






# Deep learning model to predict visual field in central 10° from optical coherence tomography measurement in glaucoma

Yohei Hashimoto <sup>1</sup>, Ryo Asaoka <sup>2,3,4</sup>, Taichi Kiwaki,<sup>5</sup> Hiroki Sugiura,<sup>6</sup> Shotaro Asano,<sup>5</sup> Hiroshi Murata,<sup>1</sup> Yuri Fujino <sup>3,7</sup>, Masato Matsuura <sup>8</sup>, Atsuya Miki,<sup>9</sup> Kazuhiko Mori,<sup>10</sup> Yoko Ikeda,<sup>10</sup> Takashi Kanamoto,<sup>11</sup> Junkichi Yamagami,<sup>12</sup> Kenji Inoue,<sup>13</sup> Masaki Tanito <sup>14</sup>, Kenji Yamanishi<sup>6</sup>

For numbered affiliations see end of article.

## Correspondence to

Ryo Asaoka, Department of Ophthalmology, The University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan; ryoa.sa0120@mac.com and Kenji Yamanishi, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan; yamanishi@mist.i.u-tokyo.ac.jp

Received 25 November 2019

Accepted 15 May 2020

Revised 31 March 2020

## ABSTRACT

**Background/Aim** To train and validate the prediction performance of the deep learning (DL) model to predict visual field (VF) in central 10° from spectral domain optical coherence tomography (SD-OCT).

**Methods** This multicentre, cross-sectional study included paired Humphrey field analyser (HFA) 10-2 VF and SD-OCT measurements from 591 eyes of 347 patients with open-angle glaucoma (OAG) or normal subjects for the training data set. We trained a convolutional neural network (CNN) for predicting VF threshold (TH) sensitivity values from the thickness of the three macular layers: retinal nerve fibre layer, ganglion cell layer+inner plexiform layer and outer segment+retinal pigment epithelium. We implemented pattern-based regularisation on top of CNN to avoid overfitting. Using an external testing data set of 160 eyes of 131 patients with OAG, the prediction performance (absolute error (AE) and  $R^2$  between predicted and actual TH values) was calculated for (1) mean TH in whole VF and (2) each TH of 68 points. For comparison, we trained support vector machine (SVM) and multiple linear regression (MLR).

**Results** AE of whole VF with CNN was  $2.84 \pm 2.98$  (mean  $\pm$  SD) dB, significantly smaller than those with SVM ( $5.65 \pm 5.12$  dB) and MLR ( $6.96 \pm 5.38$  dB) (all,  $p < 0.001$ ). Mean of point-wise mean AE with CNN was  $5.47 \pm 3.05$  dB, significantly smaller than those with SVM ( $7.96 \pm 4.63$  dB) and MLR ( $11.71 \pm 4.15$  dB) (all,  $p < 0.001$ ).  $R^2$  with CNN was 0.74 for the mean TH of whole VF, and  $0.44 \pm 0.24$  for the overall 68 points.

**Conclusion** DL model showed considerably accurate prediction of HFA 10-2 VF from SD-OCT.

CA), because SD-OCT macular scanning area overlaps mainly with this area.<sup>6</sup>

Several studies have reported models to diagnose glaucoma from OCT using multiple linear regression (MLR),<sup>7</sup> support vector machine (SVM),<sup>8</sup> decision tree classifier,<sup>9</sup> random forest<sup>10</sup> or recently developed deep learning (DL).<sup>11</sup> However, few studies have managed to predict point-wise VF from SD-OCT accurately. DL is well known for its distinguished ability to provide accurate diagnosis and prediction.<sup>12</sup> In fact, and according to a recently published study, its usefulness in predicting glaucomatous VF damage (HFA 24-2) is efficient in the whole VF but also in small sectors.<sup>13</sup> However, the proposed prediction was not performed in a point-wise manner.

We have recently reported that a convolutional neural network (CNN) model can predict HFA 10-2 VF accurately even in a point-wise manner from the three macular retinal thicknesses (retinal nerve fibre layer (RNFL), GCL+inner plexiform layer combined (GCL+IPL) and outer segment+retinal pigment epithelium (OS+RPE)) measured with SD-OCT, in conjunction with a novel method called pattern-based regularisation (PBR) method (CNN\_PBR, Figure 1).<sup>14</sup> PBR can effectively regularise the predicted TH using the patterns of VF obtained with a lot of independent (not paired with OCT) VFs through the unsupervised learning. Because of its usefulness in avoiding the overfitting problem, this method is in particular effective when the number of paired VF and OCT data is limited. In our previous report, the CNN\_PBR model achieved a good prediction performance of VF with the mean of the root mean squared error (RMSE) value of 6.16 dB, and the mean of mean absolute error (MAE) value of 4.71 dB (not shown in the paper and with a personal communication with the authors, October 2019), but it was performed with internal cross-validation.<sup>14</sup>

Therefore, the purpose of this study was to validate the prediction performance of the CNN\_PBR model to predict HFA 10-2 VF from SD-OCT measurements in glaucoma by preparing an independent external validation data set.

