https://doi.org/10.1016/j.ogla.2020.08.002
This is the Pre-Published Version.

Improving visual field trend analysis with optical coherence tomography and deeply-regularized latent-space linear regression.

Linchuan Xu, PhD [1], Ryo Asaoka, MD, PhD[2,3,4], Hiroshi Murata, MD[4], Taichi Kiwaki, PhD[1], Yuhui Zheng, [1] Masato Matsuura, [4,5] Yuri Fujino, [2,4,5], Masaki Tanito, MD, PhD[6], Kazuhiko Mori, MD, PhD[7], Yoko Ikeda, MD, PhD,[7, 8] Takashi Kanamoto, MD, PhD,[9,10] Kenji Yamanishi, PhD[1]

1: Graduate School of Information Science and technology, The University of Tokyo, Tokyo, 113-8655, Japan
2: Department of Ophthalmology, Seirei Hamamatsu General Hospital, Shizuoka, Hamamatsu, Japan
3: Seirei Christopher University, Shizuoka, Hamamatsu, Japan
4: Department of Ophthalmology, Graduate School of Medicine and Faculty of Medicine, The University of Tokyo, Tokyo, 113-8655, Japan
5: Department of Ophthalmology, Graduate School of Medical Sciences, Kitasato University, Kanagawa, 252-0374, Japan.
6: Department of Ophthalmology, Shimane University Faculty of Medicine,
7: Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, 602-8566, Japan
8: Oike-Ikeda Eye Clinic, Kyoto, Japan
9: Department of Ophthalmology, Hiroshima Memorial Hospital, Hiroshima, Japan
10: Department of Ophthalmology, Hiroshima Prefectural Hospital

Key words: Visual Field, Glaucoma, Progression, Optical Coherence Tomography

Short title: Data augmentation for visual field trend analyses

Correspondence and reprint requests to:
Ryo Asaoka, MD, PhD,
Department of Ophthalmology, University of Tokyo Graduate School of Medicine
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan.
Phone: +81-3-3815-5411, Fax:+81-3-3817-0798, email: rasaoka-tky@umin.ac.jp

Kenji Yamanishi, PhD,
Graduate School of Information Science and Technology,

© 2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

The University of Tokyo, Tokyo 113-8656, Japan
Phone: 03-5841-6895
Fax: 03-5841-8599
Email: yamanishi@mist.i.u-tokyo.ac.jp

Financial Support: Supported in part by Grants 19H01114 (RA and KY) and 18KK0253 (RA), 25861618 (HM) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, grant AIP acceleration reserach from the Japan Science and Technology Agency (RA and KY), and grants from Suzuken Memorial Foundation and Mitsui Life Social Welfare Foundation.
(no conflicting relationship exists for any author)

Address for reprints: Department of Ophthalmology, University of Tokyo Graduate School of Medicine
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan.

Introduction

Glaucoma, currently the second leading cause of blindness in the world,[1] causes irreversible visual field (VF) damage.[2] The worsening of VF damage can be halted by reducing intraocular pressure (IOP),[3-7] however, clinical interventions can cause serious complications[8-13] and decrease patients' quality of life.[14,15] Thus, adequate and timely treatment decisions are very important in the management of glaucoma; in particular, accurate assessment of VF progression is essential to decision-making. At clinical settings, the assessment of VF progression often relies on applying simple ordinary least-squares linear regression (OLSLR) to VF measurements, such as threshold values (TH), over time. This approach is employed in the software PROGRESSOR® (Medisoft Ltd., Leeds, UK).[16] It should be noted that VF TH fluctuate in the short[17] and long-term,[18] and VF measurements are associated with considerable noise even with good reliability indices,[19,20] which hampers the accurate estimation of VF progression speed,[21] in particular with point-wise linear regression (PLR).[22] As a result, PLR is often unreliable unless a considerable number of VFs are obtained.[23-27]

Glaucomatous VF damage results from the loss of retinal ganglion cells. Optical coherence tomography (OCT) is an imaging technology enabling high-resolution measurements of the retina, and is widely used to provide an objective evaluation of glaucomatous structural change, such as the thicknesses of the macular retinal nerve fiber layer (m-RNFL) and also the macular ganglion cell layer and inner plexiform layer (GCL + IPL).[28] A merit of the OCT measurement is its high reproducibility,[29-32] in contrast to VF measurements. A recent study suggested that RNFL thinning measured with OCT[33] and also rim area reduction in the optic disc measured with Heidelberg Retina Tomograph (HRT, Heidelberg Engineering GmbH, Heidelberg, Germany)[34] are predictive of future functional damage. Garway-Heath et al. developed a two-layered Bayesian model to incorporate OCT-measured cicumpapillary RNFL (cpRNFL) thickness into a VF regression model, which yielded more accurate predictions of future VF progression.[35] The development of deep learning (DL) methods represents a revolutionary advance in imaging recognition.[36] Indeed we have recently reported that it is advantageous to apply DL to OCT measurements in diagnosing glaucoma.[37] This implies it may also be helpful to use DL methods in VF progression models with structural information from OCT integrated into the model.

