8 (12.3)				

Figure 3. Visual field (Humphrey Field Analyzer 10-2 test) threshold at each test point in the cross-sectional testing data set. The mean value ranged from 16.0 to 29.3 dB.

comparisons of the cross-sectional RMSE values among MLR, SVR, and LSLR-DL models. The RMSE with LSLR-DL (6.4 ± 3.1 dB) was significantly ($P < 0.05$, linear mixed model adjusted for multiple comparisons using Holm’s method) smaller than that of the MLR (12.2 ± 4.2 dB), SVR (8.8 ± 3.5 dB), DL (7.7 ± 3.6 dB), and CNN-TR (7.3 ± 3.6 dB) models. Figure 5 shows the cross-sectional absolute prediction error at each VF (HFA 10-2 test) point. The mean value ranged from 3.4 to 6.1 dB. Prediction error values tended to be small in the inferotemporal area. Note that no

discrimination exists between the absolute prediction error and RMSE in the pointwise prediction error. In the entire VF (HFA 10-2 test), the mean of these values was 5.0 ± 0.6 dB.

Longitudinal Prediction

Figure 6 shows the threshold of the HFA 24-2 test at each test point (eighth VF) in the longitudinal testing data set. The mean value ranged from 14.2 to 28.5 dB. Figure 7 and Table 3 show the comparisons of the longitudinal RMSE values across LSLR-DL, DLLR, and PLR models. In all VF sequences (HFA 24-2 test), the LSLR-DL model significantly ($P < 0.001$, linear mixed model adjusted for multiple comparisons using Holm’s method) outperformed PLR. The LSLR-DL model significantly ($P < 0.05$, linear mixed model adjusted for multiple comparisons using Holm’s method) outperformed DLLR for the series first and second VF test to the first through fifth VF test (HFA 24-2 test). Figure 8 shows the absolute prediction errors at each VF (HFA 24-2 test) test point with the LSLR-DL, DLLR, and PLR models. In comparison with the PLR model, the values were significantly smaller with LSLR-DL model at all test points for the series the first and second VF results to the first through fourth VF results (HFA 24-2 test), in 51 test points in the first through fifth VF results (HFA 24-2 test), in 24 test points in the first through sixth VF results (HFA 24-2 test), and in 10 test points in the first through seventh VF results (HFA 24-2 test). In comparison with DLLR, the values were significantly ($P < 0.05$, linear mixed model adjusted for multiple comparisons using Holm’s method)