## MATERIALS AND METHODS

This study was approved by the Research Ethics Committee of the Graduate School of Medicine

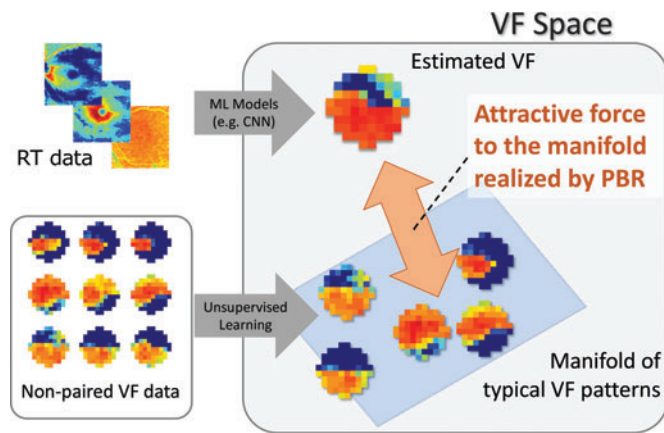
## INTRODUCTION

Glaucoma is the most common cause of irreversible blindness in the world and is characterised by progressive visual field (VF) damage.<sup>1</sup> Glaucomatous VF damage induces structural alterations such as ganglion cell death,<sup>2,3</sup> and it can now be evaluated accurately using spectral domain optical coherence tomography (SD-OCT).<sup>4,5</sup> Because of the structure–function relationship, VF sensitivity may be predicted from the retinal thickness including ganglion cell layer (GCL) in glaucoma.<sup>3,6</sup> This may be particularly feasible with the Humphrey field analyser 10-2 test (HFA 10-2, Carl Zeiss Meditec, Dublin,



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Hashimoto Y, Asaoka R, Kiwaki T, et al. *Br J Ophthalmol* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjophthalmol-2019-315600



**Figure 1** Overview of the CNN model with pattern-based regularisation. The parameters of the CNN were inherited from ResNet.<sup>15</sup> To make the best use of paired and non-paired data, we used the novel method called PBR. This method regularises CNN learning by using characteristic patterns that are obtained from VF data not paired with SD-OCT through the unsupervised learning. More details are shown in our previous report.<sup>14</sup> CNN, convolutional neural network; ML, machine learning; PBR, pattern-based regularisation; RT, retinal thickness; SD-OCT, spectral domain optical coherence tomography; VF, visual field (figure created by the authors).

and Faculty of Medicine at the University of Tokyo, Osaka University Graduate School of Medicine, Kyoto Prefectural University of Medicine, Oike-Ganka Ikeda Clinic, Hiroshima Memorial Hospital, Inouye Eye Hospital, JR Tokyo general hospital and Shimane University Faculty of Medicine. All patients provided informed consent to store their data in the hospital database for the research purposes. This study was performed according to the tenets of the Declaration of Helsinki.

### Training data set

The training paired data set (VF and OCT) comprised 591 eyes of 347 subjects. This included 86 normative eyes of 43 subjects and 505 eyes with open-angle glaucoma (OAG) of 304 patients. The unsupervised learning in PBR was performed using 7715 VF data that were not paired with SD-OCT. All subjects underwent complete ophthalmic examinations, including biomicroscopy, gonioscopy, intraocular pressure measurement, funduscopy, refraction, best-corrected visual acuity measurement and axial length (AL) measurements. Patients were enrolled during the period between April 2013 and August 2016 at either of the University of Tokyo Hospital, Inoue Eye Hospital, JR Tokyo General Hospital or Hiroshima Memorial Hospital.

Inclusion criteria for the OAG group were as follows: (1) presence of typical glaucomatous changes such as a disc rim notch and/or an RNFL defect, (2) gonioscopically wide open angles of grade 3 or 4 based on the Shaffer classification, (3) aged 20–80 years, (4) visual acuity  $\geq 0.5$  LogMAR and (5) refractive error  $< +3.0$  D. Patients with ocular diseases that could affect the results of SD-OCT examinations and VF, such as diabetic retinopathy or age-related macular degeneration, were carefully excluded; however, clinically insignificant senile cataract was included.

Inclusion criteria for the normal group were as follows: (1) no abnormal findings except for clinically insignificant senile cataract, (2) no history of ocular diseases that could affect the results of SD-OCT and VF examinations, (3) aged 20–80 years, (4) normal

VF test results according to the Anderson-Patella criteria<sup>1</sup> and (5) refractive error  $< +3.0$  D.

### VF testing

VF testing was performed using HFA 10-2 test with the SITA standard strategy within 6 months from the measurement of SD-OCT. Near refractive correction was used as necessary. All participants had previous experience in VF examinations. We excluded unreliable VFs defined as fixation losses greater than 20%, false-positive responses greater than 15%.

### SD-OCT measurement

RS 3000 (Nidek Co, Aichi, Japan) and OA-2000 (TOMEY, Aichi, Japan) were used to obtain SD-OCT and AL measurement data, respectively. All SD-OCT measurements were performed following pupil dilation with 1% tropicamide. We excluded data with apparent eye movement and involuntary blinking or saccade during the measurement, and imaging data with quality factor  $< 7$ , as recommended by the manufacturer. A square imaging area ( $9.0 \times 9.0$  mm) was centred on the fovea. Using the software supplied by the manufacturer, the macular thicknesses of the following three layers were exported as pixel images ( $512 \times 128$ ): (1) RNFL, (2) GCL + IPL and (3) OS + RPE. We included OS + RPE, because our previous findings suggested that it was beneficial to include this thickness to improve the structure–function relationship.<sup>16</sup> To inherit the parameters from ResNet,<sup>15</sup> which is one of the most popular pretrained models for image classification, the pixel images of each layer were resized to  $224 \times 224$  pixels with a bicubic interpolation<sup>17</sup> over  $4 \times 4$  neighbourhood and resampling using pixel area relation.<sup>18</sup> We further performed data augmentation via vertical flip.<sup>14</sup> We finally applied the DL method to the images.

### Testing data set

The testing data set was obtained after the completion of the training data set, during the period between September 2016 and October 2017 at Oike-Ganka Ikeda Clinic or Shimane University Hospital. The testing data set comprised 160 OAG eyes of 131 subjects, which was independent of the training data set. Inclusion and exclusion criteria as well as VF and SD-OCT measurements were identical to those of the OAG group in the training data set.

