# Glaucoma Progression Prediction Using Retinal Thickness via Latent Space Linear Regression

Yuhui Zheng
Grad. Sch. of Inf. Sci. & Tech., The University of Tokyo,
Tokyo, Japan
zheng.yuhui@ci.i.u-tokyo.ac.jp

Linchuan Xu
Grad. Sch. of Inf. Sci. & Tech., The University of Tokyo,
Tokyo, Japan
linchuan_xu@mist.i.u-tokyo.ac.jp

Taichi Kiwaki
Grad. Sch. of Inf. Sci. & Tech., The University of Tokyo,
Tokyo, Japan
kiwaki@mist.i.u-tokyo.ac.jp

Jing Wang
Grad. Sch. of Inf. Sci. & Tech., The University of Tokyo,
Tokyo, Japan
jing_wang@mist.i.u-tokyo.ac.jp

Hiroshi Murata
Department of Ophthalmology, The University of Tokyo,
Tokyo, Japan
hmurata-tky@umin.net

Ryo Asaoka
Department of Ophthalmology, The University of Tokyo,
Tokyo, Japan
rasaoka-tky@umin.ac.jp

Kenji Yamanishi
Grad. Sch. of Inf. Sci. & Tech., The University of Tokyo,
Tokyo, Japan
yamanishi@mist.i.u-tokyo.ac.jp

## ABSTRACT

Prediction of glaucomatous visual field loss has significant clinical benefits because it can help with early detection of glaucoma as well as decision-making for treatments. Glaucomatous visual loss is conventionally captured through *visual field sensitivity* (*VF*) measurement, which is costly and time-consuming. Thus, existing approaches mainly predict future VF utilizing limited VF data collected in the past. Recently, optical coherence tomography (OCT) has been adopted to measure *retinal layers thickness* (*RT*) for considerably more low-cost treatment assistance. There then arises an important question in the context of ophthalmology: *are RT measurements beneficial for VF prediction?* In this paper, we propose a novel method to demonstrate the benefits provided by RT measurements. The challenge is management of the two heterogeneities of VF data and RT data as RT data are collected according to different clinical schedules and lie in a different space to VF data. To tackle these heterogeneities, we propose *latent progression patterns* (*LPPs*), a novel type of representations for glaucoma progression. Along with LPPs, we propose a method to transform VF series to an LPP based on matrix factorization and a method to transform RT series to an LPP based on deep neural networks. Partial VF and RT information is integrated in LPPs to provide accurate prediction. The proposed framework is named *deeply-regularized latent-space linear regression* (*DLLR*). We empirically demonstrate that our proposed method outperforms the state-of-the-art technique by 12% for the best case in terms of the mean of the root mean square error on a real dataset.

## CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Computer systems organization** → *Neural networks*; • **Mathematics of computing** → *Regression analysis*.

## KEYWORDS

Glaucoma Progression Prediction, Coupled Matrix Factorization, Convolutional Neural Networks, Regularization, Regression, Multi-view Learning

**ACM Reference Format:**
Yuhui Zheng, Linchuan Xu, Taichi Kiwaki, Jing Wang, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2019. Glaucoma Progression Prediction Using Retinal Thickness via Latent Space Linear Regression. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4–8, 2019, Anchorage, AK, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3292500.3330757

Linchuan Xu, Taichi Kiwaki, and Kenji Yamanishi are corresponding authors.


*KDD '19, August 4–8, 2019, Anchorage, AK, USA*

ACM ISBN 978-1-4503-6201-6/19/08…$15.00
https://doi.org/10.1145/3292500.3330757

## 1 INTRODUCTION

### 1.1 Background and Motivation

Glaucoma is a progressive optic neuropathy that is the second leading cause of blindness in the world [14]. With disease progression, the retinal layer thickness is reduced, which eventually causes irreversible damage to the sight of glaucomatous patients [10]. Accurate prediction of glaucoma progression has attracted intensive research interests because of the irreversible and progressive nature of glaucoma [6, 9, 10, 17, 20, 21, 29]. Progression prediction can be used to

estimate the risk of future loss of sight and is, therefore, at the core of clinical decision-making for medical treatments [6, 9, 10, 29].

The current standard prediction framework requires a sufficient number of visual field sensitivity (VF) measurements [6, 29]. The visual field test is the standard basis for identifying glaucomatous progression, because it is the only method to directly assess the quality of vision of a glaucoma patient. However, the visual field test, which is conducted using the Humphrey field analyzer (HFA) [23], is very time-consuming and labor-intensive and requires considerable patient efforts. Furthermore, as a result of irregular diagnosis scheduling and variation in clinical conditions in different world regions, it is unreasonable to expect that every patient will receive a sufficient number of qualified VF measurements. To tackle this issue, intensive research has been performed to enhance the prediction accuracy achievable with a small number of VF measurements [17, 19–21]. However, the prediction accuracy remains far from satisfactory for clinical practice, especially when only two or three VF measurements are performed. Note that this scenario is common in early glaucoma diagnoses, which is of particular importance.

Recently, a new ophthalmological measure, the *retinal layers thickness* (RT), has been introduced to glaucoma diagnosis. Because glaucoma often induces structural changes in the retina, RT conveys valuable information on the pathological state of a glaucomatous eye. The RT measurement is obtained via *optical coherence tomography* (OCT) [11], which is much less time-consuming and labor-intensive than HFA and does not require patient efforts. In the field of ophthalmology, a large volume of evidence on the quantitative relationship between VF and RT measurement has been accumulated [10, 12, 28, 35]. And it has been reported that VF can be estimated from a single RT measurement [26, 31]. A natural question arises here: *Can we leverage glaucoma progression prediction accuracy with a series of RT measurements in addition to a small number of VF measurements?* The problem of interest is illustrated in Figure 1. A positive answer to this question would have a substantial impact on ophthalmology, not only because visual field test costs could be reduced significantly, but also because early and accurate medical decision-making could be realized before the occurrence of considerable glaucoma-induced damage to eyesight.

RT data cannot be directly utilized by conventional prediction methods, because of the heterogeneities of VF and RT data. First, the dimensionality of RT measurements is much higher than that of VF measurements. Moreover, each RT measurement can have several channels, whereas each VF measurement has only one. This type of heterogeneity is *data space heterogeneity*. Second, RT measurements are obtained at different timestamps to VF measurements. Moreover, both types of measurements are performed according to irregular schedules. This type of heterogeneity is *temporal heterogeneity*.

This paper, therefore, proposes a novel progression prediction framework for enhanced prediction accuracy incorporating RT measurements, which is achieved by handling the above heterogeneities. This method is named *deeply-regularized latent-space linear regression* (*DLLR*).

