Combining Optical Coherence Tomography measurements using the 'Random Forest' decision tree classifier improves the diagnosis of glaucoma

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003114
Article Type:	Research
Date Submitted by the Author:	24-Apr-2013
Complete List of Authors:	Sugimoto, Koichiro; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Murata, Hiroshi; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Aihara, Makoto; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Mayama, Chihiro; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Asaoka, Ryo; University of Tokyo Graduate School of Medicine, Department of Ophthalmology
Primary Subject Heading:	Ophthalmology
Secondary Subject Heading:	Ophthalmology
Keywords:	Glaucoma < OPHTHALMOLOGY, Optical Coherence Tomography, Visual Field, Random Forest

Title: Combining Optical Coherence Tomography measurements using the 'Random Forest' decision tree classifier improves the diagnosis of glaucoma

Koichiro Sugimoto[1], Hiroshi Murata[1], Makoto Aihara[1, 2], Chihiro Mayama[1], Ryo Asaoka[1]

Institutions: [1] Department of Ophthalmology, University of Tokyo Graduate School of Medicine, Tokyo , Japan; [2] Shirato Eye Clinic, Tokyo, Japan

Key words: Visual Field, Glaucoma, Optical Coherence Tomography, Random Forest

Word count: 2987

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

Summary

1) Article Focus

- Interpreting optical coherence tomography parameters simultaneously improves the discrimination between glaucoma and glaucoma suspect.
- It is beneficial to use a machine learning algorithm of Random Forest to combine the optical coherence tomography parameters.

2) Key Messages

- It was beneficial to combine the optical coherence tomography parameters with the Random Forest method to combine the optical coherence tomography parameters for discriminating between glaucoma and glaucoma suspect.

3) Strengths and Limitations

- Strengths: This method will lead to the improvement of the diagnosis of glaucoma, at the clinical settings.
- Limitations: No inclusion of normal subjects, because it was not our purpose to discriminate between glaucoma and normal subjects.

Abstract

Purpose

To develop a classifier to diagnose glaucoma based on measurements of these structures using the machine learning method known as the 'Random Forest' algorithm.

Methods

Spectral domain optical coherence tomography (Topcon 3D OCT-2000) and perimetry (Humphrey Field Analyzer, 24-2 or 30-2 SITA standard) measurements were conducted in 293 eyes of 179 subjects with open angle glaucoma (OAG) or suspected OAG. Visual field (VF) damage (OHTS criteria (2002)) was used as a 'gold-standard' to classify glaucomatous eyes. The 'Random Forest' method was then used to analyze the relationship between the presence/absence of glaucomatous VF damage and the following variables: age, gender, right or left eye, axial length plus 237 different OCT measurements. The area under the receiver operating characteristic curve (AROC) was then derived using the probability of glaucoma as suggested by the proportion of votes in the Random Forest classifier. For comparison, five AROCs were derived based on: (i) macular retinal nerve fiber layer (m-RNFL) alone, (ii) circumpapillary (cp-RNFL) alone, (iii) ganglion cell layer and inner plexiform layer (GCL + IPL) alone, (iv) rim area alone, and (v) a decision tree method using the same variables as the Random Forest algorithm.

Results

The AROC from the combined Random Forest classifier (0.90) was significantly larger than the AROCs based on individual measurements of m-RNFL (0.86), cp-RNFL (0.77), GCL + IPL (0.80), rim area (0.78), and the decision tree method (0.75) ($p < 0.05$).

Conclusions

Evaluating OCT measurements using the Random Forest method provides an accurate diagnosis of glaucoma.

Introduction

Glaucoma is the second most common cause of blindness. As glaucomatous visual field (VF) damage is irreversible, early diagnosis of glaucoma is essential. Structural changes at the optic nerve head[1] and retinal nerve fiber layer (RNFL) around the optic disc[2] can also indicate glaucomatous damage, and may precede measurable VF loss.