### Prediction models

#### CNN\_PBR

Using the training data set, we trained CNN\_PBR models to predict quantitatively 68 threshold (TH) values of HFA 10-2 from the three retinal thicknesses of RNFL, GCL + IPL and OS + RPE. All CNN parameters, except for the last layer which was randomly initialised, were inherited from ResNet. The ResNet is an enhanced DL algorithm based on a CNN.<sup>15</sup> Deep training networks enable the extraction of more complex and detailed features from images; however, a deep CNN often involves low diagnostic performance due to the vanishing of the gradient problem and the gradient divergence problem. Thus, the transmission of information from shallow layers to deep layers can be hampered. To facilitate a deeper and larger network, ResNet skips one or more layers, and the features are propagated to succeeding layers. In our previous study, we used VGG16,<sup>14</sup> a popular DL model, and in the present study, we have used ResNet. We used stochastic gradient descent with momentum for training and weight decay. To make the best use of paired

and non-paired data, we used the novel PBR method. This method regularises CNN learning by using characteristic patterns obtained from VF data that are not paired with SD-OCT through the unsupervised learning (Figure 1). The OCT-measured thicknesses used in this model was in  $3 \times 224 \times 224$  scholars. More details are shown in our previous study.<sup>14</sup>

### SVM and MLR

Furthermore, we trained the two other classic machine learning models: SVM<sup>19</sup> and MLR, for comparison purposes. In these models, we simply applied the thickness of RNFL, GCL+IPL, and OS/RPE at each location. In other words, the explanatory variables ( $65 \ 536 \times 3$ ) included the thickness of all 65 536 ( $512 \times 128$  pixels) points in the three layers obtained using OCT, and the response variables involved the respective TH values.

### Validation with the external testing data set

#### Whole-field analysis

We validated the prediction performances of the three trained models (CNN\_PBR, SVM and MLR) using the external testing data set. We calculated the absolute error (AE;  $AE_{\text{whole}}$ ) between the mean of the 68 predicted TH values and the mean of 68 actual ones. The  $AE_{\text{whole}}$  was compared among the three methods using the linear mixed model, whereby patients and eyes were the random effects. In addition,  $R^2$  values were also calculated as an additional measure of prediction accuracy ( $R^2_{\text{whole}}$ ).

#### Point-wise analysis

The AE between the predicted and the actual TH values for all the 68 points was calculated ( $AE_{\text{PW}}$ ). Furthermore, the mean of 68  $AE_{\text{PW}}$  ( $MAE_{\text{PW}}$ ) was calculated. In other words,  $MAE_{\text{PW}}$  was defined as follows:

$$MAE_{\text{PW}} = \frac{\sum_{i=1}^{68} |\text{predicted visual sensitivity of the } i\text{th point} - \text{actual visual sensitivity of the } i\text{th point}|}{68}$$

where  $i$  is the number of the 68 predicted test points.

$MAE$  values were compared among the three methods using the linear mixed model as mentioned earlier. In addition,  $R^2$  values were calculated ( $R^2_{\text{PW}}$ ).

#### Sector-wise analysis

We sectorised HFA 10-2 VF using the cluster map that was proposed by de Moraes (Figure 2).<sup>20</sup> The AE ( $AE_{\text{sector}}$ ) between the mean of the predicted and the actual TH values was calculated for each sector. The  $AE_{\text{sector}}$  was compared among the three methods using the linear mixed model as mentioned earlier. In addition,  $R^2$  values between the predicted and the actual TH values were calculated ( $R^2_{\text{sector}}$ ).

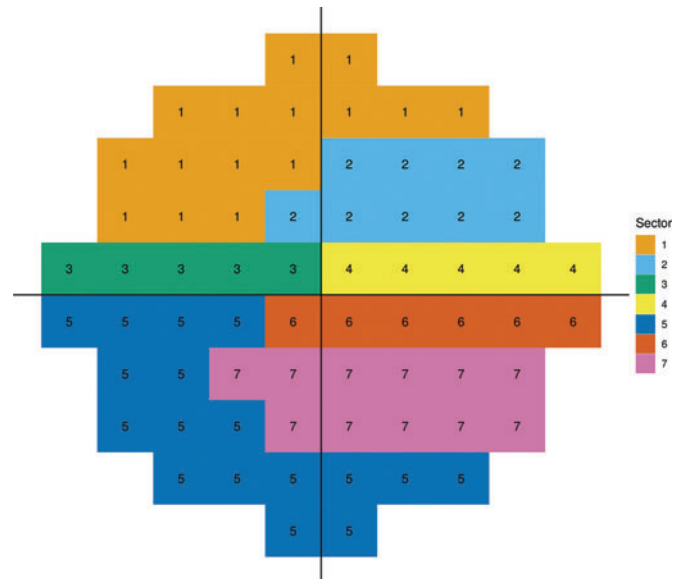
#### Location-wise analysis

Finally, the  $AE_{\text{PW}}$  of each location for all eyes in the testing data set was calculated for the CNN\_PBR.

Statistical analysis was performed with Python (version 3.7.6; Python Software Foundation) and the statistical programming language R language (version 3.6.0; R Foundation for Statistical Computing, Vienna, Austria).

## RESULTS

Table 1 shows the background of the training and test data sets.



**Figure 2** Cluster map for 10-2 visual field. HFA 10-2 VF was clustered into seven sectors, which was proposed by de Moraes *et al*<sup>20</sup> (left eyes). Right eyes were mirror imaged. HFA, Humphrey field analyser.

Figure 3 shows actual TH at each VF point. In general, visual sensitivity was higher in the infero-temporal area.