## 1.2 Novelty and Significance

The novelty and significance of this paper are summarized as follows:

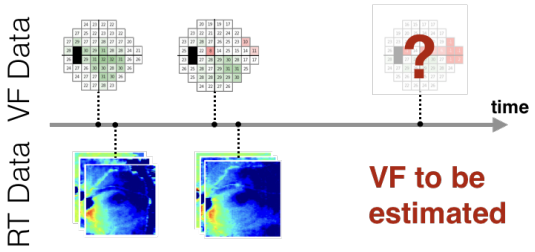



**Figure 1: An overview of the problem where a time series of VF measurements and a different time series of RT measurements are utilized for the prediction of a VF measurement at a future timestamp of interest.**

- **We propose a novel framework for utilizing RT time series to improve VF prediction.** Although previous studies have revealed relationships between VF and RT, no studies have provided evidence that RT time series can be beneficial for VF prediction. In this paper, we propose a novel framework of DLLR that utilizes RT time series for VF prediction. Experiments on a real glaucoma dataset show that DLLR works very effectively and outperforms the previous state-of-the-art technique by 12% in terms of the mean of the *root mean square error* for the best case, where the number of known VF measurements is two. Such a case is of particular clinical importance as mentioned earlier. We expect this result could have a substantial impact on ophthalmology in a sense that RT measurements could provide reliable auxiliary information for glaucoma progression prediction.
- **We propose a novel type of representations for glaucoma progression to integrate VF and RT features.** To handle the temporal and data space heterogeneities between VF and RT data, we propose *latent progression patterns* (*LPPs*), an abstract and temporally-linear representation for glaucoma progression that can simultaneously describe outcomes of glaucoma progression in both domains: thinning of RT and decline of VF. Moreover, we propose methods to transform RT and VF time series to LPPs: a deep *fully-convolutional neural network* (FCN) for RT series and a matrix factorization (MF) algorithm for VF series. Both of the methods are novel in the sense that they are designed to produce aforementioned LPPs.

## 1.3 Related Work

Initially, linear regression analysis of VF time series for each eye was used to predict glaucomatous progression [9]. However, application of simple linear regression to the time series of an individual eye with a point-wise estimator is insufficient and usually yields overfitting when the number of data points is small; this is often the case with clinical data. Although various regression models have been used to obtain prediction based on data from the target eye, the prediction accuracy has been very unsatisfactory because of the lack of data [6, 29].

To overcome this data insufficiency, some recent studies have proposed to exploit VF data from eyes of other patients to predict the

glaucomatous progression in the target eye. For example, Liang et al. previously proposed a spatio-temporal clustering-based method [17]. By following a two-step methodology, those researchers first identified eyes similar to the target eye (in terms of the spatial and temporal features of the disease progression), and then utilized data from those eyes for the prediction. As the performance of Liang et al.'s method largely depends on the employed clustering, Maya et al. proposed a hierarchical minimum description length (MDL)-based clustering method [19] to determine the optimal clustering, and substituted the resultant clusters for those used in the first step of Liang et al.'s method. However, as eyes are grouped into clusters, the above two methods cannot utilize all available data. As all the considered data pertain to glaucomatous eyes, it is reasonable to exploit the full dataset, even though different eyes have different degrees of similarity with the target eye. Maya et al. later proposed a multi-task matrix factorization method in which each task is performed for each visual field point [20]. Each pair of tasks are related via the similarity of the two corresponding points. Furthermore, Murata et al. used Bayesian linear regression to utilize measurements for other eyes [21] and evaluated the performance of this method on multi-central datasets [22]. In contrast to all existing methods described so far, the present study utilizes RT data for VF prediction.

The problem of utilizing the retinal information to support glaucoma treatment has been intensively studied. Previously, Hood et al. and Ajtony et al. showed a statistically significant correlation between VF and RT [2, 10]. The relationship between the averaged thickness of the *ganglion cell and inner plexiform layer* (GCIPL) and VF was demonstrated by Eura et al. [5]. Furthermore, detection of glaucomatous eyes based on RT measurement using several machine learning methods has been investigated [35]. Notably, Asaoka et al. evaluated a random forest classifier for identification of glaucomatous eyes using RT data from multiple imaging instruments, and significantly improved early stages glaucoma diagnosis performance [3, 4, 34]. However, the above previous studies did not indicate any method of numerically estimating multidimensional VF values. Uesaka et al. performed the first work on VF estimation from RT data [31], employing two different approaches: multi-view learning with Nonnegative MF and transformation with convolutional neural networks (CNNs). Sugiura et al. improved the CNN-based method with pattern-based regularization to overcome the imbalanced non-paired data problem [26]. Again, the framework presented in this work is distinct from these two reports in that we improve prediction of future VF based on RT time series.

### 1.4 Organization of This Paper

The remainder of this paper is organized as follows. Section 2 introduces the glaucoma data, i.e., the VF and RT data studied in this paper. We present the proposed model in Section 3 and report experiments involving a real dataset in Section 4. Section 5 concludes the paper and introduces future directions.

## 2 GLAUCOMA DATA

### 2.1 VF Data

Figure 2 shows a sample visual field test result for a left eye obtained via HFA in the SITA-Standard mode 24-2. Note that we consider

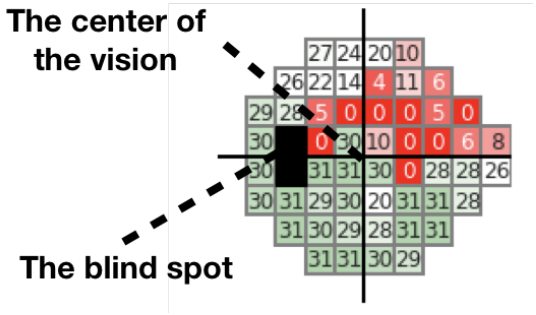
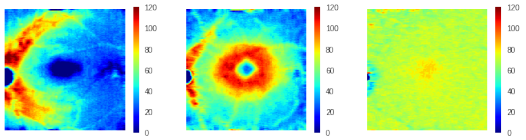
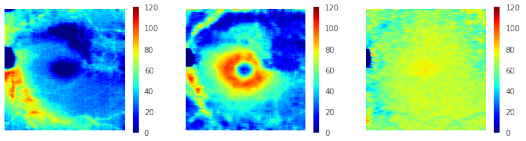



**Figure 2: An example VF measurement in 24-2 mode**



**(a) Thickness of RNFL, GCIPL, and RCL of normal eye.**



**(b) Thickness of RNFL, GCIPL, and RCL of glaucomatous eye.**

**Figure 3: RT measurements. Color indicates the thickness of retinal layers. The numbers along the color bars are in micrometers.**

24-2 mode data in this work, because 24-2 mode VF measurements are common in the clinical fields [33]. Each box in the visual field represents an angular location (i.e., a visual field point) in the sight of the eye. The integers in boxes are called threshold (TH) values. TH values range from 0 to 40 and indicate the photosensitivities of the eye at the visual field points; a larger value indicates a better sense of sight. Details of SITA-Standard can be found in [23]. Throughout this paper, a VF measuremnet is represented as a collection of THs at all the visual field points. Because the VF of a glaucomatous eye decreases on average as the disease progresses, a patient is usually subject to multiple VF measurements over a number of years. Thus, our data consisted of time series of VF measurements for a number of eyes. The two black grids on the left side correspond to the blind spot within the eye.