We recently developed a novel model to predict VF progression, which used OCT-measured m-RNFL and GCC thicknesses. The model was a deep learning method

known as deeply-regularized latent-space linear regression: DLLR. This new approach to measure VF progression significantly improved the prediction of future VF damage.[38] However this model was trained with a relatively small amount of data (253 eyes were used in five-fold-cross validation) and prediction performance was measured using internal cross-validation. The purpose of the current study was to generate a new DLLR model using a larger training dataset and validate its performance using an external 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 also Hiroshima Memorial Hospital. The study complied with the tenets of the Declaration of Helsinki. Written consent was given by patients for their information to be stored in the hospital database and used for research, otherwise, based on the regulations of the Japanese Guidelines for Epidemiologic Study 2008 (issued by the Japanese Government) the study protocols did not require that each patient provide written informed consent, instead the protocol was posted at the outpatient clinic to notify study participants.

Data collection

All of the data collected in the current study were obtained from Tokyo University Hospital, Osaka University Hospital, Hospital of Kyoto Prefectural University of Medicine, Oike-Ikeda Eye Clinic, Shimane University Hospital and Hiroshima Memorial Hospital. Inclusion criteria were: 1) glaucoma was the only disease causing VF damage; 2) each patient had at least eight VF measurements with the 24-2 or 30-2 Humphrey Field Analyzer (HFA, Carl Zeiss Meditec Inc, Dublin, CA). All of the studied patients were those with primary open angle glaucoma, which was defined as: 1) presence of typical glaucomatous changes in the optic nerve head such as a rim notch with a rim width $\leq$0.1 disc diameters or a vertical cup-to-disc ratio of $>$ 0.7 and/or 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; 2) gonioscopically wide open angles of grade 3 or 4 based on the Shaffer classification. Exclusion criterion were: 1) age below 20 years old; 2) possible secondary ocular hypertension in either eye. All patients had prior experience of VF measurement. These criteria were applied to both the training and testing datasets.

VF measurement

VF measurements were performed using the HFA with either the 30-2 or 24-2 program (Swedish Interactive Threshold Algorithm Standard), but only the central 52 test locations overlapping with the 24-2 test pattern were used in subsequent analyses. Only reliable VFs were included, defined as Fixation loss (FL) rate <33%, False positive (FP) rate <33%, and also False negative (FN) rate <33%.

OCT measurement

OCT data were obtained using the RS 3000 (Nidek Co ltd., Aichi, Japan). Axial length (AL) measurements were obtained using the OA-2000 (TOMEY, Aichi, Japan). All SD-OCT measurements were carried out after pupil dilation with 1% tropicamide and OCT imaging was performed using the laser scan protocol. Data with apparent eye movement or involuntary blinking or saccade during the measurement were carefully excluded. Following the manufacturer's recommendation, imaging data with quality factor < 7 were also excluded. Similar to our previous report analyzing OCT data,[39] the fovea was automatically identified as the pixel with thinnest retinal thickness close to the fixation point, and a square imaging area (30 x 30 degrees) was centered on the fovea, excluding the area of the optic disc and parapapillary atrophy. The magnification effect was corrected according to the manufacturer-provided formula based on Littman's equation, [40,41] using the measured AL value. Using software supplied by the manufacturer, thicknesses of i) m-RNFL, ii) GCL IPL, and iii) OS RPE (see **Figure 1**) were calculated. These were exported as 512 × 128 pixel images, and the mean thickness values in the whole analysis area (30 x 30 degrees) were calculated. The thickness of the outer segment and retinal pigment epithelium (OS RPE) was included in the DLLR model, because the structure-function relationship becomes stronger by including this layer, in addition to the m-RNFL and GCL IPL layers.[42]

Training and testing datasets

Eyes from Hiroshima Memorial Hospital, Osaka University Hospital, and Shimane University Hospital were employed as training data (998 eyes of 592 patients), whereas those from Hospital of Kyoto Prefectural University of Medicine were used as testing data (148 eyes of 84 patients).

Prediction models

The standard PLR model was carried out; for each VF test point, the TH was regressed against time using simple ordinary least-squares linear regression (OLSLR).

In brief, the DLLR model works as follows:

1) all VF test points of a single eye were forced to share regression parameters in order to take into account the correlations among VF test points.
2) eyes with similar VF values were forced to have similar regression parameters to avoid overfitting based only on their own data.
3) both OCT measurements and VF TH were transformed into a common latent space such that the progression of VF TH can be regularized by that observed in OCT sequences.

