Title: Validating the usefulness of the 'Random Forests' classifier to diagnose early glaucoma with optical coherence tomography.

Ryo Asaoka ¹, Kazunori Hirasawa ², Aiko Iwase ³, Yuri Fujino ¹, Hiroshi Murata ¹, Nobuyuki Shoji ², Makoto Araie ⁴

1: Department of Ophthalmology, The University of Tokyo,

2: Orthoptics and Visual Science, Department of Rehabilitation, School of Allied Health Sciences, Kitasato University, Kanagawa, Japan.

3: Tajimi Iwase Eye Clinic, Tajimi, Japan

4: Kanto Central Hospital of the Mutual Aid Association of Public School Teachers, Tokyo, Japan

Key words: Glaucoma, Optical Coherence Tomography, Random Forests method, Area Under the Receiver Operating Characteristic Curve

Short title: Random Forests to diagnose glaucoma

Correspondence and reprint requests to: Ryo Asaoka, MD, PhD,

Department of Ophthalmology, University of Tokyo Graduate School of Medicine

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan.

Phone: +81-3-3815-5411, Fax: +81-3-3817-0798, email:

rasaoka-tky@umin.ac.jp

Introduction

In glaucoma, measureable structural alterations, such as structural changes at the optic nerve head¹⁻³ and damage to the retinal nerve fiber layer (RNFL) around the optic disc⁴⁻⁶, can precede detectable VF damage. Indeed histological studies have shown that a large proportion of retinal ganglion cells – as much as 20% – needs to be lost before statistically significant abnormalities are apparent on standard automated perimetry (SAP).⁷⁻¹⁰ Glaucomatous structural change can now be evaluated with spectral domain optical coherence tomography (SD-OCT) by measuring the thickness of the macular ganglion cell complex (macular ganglion cell layerinner plexiform layer combined: GCC) and the circumpapillary RNFL (cpRNFL), both of which are reported to be damaged early on in the disease process.¹¹⁻¹⁸ The damage to these layers usually occurs concurrently and in characteristic patterns,^{19,20} and as a result, it is beneficial to analyze the multitude of structural measurements of SD-OCT in combination for the purpose of accurately diagnosing glaucoma.²¹⁻²³ Several mathematical models have been proposed for this purpose; Mwanza et al. analyzed multiple SD-OCT parameters using a logistic regression model²¹, Burgansky-Eliash et al. adopted a support vector machine classifier²² and Baskaran et al. have applied a decision tree classifier.²³

The decision tree classifier is a tree-like representation of a finite set of if-then-else rules²⁴. Decision trees are effective at identifying non-linear patterns but can suffer from the problem of ‘over-fitting’, which leads to a decrease in diagnostic accuracy when applied to independent datasets.²⁵ The Random Forests method^{26,27} was proposed by Breiman in 2001 to overcome this problem by creating numerous ‘bootstrapped’ subsets from the original dataset and composing a decision tree for each of the developed subsets. Furthermore, unlike standard regression approaches, predictor variables do not ‘compete’ against correlated predictors due to the node and branch decision tree structure.²⁸ Thus, the method can cope with inter-correlated explanatory variables and indeed the usefulness of this approach has been recognized in many research fields,²⁹⁻³¹ in particular when there are interactions between different explanatory variable.^{28,32,33} These properties make a Random Forests classifier appealing for the diagnosis of glaucoma using OCT measurements; first, OCT measured thicknesses of cpRNFL and GCC are closely correlated and second, there is considerable overlap between the distributions of OCT measurements in normal and glaucoma eyes^{34,35}, which makes a simple comparison of a measured OCT parameter against a normative database problematic.

We have recently demonstrated the usefulness of a RF classifier to diagnose glaucoma using OCT measurements;^{36,37} however, a limitation of our prior studies was the lack of external validation in an independent dataset, instead leave-one-out cross-validation was used to investigate diagnostic performance. Further, the study population consisted of wide range of glaucoma patients with early to advanced-stage disease, but the merit of using OCT to diagnose glaucoma may lie in early stage detection. Glaucoma is the second leading cause of blindness worldwide,³⁸ and early detection is essential. The development of glaucoma is a gradual process and indeed, there is a wide variation in OCT parameters between from early to advanced stage. However, at the clinical settings, difficulty lies in diagnosing glaucoma in early stage. Hence we collected independent training and testing datasets consisted of only early glaucomatous eyes, as well as normal eyes, for the purpose of investigating the usefulness of the Random Forests method to diagnose glaucoma in early stage.

Materials and Methods

This Cross-sectional study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at the University of Tokyo and School of Allied Health Sciences at the Kitasato University. Written consent was given by the patients for their information

to be stored in the hospital database and used for research. This study was performed according to the tenets of the Declaration of Helsinki.