### 2.2 RT Data

Our RT data were obtained via OCT, which is a non-contact and non-invasive medical imaging technology for measuring the structure of the optic disc and the thickness of the surrounding retina [10]. They consisted of OCT measurements of three different retinal layers: the GCIPL, retinal nerve fiber layer (RNFL), and rod and cone layer (RCL), as illustrated in Figure 3. Each image of each layer contained 512×128 pixels. In glaucomatous eyes, malfunction of the retinal

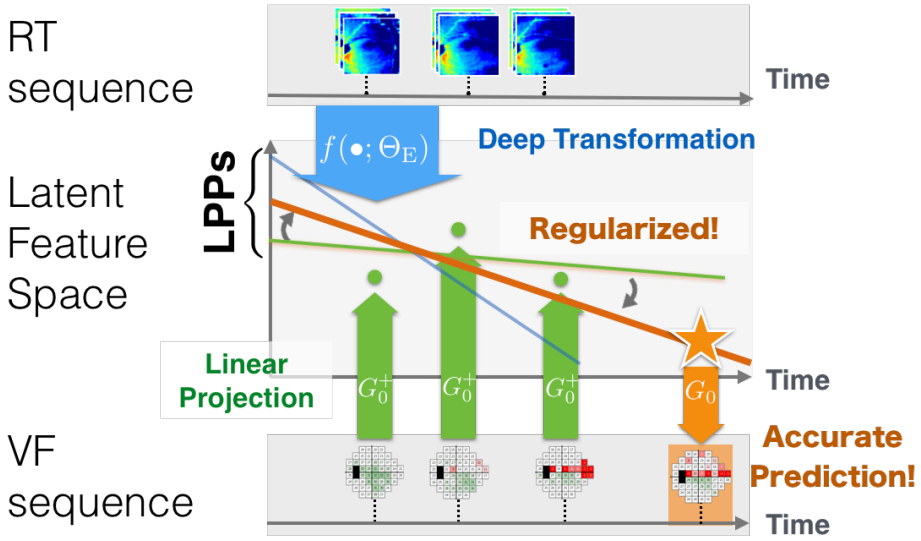



**Figure 4: Overview of proposed model.**

nerve cells reduces the retinal layers thickness, as apparent from the comparison of the RT data for a normal eye and a glaucomatous eye presented in Figure 3a and Figure 3b, respectively.

## 3 PROPOSED FRAMEWORK

### 3.1 Overview

We present the overview of our proposed framework in Figure 4. For a target eye, our objective is to accurately predict VF of this eye at a future time of interest utilizing both VF and RT time series. To resolve the data space and temporal heterogeneities between the domains, we propose LPPs, a novel type of representations for glaucoma progression for individual eyes. Moreover, we propose transformation methods from VF series to LPPs and from RT series to LPPs. Finally, we propose to make use of VF and RT series of other eyes to enhance prediction for the target eye. With LPPs and these methods, we formulate our proposed framework, DLLR. Little more details are sketched as follows:

- **Modeling glaucoma progression with LPPs.** LPPs consist of linear functions from time to low dimensional latent features of glaucoma as illustrated in Figure 4. This type of representations is the key to resolving the data space and temporal heterogeneities between RT and VF series. As for the data space heterogeneity, abstract and low-dimensional nature of latent features is advantageous in absorbing this type of heterogeneity. As for the temporal heterogeneity, LPPs are independent of timestamps of original RT and VF series because LPPs are functions of time.
- **MF for transforming VF series to LPPs.** As indicated in Figure 4, we relate LPPs and VF measurements via low-rank linear projections: $G_0$ and its pseudoinverse $G_0^+$. $G_0^+$ projects VF series into low-dimensional latent feature space. An LPP of the VF series is the regression line for these projected VF measurements. $G_0$ projects LPPs back into the space of VF. To simultaneously optimize $G_0$ and LPPs, we propose an MF method in Section 3.3.
- **A deep FCN for transforming RT series to LPPs.** An MF method is not suitable for transforming RT series because of the high-dimensionality of RT measurements and non-linearity between RT and VF [10]. Furthermore, each eye may have a different number of RT measurements due to a different clinical schedule from each other, which means that the size of an input can vary from eye to eye. Therefore, we propose an FCN that can deal with inputs with variable sizes in Section 3.4.
- **Making use of measurements from other eyes to regularize prediction of the target eye.** Previous studies [17, 20] have demonstrated that prediction accuracy can be enhanced by making use of information from other eyes than the target. We also employ this idea to leverage the prediction performance.
- **Glaucoma prediction with DLLR.** The methods to relate RT and VF series to LPPs are integrated to give our proposed framework DLLR. DLLR offers a natural way to predict glaucoma progression utilizing both RT and VF time series.

### 3.2 LPPs

We here formally define LPPs. An LPP is linear function from time $t$ to an $R$-dimensional latent space for describing glaucoma progression of a single eye. An LPP can be expressed as:

$$\boldsymbol{A}t + \boldsymbol{B} = \boldsymbol{W}[t, 1]^\top, \tag{1}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are slopes and intercepts of an LPP, $\boldsymbol{W} \stackrel{\text{def}}{=} [\boldsymbol{A}\,\boldsymbol{B}]$ is an $R \times 2$ matrix that contains parameters of the LPP. $R$ is expected to be a small integer. Because $\boldsymbol{W}$ solely determines an LPP, $\boldsymbol{W}$ can be identified with the LPP.

Although the linear functions over time may seem to be simple, it has been demonstrated to be very effective in modeling glaucoma

progression [9, 17, 20, 21]. We therefore expect that LPPs would be rich enough for describing effects of glaucoma on both RT and VF.