A more technical descriptions of DLLR follows. The transformation of VF TH into a latent space and the latent-space linear regression was realized by matrix factorization denoted:

$$F_i = G_i W_i P_i^T + \varepsilon, (1)$$

where $F_i \in \mathbb{R}^{D \times T_i}$ was a matrix consisting of the sequence of VF TH for the i-th eye, D was the number of VF points and T_i was the number of timestamps at which the VF tests were conducted, $G_i \in \mathbb{R}^{D \times R}$ and $W_i \in \mathbb{R}^{R \times 2}$ were two factor matrices to be estimated, R was a pre-defined hyper-parameter, P_i^T was the transpose of $P_i \in \mathbb{R}^{T_i \times 2}$ which was defined as $P_i = [t_i, 1_i]$. Since $t_i \in \mathbb{R}^{T_i \times 1}$ was a column vector of timestamps and $1_i \in \mathbb{R}^{T_i \times 1}$ was a vector of ones, $W_i P_i^T$ realized the linear regression of VF TH on time in the latent space. Then G_i can be interpreted as a projection matrix that realizes the transformation of VF TH, and ε contains the errors of the transformation.

In Eq. (1), the regression parameters in W_i were shared by all VF test points to force all VF test points the share the same progression pattern in the latent space. The difference among VF test points in the raw data space were preserved in the projection matrix G_i.

The transformation of the OCT measurement into the latent space was realized by a convolutional neural network (CNN), considering the spatial relations among retina voxels and the non-linear relation between the OCT-measured retinal layers' thicknesses and VF TH. CNN is currently the state-of-the-art model in capturing spatially-related information and non-linear transformation from one data space to another data space. To

regularize the progression pattern of VF TH, a CNN was employed to transform OCT-measured retinal layers' thicknesses directly to the parameters of the latent-space linear regression, which was realized by the following equation:

$$f(C_i;\theta) = W_i + \varepsilon \quad (2)$$

where $C_i \in \mathbb{R}^{3\times S_i\times H\times W}$ was a four-dimensional array representing an OCT-measured retinal layers' thicknesses, H and W were the height and width of the measurement, respectively, S_i was the number of OCT-measured retinal layers' thicknesses that were obtained on the i-th eye, and $f(\cdot;\theta)$ was a function parameterized by a CNN model and θ contained the parameters in the CNN model to be estimated. The regularization realized by Eq. (2) is motivated by the fact that functional damage is the result of structural damage in glaucoma.

Figure 1 shows the summarized structure of the proposed DLLR. The time series of OCT measurements were concatenated to be transformed into the latent space by a CNN while each VF TH was transformed into the latent space by linear projection. Linear regression of VF TH on time was performed in the latent space. OCT information was utilized to regularize the coefficient and intercept of the linear regression.

Since the number of OCT measurements varied from patient to patient, the size of the input, i.e., C_i, to the CNN model was not fixed. To tackle this issue, the CNN was designed to have convolutional layers only because convolutional layers can process data of varied values in the data's dimensions. The architecture of the CNN is illustrated in **Figure 2**. Besides, the convolution and max pooling, batch normalization and Rectified Linear Unit (ReLU) activation were associated with each layer except for the output layer.

To avoid overfitting on each eye's own data, the regression parameters of eyes with similar VF measurements were also forced to be similar. The relation between one eye and other eyes was defined as follows:

$$W_i = \sum_{j\neq i}^{N} z_{i,j} * W_j + \varepsilon, \quad (3)$$

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

$$z_{i,j} = \exp\{-\frac{(\sum_{k=1}^{T_i}\sum_{d=1}^{D}(F_i^{(d,k)}-F_j^{*(d,k)}))^2}{(T_i D\sigma_i)^2}, \quad (4)$$

where $F_i^{(d,k)}$ was the value at the *d*-th low and the *k*-th column of the VF matrix of the i-th eye, and F_j^* contained the VF THs interpolated by PLR, because the j-th eye may have different time stamps on which VF tests were conducted from those of the i-th eye. σ_i was the median of values $(\sum_{k=1}^{T_i}\sum_{d=1}^{D}(F_i^{(d,k)} - F_j^{*(d,k)}))^2/(T_i D)^2$ computed from the combinations of the i-th eye and all the other eyes. The definition of $z_{i,j}$ followed the Gaussian kernel, and was considered to be reasonable in the field of kernel regression.[43]

To jointly consider the regularizations from the OCT-measured retinal layers' thicknesses and other patients' information, the learning objective of DLLR for a target eye indexed by 0 was quantified as follows:

$$\mathcal{L} = \sum_{i=0}^{N}||F_i - G_i W_i P_i^T||_F^2 + \lambda_1\sum_{j=1}^{N} z_{0,j}||W_0 - W_j||_F^2 + \lambda_2\sum_{i=0}^{N}||f(C_i;\theta) - W_i||_F^2 , \quad (5)$$

where the eyes of indices from *j = 1 to N* were utilized as auxiliary information for the learning of linear regression for the target eye, $||\cdot||_F^2$ the square of Frobenius norm, and $\lambda_1, \lambda_2 \in \mathbb{R}$ are hyper-parameters.