Subjects

Training dataset

This dataset was inherited from our prior study,³⁷ however only eyes with a mean deviation (MD) greater than -5 dB were used (Humphrey Field Analyzer [HFA], Carl Zeiss Meditec, Dublin, CA). As a result, the glaucoma group comprised 94 eyes from 94 subjects with open angle glaucoma (OAG) and the control group was made up of 84 eyes from 84 normal subjects. Study participants were enrolled between January 2009 and March 2010 at the University of Tokyo Hospital or the Tajimi Iwase eye clinic. All subjects underwent complete ophthalmic examinations, including biomicroscopy, gonioscopy, intraocular pressure measurement, funduscopy, refraction, best-corrected visual acuity measurements and axial length (AL) measurements (IOL Master; Carl Zeiss Meditec), as well as SD-OCT imaging and visual field (VF) testing.

Full details of the glaucoma group are described elsewhere.³⁷ In short, a diagnosis was made when ophthalmoscopically apparent glaucomatous optic disc change was confirmed by a panel of glaucoma specialists and

glaucomatous VF change, defined by the ‘Anderson-Patella criteria’,³⁹ was present. Subjects aged 20 years or older and eyes with refractive error ≥ -6.0 D and < 3.0 D were included. Those with other systemic or ocular disorders were carefully excluded. Eyes with anomalous discs including tilted discs⁴⁰ were also carefully excluded.

Inclusion criteria for normal group were also inherited from our previous report;³⁷ no abnormal findings except for clinically insignificant senile cataract on biomicroscopy, gonioscopy and funduscopy, and no history of ocular diseases that could affect the results of SD-OCT examinations, such as diabetic retinopathy or age-related macular degeneration. Other inclusion criteria were age ≥ 20 years old, spherical equivalent refractive error ≥ -6.0 D and < 3.0 D, and normal VF test results according to Anderson-Patella’s criteria.³⁹ Eyes with anomalous discs including tilted discs [35] were cautiously excluded.

Testing dataset

This dataset consisted of 114 eyes of 114 subjects with open angle glaucoma (OAG) with a MD greater than -5 dB and 82 eyes of 82 normal subjects. Participants were enrolled prospectively after the training dataset was established, also at the Tokyo University Hospital, Kitasato University

Hospital or the Tajimi Iwase eye clinic. The inclusion and exclusion criterion are identical to those applied in the training dataset.

VF measurement

VF testing was performed, within 3 months of the SD-OCT examination, using the HFA with the SITA Standard strategy and the Goldmann size III target. Normative eyes in both training and testing datasets and glaucomatous eyes in the testing dataset were measured with the 24-2 program while glaucomatous eyes in the training dataset were measured either with the 24-2 or 30-2 test program. Near refractive correction was used as necessary. All of the participants had previous experience in VF examinations and unreliable VFs defined as fixation losses greater than 25 %, or false-positive responses greater than 15 % were excluded.⁴¹

SD-OCT data acquisition

SD-OCT data were obtained using either the 3D OCT-1000 or 3D OCT-2000 (Topcon Corp., Tokyo, Japan). The results obtained with 3D OCT-1000 are completely interchangeable with those from 3D-OCT 2000, but scanning speed in the latter instrument is 90 % faster. The details of the OCT image acquisition have been described elsewhere.³⁷ In short, all

SD-OCT measurements were performed after pupil dilation with 1 % tropicamide and SD-OCT imaging was performed using the raster-scan protocol. Data obtained during apparent eye movements and influenced by involuntary blinking or saccade or those with quality factor < 30% were excluded, as recommended by the manufacturer.

Analysis of SD-OCT data

Similar to our previous report,³⁷ in the macular area, the fovea was automatically identified as the pixel with the thinnest retinal thickness close to the fixation point, and a square area (6.0×6.0 mm) centered on the fovea was divided into 10×10 grids with an equal size. Then, to avoid the possible situation that the analysis area runs off the edge of the 6.0×6.0 mm data-acquisition area, the outermost units were excluded which resulted in an inner 8×8 grids. The thickness of GCC in each grid was calculated as the mean value of the thickness in each grid, and also the mean thickness over the whole analysis area (4.8×4.8 mm area centered on the fovea: Total GCC), superior hemiretina (2.4×4.8 mm area: Superior GCC) and inferior hemiretina (Inferior GCC) were calculated.

cpRNFL was measured as the RNFL thickness along a 3.4 mm diameter circle around the disc and the thicknesses of the whole circumference of the circle (Total cpRNFL), the four cpRNFL sectors each

accounting for four 90° (temporal, superior, nasal and inferior cpRNFL, respectively) and twelve 30°cpRNFL sectors (temporal, temporosuperior, superotemporal, superior, superonasal, nasosuperior, nasal, nasoinferior, inferonasal, inferior, inferotemporal and temporoinferior, in the clockwise manner from the papillomacular bundle in the right eye. Left eyes were mirror-imaged to a right eye orientation.