### 3.3 Transforming VF series to LPPs

Recall that our goal is to predict the VF measurement at a future timestamp for a target eye, which is indexed by 0 throughout this paper. To perform this task with LPPs, we construct a matrix from the VF time series of the target eye as $\boldsymbol{F_0} \in \mathbb{R}^{D\times T_0}$, where $D \in \mathbb{R}$ is the number of visual field points, and $T_0 \in \mathbb{R}$ is the number of VF measurements of the target eye plus one. The last column of $\boldsymbol{F}_0$ is to be estimated and is filled with zeros. In addition to $\boldsymbol{F_0}$, we also have $\boldsymbol{t}_0 \in \mathbb{R}^{T_0}$, a column vector containing timestamps of VF measurements.

Letting $\boldsymbol{G}_0 \in \mathbb{R}^{D\times R}$ be a projection matrix in Figure 4, we can approximate the VF time series with an LPP $\boldsymbol{W}_0$ of the target as:

$$\boldsymbol{F_0} = \boldsymbol{G}_0\boldsymbol{W}_0\boldsymbol{P}_0^\top + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{P}_0$ is defined to be $[\boldsymbol{t}_0\ \boldsymbol{1}_{T_0}]$, $\boldsymbol{\varepsilon}$ is the error matrix expected to capture the noises, and $\boldsymbol{1}_{T_0} \in \mathbb{R}^{T_0}$ is a column vector of which all elements are one.

To transform VF time series to an LPP, we need to simultaneously optimize the projection matrix $\boldsymbol{G}_0$ and an LPP $\boldsymbol{W}_0$ that best approximates $\boldsymbol{F}_0$. This problem falls into the following matrix completion problem that can be expressed as follows:

$$\underset{\boldsymbol{G}_0,\boldsymbol{W}_0}{\arg\min}\ ||\boldsymbol{M}_0 \odot (\boldsymbol{F}_0 - \boldsymbol{G}_0\boldsymbol{W}_0\boldsymbol{P}_0^\top)||_F^2, \tag{3}$$

where $\boldsymbol{M}_0$ is a mask matrix of which all elements take the value of one except for the last column of all zeros, $||\bullet||_F$ denotes the Frobenius norm, and $\odot$ denotes the element-wise product.

It is worth noting that the state-of-the-art method proposed by Maya et al. [20] also solves a matrix completion problem. However, in [20], the matrix is constructed for each visual field point instead of each eye as in our method. This difference is the key to obtaining LPPs.

*3.3.1 Making Use of Information from Other Eyes.* To stabilize the prediction, we introduce relationships among different eyes via introducing a potential term over $\{\boldsymbol{W}_i\}_i$. In particular, we perform coupled matrix factorization [1] for all eyes:

$$\underset{\boldsymbol{G}_i,\boldsymbol{W}_i}{\arg\min} \sum_{i=0}^{N} ||\boldsymbol{M}_i \odot (\boldsymbol{F}_i - \boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{P}_i^\top)||_F^2 + \lambda_0 \sum_{j=1}^{N} z_{0,j} \left\|\boldsymbol{W_0} - \boldsymbol{W_j}\right\|_F^2, \tag{4}$$

where $\lambda_0$ is a hyper parameter, $N$ is the number of eyes in a dataset plus one, and we define bases, LPPs, and timestamp matrix $\{\boldsymbol{G_i}, \boldsymbol{W_i}, \boldsymbol{P}_i\}_i$ for all eyes other than the target. $z_{0,j} \in \mathbb{R}$ is the similarity between the target eye and eye $j$ quantified as follows:

$$z_{0,j} = \exp\left(-\frac{\left(\sum_{k=1}^{T_0}\sum_{d=1}^{D}\left(\boldsymbol{F}_0^{(d,k)} - \boldsymbol{F}_j^{*(d,k)}\right)\right)^2}{(TD\sigma)^2}\right). \tag{5}$$

Here, $\boldsymbol{F}_j^*$ is a VF matrix interpolated by patient-wise linear regression as in [20]. Further, $\sigma^2$ is the median of values $\left(\sum_{k=1}^{T}\sum_{d=1}^{D}\left(\boldsymbol{F_i}^{*(d,k)} - \boldsymbol{F_j}^{*(d,k)}\right)\right)^2/(TD)^2$ computed for all possible combinations of $i$ and $j$. The formulation of $z_{0,j}$ is similar to that of the Gaussian kernel. This approach is considered to be reasonable in the field of kernel regression [25]. Note that for the sake of appearance, we also add a mask for each eye. However, the mask matrices for all other eyes are matrices only containing entries of one; thus they can be omitted from the optimization.

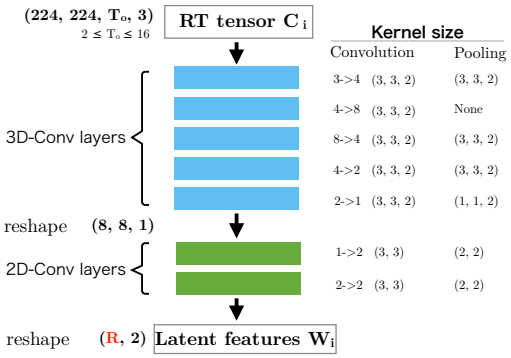



**Figure 5: The architecture of fully convolutional network for $R = 4$**

### 3.4 Transforming RT series to LPPs

RT series should also be transformed to LPPs for being utilized in VF prediction. Linear projections will not be effective for this data domain because the relation between the RT and VF of an eye is known to be nonlinear [10]. Moreover, simple linear projections do not care spatial alignment of the RT data, which is expected to convey valuable information of glaucomatous eyes. All these facts motivate us to employ CNNs because they can approximate any transformation with adequate number of parameters and can be aware of topology of the input [15, 16].

The size of each layer of RT measurements in our dataset after preprocessing is $224\times224$; thus, each input channel of the CNN has a size of $224\times224\times T_i\times3$. As different numbers of RT measurements, i.e., $T_i$, may be available for different eyes, the input size varies from eye to eye. To tackle this issue, we adopt an FCN, which consists of convolutional layers only [7, 18]. Because convolutional layers can process data of any size in the temporal or spatial domains, FCNs have the same capability.

Our FCN architecture is illustrated in Figure 5. The FCN layers are partitioned into two parts. The first part consists of three-dimensional convolutional layers for transforming four-dimensional inputs of different sizes into two-dimensional arrays of a fixed size. The second part consists of two-dimensional convolutional layers for transforming two-dimensional inputs of the fixed size into two-dimensional tensors.