Optical coherence tomography (OCT) is an imaging technology widely used in the diagnosis of glaucoma, enabling high-resolution measurements of the retina. [3] The recent advancement of OCT from the time domain to the spectral domain (SD-OCT) has greatly improved the imaging speed and resolution of the device, [4] and has enabled imaging scans of the macular retinal nerve fiber layer (m-RNFL) and the macular ganglion cell layer and inner plexiform layer (GCL + IPL). It has been reported that these retinal layers are damaged early in the glaucoma disease process[5, 6] and many studies have investigated the diagnostic performance of thickness measurements of these structures to discriminate between healthy and glaucomatous eyes[7-14]. However, in these previous studies, the different measurements were interpreted independently, yet damage to these structures does not necessarily occur in parallel.[15, 16] The purpose of this study was to improve the structural diagnosis of glaucoma (using VF damage as a gold-standard classifier) by analyzing multiple OCT measurements concurrently using the machine learning method known as the 'Random Forest' algorithm.

The Random Forest method is a decision support tool which consists of many decision trees. Decision trees have previously been used to diagnose glaucoma[27]; however, decision trees suffer from the problem of 'over-fitting', which influences the diagnostic accuracy. [28] In contrary, the Random Forest classifier overcomes this problem by summarizing the results of many decision trees. Another noteworthy advantage of the Random Forest algorithm over traditional methods, such as logistic regression, is that any interaction or correlation between variables does not adversely affect the classification since it is capable of representing high order interactions. [29] Furthermore, predictors that might otherwise be masked by their correlation with

other variables, using other classification methods, can contribute to the Random Forest classifier. In this study we have employed the Random Forest algorithm to explore multiple OCT parameters concurrently in order to build an unbiased glaucoma classifier.

Materials and Methods

The study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at the University of Tokyo. 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.

This retrospective study comprised 293 eyes of 179 consecutive patients referred to the University of Tokyo Hospital for glaucoma or suspected glaucoma between August 2010 and July 2012. Patients were referred based on optic disc damage: focal or diffuse neuroretinal rim thinning, localized notching or nerve fiber layer defects. Subjects underwent complete ophthalmic examinations, including slit lamp biomicroscopy, gonioscopy, intraocular pressure (IOP) measurement, funduscopy, and axial length (AL) measurement (IOL Master; Carl Zeiss Meditec, Dublin, CA), as well imaging with SD-OCT and VF testing. Criteria for inclusion were visual acuity better than 6/12; no previous ocular surgery, except cataract extraction and intraocular lens implantation; open anterior chamber angle (patients with angle closure glaucoma were excluded); no other anterior and posterior segment eye disease. AL was not used for the inclusion / exclusion criteria.

VF testing was performed using the Humphrey Field Analyzer (HFA, Carl Zeiss Meditec); 24-2 or 30-2 test pattern and the SITA Standard strategy, with the Goldmann size III target. Near refractive correction was used as necessary, calculated according to the subject's age by the HFA software. Unreliable VFs were excluded according to HFA criteria (fixation losses greater than 25%, or false-positive responses greater than 15%). False negative rate was not used as an indicator of test reliability following a previous report [17]. A glaucomatous VF was defined as a pattern standard deviation (PSD) value beyond the normal limit ($P < 0.05$), or a Glaucoma Hemifield Test (GHT) result outside normal limits following the criteria in [18]. All glaucoma patients had previous experience in visual field testing.

SD-OCT (3D OCT-2000; Topcon Corp., Tokyo, Japan) was used to obtain tomographic images of the parapapillary fundus with the 3D Disc scan and 3D Macula scan (128 horizontal scan lines comprised of 512 A-scans for an image area of 6×6 mm). SD-OCT uses a superluminescent diode laser with a center wavelength of 840nm and a bandwidth of 50nm as the light source. The transverse and axial resolutions are less than 20μm and 5μm, respectively. The acquisition speed is 50,000 A scans per second. In the selected eye, the macula was imaged by 6 radial lines centered at the fovea spaced 30° apart. All of the measurements were performed after pupil dilation with 1% Tropicamide and all of the images had signal strength of at least 60, as recommended by the manufacturer.