In the whole-field analysis,  $AE_{\text{whole}}$  with CNN\_PBR was  $2.84 \pm 2.98$  (mean  $\pm$  SD) dB, which was significantly smaller than that with SVM ( $5.65 \pm 5.12$  dB) and MLR ( $6.96 \pm 5.38$  dB) (both,  $p < 0.001$ , Figure 4).  $R^2_{\text{whole}}$  value with CNN\_PBR, SVM and MLR was 0.74, 0.36 and 0.26, respectively.

In the point-wise analysis,  $MAE_{\text{PW}}$  with CNN\_PBR, SVM and MLR was  $5.47 \pm 3.05$  dB,  $7.96 \pm 4.63$  dB and  $11.71 \pm 4.15$  dB, respectively.  $MAE_{\text{PW}}$  with CNN\_PBR was significantly smaller than those with SVM and MLR (both,  $p < 0.001$ , Figure 5).  $R^2_{\text{PW}}$  with CNN\_PBR, SVM and MLR was  $0.44 \pm 0.24$ ,  $0.31 \pm 0.24$  and  $0.16 \pm 0.18$ , respectively.

In the sector-wise analysis, the  $AE_{\text{sector}}$  with CNN\_PBR was  $4.31 \pm 4.14$  dB,  $4.88 \pm 4.96$  dB,  $4.26 \pm 4.78$  dB,  $4.91 \pm 4.58$  dB,  $3.36 \pm 3.31$  dB,  $4.10 \pm 4.12$  dB and  $3.96 \pm 4.18$  dB in Sector 1,

**Table 1** Characteristics of the subjects and eyes in the training and testing data sets

	Training (paired data)	Testing
Subjects (n)	347*	131**
Age, years	55.1 $\pm$ 14.8	65.8 $\pm$ 12.2
Sex, female	191 (55%)	70 (53%)
Eyes (n)	591	160
Laterality, left	289 (49%)	89 (56%)
Axial length, mm	25.4 $\pm$ 2.7	24.6 $\pm$ 1.7
Mean threshold value, dB	24.1 $\pm$ 9.3	21.9 $\pm$ 7.8
Mean deviation, dB	-8.8 $\pm$ 9.4	-10.4 $\pm$ 8.1
RNFL, $\mu$ m	30.5 $\pm$ 9.0	26.9 $\pm$ 8.4
GCL + IPL, $\mu$ m	39.7 $\pm$ 9.0	38.7 $\pm$ 7.5
OS/RPE, $\mu$ m	67.1 $\pm$ 3.8	65.6 $\pm$ 5.1

Data are presented as n (%) or mean  $\pm$  SD.

\*Consisted of 43 normative subjects and 304 patients with open glaucoma.

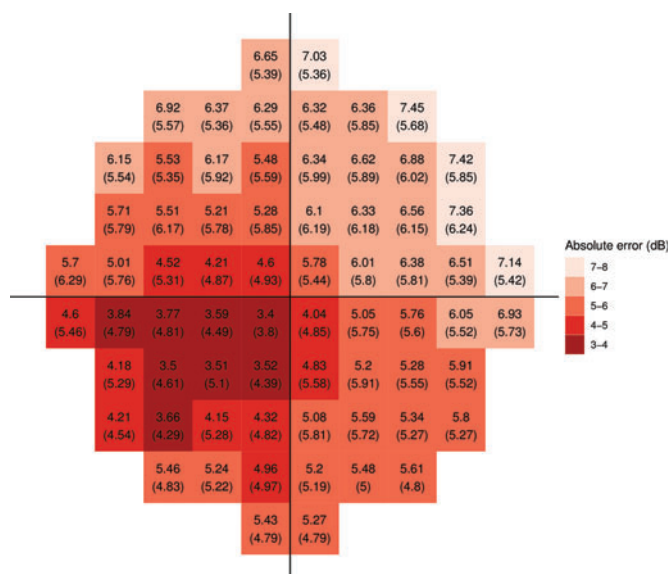
\*\*All the subjects were patients with open glaucoma.

GCL, ganglion cell layer; IPL, inner plexiform layer; OS, outer segment; RNFL, retinal nerve fibre layer; RPE, retinal pigment epithelium.



				14.37 (11.79)	15.34 (11.93)				
		15.85 (11.88)	16.18 (11.98)	15.84 (11.92)	15.94 (12.08)	16.12 (12.42)	16.07 (12.47)		
	17.31 (12.23)	19.33 (12.13)	19.42 (12.53)	18.29 (12.75)	17.67 (12.68)	17.48 (12.69)	16.63 (12.71)	16.04 (12.66)	
	21.41 (11.93)	23.58 (11.41)	23.44 (11.77)	23.49 (11.39)	21.32 (12.41)	18.8 (13.09)	18.36 (12.83)	17.08 (13.04)	
22.76 (10.8)	25.38 (9.95)	26.63 (9.48)	28.01 (8.67)	27.43 (8.45)	24.86 (10.67)	22.23 (11.99)	19.98 (12.64)	19.01 (12.33)	17.55 (12.31)
25.34 (9.03)	27.52 (7.85)	28.55 (7.94)	29.39 (7.54)	29.12 (6.87)	28.27 (7.81)	25.85 (10.61)	23.43 (11.69)	21.43 (11.98)	19.31 (12.73)
	27.2 (8.37)	28.26 (7.85)	28.48 (8.42)	28.48 (7.64)	26.64 (9.98)	24.99 (11.87)	23.19 (12.02)	20.17 (12.88)	
	25.18 (9.41)	26.26 (9.33)	25.77 (10.49)	25.52 (10.02)	23.96 (11.58)	23.45 (12.35)	21.49 (12.68)	19.06 (12.79)	
		22.43 (11.72)	22.54 (11.82)	21.81 (12.29)	21.76 (12)	20.26 (12.72)	19.16 (12.62)		
				18.88 (12.58)	18.86 (12.38)				