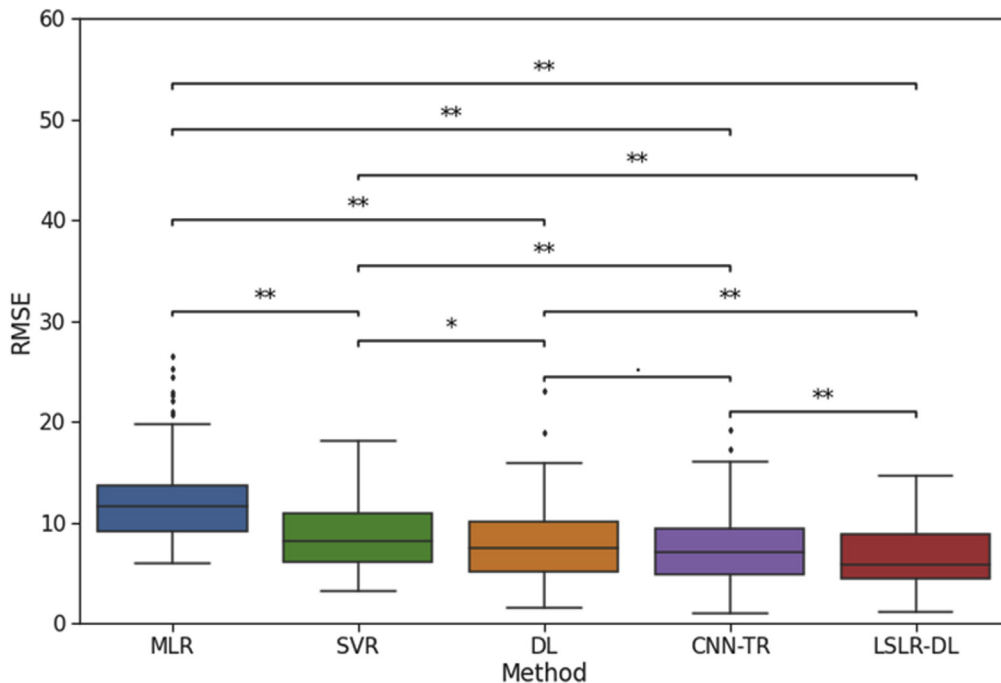


Figure 4. Box-and-whisker plot comparing cross-sectional root mean square error (RMSE) values across the multiple linear regression (MLR), support vector regression (SVR), deep learning (DL), convolutional neural network and tensor regression (CNN-TR), and latent space linear regression and deep learning (LSLR-DL) values. The RMSE with LSLR-DL values were significantly smaller than with the MLR, SVR, DL, and CNN-TR models. * $P < 0.05$. ** $P < 0.01$. · $P > 0.05$.

				5.4	5.2					
				(4.9)	(4.7)					
		6.1	5.5	5.2	5.3	5.3	5.3			
		(5.2)	(4.9)	(5.2)	(5.2)	(5.2)	(5.1)			
	5.7	5.2	5.7	5.2	5.6	5.6	5.3	5.2		
	(5.3)	(5.2)	(5.4)	(5.2)	(5.3)	(5.4)	(5.2)	(4.7)		
	5.2	5.3	5.2	4.9	5.8	5.9	5.9	5.8		
	(5.4)	(5.6)	(5.2)	(5.3)	(6.0)	(5.6)	(5.5)	(5.6)		
5.6	4.9	4.5	4.5	4.6	5.4	5.6	6.0	5.1	5.4	
(6.3)	(5.6)	(5.3)	(5.0)	(4.7)	(5.0)	(5.8)	(5.5)	(4.7)	(4.9)	
4.9	3.9	3.9	3.8	3.4	3.9	4.9	5.4	5.5	6.0	
(5.4)	(4.3)	(4.3)	(4.0)	(3.7)	(4.3)	(5.1)	(5.6)	(5.9)	(6.1)	
	4.3	3.8	3.9	3.4	4.5	5.2	5.0	5.3		
	(5.0)	(4.2)	(4.7)	(4.2)	(5.0)	(5.1)	(5.3)	(5.8)		
	4.5	3.9	4.6	4.5	5.0	5.4	5.0	4.8		
	(4.6)	(4.5)	(5.4)	(4.4)	(5.2)	(5.2)	(5.0)	(4.9)		
		5.0	4.8	4.8	4.8	4.9	4.3			
		(4.8)	(4.9)	(4.7)	(4.6)	(4.8)	(4.4)			
				4.8	4.5					
				(4.5)	(4.0)					

Figure 5. Cross-sectional absolute prediction error at each visual field (Humphrey Field Analyzer 10-2 test) point. The mean value ranged from 3.4 to 6.1 dB. Prediction error values tended to be small in the infero-temporal area.

smaller with the LSLR-DL model at 17 test points in the first and second VF results (HFA 24-2 test), in 9 test points in first through third VF results (HFA 24-2 test), in 5 test points in the first through fourth VF results (HFA 24-2 test), and in 2 test points from the first through fifth VF results to the first through sixth VF results (HFA 24-2 test).

Discussion

In the present study, we constructed a DL-based model that simultaneously predicts the HFA 10-2 test results in a cross-sectional manner, along with a longitudinal prediction of the progression of the HFA 24-2 test results, via sharing the DL component such that both predictions were used in an auxiliary manner for both tasks. This model was trained using a cross-sectional training data set from 591 eyes of 351 healthy participants or patients with OAG and a longitudinal training data set from 7984 VF results (HFA 24-2 test) from 998 eyes of 592 patients with OAG. Our results indicate that the mean RMSEs with the LSLR-DL model were 6.4 dB in the cross-sectional prediction of the HFA 10-2 test and