The 'Random Forest' algorithm is an ensemble machine learning classifier proposed by Breiman in 2001. [19, 20] The Random Forest consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. Thus the Random Forest is an ensemble classifier, which has been reported to improve the prediction accuracy of decision tree.[21] Indeed there are many reports which suggested the Random Forest gives best prediction accuracy among various machine learning methods and this method has been used in many research fields, including gene selection and cancer classification.[22-25] In the random Forest method, when classifying a new object from an input vector, the input vector is classified by each of the trees in the forest, and the tree "votes" for that class. The forest then chooses the classification having the most votes over all the trees in the forest. Each tree is constructed using a different bootstrap sample from the original data. Thus, cross-validation is performed internally and there is no need for a separate cross-validation data set to obtain an unbiased estimate of the test set error. For classification, node impurity was measured using the Gini index [26].

The Random Forest method was used to classify the presence or absence of glaucomatous VF damage using: OCT measurements (237 different measurements in total were analyzed), age, gender, AL and right/left eye (see **Table 2)**. In this procedure, 10,000 trees were

grown and five among the 241 parameters were used at each node. The area under the receiver operating characteristic curve (AROC) was derived from the probability of glaucoma (the proportion of votes) as suggested by the method; for each individual, only the data from all other subjects (n=178) was used (leave-one-out cross validation) so that right and left eyes of a subject are not used for both training and testing simultaneously. For comparison, the AROCs were also derived using only individual raw thickness measurements of: m-RNFL, or cp-RNFL, or GCL + IPL, or rim area and the prediction with the decision tree method. The diagnostic sensitivity and specificity was also calculated for the age-matched normative limits of the different measurements (P ≤ %5, or, P ≤ 1%): m-RNFL, and GCL + IPL, as shown on the instrument's print out.

Finally, variable importance was calculated by randomly permuting a variable at each decision tree and observing whether the number of correct decisions decreased[20].

All statistical analyses were carried out using the statistical programming language R (ver. 2.14.2, The R Foundation for Statistical Computing, Vienna, Austria) and Medcalc version 11.4.2.0; MedCalc statistical software, Mariakerke, Belgium). The R package "randomForset'' and "rpart" was used to carry out the analysis of the Random Forest method and decision tree method, respectively.

Results

Subject characteristics are given in **Table 1**. VFs of 224 eyes in 150 patients were diagnosed as glaucomatous while the remaining 69 eyes of 57 patients were judged as normal. The average total m-RNFL thickness, cp-RNFL thickness, GCL + IPL thickness and rim area were significantly smaller in the glaucomatous group compared with the normal group ($P < 0.05$, non-paired t-test).

As shown in the **Figure 1**, The AROC of the Random Forest method utilizing all measurements (0.90) was significantly larger than that with m-RNFL alone (0.86), cp-RNFL alone (0.77), GCL-IPL (0.80) and rim area alone (0.78) ($p < 0.05$). Furthermore, the diagnostic performance (sensitivity and specificity) of the age-matched normative database (as shown on the OCT printout) were also plotted in **Figure 1**. The sensitivity and specificity for thickness values outside normal limits were: m-RNFL ($P < 5\%$): 0.74 and 0.93; m-RNFL ($P < 1\%$): 0.61 and 0.96; GCL-IPL ($P < 5\%$): 0.48 and 0.88; GCL + IPL ($P < 1\%$): 0.42 and 0.90 (sensitivity and specificity, respectively).