The Random Forests method was used to discriminate glaucoma eyes from normal eyes using the 84 OCT measurements listed in **Table 1** and also axial length. Using the training dataset, 10,000 decision trees were grown in the Random Forests and the classification error rates were calculated by varying the number of the variables randomly sampled as candidates at each split and the minimum size of terminal nodes. The optimum values for these parameters were decided when the classification error rate was minimized. Significant ($p < 0.05$) OCT parameters in the Random Forests were calculated by randomly permuting the variable at each decision tree and observing the decrease in the number of correct classifications.²⁷ Then the developed Random Forests classifier was applied to the testing dataset, and the area under the receiver operating characteristic curve (AROC) was calculated using the proportion of 10,000 decision tree votes in the Random Forests.⁴² For comparison, AROCs were also derived using a series of logistic

regression models. Backward model selection was used to generate one model, following a previous study by Mwanza et al.,²¹ while bidirectional model selection was applied to generate a further model.

Backward model selection begins with all candidate variables included, each variable is then removed, one at a time, and model fit is assessed. If deletion improves the model then the variable is dropped, the process is repeated until no further improvement is observed. Forward model selection starts with no variables in the model, and tests whether the addition of a variable improves the model, again the process is repeated until the model cannot be further improved. Bidirectional selection is a combination of both approaches, testing variables to be included or excluded at each step. In this study, the corrected AIC (AICc) was used to assess model fit as it gives an accurate estimation even when the sample size is small.⁴³ Statistical models, such as linear or logistic regression are frequently used as a means to answer scientific questions, but it is also widely acknowledged that these models may be over-fitted to the original sample, particularly when the number of predictor variables is large. Model selection methods have been proposed to try to overcome this problem, by adequately reducing the number of predictor variables. The final models obtained from the training dataset using the backward and bidirectional selections with AICc

were applied to each eye in the testing dataset to calculate the estimated possibility of the diagnosis of glaucoma. Then AROCs were derived using the obtained possibility of glaucoma.

Alternatively, Tibshirani et al. proposed over-fitting in linear modeling could be mitigated by applying a shrinkage method in which the sum of the absolute values of the regression coefficients is constrained or penalized, known as least absolute shrinkage and selection operator regression (LASSO).^{44,45} More exactly, in the case of ordinary least squared linear regression, let $\mathbf{x} \in \mathbb{R}^p$ denote the variables and let $\mathbf{y} \in \mathbb{R}$ denote the response (please note $\mathbf{x}_{ij}$ are normalized and $\mathbf{y}$ has mean zero). The Lasso algorithm solves the following problem:

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \mathbf{x}_i^T \beta)^2 + \lambda P_\alpha(\beta) \right], \quad (1)$$

$$P_\alpha(\beta) = (1 - \alpha) \frac{1}{2} \|\beta\|_{l_2}^2 + \alpha \|\beta\|_{l_1} \quad (2)$$

$$= \sum_{j=1}^p \left[\frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right] \quad (3)$$

In Equation (1), $\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \mathbf{x}_i^T \beta)^2 \right]$ is identical to ordinary least squares linear regression and $\lambda P_\alpha(\beta)$ is the penalty term for the shrinkage.^{44,45} Thus the λ (Lambda) value represents the degree of penalty in Lasso. This constraint can be applied in multiple logistic regression. The equation (1) is Lasso regression when $\alpha = 1$ and Ridge

regression when $\alpha = 0$. The LASSO regression has been used in many different fields, from the analysis of human perception to genetic analysis,^{46,47} and we have recently shown the usefulness of this approach in glaucoma.^{48,49}

All statistical analyses were carried out using the statistical programming language R (ver. 3.1.3, The R Foundation for Statistical Computing, Vienna, Austria). The R packages “randomForest” and “glmnet” were used to carry out the Random Forests and LASSO analyses. The AROC was used to evaluate the clinical usefulness of each classifier, as reported in a previous paper.⁵⁰ Comparison of multiple AROCs was carried out using DeLong’s method.⁵¹ Holm’s method^{52,53} was used to correct P values for the problem of multiple testing. The optimum cut-off point in the ROC was calculated using Youden’s method.⁵⁴

Results

Subject characteristics are given in **Table 2**. Thickness of the Total cpRNFL and Total GCC, and MD and MD were significantly smaller in the glaucoma group compared with the normal group, both in the training and testing datasets ($p < 0.001$, unpaired Wilcoxon-test). Axial length was significantly smaller in the glaucoma group compared with the normal group,

both in the training dataset ($P < 0.001$, unpaired Wilcoxon-test), but this was not the case in the testing dataset.

Using the testing dataset, the AROCs obtained with the raw values of the thickness of 360°-cpRNFL, superior and 90°-cpRNFL, and 12 30°-cpRNFL, total, superior and inferior hemiretina GCC are shown in **Table 3**. Among these parameters, the largest AROC was associated with inferior quadrant cpRNFL, and was equal to 89.5 [CI: 85.1 – 93.9] %; see **Table 3**.

