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 Yari 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, PhD5

**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/).
spectral-domain OCT macular scanning area overlaps primarily with the retina within the central 10°. 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, 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. We recently reported the benefits of applying DL to OCT measurements regarding various aspects of VF, such as the diagnosis of glaucoma and the cross-sectional prediction of the HFA 10-2 test. 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. 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 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. 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 nerve fiber bundle; (2) a retinal nerve fiber bundle converging 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, 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, 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, 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 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, 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 pretrained 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:

$$f_{2}(C_{t}; \theta_{t}) = V_{t} + \varepsilon$$  \hspace{1cm} (1)

where $C_{t} \in \mathbb{R}^{3} \times H \times W$ 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 $t$th eye; $H$ and $W$ are the height and width of the measurement, respectively; $V_{t} \in \mathbb{R}^{H}$ is the VF in the 10-2 test mode; $f_{2}(\cdot; \theta_{t})$ is a function parameterized by VGG16; $\theta_{t}$ contains the parameters to be estimated; and $\varepsilon$ 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:

$$F_{t} = G_{f} \cdot W_{t}^{P_{t}} + \varepsilon$$  \hspace{1cm} (2)

where $F_{t} \in \mathbb{R}^{D} \times T$, is a matrix consisting of the sequence of VF results (HFA 24-2 test) for the $t$th eye, $D$ is the number of the HFA 24-2 test points, $T$ is the number of time stamps at which the HFA 24-2 tests were conducted, $G_{f} \in \mathbb{R}^{D} \times q$ and $W_{t} \in \mathbb{R}^{q} \times 2$ are 2-factor matrices to be estimated, $S$ is a predefined hyperparameter, and $P_{t}$ is the transpose of $P_{t} \in \mathbb{R}^{T} \times 2$, which was defined as $P_{t} = [t_{1}, 1]$. Because $t_{1} \in \mathbb{R}^{T} \times 1$ was a column vector of time stamps and $1 \in \mathbb{R}^{T} \times 1$ was a vector of ones, $W_{t}^{P_{t}}$ 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 $W_{t}$ correspond to the coefficient and intercept, respectively. $G_{f}$ then can be interpreted as a projection matrix that realizes the transformation of VF (HFA 24-2 test) threshold, and $\varepsilon$ contains the errors of the projection. In equation (2), the regression parameters in $W_{t}$ 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 $G_{f}$.

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:

$$f_{1}(C_{t}; \theta_{l}) = V_{l} + \varepsilon$$  \hspace{1cm} (3)

where $C_{t} \in \mathbb{R}^{3} \times H \times W$ is a 3-dimensional array representing the $t$th thicknesses of the OCT-measured retinal layer of the $t$th eye and $f_{1}(\cdot; \theta_{l})$ is a function parameterized by VGG16, and $\theta_{l}$ 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.](image-url)
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, \( f_i(\cdot; \theta_i) \) shares parameters in the convolutional layers with \( f_j(\cdot; \theta_j) \), which was for the cross-sectional prediction in equation (1). \( f_i(\cdot; \theta_i) \) is not exactly the same as \( f_j(\cdot; \theta_j) \), 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 \( f_i(\cdot; \theta_i) \) and \( f_j(\cdot; \theta_j) \). 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 10-2 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:

\[
W_i = \sum_{j \neq i}^N z_{ij} \cdot W_j + \epsilon
\]  

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

\[
z_{ij} = \exp \left\{ -\frac{\sum_{k=1}^{T_i} \sum_{d=1}^{D} (F_{ij}^{(d,k)} - F_{ij}^{(d,k)})^2}{(T_i D_i)^2} \right\},
\]  

where \( F_{ij}^{(d,k)} \) 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 \( F_{ij}^{(d,k)} \) 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. \( \sigma \) was the median of values \( \left( \sum_{k=1}^{T_i} \sum_{d=1}^{D} (F_{ij}^{(d,k)} - F_{ij}^{(d,k)})^2 / (T_i D_i)^2 \right) \) computed from the combinations of the \( i \)th eye and all other eyes. The definition of similarity between 2 eyes, \( z_{ij} \), followed the Gaussian kernel, was considered to be reasonable in the field of kernel regression, and worked well as demonstrated by our previous study.