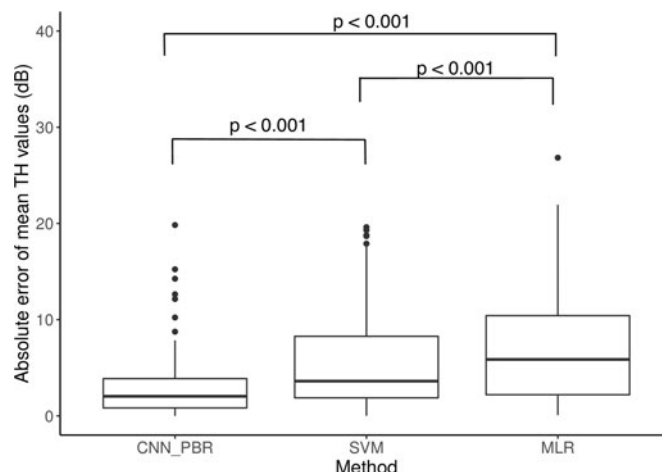
**Figure 3** Actual threshold values. Mean (upper row) and SD (lower row) values at each test point are illustrated (left eye). Right eyes were mirror imaged.



**Figure 4** Absolute error between the mean of the whole 68 predicted threshold values and the mean of whole 68 actual ones. The CNN model showed smaller absolute prediction error than SVM and MLR (both,  $p < 0.001$ ; linear mixed model). CNN, convolutional neural network; MLR, multiple linear regression; PBR, pattern-based regularisation; SVM, support vector machine.

Sector 2, Sector 3, Sector 4, Sector 5, Sector 6 and Sector 7, respectively (Table 2).

CNN\_PBR showed significantly smaller error than SVM (range of mean  $AE_{\text{sector}}$ : 5.47–8.66 dB) and MLR (range of mean  $AE_{\text{sector}}$ : 8.19–12.00 dB) (all,  $p < 0.001$  except for the following; CNN\_PBR vs SVM in Sector 3,  $p = 0.035$ ).  $R^2_{\text{sector}}$  for CNN\_PBR was 0.65, 0.60, 0.49, 0.60, 0.70, 0.58 and 0.66 in Sector 1, Sector 2, Sector 3, Sector 4, Sector 5, Sector 6 and



**Figure 5** Mean of absolute error between the predicted 68 threshold values and the actual 68 ones. The CNN model showed smaller absolute prediction error than SVM and MLR (both,  $p < 0.001$ ; linear mixed model). CNN, convolutional neural network; MLR, multiple linear regression; PBR, pattern-based regularisation; SVM, support vector machine

Sector 7, respectively. Moreover, smaller  $R^2_{\text{sector}}$  values were obtained with SVM and MLR (Table 2). Sector 5 of the CNN\_PBR model exhibited the smallest  $AE_{\text{sector}}$  and the largest  $R^2_{\text{sector}}$  among the seven sectors, followed by Sector 7. In terms of location-wise analysis,  $AE_{\text{PW}}$  were smaller in the infero-temporal region than in other regions (Figure 6).

## DISCUSSION

In the current study, we trained a CNN model with PBR to predict HFA 10-2 VF from the retinal thicknesses, measured by SD-OCT, and validated the prediction performance, using an external data set. Consequently, the whole-field analysis resulted in an MAE value of 2.84 dB, with  $R^2$  value of 0.74. In the point-wise analysis, the mean of all MAEs was 5.47 dB and the mean of  $R^2$  value was 0.44. Moreover, in the sector-wise analysis, the mean of AE ranged from 3.36 dB to 4.91 dB depending on the sectors, with the mean of  $R^2$  values ranging between 0.49 and 0.70. In all analyses, the CNN model showed significantly higher prediction performance compared with the SVM and MLR. Sector-wise and location-wise analyses revealed that prediction accuracy was better in the infero-temporal area compared to other areas.

We have recently reported that the application of CNN to OCT images is advantageous for discriminating between glaucoma and normal eyes.<sup>11</sup> In contrast, the challenge in this study was to quantitatively predict TH values in a point-wise manner. A similar approach was implemented in a recent study by Christopher *et al.*<sup>13</sup> They trained a DL model to predict sectorial VF (HFA 24-2 VF) using circumpapillary RNFL (cpRNFL) captured by SD-OCT.<sup>13</sup> The predicted mean deviation had a mean MAE value of 2.5 dB with  $R^2$  of 0.70 in their study, which is similar to our results (mean of  $AE_{\text{whole}}$ : 2.84 dB and  $R^2_{\text{whole}}$ : 0.74). However, their prediction was not performed in the point-wise manner. Point-wise prediction is especially important when considering the future application of DL models to real-world clinical settings. In our study, the point-wise prediction performance was  $MAE_{\text{PW}}$  of  $5.47 \pm 3.05$  dB. Considering that the reported test–retest variabilities are between 1 and 2 dB (central area) and 4–6 dB (at 27°),<sup>21–22</sup> our results can be regarded as fairly good.

**Table 2** Accuracy performance of the three models (CNN\_PBR, SVM and MLR) for predicting sectoral mean TH values

		Method proposed by de Moraes <i>et al</i>						
		Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6	Sector 7
Absolute error (dB)	CNN_PBR	4.31±4.14	4.88±4.96	4.26±4.78	4.91±4.58	3.36±3.31	4.10±4.12	3.96±4.18
	SVM	7.64±5.59	8.66±5.48	5.47±7.23	7.67±6.17	5.91±6.44	6.07±6.50	6.74±7.57
	MLR	10.82±8.33	12.00±7.76	8.19±6.65	9.97±7.22	8.61±5.20	8.52±5.49	10.78±6.34
R <sup>2</sup>	CNN_PBR	0.65 (0.55–0.73)	0.60 (0.50–0.69)	0.49 (0.38–0.60)	0.60 (0.49–0.69)	0.70 (0.61–0.77)	0.58 (0.47–0.67)	0.66 (0.57–0.74)
	SVM	0.19 (0.09–0.31)	0.17 (0.07–0.28)	0.12 (0.04–0.23)	0.27 (0.16–0.39)	0.20 (0.10–0.31)	0.25 (0.14–0.37)	0.17 (0.08–0.29)
	MLR	0.04 (0.00–0.11)	0.05 (0.01–0.14)	0.05 (0.01–0.14)	0.15 (0.06–0.26)	0.13 (0.05–0.24)	0.17 (0.07–0.28)	0.08 (0.02–0.18)

Data of absolute error are presented as mean±SD. R<sup>2</sup> data are presented as estimates (95% CI).