			20.0	19.5	19.3	19.5				
			(8.4)	(8.4)	(8.9)	(8.7)				
		22.4	22.3	21.3	21.1	22.0	20.2			
		(8.0)	(8.4)	(9.2)	(9.5)	(8.7)	(9.8)			
24.1	23.3	21.8	21.0	22.2	22.8	21.2	10.9			
(6.7)	(8.4)	(9.9)	(11.2)	(10.7)	(9.8)	(10.3)	(10.5)			
25.3		23.1	25.5	23.2	22.6	20.9	18.3	14.2		
(6.4)		(10.1)	(10.0)	(11.2)	(11.2)	(11.3)	(11.0)	(11.0)		
25.7		27.2	28.5	27.7	24.4	22.5	20.3	17.5		
(5.7)		(6.4)	(7.0)	(8.1)	(11.1)	(11.1)	(10.4)	(10.1)		
26.6	26.1	25.1	25.4	24.2	23.9	23.2	21.1			
(4.1)	(5.8)	(8.6)	(9.3)	(10.6)	(10.5)	(10.2)	(9.8)			
	26.6	26.8	25.5	24.8	24.4	22.9				
	(5.5)	(5.6)	(7.1)	(8.1)	(8.3)	(8.8)				
		23.0	25.4	23.8	22.9					
		(5.1)	(5.6)	(6.5)	(7.7)					

Figure 6. Visual field (Humphrey Field Analyzer 24-2 test) threshold at each test point in the longitudinal testing data set. The mean value ranged from 10.9 to 28.5 dB.

between 4.5 dB (first and second VF test) and 3.6 dB (first through seventh VF test) in the longitudinal prediction of the HFA 24-2 test. These values were significantly smaller than those of other methods, including the MLR, SVR, DL only, and CNN-TR models for the cross-sectional prediction and the PLR model for the longitudinal prediction.

The results of our study suggest that in the cross-sectional prediction of VF (HFA 10-2 test), the largest prediction error is observed with the MLR model (12.1 dB on average; Fig 5).^{20,36} This may be because the relationship between structures, such as the thickness of the RNFL and GCC, and function has been known to be nonlinear.^{13,37} Conversely, in the SVR model, regression is performed in a latent space (kernel plane); hence, no discrimination exists between linear and nonlinear. Indeed, the SVR model yielded a significantly more accurate prediction error (8.8 dB on average). The application of DL only and the CNN-TR model resulted in even smaller RMSE values (7.7 and 7.3 dB on average, respectively). This is because CNN is currently the state-of-the-art model for capturing spatially related information and nonlinear transformation from one data space to another. Overfitting is a common phenomenon in DL. The usefulness of a DL model cannot be generalized to outside data, and overfitting often is observed with DL when the size of the training data is small, such as in the present study (591 eyes). For small training data sets, the augmentation method has been known to be particularly useful for avoiding this overfitting problem,^{38,39} and indeed, we have shown the usefulness of this technique for the diagnosis of glaucoma using a fundus photograph and DL⁴⁰ as well as longitudinal VF prediction using PLR.⁴¹ In the LSLR-DL model, the cross-sectional prediction of VF (HFA 10-2 test) was performed simultaneously with the longitudinal prediction, which serves as a kind of data augmentation. This augmentation method in the LSLR-DL model may be more useful than the simple application of an ordinary augmentation (such as rotation, scale up and scale down, and change brightness) to OCT images, because the training data in the longitudinal prediction are real data instead of manipulated data. Consequently, the cross-sectional prediction of VF (HFA 10-2 test) enabled a significantly smaller prediction error in comparison with the simple application of DL.

In this study, prediction errors tended to be smaller in the central inferotemporal area than in other areas (Fig 6), which corresponds to the preserved central isle of the VF (HFA 10-2 test) in patients with advanced glaucoma as proposed by Weber et al⁴² and Hood et al.⁴³ Indeed, in the present study, the values of visual sensitivity in this area were higher than in other regions in general (Fig 4). The smaller absolute prediction error values in this area (Fig 6) may be attributed to the smaller variation of visual sensitivity in this area; otherwise, the predominant usefulness of OCT is in early to moderate glaucoma, rather than in advanced glaucoma.^{13,44} This tendency was in agreement with our previous studies.^{20,36} Sequentially, further improvement is required in the prediction accuracy in other regions.