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

\[
L = \sum_{i=0}^{N} ||F_i - G_i W_i P_i^T||_F^2 + \lambda_1 \sum_{j=1}^{N} ||W_0 - W_j||_F^2 + \lambda_2 \sum_{i=0}^{N} \sum_{l} ||f_l(C_i; \theta) - G_i W_i [f_l(1), 1]^T||_F^2,
\]
where the eyes of indices from \( j = 1 \) to \( N \) were used as training information for the learning of linear regression for the test eyes. \( \cdot \|_{F}^{2} \) is the square of the Frobenius norm, and \( \lambda_1, \lambda_2 \in \mathbb{R} \) 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:

\[
L = \sum_{i=0}^{N} \| F_i - G_i W_i P_i^T \|_{F}^{2} + \lambda_1 \sum_{j=1}^{N} z_{0,j} \| W_0 - W_j \|_{F}^{2} + \lambda_2 \left( \sum_{i=0}^{N} \sum_{l} \| f_l(C_i; \theta) - G_i W_i [r_i^T, 1]^T \|_{F}^{2} + \sum_{i=0}^{M} \| f_e(C_i; \theta) - V_i \|_{F}^{2} \right),
\]

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., Adam\(^{31}\)) 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 \( \lambda_1 \) and \( \lambda_2 \) were determined by grid search, and the value of the hyperparameter \( S \) was chosen as 4 because our previous study\(^{21}\) 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 × 224 × 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.\(^7\) At each test point, the VF (HFA 10-2 test) threshold was predicted by applying SVR to 150,528 (224 × 224 × 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 OCT measurements to regularize the linear regression of VF measurements against time in a latent space. 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 short\(^4\) and long\(^5\) terms. Moreover, VF measurements are associated with considerable noise,\(^6,7\) which hampers the accurate estimation of the speed of VF progression,\(^3\) whereas OCT measurements are highly reproducible.\(^14-17\)
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-TR\(^{30}\) to the GCC, RNFL, and OS+RPE thicknesses (224 × 224 pixels with 3 channels). Convolutional neural network and tensor regression is a recent method that combines a CNN with tensor regression.

\[
\text{RMSE} = \sqrt{\frac{1}{52} \sum_{i=1}^{52} (\text{predicted VF threshold of the ith point} - \text{actual VF threshold of the ith point})^{2}},
\]

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

\[
\text{RMSE} = \sqrt{\frac{1}{52} \sum_{i=1}^{52} (\text{predicted VF threshold of the ith point} - \text{actual VF threshold of the ith point})^{2}},
\]

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
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 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
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 \( (P < 0.05, \text{linear mixed model adjusted for multiple comparisons using Holm's method}) \) 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 \( (P < 0.001, \text{linear mixed model adjusted for multiple comparisons using Holm's method}) \) outperformed PLR. The LSLR-DL model significantly \( (P < 0.05, \text{linear mixed model adjusted for multiple comparisons using Holm's method}) \) 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 fourth 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 \( (P < 0.05, \text{linear mixed model adjusted for multiple comparisons using Holm's method}) \)
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 healthy participants or patients with OAG and a longitudinal 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 healthy participants or patients with OAG.

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). The application of DL only and the CNN-TR model resulted in even smaller RMSE values (7.7 and 7.3 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).

The sensitivity of VF fluctuates in both the short4 and long2 terms, and accurately assessing VF progression is hampered by VF variability.5 The reliability of VF measurements is affected inherently by the patient’s
concentration, and previous reports have suggested that measurement noise cannot be avoided, even when reliability indices are sufficient. Consequently, a considerable number of VF results must be required to obtain reliable PLR results, as widely discussed in previous studies.

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. 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 enabled a much more accurate prediction of VF progression. 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 information was found to be crucial for improving the prediction accuracy.

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.
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 problem 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 ofVF (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. 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, to predict VF findings in a different region. Further study is needed.
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, 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 Invoking 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-tky@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
44. Swanson WH, Felius 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.