Results with the training dataset: the parameters selected in the backward selection model and the bidirectional model were identical. The parameters and their coefficients are shown in **Table 4**. When applied to the training dataset, the AROC of this model was 100.0 [CI: 100.0 – 100.0] %. The optimum penalty with the LASSO and Ridge models were 0.014 and 0.012, respectively; the AROCs were 98.6 [CI: 97.1 – 100.0] %, and 98.7 [CI: 97.1 – 100.0] %, respectively. The AROC associated with the Random Forests method was 99.7 [CI: 99.3 – 100.0] %. For the Random Forests method, 23 of the 84 OCT measurements and axial length were significant predictors. Significant predictors included total cpRNFL, total, superior and inferior hemiretina GCC, 16 grid GCC, as well as superotemporal, nasosuperior, nasal, nasoinferior and inferotemporal 30°-cpRNFL sectors and inferior

90°-cpRNFL sectors (**Figure 1**).

Results with the testing dataset: the ROCs and AROCs obtained using the different classifiers in the testing dataset are shown in **Figure 2**. Among these, the AROC associated with the Random Forests method (93.0 [CI: 89.4 – 96.6] %) was significantly larger than AROC obtained with any other model; bidirectional/backward model (71.9 [CI: 65.4 – 78.4] %), LASSO (89.6 [CI: 84.2 – 93.1] %) and also Ridge (89.2 [CI: 84.8 – 93.6] %); $p < 0.05$, DeLong's method after correction for multiple testing. The AROC associated with the Random Forest method was significantly larger than the AROC associated with inferior quadrant cpRNFL ($p = 0.048$, DeLong's method after correction for multiple testing). The optimum discrimination of the Random Forest method was at a sensitivity of 96.3 % and specificity of 78.9 %, using Youden's method.⁵⁴

Discussion

In the current study, a Random Forests classifier was developed to diagnose early glaucoma using multiple OCT parameters. The developed classifier was then applied to an independent dataset and as a result, a significantly larger AROC was obtained using the classifier compared with

any single OCT parameter, and also a series of multiple logistic regression models.

The diagnostic performance of glaucoma classifiers tends to be poor in early stage glaucomatous eyes.^{13,18,34} Reviewing relevant previous reports, Tan et al. reported an AROC of 90% with mean GCC thickness over the macular area in discriminating early (average MD was equal to -4.6 dB) glaucoma eyes from normal eyes.¹¹ Kim et al. reported AROCs of 83.6 and 86.9 % with total, and superior and inferior cpRNFL thickness, and AROCs between 82.6 and 89.5 % with total, and superior and inferior GCC thickness, in glaucoma eyes with an average MD of -8.49 dB.¹⁵ In the current study, the average MD of glaucomatous eyes in the testing dataset was just -2.5 dB, and the AROCs obtained were comparable, albeit slightly smaller, to those in these other studies: 85.5, 80.3 and 84.3 % with total, superior and inferior GCC, respectively, and 83.3, 82.5, 89.5 % with total, superior and inferior cpRNFL, respectively.

We have previously reported a very large AROC of 98.5 % in discriminating glaucoma and normal eyes using the Random Forests method, but the study data included eyes with a wide range of glaucoma severity (MD was -23.3 to 1.8 dB).³⁷ At the clinical setting, OCT is useful to give a

diagnosis of early stage glaucoma. In the current study, only eyes with early glaucoma (MD greater than -5.0 dB) were included in the testing dataset, and, as a result, a considerable decrease in the AROC was observed. In addition, all of the AROC values associated with single OCT parameters were much lower than those in our previous study.

As discussed in our previous study,³⁷ summary measures such as average or hemiretinal index and also OCT parameters obtained from smaller regions have limitations. Early local structural change may not be reflected in the summary measures but making a diagnosis based on OCT parameters obtained from smaller regions is also difficult as these measurements are more susceptible to measurement noise. Indeed, the regression model decided by backwards/bidirectional model selection methods classified glaucoma and normal eyes perfectly (AROC = 100.0 [CI: 100.0 – 100.] %) in the training dataset, however, the AROC was greatly reduced in the independent testing dataset: 71.9 [CI: 65.4 – 78.4] %. Thus, these model selection approaches still suffer from the ‘over-fitting’ problem. LASSO and Ridge methods have been proposed to overcome this problem, but even they are not free from this problem as suggested by the difference in AUCs between the training and testing datasets: 98.6 [CI: 97.1 – 100.0] % and 98.7 [CI: 97.1 – 100.0] % when LASSO and Ridge were applied to the

training dataset, which reduced to 89.6 [CI: 84.2 – 93.1] % and (89.2 [CI: 84.8 – 93.6] %) when LASSO and Ridge were applied to the testing dataset.

Indeed, these AROCs were not larger than that with the inferior quadrant cpRNFLT (89.5 [CI: 85.1 – 93.9] %). The AROC associated with the Random Forests classifier was also smaller in the testing dataset (93.0 [CI: 89.4 – 96.6] %) than in the training dataset (99.7 [CI: 99.3 – 100.0] %), but remained significantly larger than any other method. The greater performance is probably because the Random Forest classifier can concurrently analyze all measurements and recognize global patterns.^{55,56} In addition, the Random Forests method is an ensemble classifier consisting of numerous decision trees; we applied a simple decision tree method in our previous study and the AROC was equal to just 88.4 % despite the fact the dataset included many advanced cases. Other studies have investigated different methods to classify glaucoma using SD-OCT measurements.