The sensitivity of VF fluctuates in both the short⁴ and long⁵ terms, and accurately assessing VF progression is hampered by VF variability.⁸ The reliability of VF measurements is affected inherently by the patient's

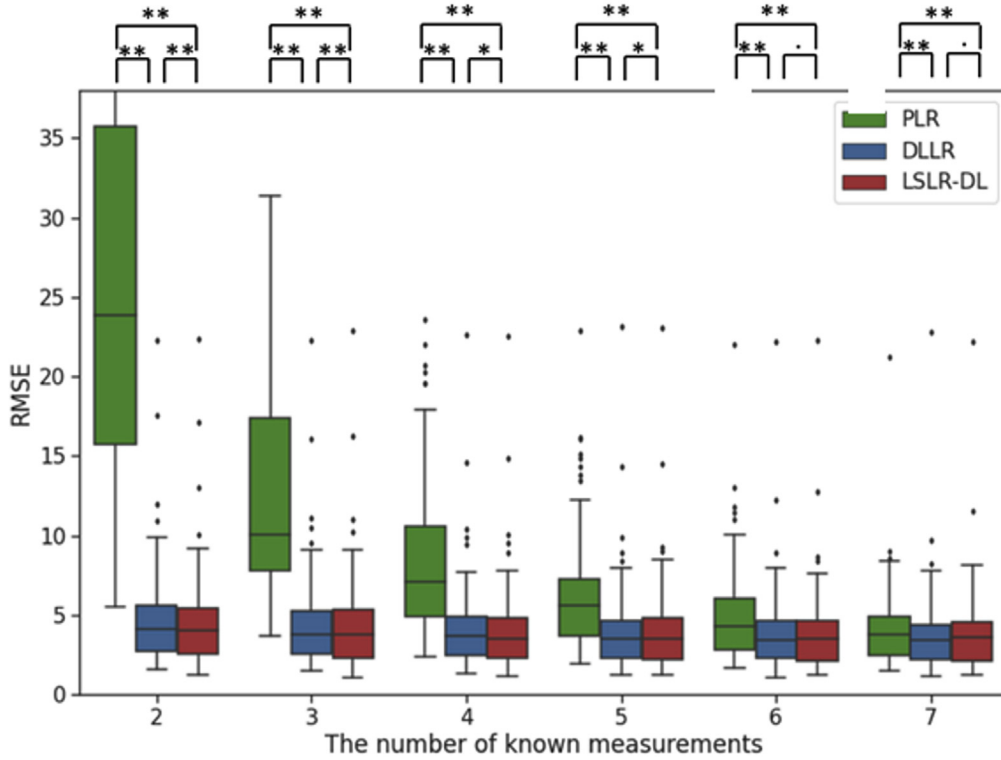


Figure 7. Box-and-whisker plot comparing the longitudinal root mean square error (RMSE) values between latent space linear regression and deep learning (LSLR-DL), deeply regularized latent space linear regression (DLLR), and pointwise linear regression (PLR) values. The LSLR-DL model significantly outperformed the PLR model with all sequences of visual field (VF; Humphrey Field Analyzer [HFA] 24-2 test). The LSLR-DL model significantly outperformed the DLLR model from the first and second VF tests to the first through fifth VF tests (HFA 24-2 test). * $P < 0.05$. ** $P < 0.01$. $\cdot P > 0.05$.

concentration, and previous reports have suggested that measurement noise cannot be avoided, even when reliability indices are sufficient.^{6,7} Consequently, a considerable number of VF results must be required to obtain reliable PLR results, as widely discussed in previous studies.^{45–48} By contrast, as shown in the present study, a much more accurate prediction of the longitudinal progression of VF (HFA 24-2 test) was achieved via the LSLR-DL model in comparison with the PLR model through the use of additional information on the thickness of the retinal layers (see Fig 8). The accuracy of LSLR-DL using the initial 2 VF results (HFA 24-2 test) was almost similar to that with an initial 6 or 7 VF results (HFA 24-2 test) with the PLR model when predicting the eighth VF

result (HFA 24-2 test). This is in agreement with our previous studies suggesting that the application of L1 or L2 regularization to the PLR model using the least absolute shrinkage and selection operator regression^{49,50} enabled a much more accurate prediction of VF progression.^{51,52} In comparison with the application of least absolute shrinkage and selection operator regression, the LSLR-DL model has possible merit in that the regularization in the longitudinal prediction was performed using related clinical information (cross-sectional prediction). In the LSLR-DL model, the longitudinal prediction of VF (HFA 24-2 test) progression was performed by further using the cross-sectional prediction of VF findings (HFA 10-2 test) as auxiliary information. The use of the cross-sectional

Table 3. Comparisons of the Longitudinal Root Mean Square Error Values across Pointwise Linear Regression, Deeply Regularized Latent Space Linear Regression, and Latent Space Linear Regression and Deep Learning Models