Figure 2 illustrates the OCT measurements analyzed. Among 237 measurements, 76 had a significant variable importance measure including: total and inferior m-RNFL thickness, total and inferior GCL + IPL thickness, an m-RNFL thickness value outside normal limits ($P < 5\%$), various sectorial m-RNFL thickness values (**Figure 2a**), various GCL + IPL thickness values (**Figure 2b**), and two cp-RNFL thickness values (**Figure 2c**). Age, AL, gender, and right or left eye were not significant.

Discussion

Glaucoma can be diagnosed and monitored using structural measurements from OCT; to date, cp-RNFL thickness measurements have generally been used to quantify glaucomatous damage using time-domain OCT [27-31]. Another "traditional" way to measure structural glaucomatous damage is to use scanning laser tomography (HRT; Heidelberg retina tomography, Heidelberg engineering, Heidelberg, Germany) to measure characteristics of the optic disc such as size and shape. HRT works on the principle of confocal scanning laser ophthalmoscopy, and is long established as a diagnostic tool for glaucoma. HRT measurements of rim area, among the various optic disc parameters, have been reported as most clinically meaningful, repeatable and reliable[32-34].

The development of SD-OCT has improved the scan speed, resolution[35] and repeatability[36] of captured images. These improvements may, in turn, strengthen the association between structural SD-OCT measurements and functional VF measurements in glaucoma patients, which is referred to as the 'structure-function' relationship [37, 38]. The advent of SD-OCT has also enabled the measurement of the m-RNFL and GCL + IPL layers. Recent studies have investigated the structure-function relationship in glaucoma patients using measurements of the macular ganglion cell complex (GCC), which is the total thickness of the GCL + IPL and m-RNFL layers[7-11, 13, 14, 39-42]. This research suggests that structural measurements of the cp-RNFL layer, or the GCC, give rise to an analogous structure-function relationship and diagnostic ability to detect glaucoma, and there is no consensus on which structure is optimum for diagnosing glaucoma. Indeed, specific structures may be preferentially damaged in any given patient. For example, Cordeiro et al. reported that the diagnostic performance of cp-RNFL thickness measurements tended to be better in patients with a small optic disc, and an inverse effect was observed using the GCC measurement. [43] Conversely, GCC may be preferential to detect glaucomatous change in high myopic patients. [44] Thus, it

appears that no single structural measurement is best for diagnosing glaucoma.

Chen et al. used a logistical diagnostic model to diagnose glaucoma; the model analyzed a patient's optic cup:optic disc vertical ratio, cp-RNFL thickness and rim area simultaneously, but the authors found that diagnostic performance was not significantly improved compared with using individual measurements. [40] On the other hand, Burgansky-Eliash et al. used a support vector machine classifier of multiple Stratus OCT parameters to diagnose glaucoma, and showed that the AROC was significantly larger. [45] Also, a recent study suggested the decision tree method is useful to discriminate between glaucoma and normal subject.[46] In contrary, in the current study, the decision tree failed to show the benefit in discriminating glaucoma and glaucoma subject, however it was beneficial to use the Random Forest method which is the ensemble classifier of decision trees. There are other studies which suggested the merit of combining multiple structural measurements to diagnose glaucoma, [47, 48] yet none of these studies have analyzed the m-RNFL and GCL + IPL layers simultaneously with cp-RNFL, optic disc shape parameters as well age and AL.

A merit of the Random Forest method is that the importance of each parameter for its classification can be tested. The variable importance measure suggested that total m-RNFL thickness, total GCL + IPL thickness, and m-RNFL thickness outside normal limits (P < 5%) significantly contributed to the diagnosis of glaucoma. In contrast, age, AL, gender, eye (right / left), and optic disc measurements such as rim area, were not significant. Reports have suggested that optic disc shape parameters are useful for classifying glaucomatous eyes, but are less useful compared to RNFL parameters[16, 50]. However, previous results have been based on HRT measurements of the optic disc, and there are notable differences between the corresponding measurements in SD-OCT. For instance, the margin of the optic disc and cup is automatically identified in SD-OCT, whereas it is manually drawn by the examiner in HRT. Furthermore, it has been reported that HRT measurements of optic disc shape detect a different population of

glaucoma patients to OCT measurements of the RNFL[16]. Accordingly, the diagnostic performance of the Random Forest classifier may be further improved by also including various optic disc shape parameters derived from HRT. We intend to investigate this hypothesis in a future study.