Mwanza et al. analyzed multiple disc and SD-OCT parameters using logistic regression and reported an AROC of 97.4% in glaucoma eyes with an average MD of –3.2 dB.²¹ This model was applied to the training dataset and, as a result, eyes were successfully discriminated with 100% accuracy.

Nonetheless, this same model obtained an AROC of merely 71.9 ± CI: [65.4 to 78.4 %] in the testing dataset.

The Random Forests method is just one of a large group of machine learning methods. Burgansky-Eliash et al. used a Support Vector Machine Classifier to analyze multiple TD-OCT parameters and reported a high AROC to discriminate normal and glaucoma eyes (98.1 %).²² It is not known whether the Random Forests method outperforms other machine learning methods in discriminating glaucoma from normal eyes, but we were less concerned with this aspect, because of the strength of the Random Forests method, including its interpretability. The contribution of each parameter in the diagnosis process, as shown in **Figure 1**, cannot be obtained with other machine learning methods, such as the support vector machine or gradient boosting classifier. In the current study, as shown in **Figure 1b**, there were many contributing GCC grids in the superior and inferior-temporal areas. These areas are likely to be involved in the early stages of glaucoma^{19,20} and were also identified in our previous report.³⁷ However, in contrast to our previous report, in general, the grids near the edge of the analyzed plane were included in the current classifier. This may be because the current study included only early stage glaucoma patients. Glaucomatous VF damage usually starts from outer regions in the 10-2 HFA VF which corresponds to the contributed grids in the current study, and spreads toward the inner area in the later stages of the disease.

One limitation of the current study is the exclusion of eyes with high myopia (spherical refraction error < -6 diopters). Myopia is an apparent risk factor for developing glaucoma,⁵⁷⁻⁶¹ nonetheless making an accurate diagnosis of glaucoma in this type of eyes is a challenge, even with imaging devices such as the Heidelberg Retina Tomograph (Heidelberg Engineering GmbH, Heidelberg, Germany) or OCT.⁶²⁻⁶⁵ This should be further pursued in a future study.

In conclusion, we have further validated that the combined analysis of cpRNFL- and GCC-related thickness measurements using a Random Forests classifier is useful to diagnose glaucoma, especially in early stage glaucomatous eyes.

Acknowledgement

Funding/Support: supported in part by the Japan Science and Technology Agency (JST)-CREST and Grant 26462679 from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Financial disclosure: none

Figure and Table Legends

Table 1

Parameters used in the Random Forests method and other multiple logistic regression models.

cpRNFL: circumpapillary retinal nerve fiber layer (RNFL), GCC: ganglion cell complex

Table 2

Characteristics of the study participants.

MD was calculated wither with 24-2 and 30-2 Humphrey visual field

sd: standard deviation, MD: mean deviation, cpRNFL: circumpapillary retinal nerve fiber layer (RNFL), GCC: ganglion cell complex

Table 3

Area under the receiver operating characteristic curves with summary spectral domain optical coherence tomography measures.

AROC: Area Under the Receiver Operating Characteristic Curve, SD-OCT: spectral domain optical coherence tomography, CI: 95 % confidence interval, cpRNFL: circumpapillary retinal nerve fiber layer (RNFL), GCC: ganglion cell complex.

Table 4

The parameters selected in the backward selection model and the bidirectional model

The parameters selected in the backward selection model and the bidirectional model were identical. The locations of the grids can be seen in the **Figure 1b**.

cpRNFL: circumpapillary retinal nerve fiber layer (RNFL), GCC: ganglion cell complex.

Figure 1

CpRNFL and GCC-related parameters that significantly contributed to the discrimination of glaucoma eyes from normal eyes (highlighted in red).

Left figures: CpRNFL -related parameters that significantly contributed to the discrimination of glaucoma eyes from normal eyes, and right figure:

GCC-related parameters that significantly contributed to the discrimination of glaucoma eyes from normal eyes

cpRNFL: circumpapillary retinal nerve fiber layer (RNFL), GCC: ganglion cell complex

T: temporal, TS: temporosuperior, ST: superotemporal, S: superior, SN: superonasal, NS: nasosuperior, N: nasal, NI: nasoinferior, IN: inferonasal, I:

inferior, IT: inferotemporal and TI: temporoinferior

Figure 2

Receiver operating characteristic (ROC) curves in the testing dataset, obtained with the Random Forests method and the best performing cpRNFL- and GCC-related parameters.