Considering the small volume of data, regularizations on the model learning objective can help to avoid overfitting. In particular, auto-encoding regularizer on the CNN and L_2 regularization were employed. As a result, the final learning objective was obtained as follows:

$$\mathcal{L}_{final} = \mathcal{L} + \lambda_3\sum_{i=0}^{N}||C_i - g(f(C_i;\theta);\theta')||_F^2 + \lambda_4\sum_{i=0}^{N}(||G_i||_F^2 + ||W_i||_F^2), \quad (6)$$

where $\lambda_3, \lambda_4 \in \mathbb{R}$ were hyper-parameters, $g(\cdot;\theta')$ was the decoder function with parameters being of a CNN that had a structure symmetric to that of $f(\cdot;\theta)$.

DL optimization algorithms, e.g., Adam [44], can be utilized to solve the learning objective. After the learning, the prediction of the VF THs at a future timestamp of interest can be obtained as follows:

$$\text{predicted VF sesitivity} = G_0 W_0 [\text{time}, 1]^T, \quad (7)$$

where $[\text{time}, 1]^T$ is a matrix with two rows and a single column where time is the timestamp of interest for the target eye. Since valid VF TH values range from 0 to 40, the predicted values out of the range were clipped to be the floor and the ceiling values, respectively.

The original OCT-measured thicknesses of i) m-RNFL, ii) GCL IPL, and iii) outer segment (OS) + retinal pigment epithelium (RPE) were stored in images that were 512 × 128 pixels in dimension. We resized the images to be consistent with the more commonly used image size: 224 × 224 pixels. We achieved resizing by resampling using pixel area relation which is a preferred method for image decimation.

Statistical analysis

The DLLR model was implemented in Pytorch, and Adam was employed as the learning algorithm. The values of hyper-parameters λ_1, λ_2, λ_3 and λ_4 were determined by grid search, and the value of the hyper-parameter R was chosen as 4.

Using TH at each of the 52 test points in VF1-2, the values in the eighth VF were predicted with standard PLR and the prediction error was calculated as the root mean squared error ($RMSE_{PLR}$_1-2), defined as follows:

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

This was iterated for other series lengths: from VF1-3, calculating $RMSE_{PLR}$_1-3 to VF1-7, calculating $RMSE_{PLR}$_1-7.

Prediction errors were also calculated using mean absolute error (MAE) in order to assess clinical relevance.

RMSE values were also calculated for DLLR (from $RMSE_{DLLR}$_1-2 to $RMSE_{DLLR}$_1-7). Only OCT measurements within the VF observation period were used in the model.

RMSE values with PLR and DLLR were compared using the linear mixed model whereby values were nested within patients and test points and also the test points of the VF. 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 makes the assumption that all observations are independent of each other. In the current study, measurements were nested within subjects and also test points, and hence, dependent of each other. Ignoring this grouping of the measurements will result in the underestimation of standard errors of regression coefficients. The linear mixed model adjusts for the hierarchical structure of the data, modeling in a way in which measurements are grouped within subjects to reduce the possible bias derived from the nested structure of data.[45,46]

Results

Demographic information for the training and testing datasets is shown in **Table 1**. The age of the patients was 60.73 ± 13.47 (mean ± standard deviation: SD) and 61.21 ± 10.38 years, in the training and testing dataset, respectively. VFs were obtained over a period of 5.87 ± 1.93 and 5.39 ± 1.14 years, in the training and testing dataset, respectively. The mTD value in the initial VF was -6.21 ± 7.14 and -4.88 ± 4.55 dB, in the training and testing dataset respectively. The mTD progression rate with VF1-10 was -0.33 ± 0.77 and -0.30 ± 0.71 dB/year, in the training and testing dataset respectively.

Figure 3 shows the VF TH at each test point in the testing dataset.

RMSE values for PLR and DLLR are shown in **Figure 4** and **Table 2**. The results show that DLLR significantly ($p < 0.001$) outperformed PLR for with all sequences of VF.

Table 3 shows the MAE values with PLR and DLLR. The results also show that DLLR significantly ($p < 0.001$) outperformed PLR for all VF sequences.

Figure 5 shows the absolute prediction errors at each VF test point for PLR and DLLR. The values were significantly smaller with DLLR at all test points for series VF1-2 to VF1-5, in 46 test points in VF1-6 and nine test points in VF1-7.

Discussion

In the current study, the usefulness of DLLR, which incorporates the thicknesses of m-RNFL, GCL IPL, and OSL + RPE layers, into VF trend analyses was explored. As a result, a significant improvement in predicting VF progression was observed compared to PLR. In our previous study,[38] the DLLR model was trained with a relatively small number of eyes (202 or 203 eyes) which resulted in the RMSE values between 4.33 dB (prediction with VF1-7) and 4.96 dB (prediction with VF1-2) when the eighth VF was predicted. In contrast, the DLLR model was trained using a much larger dataset (998 eyes) in the current study, and as a result, the RMSE values were smaller; RMSEs ranged between 3.65 dB (prediction with VF1-7) and 4.57 dB (prediction with VF1-2) when the eighth VF was predicted, although it is not appropriate to compare these results, because of the different study designs (five-fold-cross validation was used in the previous study). The DLLR model uses OCT measurements and VF measurements; there are two broad differences between the two sets of measurements. First, the OCT measurements lie in a completely different data space to the VF TH, and, second, the OCT measurements may be obtained at different points in time to the VF measurements. DLLR address these two heterogeneities through the latent-space linear regression. For the data space heterogeneity, DLLR transforms both the OCT and VF measurements into a common latent space. For the time heterogeneity, DLLR focuses on the coefficient and the intercept of the latent-space linear regression of the sets of measurements.