Interestingly, our results question the validity of SD-OCT's normal limits to discriminate glaucoma. For example, the blue cross in **Figure 1** indicates that GCL + IPL measurements outside normal limits at the P < 1% level have a specificity of 90%. The normal limits of the SD-OCT are derived by testing 'normal' subjects without ocular disease; Rao et al. have reported that cp-RNFL thickness measurements from normal subjects and patients with glaucoma overlap considerably [51]. A significant advantage of the Random Forest classifier is that normal limits could be established based on results from normal subjects and glaucoma patients; these would be expected to better reflect the 'true' specificity of the test result. Another merit of the Random Forest method, in comparison to the current standard, is that the method gives an exact probability of glaucoma, rather than a binary classification (glaucoma or not at P < 1%, or P < 5%); such a value could be interpreted in a manner similar to that of the 'Nerve Fiber Index' (NFI) score in the nerve fibre analyzer imaging instrument (GDx, Carl Zeiss Meditec), which is a continuous numeric score from 0 to 99.

In our Random Forest classifier, many sectorial thickness measurements of the m-RNFL, GCL + IPL and cp-RNFL layers were deemed significant for the diagnosis of glaucoma. Significant sectors were generally located in the inferior hemi-retina, although a few sectors were also situated in the superior hemi-retina (see **Figure 2**). Previous studies have suggested that glaucomatous VF damage preferably affects the superior hemifield[52, 53]. Interestingly, the significant m-RNFL, GCL + IPL and cp-RNFL sectors in our classifier were principally distributed along the inferotemporal RNFL bundle, which likely corresponds to an arcuate defect in the superior VF[54]. Thus, these results also suggest that glaucomatous RNFL / GCL + IPL damage tends to occur in the inferior hemi-retina.

OCT structural measurements are influenced by ageing; cp-RNFL [55-57], rim area [58], m-RNFL, and GCL + IPL all become thinner with age [59]. In addition, studies suggest that AL may have an effect on measurements of the cp-RNFL[58, 60], rim area [58, 60], m-RNFL [59], and GCL + IPL [59]; however any such effects remain contentious[61-63]. In our study, removing age and AL factors did not affect the AROC of the Random Forest classifier.

Other machine learning methods, such as support vector machines, boosting and bagging classifiers could also be used to diagnose glaucoma. Previous reports suggest that the Random Forest method outperforms most other methods [24, 64, 65]; hence the Random Forest algorithm was used in the current study. Nevertheless, in a future study, we intend to investigate the performance of machine learning methods for discriminating perimetric and preperimetric glaucoma.

In conclusion, we have shown that combining SD-OCT measurements of the m-RNFL, cp-RNFL, GCL + IPL layers, using the Random Forest method, is hugely beneficial for diagnosing glaucoma, especially when compared with the current OCT reference-standard of comparing these measurements to an age-matched normative database.

Figure and Table Legends

Table 1

Characteristics of the study participants.

MD: mean deviation, m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Table 2

The variables used in the analysis, including 237 optical coherence tomography parameters.

m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Figure 1

Receiver operating characteristic (ROC) curves with the probability of glaucoma suggested by the Random Forest classifier and raw thickness measurements of: m-RNFL alone, cp-RNFL alone, and GCL + IPL alone, and decision tree method.

The area under the ROC with the Random Forest method was significantly larger than those of individual measurements and decision tree method ($P < 0.05$). The colored "X" represent the sensitivity and specificity of the SD-OCT normative database (red: m-RNFL ($P < 5\%$), orange: m-RNFL ($P < 1\%$), green: GCC ($P< 5\%$), blue: GCC ($P < 1\%$)).

m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Figure 2

Variables in the Random Forest classifier having a significant effect on the diagnosis of glaucoma.