Through dimensionality reduction of the pooling operation, this network is designed to output a $4\times2$ tensor; its size is exactly the same as those of parameter matrices of LPPs. Thus, this network $f(\bullet;\Theta_\mathrm{E})$ can be trained to estimate LPPs as

$$\boldsymbol{W_i} \approx f(\boldsymbol{C_i};\Theta_\mathrm{E}), \tag{6}$$

where $\boldsymbol{C_i} \in \mathbb{R}^{224\times224\times T_i\times3}$ is the RT input, and $\Theta_\mathrm{E}$ is a set of all network parameters. To the best of our knowledge, there are no existing

CNNs in glaucoma data mining designed for directly estimating parameters of latent representations of glaucoma progression.

### 3.5 DLLR

By combining the matrix factorization and the FCN, we define our proposed DLLR Framework. To be specific, the objective function of DLLR is quantified as follows:

$$\mathcal{L}_{\mathrm{DLLR}} = \sum_{i=0}^{N} ||\boldsymbol{M_i} \odot (\boldsymbol{F_i} - \boldsymbol{G_i}\boldsymbol{W_i}\boldsymbol{P_i}^{\top})||_F^2 + \lambda_0 \sum_{j=1}^{N} z_{0,j}||\boldsymbol{W_0} - \boldsymbol{W_j}||_F^2 + \lambda_1 \sum_{i=0}^{N} ||f(\boldsymbol{C_i}; \Theta_{\mathrm{E}}) - \boldsymbol{W_i}||_F^2, \tag{7}$$

where $\lambda_1$ is a hyperparameter for tuning the importance of the term with RT series. Note that DLLR conforms to multi-view learning paradigm [27] where learning on different domains can result in a better regularization. We thus expect the usage of RT series in addition to VF series to improve prediction accuracy.

*3.5.1 DLLR Objective for Practice.* We can improve the objective in Eq. (7) by introducing conventional regularization schemes for practical usage. We here think of L2 regularizers and auto-encoding regularizer [8, 32] to compose an improved objective as:

$$\mathcal{L}'_{\mathrm{DLLR}} = \mathcal{L}_{\mathrm{DLLR}} + \lambda_2 \sum_{i=0}^{N} ||\boldsymbol{C_i} - g(f(\boldsymbol{C_i}; \Theta_{\mathrm{E}}); \Theta_{\mathrm{D}})||_F^2 + \lambda_3 \sum_{i=0}^{N} (||\boldsymbol{G_i}||_F^2 + ||\boldsymbol{W_i}||_F^2), \tag{8}$$

where $\lambda_2$ and $\lambda_3$ are hyperparameters for controlling the component importance. The fourth term introduces auto-encoding regularization [8, 32] with $g(\cdot; \Theta_{\mathrm{D}})$ being a decoder function with parameters $\Theta_{\mathrm{D}}$ that are implemented with an FCN having a symmetric structure to that associated with $f(\cdot; \Theta_{\mathrm{E}})$. This term is useful because our glaucoma data are small-sized. In experiments with the real glaucoma dataset, we use this objective.

In the proposed framework, we optimize the objective with Adam [13], which is a first-order gradient-based optimizer. Adam works well in practice and compares favorably to other adaptive stochastic learning methods [24]. The complexity of minimizing the loss function is thus the complexity of computing the loss function, which is dominated by performance of the convolution operations.

### 3.6 Details of DLLR Algorithm and Implementation

Implementations of DLLR and baseline methods are available at https://goo.gl/QDP1Cr. Algorithm 1 shows the pseudo-codes of DLLR. The algorithm is implemented with Pytorch V1.0.0 and can be executed on a machine with two TITAN X GPUs (Pascal). The training time for each patient is about 71 minutes.

**Algorithm 1:** The training procedure

**Input :** The preprocessed VF training data, the preprocessed VF test data, RT data, $R$, $\lambda_0$, $\lambda_1$, $\lambda_2$, $\lambda_3$
**Output :** $G$, $W$

```
Align timestamps of VF and RT according to their first measurement respectively;
For each point of each training patient, use all the measurements to fit a linear function of VF sensitivity against time;
for each test patient do
    for S in the range(2,8) do
        Construct a VF matrix F_0;
        Construct a VF matrix for each training patient by linear interpolation where the timestamps are the first S timestamps and the last timestamp of the current test patient;
        All the VF matrix values are rescaled into the range [0,1] by dividing themselves by 40;
        Compute the pair-wise similarity z_{0,i} by Eq. (5);
        For each patient, select all RT measurements whose timestamps are not larger than the S_th timestamp, and concatenate them into a four-dimensional tensor;
        Pre-train the model as indicated in Section 4.3.3;
        Use Adam to solve optimization problem Eq. (8);
return G, W
```

## 4 EXPERIMENTS

### 4.1 Description of Data

The dataset employed in this study was provided by the Department of Ophthalmology, The University of Tokyo Hospital. The dataset contained diagnosis information for 254 glaucomatous eyes in various pathological stages, collected from Hiroshima Memorial Hospital and Osaka University Hospital. For each glaucomatous eye, there were two diagnosis time series, VF measurements and RT measurements.

We studied VF measurements in the 24-2 mode (containing 52 visual field points). The VF data range was $[0, 40]$ measured in terms of TH. The lengths of VF time series ranged from 3 to 25. The mean length was 9.2.

The RT measurements were taken in the vicinity of the macula, and had three channels corresponding to three different retinal layers: GCPL, RNFL and RCL. Each measurement channel was a two-dimensional array with size $512 \times 128$. The range of the RNFL data was $[0, 1848]$, the range of the GCPL data was $[0, 1673]$, and the range of the RCL data was $[0, 394.3]$. All measurements were in micrometers ($\mu m$). The lengths of RT time series ranged from 1 to 13. The mean length was 5.3.

### 4.2 Data Preprocessing

We horizontally flipped the VF data of the right eye to resolve the horizontal symmetry. Furthermore, we flipped the RT data both vertically and horizontally for the optical image flip in the eyes to be taken into account. We thus maintained the natural relationship between the RT and VF [30]. All channels of the RT data were resized to $224 \times 224$, and the values were rescaled to $[0, 255]$.

When feeding RT series to the FCN, we aligned RT series into a four-dimensional tensor so that the index for the third axis of the tensor roughly corresponded to RT timestamps. If we encounter

a missing value along this axis, we fill the missing values with estimates from linear regression. Further details of preprocessing can be found in the supplementary material.