Method	No. of Known Visual Field (Humphrey Field Analyzer 24-2 Test) Measurements					
	2	3	4	5	6	7
PLR	27.5 (16.1)	12.9 (7.0)	8.2 (4.4)	6.0 (3.3)	4.7 (2.6)	4.0 (2.3)
DLLR	4.6 (2.7)	4.4 (2.7)	4.1 (2.6)	4.0 (2.6)	3.8 (2.4)	3.7 (2.3)
LSLR-DL	4.4 (2.7)* [†]	4.2 (2.7)* [†]	4.0 (2.6)* [†]	3.9 (2.6)* [†]	3.8 (2.4)*	3.7 (2.3)*

DLLR = deeply regularized latent space linear regression; LSLR-DL = latent space linear regression and deep learning; PLR = pointwise linear regression.

* $P < 0.01$, PLR vs. LSLR-DL.

[†] $P < 0.05$, DLLR vs. LSLR-DL.



Figure 8. Absolute prediction errors at each visual field (VF; Humphrey Field Analyzer [HFA] 24-2 test) test point with the latent space linear regression and deep learning (LSLR-DL), deeply regularized latent space linear regression (DLLR), and pointwise linear regression (PLR) models. The values were significantly smaller with the LSLR-DL model than the PLR model at all test points for series of the first and second VF tests to the first through fourth VF tests (HFA 24-2 test), in 51 test points in the first through fifth VF tests (HFA 24-2 test), in 24 test points in the first through sixth VF tests (HFA 24-2 test), and in 10 test points in the first through seventh VF tests (HFA 24-2 test). The values were significantly smaller with the LSLR-DL model than the DLLR model at 17 test points in the first and second VF test (HFA 24-2 test), in 9 test points in the first through third VF tests (HFA 24-2 test), in 5 test points in the first through fourth VF tests (HFA 24-2 test), and in 2 test points from the first through fifth VF tests to the first through sixth VF tests (HFA 24-2 test).

prediction of VF findings (HFA 10-2 test) can be regarded as a kind of additional regularization on the DL parameters in the longitudinal prediction of VF (HFA 24-2 test) progression. Moreover, the joint prediction of VF findings (HFA 10-2 test) and the prediction of VF (HFA 24-2 test) progression is a typical multitask learning problem²² in which individual tasks have relatedness and can lend strengths to each other. Indeed, accurate cross-sectional and longitudinal predictions are related to each other: an accurate cross-sectional prediction is beneficial for longitudinal prediction, which implies that pursuing each task provides useful information for another task. Consequently, an even more accurate prediction of the longitudinal progression of VF (HFA 24-2 test) was achieved via the LSLR-DL model in comparison with our previous model of DLLR (Fig 8).

However, we observed a significant difference only with short VF sequences (up to the first through fifth VF test) and not at all test points (between 17 and 32 test points). This may be because, first, DLLR already has achieved a considerably accurate prediction performance with sufficiently long VF sequences (RMSE, 3.8 dB with the first through sixth VF tests and 3.7 dB with the first through seventh VF tests). Second, in the cross-sectional prediction, the number of training OCTs was

significantly less than in the longitudinal prediction, with a ratio of approximately 1:10. Therefore, because of the significant relatively smaller number of OCT results, the cross-sectional prediction may provide only limited information on the longitudinal prediction. We used the OS+RPE thickness because the structure–function relationship becomes stronger by including this layer, in addition to the macular RNFL and GCL+IPL.²⁸ In our study with DLLR, we investigated the influence of omitting this layer, and as a result, no significant change was found in the prediction accuracy, and hence only a negligible difference would result by including or removing this layer.

A limitation of the present study is that in the cross-sectional prediction task, we studied the values of VF sensitivity only within the central 10°. The HFA 24-2 VF test usually is used in the clinical setting, and predictions using the HFA 24-2 test also would be useful clinically. We recommend a future study to investigate the current approach to predicting the HFA 24-2 test. The macular OCT scan does not cover such a wide area, and the current approach may not be used directly for that purpose. Nevertheless, we suggest the usefulness of the application of mathematical methods, such as SVR or DL,^{53,54} to predict VF findings in a different region. Further study is

necessary to investigate whether such an approach, in conjunction with the current method, is useful.