Sectors of the cp-RNFL, m-RNFL, and GCL + IPL were superimposed onto a fundus photograph[54]; significant sectors are highlighted in red. If a subject's left eye was tested, the recorded data were mapped to a right eye format for analysis. Figure 2a: cp-RNFL, Figure 2b: m-RNFL,

Figure 2c: GCL-IPL
m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Funding

None

Competing Interests

None

Contributorship

Conceived and designed the experiments: KS, HM, RA
Performed the experiments: KS, HM, RA
Analyzed the data: KS, HM, RA
Contributed reagents/materials/analysis tools: KS, HM, RA
Wrote the manuscript: KS, HM, RA
Gave advice from the viewpoint of glaucom specialist: MA, CM

Data sharing

No additional data available

References

1. 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.
2. 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.
3. Chang R, Budenz DL. New developments in optical coherence tomography for glaucoma. Curr Opin Ophthalmol. 2008;19(2):127-135.
4. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. Science. 1991;254(5035):1178-1181.
5. 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.
6. Nakano N, Ikeda HO, Hangai M, et al. Longitudinal and simultaneous imaging of retinal ganglion cells and inner retinal layers in a mouse model of glaucoma induced by N-methyl-D-aspartate. Investigative ophthalmology & visual science. 2011;52(12):8754-8762.
7. 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. Investigative ophthalmology & visual science. 2010;51(12):6401-6407.
8. 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. 2011;25(1):57-65.
9. 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. Investigative ophthalmology & visual science. 2010;51(9):4646-4651.
10. 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. Canadian journal of ophthalmology Journal canadien d'ophtalmologie. 2011;46(6):543-547.
11. 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(1):133-139.
12. Rao HL, Kumbar T, Addepalli UK, et al. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. Investigative

ophthalmology & visual science. 2012;53(2):1058-1065.

13. 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.

14. 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.

15. Tuulonen A, Lehtola J, Airaksinen PJ. Nerve fiber layer defects with normal visual fields. Do normal optic disc and normal visual field indicate absence of glaucomatous abnormality? Ophthalmology. 1993;100(5):587-597; discussion 597-588.

16. Leung CK, Choi N, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: pattern of RNFL defects in glaucoma. Ophthalmology. 2010;117(12):2337-2344.

17. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? Investigative ophthalmology & visual science. 2000;41(8):2201-2204.

18. Gordon MO, Beiser JA, Brandt JD, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. Archives of ophthalmology. 2002;120(6):714-720; discussion 829-730.

19. Breiman L. Random Forests. Machine Learning. 2001;45:5-32.

20. Breiman L, Cutler A. Random Forests Available: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm. Accessed 2/9/2012.

21. Dietterich TG. Ensemble learning. In The Handbook of Brain Theory and Neural Networks, 2nd ed. Cambridge: The MIT Press 2002.

22. Palmer DS, O'Boyle NM, Glen RC, et al. Random forest models to predict aqueous solubility. Journal of chemical information and modeling. 2007;47(1):150-158.

23. Wu B, Abbott T, Fishman D, et al. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics. 2003;19(13):1636-1643.

24. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. BMC bioinformatics. 2006;7:3.

25. Svetnik V, Liaw A, Tong C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. Journal of chemical information and computer sciences. 2003;43(6):1947-1958.

26. Gini C. 1909 Concentration and dependency ratios (in Italian). English translation in Rivista di Politica Economica. 1997;87:769-789.