CNN, convolutional neural network; MLR, multiple linear regression; PBR, pattern-based regularisation; SVM, support vector machine; TH, threshold.

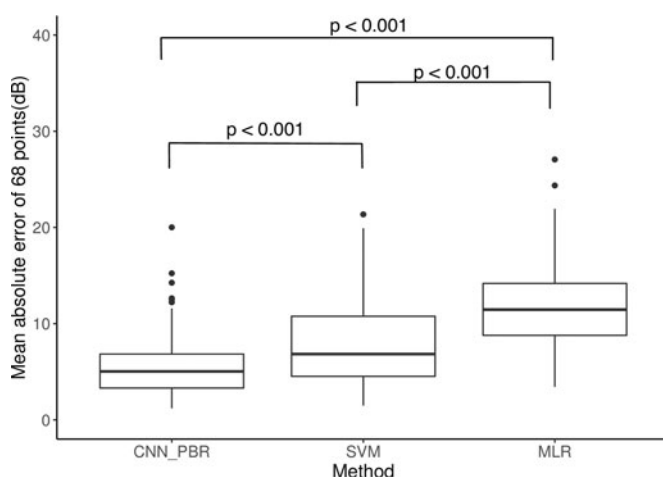
Moreover, it has been suggested that retinal thickness reduces in myopic eyes; thus, we excluded those with AL≥26.5 mm.<sup>23</sup> Nonetheless, only marginal difference was observed in the prediction error; MAE<sub>PW</sub> was 5.67±3.10 dB (not shown in the ‘Results’ section). This suggests that the usefulness of CNN\_PBR model was not largely influenced by the myopia status.

Christopher *et al* have also predicted sectoral pattern deviation of HFA 24–2 VF, with R<sup>2</sup> of 0.67 (superior nasal), 0.60 (inferior nasal), 0.35 (superior), 0.26 (inferior), 0.15 (central) and 0.12 (temporal). In contrast, higher R<sup>2</sup><sub>sector</sub> values (0.49–0.70) were observed in the current study. It may be clinically important that the model developed by Christopher *et al* had relatively poor prediction performance in the central sector (R<sup>2</sup> of 0.15), compared to when almost identical sectorisation was used with our model (R<sup>2</sup> of 0.71, not shown in the ‘Results’ section). This difference may be attributed to the differences in the DL model itself and also to the difference between the input variables: they used cpRNFL thickness, whereas we used macular thickness. The latter may be a more effective approach, because the macular thickness of a certain point directly reflects the thickness of ganglion cells on the corresponding VF point. On the other hand, cpRNFL thickness merely reflects the collection of the axons arising from ganglion cells.<sup>24</sup> Considering that the underlying disease process of glaucoma is the loss of

retinal ganglion cell,<sup>25</sup> prediction using the corresponding retinal thickness is more straightforward. Second, the measurement of cpRNFL thickness is more variable than that of macular thickness.<sup>24</sup> This is because cpRNFL thickness is influenced by the variation of optic nerve head in shape and size, refractive error and AL. For example, it has been shown that the refractive error was significantly associated with cpRNFL thickness but not with macular thickness in high myopic eyes.<sup>26</sup> This variation of cpRNFL would be disadvantageous for predicting visual sensitivities.

Sector-wise (Table 2) and point-wise (Figure 6) analyses revealed that the prediction accuracy was relatively accurate in the infero-temporal area. Weber *et al* reported that there is a preserved ‘central isle’ of the VF in patients with advanced glaucoma,<sup>27</sup> which corresponds to this infero-temporal area. Previous studies have shown the limited usefulness of OCT in advanced glaucoma compared to in early-to-moderate glaucoma.<sup>28–29</sup> The smaller absolute prediction error values in this area may be due to the relatively preserved visual sensitivity in this region. Moreover, the smaller variation of visual sensitivity in this area may be another possible reason.

The relation between retinal function and structure, such as RNFL thickness, has been reported to be non-linear.<sup>30–31</sup> For instance, Hood *et al* have proposed that the deterioration of visual sensitivity is not observed until RNFL thickness reaches a critical thickness (floor effect).<sup>28</sup> The potential advantage of machine learning methods, in particular with DL, is that the assumption such as linear relationship is not needed. Machine learning models predict VF using characteristic patterns of the glaucomatous damage to retinal layers.<sup>32–33</sup> Indeed, previous studies have revealed the usefulness of machine learning in diagnosing glaucoma, using the MLR,<sup>7</sup> SVMs,<sup>8</sup> decision tree classifiers<sup>9</sup> and random forests.<sup>10</sup> However, very few methods have proposed a model to predict quantitative VF sensitivities from OCT. This may stem from the fact that the relationship between GCC thickness and VF sensitivity is usually not straightforward. The current results suggested that the machine learning approach is also useful in predicting VF sensitivity values. Another possible advantage of this machine learning approach is that the prediction is not influenced by the displacement of GCC near fovea. This is in particular true because several adjustment methods have been proposed, and there has been no study to compare the accuracy of these models.<sup>34</sup> For instance, CNNs use all thicknesses by considering the pattern of GCC thinning. In general, DL models can automatically identify, extract and learn features in highly structured data with multiple data sets, in contrast to other machine learning



**Figure 6** Point-wise absolute prediction error with the convolutional neural network model. Mean (upper row) and SD (lower row) at each location is illustrated. Smaller error is shown with dark red, and larger error is shown with light red. The infero-temporal region showed better accuracy performance than other regions. Right eyes were mirror imaged.

techniques.<sup>35</sup> Therefore, DL in our study facilitated an improved accurate prediction compared with SVMs.