To conclude, we constructed a novel model of LSLR-DL that jointly performs a cross-sectional prediction of VF

(HFA 10-2 test) findings and longitudinal prediction of VF (HFA 24-2 test) findings, sharing a DL component such that either one can lend strengths to the other. Consequently, we found that an accurate prediction was achieved in both tasks.

Footnotes and Disclosures

Originally received: May 19, 2021.

Final revision: August 3, 2021.

Accepted: August 30, 2021.

Available online: September 7, 2021. Manuscript no. D-21-00086.

¹ Department of Ophthalmology, Seirei Hamamatsu General Hospital, Shizuoka, Hamamatsu, Japan.

² Seirei Christopher University, Shizuoka, Hamamatsu, Japan.

³ Nanovision Research Division, Research Institute of Electronics, Shizuoka University, Shizuoka, Japan.

⁴ The Graduate School for the Creation of New Photonics Industries, Shizuoka, Japan.

⁵ Department of Ophthalmology, Graduate School of Medicine and Faculty of Medicine, The University of Tokyo, Tokyo, Japan.

⁶ Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan.

⁷ Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

⁸ Department of Ophthalmology, Graduate School of Medical Sciences, Kitasato University, Kanagawa, Japan.

⁹ Department of Ophthalmology, Shimane University Faculty of Medicine, Izumo, Japan.

¹⁰ Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan.

¹¹ Oike-Ikeda Eye Clinic, Kyoto, Japan.

¹² Department of Ophthalmology, Hiroshima Memorial Hospital, Hiroshima, Japan.

¹³ Department of Ophthalmology, Hiroshima Prefectural Hospital, Hiroshima, Japan.

¹⁴ Inouye Eye Hospital, Tokyo, Japan.

¹⁵ Department of Ophthalmology, JR General Hospital, Tokyo, Japan.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Supported in part by the Ministry of Education, Culture, Sports, Science and Technology of Japan (grant nos.: 19H01114 [R.A. and K.Y.] and 18KK0253 [R.A.], 20K09784 [R.A.], 25861618 [H.M.], and 20K18337 [Y.F.]); and the Japan Science and Technology Agency (JPMJCR19U4:

AIP acceleration research [R.A. and K.Y.]). The sponsor or funding organization had no role in the design or conduct of this research.

HUMAN SUBJECTS: Human subjects were included in this study. The human ethics committees at University of Tokyo, Osaka University, Kyoto Prefectural University of Medicine, Shimane University, and Hiroshima Memorial Hospital approved the study. All research adhered to the tenets of the Declaration of Helsinki. All participants provided informed consent, except where the study protocols did not require that each patient provide written informed consent (based on the Ethical Guidelines for Medical and Health Research Involving Human Subjects issued by the Japanese government), in which cases, the protocol instead was posted at the outpatient clinic to notify study participants.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Asaoka, Xu, Murata, Kiwaki, Yamanishi

Analysis and interpretation: Asaoka, Xu, Murata, Kiwaki, Yamanishi

Data collection: Asaoka, Matsuura, Fujino, Tanito, Mori, Ikeda, Kanamoto, Inoue, Yamagami

Obtained funding: N/A; Study was performed as part of the authors' regular employment duties. No additional funding was provided.

Overall responsibility: Asaoka, Xu, Yamanishi

Abbreviations and Acronyms:

CNN = convolutional neural network; **CNN-TR** = convolutional neural network and tensor regression; **DL** = deep learning; **GCL** = ganglion cell layer; **DLR** = deeply regularized latent space linear regression; **HFA** = Humphrey Field Analyzer; **IPL** = inner plexiform layer; **LSLR-DL** = latent space linear regression and deep learning; **MLR** = multiple linear regression; **mTD** = mean total deviation; **OAG** = open-angle glaucoma; **OS** = outer segment; **PLR** = pointwise linear regression; **RMSE** = root mean square error; **RNFL** = retinal nerve fiber layer; **RPE** = retinal pigment epithelium; **SVR** = support vector regression; **VF** = visual field.

Key words:

Glaucoma, OCT, Progression, Visual field.

Correspondence:

Ryo Asaoka, MD, PhD, Department of Ophthalmology, University of Tokyo Graduate School of Medicine, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan. E-mail: rasaoka-ky@umin.ac.jp.