In the current study, VF data were obtained from real world clinics at multiple institutes in Japan. The average mTD progression rate was -0.30 ± 0.71 dB/year with the average (mean) baseline mTD value of -4.88 ± 4.55 dB, in the testing dataset. Heijl et al. reported a VF progression rate of -0.80 dB/year with an average baseline MD value of -10.0 dB (median), in 583 patients with open angle glaucoma, also from a read world clinic.[47] In 587 patients with glaucoma, De Moraes et al. reported a -0.45 dB/year VF progression rate when the baseline MD value was equal to -7.1 dB (mean).[48] We previously collected VF data from 710 eyes in 490 patients with open angle glaucoma at multiple clinics in Japan, and the VF progression rate and the average baseline MD value were very similar to the current study: -0.26 dB/year and -6.9 dB.

In the current study, with ordinary PLR, the prediction error was smallest with the maximum VF sequence (VF1-7, 3.98 ± 2.25 dB), as shown in **Figure 4**. This prediction error increased as the number of VFs used in the prediction decreased, up to 27.48 ± 16.14 dB with VF1-2. This tendency was also observed with DLLR, however, the

magnitude of the prediction error was much smaller (from 3.65 ± 2.27 dB with VF1-9 to 4.57 ± 2.71 with VF1-2). For instance, the mean RMSE value with VF1-2 using DLLR was almost identical to that with VF1-6 using PLR; VFs were measured approximately every six monthsso this suggests that DLLR can achieve similar accuracy to PLR, but two years earlier.

We previously developed a variational Bayes linear regression (VBLR) model, in which PLR is optimized using a Bayesian technique that considers the spatial and temporal patterns of VF damage.[49,50] This model achieved much smaller prediction errors compared to PLR.Garway-Heath et al. also developed an approach to improve VF regression modelling; the approach, coined Non-Stationary Weibull Error Regression and Spatial enhancement (ANSWERS), controls for heteroskedasticity of residuals (non-normally distributed residuals) using a mixture of Weibull distributions.[51] It would be interesting to compare the prediction performance of DLLR with these previous models. In addition, it would be helpful to investigate whether the VBLR and ANSWERS models can achieve improved prediction by incorporating OCT measurements.

Numerous studies have suggested the usefulness of structural measurements, such as OCT, for the diagnosis of glaucoma, such as [37,52-58]. Much fewer studies have reported the utility of structural measurements to improve the estimates of future VF progression. Miki et al compared the rates of OCT-measured RNFL loss in patients suspected of having glaucoma who developed VF damage to those who did not develop VF damage. [33] As a result, it was suggested that the rate of global RNFL loss was more than twice as fast in eyes that later developed VF damage compared with those that did not develop VF damage; a 1μm/year faster rate of RNFL thickness loss corresponded to roughly a twofold higher risk of developing VF damage. This study did not, however, propose how measurements of RNFL thinning could be used to improve VF trend analyses. Nonetheless, it supports the current results that VF regression models can be improved by incorporating OCT-measured m-RNFL, GCL IPL, OS RPE thicknesses. Russell et al. reported that a Bayesian linear regression model, in which an MD trend analysis was reinforced using HRT-measured rim area as a prior, outperformed the standard MD trend analysis.[34] HRT requires subjective identification of the rim area and the instrument cannot measure the 3-dimensional structure of the optic disc, hence it is no longer frequently used in the clinical setting. Garway-Heath also reported that incorporating OCT-measured cpRNFL thickness into the ANSWERS model using a

two-layered Bayesian model (SANSWERS) further improved the prediction of future VF progression.[35] The DLLR model proposed in the current study is somewhat similar to the SANSWERS model, however, the only structural information used in SANSWERS is cpRNFL thickness (one value for each VF test point), whereas the DLLR models uses m-RNFL, GCL + IPL, and RPE + PhR measurements. In DLLR, each of the three layers of m-RNFL, GCL + IPL, and RPE + PhR data consisted of 512 x 128 pixels. One of the merits of machine learning methods, including DL, is that such dense data can be compressed to low dimensional features that effectively summarize the most essential information. The CNN is one such method to compress the information into low dimensional data through the convolutional technique.[59] We recently reported that analyzing raw OCT parameters at all locations is more useful than taking the averaged data of the scanned field when diagnosing glaucoma.[37,52] However, a weakness of this approach is that it is usually applied to cross-sectional image data and additional modeling is needed to deal with time series data. In particular, VF data are usually acquired with irregular intervals which makes the direct application of CNNs difficult. Direct application of conventional time series DL methods, such as Long short-term memory (LSTM)[60] is not appropriate in this condition. In the current study, this problem was overcome by applying the deep transformation approach in a latent space.