The AROC obtained using the Random Forests method (93.0 [CI: 89.4 – 96.6] %) was significantly larger than any of the AROC associated with backward and bidirectional stepwise model selections (71.9 [CI: 65.4 – 78.4] %), LASSO (89.6 [CI: 84.2 – 93.1] %), Ridge (89.2 [CI: 84.8 – 93.6] %) ($p < 0.05$, DeLong's method after correction for multiple testing).

LASSO: least absolute shrinkage and selection operator regression, Step backward and bidirectional: backward and bidirectional step model selection, GCC: ganglion cell complex.

References

1. Sommer A, Pollack I, Maumenee AE. Optic disc parameters and onset of glaucomatous field loss. I. Methods and progressive changes in disc morphology. *Arch Ophthalmol* 1979;97(8):1444-1448.
2. Pederson JE, Anderson DR. The mode of progressive disc cupping in ocular hypertension and glaucoma. *Arch Ophthalmol* 1980;98(3):490-495.
3. Quigley HA, Katz J, Derick RJ, et al. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. *Ophthalmology* 1992;99(1):19-28.
4. Sommer A, Miller NR, Pollack I, et al. The nerve fiber layer in the diagnosis of glaucoma. *Arch Ophthalmol* 1977;95(12):2149-2156.
5. Sommer A, Katz J, Quigley HA, et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol* 1991;109(1):77-83.
6. Sommer A, Quigley HA, Robin AL, et al. Evaluation of nerve fiber layer assessment. *Arch Ophthalmol* 1984;102(12):1766-1771.
7. 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(3):741-748.
8. Harwerth RS, Carter-Dawson L, Shen F, et al. Ganglion cell losses underlying visual field defects from experimental glaucoma. *Invest Ophthalmol Vis Sci* 1999;40(10):2242-2250.
9. Harwerth RS, Carter-Dawson L, Smith EL, 3rd, et al. Neural losses correlated with visual losses in clinical perimetry. *Invest Ophthalmol Vis Sci* 2004;45(9):3152-3160.
10. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol* 1989;107(5):453-464.
11. Tan O, Chopra V, Lu AT, et al. Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology* 2009;116(12):2305-2314 e2301-2302.
12. Garas A, Vargha P, Hollo G. Diagnostic accuracy of nerve fibre layer, macular thickness and optic disc measurements made with the RTVue-100 optical coherence tomograph to detect glaucoma. *Eye (Lond)* 2011;25(1):57-65.
13. Moreno PA, Konno B, Lima VC, et al. Spectral-domain optical coherence tomography for early glaucoma assessment: analysis of macular ganglion cell complex versus peripapillary retinal nerve fiber layer. *Can J Ophthalmol* 2011;46(6):543-547.
14. 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 (Lond)* 2012;26(1):133-139.
15. Kim NR, Lee ES, Seong GJ, et al. Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT

in glaucoma. *Invest Ophthalmol Vis Sci* 2010;51(9):4646-4651.

16. Rao HL, Kumbar T, Addepalli UK, et al. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. *Invest Ophthalmol Vis Sci* 2012;53(2):1058-1065.
17. Cho JW, Sung KR, Lee S, et al. Relationship between visual field sensitivity and macular ganglion cell complex thickness as measured by spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci* 2010;51(12):6401-6407.
18. Schulze A, Lamparter J, Pfeiffer N, et al. Diagnostic ability of retinal ganglion cell complex, retinal nerve fiber layer, and optic nerve head measurements by Fourier-domain optical coherence tomography. *Graefes Arch Clin Exp Ophthalmol* 2011;249(7):1039-1045.
19. Shields MB. *TEXTBOOK OF GLAUCOMA*. Maryland, USA: William & Wilkins, 1997.
20. Zimmerman TJ, KS K. *Clinical Pathways in Glaucoma*. NY: Thieme, 2001.
21. Mwanza JC, Warren JL, Budenz DL, et al. Combining spectral domain optical coherence tomography structural parameters for the diagnosis of glaucoma with early visual field loss. *Invest Ophthalmol Vis Sci* 2013;54(13):8393-8400.
22. Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci* 2005;46(11):4147-4152.
23. Baskaran M, Ong EL, Li JL, et al. Classification algorithms enhance the discrimination of glaucoma from normal eyes using high-definition optical coherence tomography. *Invest Ophthalmol Vis Sci* 2012;53(4):2314-2320.
24. Han J, M. K. *Data Mining: Concepts and Techniques*. . Burlington, Massachusetts: Morgan Kaufmann Publishers, 2006.
25. Mitchell TM. *Machine learning*. NY: McGraw-Hill Higher Education, 1997.
26. Breiman L. Random Forests. *Machine Learning* 2001;45(5-32).
27. Breiman L, Cutler A. Random Forests, 2004.
28. Strobl C, Boulesteix AL, Kneib T, et al. Conditional variable importance for random forests. *BMC Bioinformatics* 2008;9(307).
29. Dietterich TG. *Ensemble learning*. In *The Handbook of Brain Theory and Neural Networks, 2nd ed*. Cambridge: The MIT Press, 2002.
30. Maroco J, Silva D, Rodrigues A, et al. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res Notes* 2011;4(299).
31. Douglas PK, Harris S, Yuille A, et al. Performance comparison of machine learning algorithms and number of independent components used in fMRI decoding of belief vs. disbelief. *Neuroimage* 2011;56(2):544-553.
32. Lunetta KL, Hayward LB, Segal J, et al. Screening large-scale association

study data: exploiting interactions using random forests. *BMC Genet* 2004;5(32).