Previous studies have suggested that central VF cannot be accurately assessed unless VF is monitored using a specified program for the central 10°, such as the HFA 10-2 VF.<sup>4,5</sup> One of the most challenging obstacles in clinical settings is that in addition to the central 24° VF testing, undergoing detailed measurements of the central 10° VF is an additional physical and economic burden to patients. In fact, it has been reported that HFA 24-2 or its equivalents with a sufficient frequency is a relatively heavy burden for general clinical facilities to carry out.<sup>36</sup> This implies that predicting the central HFA 10-2 VF from OCT-measured retinal thicknesses would be a beneficial approach.

There are several limitations in the current study. First, we predicted TH values only in the central 10°, not in 24° or 30°, because macular OCT scan is unable to cover such a wide area. Thus, this approach cannot be directly used for the prediction of HFA 24-2 VF. However, it would be advantageous to investigate whether the current approach yields similar results even when predicting HFA 24-2 VF.

Second, some outlier cases were observed in the CNN\_PBR (Figure 4). Among nine variables of the right/left eye, sex, age, AL, mean TH value, mean deviation, fixation loss, false negative and false positive, only higher false-negative rates and higher false-positive rates were significantly associated with the outliers (data not shown in the 'Results' section). This suggests that the outliers were attributable to the less accurate VF measurement rather than to the prediction accuracy of CNN\_PBR.

Third, a possible caveat of the current results is that the prediction with the currently proposed method is not readily applicable in the clinical setting. Therefore, it would be clinically beneficial to develop software/support tools to predict VF, as introduced in this study. In particular, only standard data are needed to apply the current methodology in the clinical setting; OCT measured three retinal thicknesses of RNFL, GCL+IPL and OS+RPE. The prediction will be calculated using the thickness of automatically extracted layers following the development of a suitable medical software/support tool.

In conclusion, we trained and validated a DL model (CNN\_PBR) to predict 10-2 VF from retinal thicknesses measured with SD-OCT. The CNN model exhibited better accuracy performance compared with classic machine learning algorithms and achieved fairly good performance even in the sectorial and point-wise manners.

#### Author affiliations

- <sup>1</sup>Department of Ophthalmology, The University of Tokyo, Tokyo, Japan
- <sup>2</sup>Department of Ophthalmology, University of Tokyo Graduate School of Medicine, Tokyo, Japan
- <sup>3</sup>Seirei Hamamatsu General Hospital, Shizuoka, Hamamatsu, Japan
- <sup>4</sup>Seirei Christopher University, Shizuoka, Hamamatsu, Japan
- <sup>5</sup>Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo, Japan
- <sup>6</sup>Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
- <sup>7</sup>Ophthalmology, The University of Tokyo Hospital, Bunkyo-ku, Japan
- <sup>8</sup>University of Tokyo, Tokyo, Japan
- <sup>9</sup>Ophthalmology, Osaka Daigaku Daigakuin Igakukei Kenkyuka Igakubu, Suita, Osaka, Japan
- <sup>10</sup>Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan
- <sup>11</sup>Ophthalmology, Hiroshima University, Higashihiroshima, Hiroshima, Japan
- <sup>12</sup>Ophthalmology, JR Tokyo General Hospital, Tokyo, Japan
- <sup>13</sup>Ophthalmology, Inouye Eye Hospital, Tokyo, Japan
- <sup>14</sup>Ophthalmology, Shimane University Faculty of Medicine, Izumo, Japan

#### Acknowledgements

Grant information.

**Contributors** YH and RA had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: YH, RA, TK, HS and KY. Acquisition, analysis and interpretation of data: YH, RA, TK, SA, HM, YF, MM, AM, KM, YI, TK, JY, KI, MT and KY. Drafting of the manuscript: YH and RA. Critical revision of the manuscript for important intellectual content: YH and RA. Statistical analysis: YH and RA.

**Funding** RA: The Ministry of Education, Culture, Sports, Science and Technology of Japan (grant numbers 18KK0253, 19H01114 and 17K11418); the Daiichi Sankyo Foundation of Life Science, Tokyo, Japan; Suzuken Memorial Foundation, Tokyo, Japan; The Translational Research Program; Strategic Promotion for practical application of Innovative medical Technology (TR-SPRINT) from Japan Agency for Medical Research and Development (AMED); JST-AIP JPMJCR19U4. HM: The Ministry of Education, Culture, Sports, Science and Technology of Japan (grant number 25861618). YF: The Ministry of Education, Culture, Sports, Science and Technology of Japan (grant number 20768254). MM: The Ministry of Education, Culture, Sports, Science and Technology of Japan (grant number 00768351). KY: JST-AIP JPMJCR19U4 and The Ministry of Education, Culture, Sports, Science and Technology of Japan (grant number 19H01114).