As described above, DLLR incorporates both OCT and VF data, for the purpose of predicting future VF progression. The construction of this model was motivated by the fact that functional damage is the result of structural damage in glaucoma. On the other hand, the progression of glaucoma (both structure and function) is largely associated with IOP.[3-7] We have recently shown the usefulness of incorporating IOP in VF regression models when predicting future VF progression. For instance, it was advantageous to regress VF against the product of time and IOP rather than IOP.[61] As another example, it was also useful to apply the segmented regression according to the status of IOP level.[62] It would be of further interest to examine whether incorporating IOP with DLLR yielded an even better prediction of future VF progression.

One of the limitations of the current study is a lack of results from the HFA 10-2 test which is the main corresponding region of the OCT scanning area. Recent studies have revealed that the HFA 24-2 test is not sufficient to assess damage in the central VF and therefore it is recommended to also measure the HFA 10-2 VF.[63-68] In addition, damage to this area of the VF is more directly associated with patients' vision related to

the quality of life[69,70] A future study should be attempted shedding light on the usefulness of VF augmentation in the HFA 10-2 test. In addition, it may be helpful to implement the DLLR algorithm in clinical decision-supportsoftware. For instance, VBLR is now available at the clinical settings, being implemented in multiple medical software of NAVIS® (Nidek, co.ltd., Gamagori, Aichi) and GlaPre® (Beeline co.ltd, Tokyo, japan).

In conclusion, we have investigated the usefulness of the DLLR method to measure VF progression. The approach appears to be much more useful than PLR, especially when the number of VFs available is small.

References

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol.* 2006;90(3):262-267.
2. 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.
3. Garway-Heath DF, Crabb DP, Bunce C, et al. Latanoprost for open-angle glaucoma (UKGTS): a randomised, multicentre, placebo-controlled trial. *Lancet.* 2015;385(9975):1295-1304.
4. The effectiveness of intraocular pressure reduction in the treatment of normal-tension glaucoma. Collaborative Normal-Tension Glaucoma Study Group. *American journal of ophthalmology.* 1998;126(4):498-505.
5. Heijl A, Leske MC, Bengtsson B, et al. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Archives of ophthalmology.* 2002;120(10):1268-1279.
6. Kass MA, Heuer DK, Higginbotham EJ, et al. The Ocular Hypertension Treatment Study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Archives of ophthalmology.* 2002;120(6):701-713; discussion 829-730.
7. Ederer F, Gaasterland DE, Sullivan EK, Investigators A. The Advanced Glaucoma Intervention Study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Controlled clinical trials.* 1994;15(4):299-325.
8. Palanca-Capistrano AM, Hall J, Cantor LB, Morgan L, Hoop J, WuDunn D. Long-term outcomes of intraoperative 5-fluorouracil versus intraoperative mitomycin C in primary trabeculectomy surgery. *Ophthalmology.* 2009;116(2):185-190.
9. Allingham RR, Damji KF, Shields MB. *Shields textbook of glaucoma.* 6th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2011.
10. Aihara M, Shirato S, Sakata R. Incidence of deepening of the upper eyelid sulcus after switching from latanoprost to bimatoprost. *Japanese journal of ophthalmology.* 2011;55(6):600-604.
11. Sakata R, Shirato S, Miyata K, Aihara M. Incidence of deepening of the upper eyelid sulcus on treatment with a tafluprost ophthalmic solution. *Japanese journal of ophthalmology.* 2014;58(2):212-217.
12. Yamada Y, Takayanagi R, Tsuchiya K, et al. Assessment of systemic adverse reactions induced by ophthalmic β-adrenergic receptor antagonists. *Journal of ocular pharmacology and therapeutics.* 2001;17(3):235-248.

13. Yamamoto T, Kuwayama Y, Kano K, Sawada A, Shoji N, Infection SGftJGSSoBr. Clinical features of bleb - related infection: a 5 - year survey in Japan. *Acta ophthalmologica.* 2013;91(7):619-624.
14. Matsuura M, Hirasawa K, Hirasawa H, et al. Developing an Item Bank to Measure Quality of Life in Individuals With Glaucoma, and the Results of the Interview With Patients: The Effect of Visual Function, Visual Field Progression Rate, Medical, and Surgical Treatments on Quality of Life. *J Glaucoma.* 2017;26(2):e64-e73.
15. Kotecha A, Feuer WJ, Barton K, Gedde SJ, Tube Versus Trabeculectomy Study G. Quality of Life in the Tube Versus Trabeculectomy Study. *American journal of ophthalmology.* 2017;176:228-235.
16. Fitzke FW, Hitchings RA, Poinoosawmy D, McNaught AI, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol.* 1996;80(1):40-48.
17. Flammer J, Drance SM, Fankhauser F, Augustiny L. Differential light threshold in automated static perimetry. Factors influencing short-term fluctuation. *Archives of ophthalmology.* 1984;102(6):876-879.
18. 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. *Archives of ophthalmology.* 1984;102(5):704-706.
19. 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.
20. 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.
21. Jansonius NM. On the accuracy of measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2010;94(10):1404-1405.
22. Jansonius NM. On the accuracy of measuring rates of visual field change in glaucoma. *Br J Ophthalmol.* 2010;94(10):1404-1405.
23. 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.
24. 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.
25. Krakau CE. A statistical trap in the evaluation of visual field decay. *Acta Ophthalmol Suppl.* 1985;173:19-21.
26. Spry PG, Bates AB, Johnson CA, Chauhan BC. Simulation of longitudinal threshold