## 4.3 Experimental Settings

We partitioned the datasets into five sets of approximately equal size and applied five-fold cross-validation. One set was regarded as the test set, and the remaining sets formed the training set. We predicted the VF from the most recent diagnosis for each patient using information from their first $S$ VF diagnoses and corresponding RT diagnoses , for which the last timestamp did not exceed that of the VF. Note that the $S$ diagnoses are also referred to as *known measurements* in this section.

*4.3.1 Baseline Methods.* Three baselines were used here: patient-wise linear regression (LR); a clustering-based method (TSLR) [17]; and the multi-task matrix completion approach (MTMC) [20], which is the state-of-the-art technique.

Linear regression is performed for each visual field point. For the TLSR, we chose to employ the SVD-EM method for spatial feature clustering part, because this choice showed generally best performance in [17]. We also followed the grid search strategy for hyperparameters given in the original paper [17]. For the MTMC, we followed the same set of hyperparameters for the model and preprocessing of the dataset reconstruction. Note that only the VF data can be utilized by these baselines, while both the VF and RT data can be utilized by the proposed model.

*4.3.2 Evaluation Metric.* The performance was quantified by the *root mean squared error* (RMSE), which is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{d=1}^{D} (\hat{y}_d - y_d)^2}{D}}, \tag{9}$$

where $y_d$ and $\hat{y}_d$ denote the actual and predicted $d$th local TH value on the final diagnosis day, respectively.

We also evaluated the methods in terms of the improvement rate compared to the patient-wise linear regression, which is defined as

$$\mathrm{IR}(S) = \frac{1}{M} \sum_{i=1}^{M} \left(1 - \frac{\mathrm{RMSE}_{\boldsymbol{f}}(i,S)}{\mathrm{RMSE}_{\boldsymbol{LR}}(i,S)}\right), \tag{10}$$

where $M$ is the number of eyes in the test dataset, $\mathrm{RMSE}_{\boldsymbol{LR}}(i,S)$ is the RMSE of the $i$th patient for patient-wise linear regression, and $\mathrm{RMSE}_{\boldsymbol{f}}(i,S)$ is that for prediction method $\boldsymbol{f}$, with $S$ known measurements. If IR equaled 0, the prediction accuracy of $\boldsymbol{f}$ was identical to that of the LR. If IR became large, $\boldsymbol{f}$ outperformed the LR and vice versa.

*4.3.3 Model Setting and Training Details.* We set the latent space dimensionality $R$ to be four, according to the prior knowledge that only a small number of alternative patterns is needed to achieve a good fit [20]. We optimize hyperparameters $\lambda_0, \lambda_1, \lambda_2$ via 5-fold cross-validation. $\lambda_3$ was fixed because the result was less sensitive to the choice. The hyperparameter tuning was performed for each of the settings of $S$. For each $\boldsymbol{W_i} = [\boldsymbol{A_i}\boldsymbol{B_i}]$, we initialized $\boldsymbol{A_i}$ and $\boldsymbol{B_i}$ by sampling from uniform distributions. We initialized $\boldsymbol{G_i}$, $\Theta_{\mathrm{E}}$, and $\Theta_{\mathrm{D}}$ via pretraining. Further details can be found in the supplementary material.

**Table 1: Comparison of the mean of RMSE for prediction of TH values.**

| Methods | $S$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| LR | 50.66 | 21.11 | 13.21 | 9.76 | 7.98 | 6.52 | 6.06 |
| TSLR | 6.27 | 6.01 | 5.65 | 5.54 | 5.52 | 5.09 | 5.02 |
| MTMC | 5.65 | 5.47 | 4.99 | 4.82 | 4.83 | 4.67 | 4.62 |
| **DLLR** | **4.96** | **4.76** | **4.61** | **4.49** | **4.48** | **4.46** | **4.43** |

**Table 2: Comparison of the median of RMSE for prediction of TH values.**

| Methods | $S$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| LR | 34.74 | 16.94 | 10.58 | 8.64 | 7.12 | 5.68 | 5.24 |
| TSLR | 6.09 | 5.75 | 5.45 | 5.10 | 5.00 | 4.70 | 4.66 |
| MTMC | 4.96 | 4.74 | 4.65 | 4.53 | 4.57 | 4.36 | 4.27 |
| **DLLR** | **4.62** | **4.37** | **4.38** | **4.20** | **4.19** | **4.15** | **4.15** |

**Table 3: Comparison of improvement rates.**

| Methods | $S$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| TSLR | 0.67 | 0.48 | 0.34 | 0.22 | 0.17 | 0.10 | 0.03 |
| MTMC | 0.76 | 0.62 | 0.48 | 0.39 | 0.30 | 0.20 | 0.15 |
| **DLLR** | **0.80** | **0.65** | **0.52** | **0.44** | **0.36** | **0.25** | **0.20** |

## 4.4 Performance Comparison

The experiment results are presented in Figure 6 and Tables 1 to 3. It is apparent that the proposed method outperformed the baselines in terms of all performance metrics for all values of $S$, especially when $S$ was small (e.g., two and three). In particular, DLLR outperformed the previous state-of-the-art technique by as much as 12% when $S = 2$. Moreover, Table 3 critically demonstrates the performance improvements, indicating that the IR values of the proposed method could be as high as 0.8. Note that a smaller number of available VF measurements constitutes a more challenging prediction task, as apparent from the increased errors with decreasing $S$ obtained in this experiment. However, even for $S = 3$, the proposed method could achieve comparable performance to the state-of-the-art technique with $S = 8$, which can be seen as acceptable accuracy for medical field. This result shows that the proposed model can greatly improve the efficiency of the current glaucoma diagnosis because the number of costly VF measurements can be drastically reduced through the use of cheaper RT measurements to maintain acceptable prediction accuracy.

To check the significance of the results, we conducted one-sided binomial tests between our method and the baselines for left eyes and right eyes, respectively. The null hypothesis is that the performance of our method was worse than that of the other methods. For S < 8, the results of these tests show that the improved performance of our method is statistically significant, with all p-values less than 0.01. For S = 8 and the test between our method and the state-of-the-art method, the p-value for the right eye is 0.018 and for

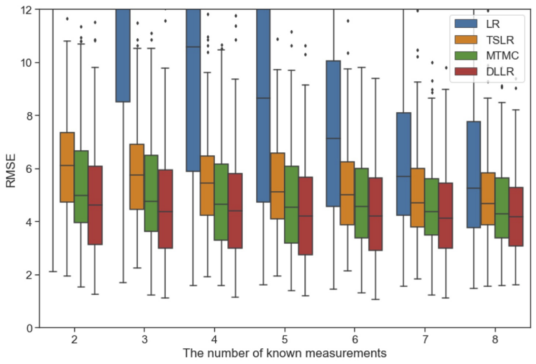



**Figure 6: Prediction error for VF TH by different methods against the number of known measurements**

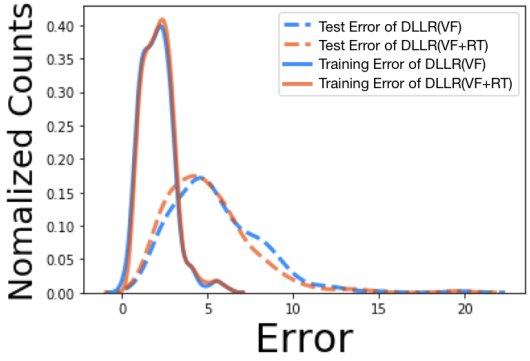



**Figure 7: Training and test errors of DLLR with and without RT data.**

the left eye is 0.111, which may be because with sufficient number of known VF measurements, the prediction without OCT data is already sufficiently accurate.