References

1. Aoki S, Murata H, Fujino Y, et al. Investigating the usefulness of a cluster-based trend analysis to detect visual field progression in patients with open-angle glaucoma. *Br J Ophthalmol*. 2017;101(12):1658–1665.
2. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90(3):262–267.
3. Fitzke FW, Hitchings RA, Poinoosawmy D, et al. Analysis of visual field progression in glaucoma. *Br J Ophthalmol*. 1996;80(1):40–48.
4. Flammer J, Drance SM, Fankhauser F, Augustiny L. Differential light threshold in automated static perimetry. Factors influencing short-term fluctuation. *Arch Ophthalmol*. 1984;102(6):876–879.
5. Flammer J, Drance SM, Zulauf M. Differential light threshold. Short- and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Arch Ophthalmol*. 1984;102(5):704–706.
6. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test

- reliability? *Invest Ophthalmol Vis Sci.* 2000;41(8):2201–2204.
7. Henson DB, Evans J, Chauhan BC, Lane C. Influence of fixation accuracy on threshold variability in patients with open angle glaucoma. *Invest Ophthalmol Vis Sci.* 1996;37(2):444–450.
 8. Jansonius NM. On the accuracy of measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2010;94(10):1404–1405.
 9. Crabb DP, Russell RA, Malik R, et al. Frequency of visual field testing when monitoring patients newly diagnosed with glaucoma: mixed methods and modelling. Southampton (UK): NIHR Journals Library; 2014 Aug (Health Services and Delivery Research, No. 2.27.) Available at: <https://www.ncbi.nlm.nih.gov/books/NBK259972/>.
 10. Malik R, Baker H, Russell RA, Crabb DP. A survey of attitudes of glaucoma subspecialists in England and Wales to visual field test intervals in relation to NICE guidelines. *BMJ Open.* 2013;3(5):e002067. <https://doi.org/10.1136/bmjopen-2012-002067>.
 11. Harwerth RS, Carter-Dawson L, Smith 3rd EL, et al. Neural losses correlated with visual losses in clinical perimetry. *Invest Ophthalmol Vis Sci.* 2004;45(9):3152–3160.
 12. Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res.* 2017;57:46–75.
 13. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res.* 2007;26(6):688–710.
 14. Wadhvani M, Bali SJ, Satyapal R, et al. Test-retest variability of retinal nerve fiber layer thickness and macular ganglion cell-inner plexiform layer thickness measurements using spectral-domain optical coherence tomography. *J Glaucoma.* 2015;24(5):e109–e115.
 15. Francoz M, Fenolland JR, Giraud JM, et al. Reproducibility of macular ganglion cell-inner plexiform layer thickness measurement with cirrus HD-OCT in normal, hypertensive and glaucomatous eyes. *Br J Ophthalmol.* 2014;98(3):322–328.
 16. Mwanza JC, Oakley JD, Budenz DL, et al. Macular ganglion cell-inner plexiform layer: automated detection and thickness reproducibility with spectral domain-optical coherence tomography in glaucoma. *Invest Ophthalmol Vis Sci.* 2011;52(11):8323–8329.
 17. Ghasia FF, El-Dairi M, Freedman SF, et al. Reproducibility of spectral-domain optical coherence tomography measurements in adult and pediatric glaucoma. *J Glaucoma.* 2015;24(1):55–63.
 18. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput.* 2006;18(7):1527–1554.
 19. Asaoka R, Murata H, Hirasawa K, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol.* 2019;198:136–145.
 20. Xu L, Asaoka R, Kiwaki T, et al. Predicting the glaucomatous central 10 degrees visual field from optical coherence tomography using deep learning and tensor regression. *Am J Ophthalmol.* 2020;218:304–313.
 21. Zheng Y, Xu L, Kiwaki T, et al. Glaucoma Progression prediction using retinal thickness via latent space linear regression. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2019:2278–2286.
 22. Zhang Y, Yang Q. A survey on multi-task learning. arXiv. 170708114. Published online July 25, 2017.
 23. Ruder S. An overview of multi-task learning in deep neural networks. arXiv. 1706.05098. Published online June 15, 2017.
 24. Xu L, Asaoka R, Kiwaki T, et al. A computational module for joint estimation and progression prediction of glaucoma. SIGKDD 2021: the Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data. *Mining.* 2021:3826–3834.
 25. Asaoka R, Murata H, Yanagisawa M, et al. The association between photoreceptor layer thickness measured by optical coherence tomography and visual sensitivity in glaucomatous eyes. *PLoS One.* 2017;12(10):e0184064.
 26. Littman H. Zur Bestimmung der wahren Größe eines Objektes auf dem Hintergrund des lebenden Auges. *Klin Monatsbl Augenheilkd.* 1982;180:286–289.
 27. Littman H. Zur Bestimmung der wahren Größe eines Objektes auf dem Hintergrund eines lebenden Auges. *Klin Monatsbl Augenheilkd.* 1988;192:66–67.
 28. Matsuura M, Fujino Y, Kanamoto T, et al. Improving the structure-function relationship in glaucomatous and normative eyes by incorporating photoreceptor layer thickness. *Sci Rep.* 2018;8(1):10450.
 29. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv. 1409155. Published online September 2014.
 30. Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. Cambridge, UK: Cambridge University Press; 2004.
 31. Kingma DP, Ba JL. Adam: a method for stochastic optimization. International Conference for Learning Representations. Published as a conference paper at ICLR 2015. arXiv. 1412.6980. 2015.
 32. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press; 2000.
 33. Xu L, Asaoka R, Murata H, et al. Improving visual field trend analysis with OCT and deeply regularized latent-space linear regression. *Ophthalmol Glaucoma.* 2021;4(1):78–88.
 34. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang.* 2008;59(4):390–412.
 35. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67(1):1–48.
 36. Hashimoto Y, Asaoka R, Kiwaki T, et al. Deep learning model to predict visual field in central 10 degrees from optical coherence tomography measurement in glaucoma. *Br J Ophthalmol.* 2021;105(4):507–513.
 37. Garway-Heath DF, Caprioli J, Fitzke FW, Hitchings RA. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Invest Ophthalmol Vis Sci.* 2000;41(7):1774–1782.
 38. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis.* 2015;15(3):211–252.
 39. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data.* 2019;6(1):60.
 40. Asaoka R, Tanito M, Shibata N, et al. Validation of a deep learning model to screen for glaucoma using images from different fundus cameras and data augmentation. *Ophthalmology Glaucoma.* 2019;2:224–231.
 41. Asaoka R, Murata H, Matsuura M, et al. Usefulness of data augmentation for visual field trend analyses in patients with glaucoma. *Br J Ophthalmol.* 2020;104(12):1697–1703.
 42. Weber J, Schultze T, Ulrich H. The visual field in advanced glaucoma. *Int Ophthalmol.* 1989;13(1–2):47–50.
 43. Hood DC, Raza AS, de Moraes CG, et al. Glaucomatous damage of the macula. *Prog Retin Eye Res.* 2013;32:1–21.