visual field data. *Invest Ophthalmol Vis Sci.* 2000;41(8):2192-2200.

27. Taketani Y, Murata H, Fujino Y, Mayama C, Asaoka R. How Many Visual Fields Are Required to Precisely Predict Future Test Results in Glaucoma Patients When Using Different Trend Analyses? *Invest Ophthalmol Vis Sci.* 2015;56(6):4076-4082.
28. Chang R, Budenz DL. New developments in optical coherence tomography for glaucoma. *Current opinion in ophthalmology.* 2008;19(2):127-135.
29. Wadhwani 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):e109-115.
30. 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.
31. Mwanza JC, Oakley JD, Budenz DL, Chang RT, Knight OJ, Feuer WJ. 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.
32. Ghasia FF, El-Dairi M, Freedman SF, Rajani A, Asrani S. Reproducibility of spectral-domain optical coherence tomography measurements in adult and pediatric glaucoma. *J Glaucoma.* 2015;24(1):55-63.
33. Miki A, Medeiros FA, Weinreb RN, et al. Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes. *Ophthalmology.* 2014;121(7):1350-1358.
34. Russell RA, Malik R, Chauhan BC, Crabb DP, Garway-Heath DF. Improved estimates of visual field progression using bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci.* 2012;53(6):2760-2769.
35. Garway-Heath DF, Zhu H, Cheng Q, et al. Combining optical coherence tomography with visual field data to rapidly detect disease progression in glaucoma: a diagnostic accuracy study. *Health Technol Assess.* 2018;22(4):1-106.
36. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural computation.* 2006;18(7):1527-1554.
37. 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. *American journal of ophthalmology.* 2019;198:136-145.
38. 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.

39. 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.
40. Littman H. Zur Bestimmung der wahren Gr¨oße eines Objektes auf dem Hintergrund des lebenden Auges. *Klin Monatsbl Augenheilkd.* 1982;180:286–289.
41. Littman H. Zur Bestimmung der wahren Gr¨oße eines Objektes auf dem Hintergrund eines lebenden Auges. *Klin Monatsbl Augenheilkd.* 1988;192:66–67.
42. 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.
43. Shawe-Taylor John, Cristianini N. *Kernel methods for pattern analysis.* Vol Cambridge, UK: Cambridge university press; 2004.
44. Kingma DP, and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference for Learning Representations (ICLR).* 2015.
45. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language.* 2008;59(4):390-412.
46. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software.* 2015;67(1).
47. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol.* 2013;91(5):406-412.
48. De Moraes CG, Juthani VJ, Liebmann JM, et al. Risk factors for visual field progression in treated glaucoma. *Archives of ophthalmology.* 2011;129(5):562-568.
49. Murata H, Araie M, Asaoka R. A new approach to measure visual field progression in glaucoma patients using variational bayes linear regression. *Invest Ophthalmol Vis Sci.* 2014;55(12):8386-8392.
50. Murata H, Zangwill LM, Fujino Y, et al. Validating Variational Bayes Linear Regression Method With Multi-Central Datasets. *Invest Ophthalmol Vis Sci.* 2018;59(5):1897-1904.
51. Zhu H, Russell RA, Saunders LJ, Ceccon S, Garway-Heath DF, Crabb DP. Detecting changes in retinal function: Analysis with Non-Stationary Weibull Error Regression and Spatial enhancement (ANSWERS). *PloS one.* 2014;9(1):e85654.
52. Asaoka R, Hirasawa K, Iwase A, et al. Validating the Usefulness of the "Random Forests" Classifier to Diagnose Early Glaucoma With Optical Coherence Tomography. *American journal of ophthalmology.* 2017;174:95-103.

53. Mwanza JC, Warren JL, Budenz DL, Ganglion Cell Analysis Study G. Combining spectral domain optical coherence tomography structural parameters for the diagnosis of glaucoma with early visual field loss. *Invest Ophthalmol Vis Sci.* 2013;54(13):8393-8400.

54. Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci.* 2005;46(11):4147-4152.

55. Baskaran M, Ong EL, Li JL, et al. Classification algorithms enhance the discrimination of glaucoma from normal eyes using high-definition optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2012;53(4):2314-2320.