**Competing interests** None declared.

**Data sharing statement** Data are available upon reasonable request.

**Provenance and peer review** Not commissioned; externally peer reviewed.

#### ORCID iDs

Yohei Hashimoto <http://orcid.org/0000-0002-7944-6923>  
 Ryo Asaoka <http://orcid.org/0000-0001-7182-1912>  
 Yuri Fujino <http://orcid.org/0000-0001-6082-0738>  
 Masato Matsuura <http://orcid.org/0000-0002-6120-1100>  
 Masaki Tanito <http://orcid.org/0000-0001-6512-7203>

#### REFERENCES

- 1 Jonas JB, Aung T, Bourne RR, *et al.* Glaucoma. *The Lancet* 2017;390:2183–93.
- 2 Kerrigan-Baumrind LA, Quigley HA, Pease ME, *et al.* Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophthalmol Vis Sci* 2000;41:741–8.
- 3 Harwerth RS, Carter-Dawson L, Smith EL, *et al.* Neural losses correlated with visual losses in clinical perimetry. *Investig Ophthalmol Vis Sci* 2004;45:3152.
- 4 Tan O, Chopra V, Lu AT-H, *et al.* Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology* 2009;116:2305–2314.e2.
- 5 Rao HL, Babu JG, Addepalli UK, *et al.* Retinal nerve fiber layer and macular inner retina measurements by spectral domain optical coherence tomograph in Indian eyes with early glaucoma. *Eye* 2012;26:133–9.
- 6 Hood DC. Improving our understanding, and detection, of glaucomatous damage: an approach based upon optical coherence tomography (OCT). *Prog Retin Eye Res* 2017;57:46–75.
- 7 Mwanza J-C, Warren JL, Budenz DL. Combining spectral domain optical coherence tomography structural parameters for the diagnosis of glaucoma with early visual field loss. *Investig Ophthalmol Vis Sci* 2013;54:8393.
- 8 Burgansky-Eliash Z, Wollstein G, Chu T, *et al.* Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Investig Ophthalmol Vis Sci* 2005;46:4147.
- 9 Baskaran M, Ong E-L, Li J-L, *et al.* Classification algorithms enhance the discrimination of glaucoma from normal eyes using high-definition optical coherence tomography. *Investig Ophthalmol Vis Sci* 2012;53:2314.
- 10 Asaoka R, Hirasawa K, Iwase A, *et al.* Validating the usefulness of the "random forests" classifier to diagnose early glaucoma with optical coherence tomography. *Am J Ophthalmol* 2017;174:95–103.
- 11 Asaoka R, Murata H, Hirasawa K, *et al.* Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol* 2019;198:136–45.
- 12 Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;18:1527–54.
- 13 Christopher M, Bowd C, Belghith A, *et al.* Deep learning approaches predict glaucomatous visual field damage from optical coherence tomography optic nerve head enface images and retinal nerve fiber layer thickness maps. *Ophthalmology* 2019; S0161642019321037.
- 14 Sugiura H, Kiwaki T, Yousefi S, *et al.* 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 - KDD'18. London, UK: ACM Press 2018. 783–92.
- 15 He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE 2016. 770–8.

- 16 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;8:10450.
- 17 Das V. A novel diagnostic information based framework for super-resolution of retinal fundus images. *Comput Med Imaging Graph* 2019;12.
- 18 Thévenaz P, Blu T, Unser M. *Image interpolation and resampling. Handbook of medical imaging, processing and analysis*. 2000;1:393–3420.
- 19 Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge; New York: Cambridge University Press 2000.
- 20 de Moraes CG, Song C, Liebmann JM, *et al*. Defining 10-2 visual field progression criteria. *Ophthalmol* 2014;121:741–9.
- 21 Parrish RK. Static and kinetic visual field testing: reproducibility in normal volunteers. *Arch Ophthalmol* 1984;102:1497–502.
- 22 Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the central visual field. *Arch Ophthalmol* 1987;105:1544–9.
- 23 Ohno-Matsui K, Lai TYY, Lai -C-C, *et al*. Updates of pathologic myopia. *Prog Retin Eye Res* 2016;52:156–87.
- 24 Sung KR, Sun JH, Na JH, *et al*. Progression detection capability of macular thickness in advanced glaucomatous eyes. *Ophthalmol* 2012;119:308–13.
- 25 Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol* 1989;107:453–64.
- 26 Shoji T, Nagaoka Y, Sato H, *et al*. Impact of high myopia on the performance of SD-OCT parameters to detect glaucoma. *Graefes Arch Clin Exp Ophthalmol* 2012;50:1843–9.
- 27 Weber J, Schultze T, Ulrich H. The visual field in advanced glaucoma. *Int Ophthalmol* 1989;13:47–50.
- 28 Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. *Prog Retin Eye Res* 2007;26:688–710.
- 29 Swanson WH, Felius J, Pan PF. Perimetric defects and ganglion cell damage: interpreting linear relations using a two-stage neural model. *Investig Ophthalmol Vis Sci* 2004;45:466.
- 30 Altangerel U, Spaeth GL, Rhee DJ. Visual function, disability, and psychological impact of glaucoma. *Curr Opin Ophthalmol* 2003;14:100–5.
- 31 Leite MT, Zangwill LM, Weinreb RN, *et al*. Structure-function relationships using the cirrus spectral domain optical coherence tomograph and standard automated perimetry. *J Glaucoma* 2012;21:49–54.
- 32 Shields M. *Textbook of glaucoma*. MD, USA: William & Wilkins. 1997
- 33 Zimmerman TJ, Kooner KS. *Clinical pathways in glaucoma*. New York: Thieme, 2001.
- 34 Drasdo N, Milligan CL, Katholi CR, *et al*. The length of Henle fibers in the human retina and a model of ganglion receptive field density in the visual field. *Vision Res* 2007;47:2901–11.
- 35 Hinton G. Deep belief networks. *Scholarpedia* 2009;4:5947.
- 36 Crabb DP, Russell RA, Malik R, *et al*. Frequency of visual field testing when monitoring patients newly diagnosed with glaucoma: mixed methods and modelling. *Health Serv Deliv Res* 2014;2:1–102.