44. Swanson WH, Feliuss J, Pan F. Perimetric defects and ganglion cell damage: interpreting linear relations using a two-stage neural model. *Invest Ophthalmol Vis Sci.* 2004;45(2):466–472.
45. Nouri-Mahdavi K, Hoffman D, Gaasterland D, Caprioli J. Prediction of visual field progression in glaucoma. *Invest Ophthalmol Vis Sci.* 2004;45(12):4346–4351.
46. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest Ophthalmol Vis Sci.* 2002;43(5):1400–1407.
47. Krakau CE. A statistical trap in the evaluation of visual field decay. *Acta Ophthalmol Suppl.* 1985;173:19–21.
48. Spry PG, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold visual field data. *Invest Ophthalmol Vis Sci.* 2000;41(8):2192–2200.
49. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B.* 1996;58(1):267–288.
50. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.
51. Fujino Y, Murata H, Mayama C, Asaoka R. Applying “lasso” regression to predict future visual field progression in glaucoma patients. *Invest Ophthalmol Vis Sci.* 2015;56(4):2334–2339.
52. Fujino Y, Murata H, Mayama C, et al. Applying “lasso” regression to predict future glaucomatous visual field progression in the central 10 degrees. *J Glaucoma.* 2017;26(2):113–118.
53. Sugisaki K, Asaoka R, Inoue T, et al. Predicting Humphrey 10-2 visual field from 24-2 visual field in eyes with advanced glaucoma. *Br J Ophthalmol.* 2020;104(5):642–647.
54. Asaoka R. Measuring visual field progression in the central 10 degrees using additional information from central 24 degrees visual fields and ‘lasso regression.’ *PLoS One.* 2013;8(8):e72199.