## 4.5 Effectiveness of RT Data for Progression Prediction

*4.5.1 Effect on Generalization.* To verify the effectiveness of the RT data as regards the progression prediction, we conducted an experiment in which we compared the degree of overfitting and the performance of the proposed model with and without the RT data. (Note that we can simply remove all components with $f(\cdot;\Theta_E)$ from Eq. 8 if we wish to exclude the RT data. ) In particular, we examined the training and test errors for these two scenarios. For each test eye, the training error was the mean of the RMSE computed on all known VF measurements, and the test error was the RMSE on the VF measurement to be predicted.

The results for $S = 2$ and all test patients are presented in Figure 7. Here, DLLR(VF) and DLLR(VF+RT) denote exclusion and inclusion of the RT data, respectively. It is apparent that with the regularization from the RT data, the training error is slightly increased; this is expected because the solution to the matrix factorization is distorted from the optimal by the regularization from the RT data. However, this regularization provides visible benefits to the prediction, as apparent from the considerably large margin between the test errors of the DLLR(VF) and DLLR(VF+RT). Collectively, we can conclude that, when the number of known measurements is small, the DLLR(VF) tends to overfit the data and, thus, does not generalize well on future measurements. However, if we further utilize the RT data, this overfitting can be effectively alleviated.

*4.5.2 A Case Study.* We further examine the linear functions fitted by MTMC and DLLR. In particular, we present the linear functions for a local TH value of a representative test patient in Figure 8. When $S = 3$, since the first two known measurements do not decrease considerably, MTMC learns a linear function with a negative slope of small magnitude. In this case, MTMC overfits the limited VF data, and achieves bad prediction. When $S = 8$, we can see that decreasing of the VF value is significant, and hence MTMC learns a negative slope with a relative larger magnitude. But even when

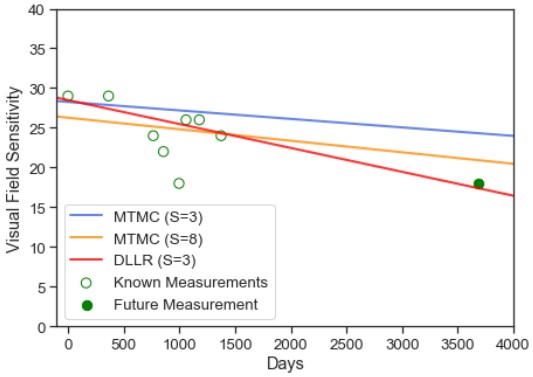



**Figure 8: Example regression lines for TH value fitted with DLLR and MTMC.**

$S = 3$, the proposed DLLR does not overfit the data and learns a negative slope with a large magnitude to give accurate prediction.

## 5 CONCLUSION AND FUTURE WORK

In this paper, we address an important open problem in ophthalmology of effective utilization of the RT time series obtained through OCT to improve glaucoma progression prediction, for the first time. To this end, we propose LPPs, latent representations for glaucoma progression to resolve the temporal and data space heterogeneities between VF and RT measurements, which inhibit effective use of RT measurements in glaucoma progression prediction. Along with methods to transform RT and VF time series to LPPs, we propose DLLR, a framework for predicting glaucoma progression from both RT and VF time series via LPPs. Experiments on a real dataset demonstrate that the proposed model works very effectively.

To our best knowledge, no previous studies had shown evidence that RT measurements are beneficial for the prediction of VF. Hence, we foresee that the proposed framework could give a substantial impact on ophthalmology.

We can seek several future directions. In the proposed model, we assume that the relationship between VF and time is linear. Firstly,

we can extend our framework to allow nonlinear relationship between VF and time because the speed of glaucoma progression can actually vary due to health condition of a patient [23]. Another important direction is evaluation of our framework with validation datasets collected in different institutes than those from which our dataset are provided. This should clarify ophthalmological impact of our proposed framework.

## ACKNOWLEDGMENTS

This work was partially supported by JSPS KAKENHI Grant Number 19H01114, JSPS KAKENHI Grant Number JP18K18121, and by JST AIP, and by Grants 25861618 (HM) and 26462679 and 18KK0253 (RA) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by Grants from Suzuken Memorial Foundation and Mitsui Life Social Welfare Foundation. We thank Mr Yuri Fujino and Mr Masato Matsuura, Department of Ophthalmology, The University of Tokyo, for providing us useful comments. We thank Dr Atsuya Miki, Osaka University Hospital, and Dr Takashi Kanamoto, Hiroshima Memorial Hospital, for providing us the data.

## REFERENCES

[1] Evrim Acar, Gozde Gurdeniz, Morten A Rasmussen, Daniela Rago, Lars O Dragsted, and Rasmus Bro. 2012. Coupled matrix factorization with sparse factors to identify potential biomarkers in metabolomics. *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)* 3, 3 (2012), 22–43.

[2] Csilla Ajtony, Zsolt Balla, Szabolcs Somoskeoy, and Balint Kovacs. 2007. Relationship between visual field sensitivity and retinal nerve fiber layer thickness as measured by optical coherence tomography. *Investigative Ophthalmology & Visual Science* 48, 1 (2007), 258–263.

[3] Ryo Asaoka, Kazunori Hirasawa, Aiko Iwase, Yuri Fujino, Hiroshi Murata, Nobuyuki Shoji, and Makoto Araie. 2017. Validating the usefulness of the "random forests" classifier to diagnose early glaucoma with optical coherence tomography. *American journal of ophthalmology* 174 (2017), 95–103.