33. Segal MR, Cummings MP, Hubbard AE. Relating amino acid sequence to phenotype: analysis of peptide-binding data. *Biometrics* 2001;57(2):632-642.

34. Rao HL, Zangwill LM, Weinreb RN, et al. Comparison of different spectral domain optical coherence tomography scanning areas for glaucoma diagnosis. *Ophthalmology* 2010;117(9):1692-1699, 1699 e1691.

35. Lisboa R, Paranhos A, Jr., Weinreb RN, et al. Comparison of different spectral domain OCT scanning protocols for diagnosing preperimetric glaucoma. *Invest Ophthalmol Vis Sci* 2013;54(5):3417-3425.

36. Sugimoto K, Murata H, Hirasawa H, et al. Cross-sectional study: Does combining optical coherence tomography measurements using the 'Random Forest' decision tree classifier improve the prediction of the presence of perimetric deterioration in glaucoma suspects? *BMJ Open* 2013;3(10):e003114.

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

38. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006;90(3):262-267.

39. Anderson DR, Patella VM. *Automated Static Perimetry*. 2nd ed. St.Louis: Mosby, 1999.

40. Apple DJ, Rabb MF, Walsh PM. Congenital anomalies of the optic disc. *Surv Ophthalmol* 1982;27(1):3-41.

41. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Invest Ophthalmol Vis Sci* 2000;41(8):2201-2204.

42. Japkowicz N. *Evaluating Learning Algorithms: A Classification Perspective*. . Cambridge, UK: : Cambridge University Press, 2011.

43. Burnham KP, DR. A. Multimodel inference: understanding: AIC and BIC in model selection. *Sociological Methods & Research* 2004;33(261-304).

44. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B* 1996;58(1):267-288.

45. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33(1):1-22.

46. Yuki K, Asaoka R, Tsubota K. The Relationship between Central Visual Field Damage and Motor Vehicle Collisions in Primary Open-Angle Glaucoma Patients. *PLoS One* 2014;9(12):e115572.

47. Asaoka D, Nagahara A, Hojo M, et al. The Relationship between H. pylori Infection and Osteoporosis in Japan. *Gastroenterol Res Pract* 2014;2014(340765).

48. Asaoka R. Measuring visual field progression in the central 10 degrees using additional information from central 24 degrees visual fields and 'lasso regression'. *PLoS One* 2013;8(8):e72199.

49. Fujino Y, Murata H, Mayama C, et al. Applying "Lasso" Regression to

Predict Future Visual Field Progression in Glaucoma Patients. *Invest Ophthalmol Vis Sci* 2015;56(4):2334-2339.

50. Demšar J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research* 2006;7(1):30.
51. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837-845.
52. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979;6(65):70.
53. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am J Public Health* 1996;86(5):726-728.
54. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3(1):32-35.
55. Jamie S, Matthew J, Roberto C. Semantic texton forests for image categorization and segmentation. *Proc IEEE Conf Computer Vision and Pattern Recognition* 2008:1-8.
56. Liang W, Yizhou W, Z D. Building Emerging Pattern (EP) Random forest for recognition. *Proc IEEE Conf on Image Processing* 2010:1457-1460.
57. Mitchell P, Hourihan F, Sandbach J, et al. The relationship between glaucoma and myopia: the Blue Mountains Eye Study. *Ophthalmology* 1999;106(10):2010-2015.
58. Suzuki Y, Iwase A, Araie M, et al. Risk factors for open-angle glaucoma in a Japanese population: the Tajimi Study. *Ophthalmology* 2006;113(9):1613-1617.
59. Xu L, Wang Y, Wang S, et al. High myopia and glaucoma susceptibility the Beijing Eye Study. *Ophthalmology* 2007;114(2):216-220.
60. Perera SA, Wong TY, Tay WT, et al. Refractive error, axial dimensions, and primary open-angle glaucoma: the Singapore Malay Eye Study. *Arch Ophthalmol* 2010;128(7):900-905.
61. Liang YB, Friedman DS, Zhou Q, et al. Prevalence of primary open angle glaucoma in a rural adult Chinese population: the Handan eye study. *Invest Ophthalmol Vis Sci* 2011;52(11):8250-8257.
62. Mayama C, Tsutsumi T, Saito H, et al. Glaucoma-Induced Optic Disc Morphometric Changes and Glaucoma Diagnostic Ability of Heidelberg Retina Tomograph II in Highly Myopic Eyes. *PLoS One* 2014;9(1):e86417.
63. Yamazaki Y, Yoshikawa K, Kunimatsu S, et al. Influence of myopic disc shape on the diagnostic precision of the Heidelberg Retina Tomograph. *Jpn J Ophthalmol* 1999;43(5):392-397.
64. 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;250(12):1843-1849.
65. Kim NR, Lee ES, Seong GJ, et al. Comparing the ganglion cell complex and retinal nerve fibre layer measurements by Fourier domain OCT to detect glaucoma in high myopia. *Br J Ophthalmol* 2011;95(8):1115-1121.