56. Tan O, Chopra V, Lu AT, et al. Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology.* 2009;116(12):2305-2314 e2301-2302.

57. Kim NR, Lee ES, Seong GJ, Kim JH, An HG, Kim CY. Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT in glaucoma. *Invest Ophthalmol Vis Sci.* 2010;51(9):4646-4651.

58. Yoshida T, Iwase A, Hirasawa H, et al. Discriminating between glaucoma and normal eyes using optical coherence tomography and the 'Random Forests' classifier. *PloS one.* 2014;9(8):e106117.

59. Schmidhuber J. Deep Learning in Neural Networks: An Overview. *Neural Networks.* 2015;61:85-117.

60. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation.* 1997;9(8):1735-1780.

61. Japanese Archive of Multicentral Database in Glaucoma construction g. A novel method to predict visual field progression more accurately, using intraocular pressure measurements in glaucoma patients. *Sci Rep.* 2016;6:31728.

62. Tomoda K, Morino K, Murata H, Asaoka R, K Y. Progression with Piecewise Regression Model from Heterogeneous Medical Data. *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies.* 2016;5:93-104.

63. De Moraes CG, Hood DC, Thenappan A, et al. 24-2 Visual Fields Miss Central Defects Shown on 10-2 Tests in Glaucoma Suspects, Ocular Hypertensives, and Early Glaucoma. *Ophthalmology.* 2017;124(10):1449-1456.

64. Grillo LM, Wang DL, Ramachandran R, et al. The 24-2 Visual Field Test Misses Central Macular Damage Confirmed by the 10-2 Visual Field Test and Optical Coherence Tomography. *Transl Vis Sci Technol.* 2016;5(2):15.

65. Park HY, Hwang BE, Shin HY, Park CK. Clinical Clues to Predict the Presence of Parafoveal Scotoma on Humphrey 10-2 Visual Field Using a Humphrey 24-2 Visual Field. *American journal of ophthalmology.* 2016;161:150-159.
66. Hangai M, Ikeda HO, Akagi T, Yoshimura N. Paracentral scotoma in glaucoma detected by 10-2 but not by 24-2 perimetry. *Jpn J Ophthalmol.* 2014;58(2):188-196.
67. Traynis I, De Moraes CG, Raza AS, Liebmann JM, Ritch R, Hood DC. Prevalence and nature of early glaucomatous defects in the central 10 degrees of the visual field. *JAMA Ophthalmol.* 2014;132(3):291-297.
68. Park SC, Kung Y, Su D, et al. Parafoveal scotoma progression in glaucoma: humphrey 10-2 versus 24-2 visual field analysis. *Ophthalmology.* 2013;120(8):1546-1550.
69. Murata H, Hirasawa H, Aoyama Y, et al. Identifying areas of the visual field important for quality of life in patients with glaucoma. *PloS one.* 2013;8(3):e58695.
70. Sumi I, Shirato S, Matsumoto S, Araie M. The relationship between visual disability and visual field in patients with glaucoma. *Ophthalmology.* 2003;110(2):332-339.

Table and Figure legend

Table 1: Demographic information of training and testing datasets

SD: standard deviation, AL: axial length, mTD: mean of total deviation, VF1-10: from 1st to 10th visual field, m-RNFL: macular retinal nerve fiber layer, GCL IPL: macular ganglion cell layer and inner plexiform layer, OS: outer segment, RPE: retinal pigment epithelium

Table 2: RMSE values with PLR and DLLR.

Values were represented as mean (SD).

RMSE: root mean squared error, PLR: point-wise linear regression, DLLR: deeply-regularized latent-space linear regression

Table 3. MAE values with PLR and DLLR.

Values were represented as mean (SD).

MAE: mean absolute error, PLR: point-wise linear regression, DLLR: deeply-regularized latent-space linear regression

Figure 1: The illustration of DLLR

For each patient, VF time series are transformed into a latent space while OCT time series are simultaneously transformed into the same latent space.

OCT: optical coherence tomography, VF: visual field, TH: threshold

Figure 2: The architecture of the convolutional neural network.

The first number in the kernel size corresponds to the number of channels, e.g., 4 in (4, 3x3x2) denoted four channels.

Figure 3. VF TH at each test point in the testing dataset.

Values were represented as mean (SD). Right eyes were mirror imaged to left eye.

VF: visual field, TH: threshold, SD: standard deviation.

Figure 4: The RMSE values for PLR and DLLR.

DLLR significantly ($p < 0.001$) outperformed PLR with all VF sequences.

RMSE: root mean squared error, PLR: point-wise linear regression, DLLR: deeply-regularized latent-space linear regression

Figure 5: The absolute prediction error at each VF teste point for PLR and DLLR.

Right eyes were mirror imaged to left eye. The values were significantly smaller with DLLR at all (from VF1-5 to VF1-7), 46 (VF1-6) or nine test points (VF1-7).

PLR: point-wise linear regression, DLLR: deeply-regularized latent-space linear regression