[4] Ryo Asaoka, Richard A Russell, Rizwan Malik, Gay Verdon-Roe, and David F Garway-Heath. 2012. Evaluating A 'Random Forest' Decision Tree Classifier To Identify Eyes With Glaucomatous Visual Field Loss Applied To Measurements From Multiple Imaging Devices. *Investigative Ophthalmology & Visual Science* 53, 14 (2012), 5619–5619.

[5] Mariko Eura. 2014. orrespondence Between Visual Field Test Results and GCL+IPL Thickness in the Maculae of Glaucomatous Eyes. *Ph.D. Thesis* (2014), 1–16.

[6] Yuri Fujino, Hiroshi Murata, Chihiro Mayama, and Ryo Asaoka. 2015. Applying "lasso" regression to predict future visual field progression in glaucoma patients. *Investigative Ophthalmology & Visual Science* 56, 4 (2015), 2334–2339.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *European Conference on Computer Vision*. Springer, 346–361.

[8] Geoffrey E Hinton and Richard S Zemel. 1994. Autoencoders, minimum description length and Helmholtz free energy. In *Advances in Neural Information Processing Systems*. 3–10.

[9] Catharina Holmin and CET Krakau. 1982. REGRESSION ANALYSIS OF THE CENTRAL VISUAL FIELD IN CHRONIC GLAUCOMA CASES: A follow-up study using automatic perimetry. *Acta Ophthalmologica* 60, 2 (1982), 267–274.

[10] Donald C Hood and Randy H Kardon. 2007. A framework for comparing structural and functional measures of glaucomatous damage. *Progress in retinal and eye research* 26, 6 (2007), 688–710.

[11] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. 1991. Optical coherence tomography. *Science* 254, 5035 (1991), 1178–1181.

[12] Akiyasu Kanamori, Makoto Nakamura, Michael FT Escano, Ryu Seya, Hidetaka Maeda, and Akira Negi. 2003. Evaluation of the glaucomatous damage on retinal nerve fiber layer thickness measured by optical coherence tomography. *American Journal of Ophthalmology* 135, 4 (2003), 513–520.

[13] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[14] Sharon Kingman. 2004. Glaucoma is second leading cause of blindness globally. *Bulletin of the World Health Organization* 82 (2004), 887–888.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. 1097–1105.

[16] Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard, and Lawrence D Jackel. 1990. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*. 396–404.

[17] Zenghan Liang, Ryota Tomioka, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2013. Quantitative prediction of glaucomatous visual field loss from few measurements. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 1121–1126.

[18] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3431–3440.

[19] Shigeru Maya, Kai Morino, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2015. Discovery of glaucoma progressive patterns using hierarchical MDL-based clustering. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1979–1988.

[20] Shigeru Maya, Kai Morino, and Kenji Yamanishi. 2014. Predicting glaucoma progression using multi-task learning with heterogeneous features. In *Big Data (Big Data), 2014 IEEE International Conference on*. IEEE, 261–270.

[21] Hiroshi Murata, Makoto Araie, and Ryo Asaoka. 2014. A new approach to measure visual field progression in glaucoma patients using variational Bayes linear regression. *Investigative Ophthalmology & Visual Science* 55, 12 (2014), 8386–8392.

[22] Hiroshi Murata, Linda M Zangwill, Yuri Fujino, Masato Matsuura, Atsuya Miki, Kazunori Hirasawa, Masaki Tanito, Shiro Mizoue, Kazuhiko Mori, Katsuyoshi Suzuki, et al. 2018. Validating Variational Bayes Linear Regression Method With Multi-Central Datasets. *Investigative Ophthalmology & Visual Science* 59, 5 (2018), 1897–1904.

[23] Alan L Robin, Donald L Budenz, Nathan Congdon, Ravilla D Thulasiraj, and Tarek M Shaarawy. 2009. Glaucoma Volume 1: Medical Diagnosis and Therapy. (2009).

[24] Sebastian Ruder. 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* (2016).

[25] John Shawe-Taylor, Nello Cristianini, et al. 2004. *Kernel methods for pattern analysis*. Cambridge university press.

[26] Hiroki Sugiura, Taichi Kiwaki, Siamak Yousefi, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2018. Estimating Glaucomatous Visual Sensitivity from Retinal Thickness with Pattern-Based Regularization and Visualization. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 783–792.

[27] Shiliang Sun. 2013. A survey of multi-view machine learning. *Neural Computing and Applications* 23, 7-8 (2013), 2031–2038.

[28] Mai Takahashi, Kazuko Omodaka, Kazuichi Maruyama, Takuhiro Yamaguchi, Noriko Himori, Yukihiro Shiga, Morin Ryu, Hiroshi Kunikata, and Toru Nakazawa. 2013. Simulated visual fields produced from macular RNFLT data in patients with glaucoma. *Current eye research* 38, 11 (2013), 1133–1141.

[29] Yukako Taketani, Hiroshi Murata, Yuri Fujino, Chihiro Mayama, and Ryo Asaoka. 2015. How many visual fields are required to precisely predict future test results in glaucoma patients when using different trend analyses? *Investigative Ophthalmology & Visual Science* 56, 6 (2015), 4076–4082.

[30] James Garden Taylor. 1962. The behavioral basis of perception. (1962).

[31] Toshimitsu Uesaka, Kai Morino, Hiroki Sugiura, Taichi Kiwaki, Hiroshi Murata, Ryo Asaoka, and Kenji Yamanishi. 2017. Multi-view Learning over Retinal Thickness and Visual Sensitivity on Glaucomatous Eyes. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2041–2050.

[32] Shen Wang, Lifang He, Bokai Cao, Chun-Ta Lu, Philip S Yu, and Ann B Ragin. 2017. Structural deep brain network mining. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 475–484.

[33] Yanfang Wang and David B Henson. 2013. Diagnostic performance of visual field test using subsets of the 24-2 test pattern for early glaucomatous field loss. *Investigative Ophthalmology & Visual Science* 54, 1 (2013), 756–761.

[34] Tatsuya Yoshida, Aiko Iwase, Hiroyo Hirasawa, Hiroshi Murata, Chihiro Mayama, Makoto Araie, and Ryo Asaoka. 2014. Discriminating between glaucoma and normal eyes using optical coherence tomography and the 'Random Forests' classifier. *PloS one* 9, 8 (2014), e106117.

[35] Siamak Yousefi, Michael H Goldbaum, Madhusudhanan Balasubramanian, Tzyy-Ping Jung, Robert N Weinreb, Felipe A Medeiros, Linda M Zangwill, Jeffrey M Liebmann, Christopher A Girkin, and Christopher Bowd. 2014. Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points. *IEEE Transactions on Biomedical Engineering* 61, 4 (2014), 1143–1154.