Table 1

Parameters used in the Random Forests method and other multiple logistic regression models.

SD-OCT parameters

	total
	4 quadrants (superior, temporal, nasal, inferior quadrant)
cpRNFL (17 in total)	12 (temporal, temporal-superior, superior-temporal, superior, superior-nasal, nasal-superior, nasal, nasal-inferior, inferior-nasal, inferior, inferior-temporal, temporal-inferior) sectors
	total
GCC (67 in total)	2 sectors (superior and inferior hemiretina) 64 grid

Table 2

Characteristics of the study participants

	Training dataset			Testing dataset		
	Glaucoma group	Normal group	p value	Glaucoma group	Normal group	p value
Eye (right / left)	50 / 44	65 / 19	0.0013	56 / 58	53 / 29	0.044
Age, y (mean $\pm$ sd)	57.9 $\pm$ 11.0	52.6 $\pm$ 15.6	0.044	61.9 $\pm$ 13.9	52.6 $\pm$ 13.9	< 0.001
Gender (male / female)	35 / 59	47 / 37	0.019	49 / 65	26 / 56	0.15
MD, dB (mean $\pm$ sd)	-2.7 $\pm$ 1.8	-0.40 $\pm$ 1.3	<0.001	-2.5 $\pm$ 1.9	-0.36 $\pm$ 1.9	<0.001
Axial length, μ m (mean $\pm$ sd)	24.5 $\pm$ 1.4	23.6 $\pm$ 0.78	<0.001	24.5 $\pm$ 1.4	24.5 $\pm$ 1.4	0.42
cpRNFL, μ m (mean $\pm$ sd)	86.8 $\pm$ 11.0	102.4 $\pm$ 8.3	<0.001	84.5 $\pm$ 14.1	100.3 $\pm$ 7.8	<0.001
GCC, μ m (mean $\pm$ sd)	88.2 $\pm$ 11.0	106.0 $\pm$ 7.3	<0.001	91.8 $\pm$ 12.2	106.2 $\pm$ 7.3	<0.001

Table 3

Area under the receiver operating characteristic curves with summary spectral domain optical coherence tomography measures

SD-OCT parameter	AROC [CI] (%)
total cpRNFL	83.3 [77.6 to 89.1]
temporal quadrant cpRNFL	67.6 [60.2 to 75.1]
superior quadrant cpRNFL	82.5 [76.7 to 88.4]
nasal quadrant cpRNFL	53.7 [45.6 to 61.9]
inferior quadrant cpRNFL	89.5 [85.1 to 93.9]
12 sector cpRNFL (temporal)	62.1 [54.3 to 69.9]
12 sector cpRNFL (temporal-superior)	71.1 [63.9 to 78.2]
12 sector cpRNFL (superior-temporal)	79.1 [72.9 to 85.2]
12 sector cpRNFL (superior)	70.3 [63.1 to 77.6]
12 sector cpRNFL (superior-nasal)	73.8 [66.7 to 80.9]
12 sector cpRNFL (nasal-superior)	55.1 [47.0 to 63.2]
12 sector cpRNFL (nasal)	49.8 [41.7 to 57.9]
12 sector cpRNFL (nasal-inferior)	53.8 [45.6 to 62.0]
12 sector cpRNFL (inferior-nasal)	74.1 [67.2 to 81.0]
12 sector cpRNFL (inferior)	87.1 [82.3 to 91.9]
12 sector cpRNFL (inferior-temporal)	84.7 [79.3 to 90.1]
12 sector cpRNFL (temporal-inferior)	64.3 [56.7 to 72.0]
total GCC	85.5 [80.2 to 90.8]
superior hemiretina GCC	80.3 [74.3 to 86.4]
inferior hemiretina GCC	84.3 [78.8 to 89.9]

Table 4

The parameters selected in the backward selection model and the
bidirectional model

selected parameters	coefficients
temporal quadrant cpRNFL	-97.2
superior quadrant cpRNFL	-137.2
12 sector cpRNFL (temporal)	86.2
12 sector cpRNFL (superior-temporal)	83.8
12 sector cpRNFL (superior)	83.4
12 sector cpRNFL (nasal-superior)	61.1
12 sector cpRNFL (inferior)	-32.3
12 sector cpRNFL (inferior-temporal)	50.7
grid 22	-90.3
grid 32	88.8
grid 35	-26.8
grid 73	-195.9
grid 75	100.0
grid 82	310.8
grid 83	-36.5

Figure 1ab

Figure 2