27. Hood DC, Kardon RH. A framework for comparing structural and functional measures of glaucomatous damage. Prog Retin Eye Res. 2007;26(6):688-710.
28. Harwerth RS, Wheat JL, Fredette MJ, et al. Linking structure and function in glaucoma. Prog Retin Eye Res. 2010;29(4):249-271.
29. Harwerth RS, Vilupuru AS, Rangaswamy NV, et al. The relationship between nerve fiber layer and perimetry measurements. Investigative ophthalmology & visual science. 2007;48(2):763-773.
30. Ferreras A, Pablo LE, Garway-Heath DF, et al. Mapping standard automated perimetry to the peripapillary retinal nerve fiber layer in glaucoma. Investigative ophthalmology & visual science. 2008;49(7):3018-3025.
31. Asaoka R, Ishii R, Kyu N, et al. Early detection of thinning of retinal nerve fiber layer in glaucomatous eyes by optical coherence tomography 3000: analysis of retinal nerve fiber layer corresponding to the preserved hemivisual field. Ophthalmic Res. 2006;38(1):29-35.
32. Strouthidis NG, White ET, Owen VM, et al. Factors affecting the test-retest variability of Heidelberg retina tomograph and Heidelberg retina tomograph II measurements. Br J Ophthalmol. 2005;89(11):1427-1432.
33. Strouthidis NG, White ET, Owen VM, et al. Improving the repeatability of Heidelberg retina tomograph and Heidelberg retina tomograph II rim area measurements. Br J Ophthalmol. 2005;89(11):1433-1437.
34. Tan JC, Garway-Heath DF, Hitchings RA. Variability across the optic nerve head in scanning laser tomography. Br J Ophthalmol. 2003;87(5):557-559.
35. Wojtkowski M, Srinivasan V, Fujimoto JG, et al. Three-dimensional retinal imaging with high-speed ultrahigh-resolution optical coherence tomography. Ophthalmology. 2005;112(10):1734-1746.
36. Arthur SN, Smith SD, Wright MM, et al. Reproducibility and agreement in evaluating retinal nerve fibre layer thickness between Stratus and Spectralis OCT. Eye (Lond). 2011;25(2):192-200.
37. Sehi M, Grewal DS, Sheets CW, et al. Diagnostic ability of Fourier-domain vs time-domain optical coherence tomography for glaucoma detection. Am J Ophthalmol. 2009;148(4):597-605.
38. Lee JR, Jeoung JW, Choi J, et al. Structure-function relationships in normal and glaucomatous eyes determined by time- and spectral-domain optical coherence tomography. Investigative ophthalmology & visual science. 2010;51(12):6424-6430.
39. Bowd C, Tafreshi A, Zangwill LM, et al. Pattern electroretinogram association with spectral domain-OCT structural measurements in glaucoma. Eye (Lond).

2011;25(2):224-232.

40. Fang Y, Pan YZ, Li M, et al. Diagnostic capability of Fourier-Domain optical coherence tomography in early primary open angle glaucoma. Chin Med J (Engl). 2010;123(15):2045-2050.

41. Garas A, Kothy P, Hollo G. Accuracy of the RTVue-100 Fourier-domain optical coherence tomograph in an optic neuropathy screening trial. Int Ophthalmol. 2011;31(3):175-182.

42. Rolle T, Briamonte C, Curto D, et al. Ganglion cell complex and retinal nerve fiber layer measured by fourier-domain optical coherence tomography for early detection of structural damage in patients with preperimetric glaucoma. Clin Ophthalmol. 2011;5:961-969.

43. Cordeiro DV, Lima VC, Castro DP, et al. Influence of optic disc size on the diagnostic performance of macular ganglion cell complex and peripapillary retinal nerve fiber layer analyses in glaucoma. Clinical ophthalmology. 2011;5:1333-1337.

44. 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.

45. Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. Investigative ophthalmology & visual science. 2005;46(11):4147-4152.

46. Baskaran M, Ong EL, Li JL, et al. Classification algorithms enhance the discrimination of glaucoma from normal eyes using high-definition optical coherence tomography. Investigative ophthalmology & visual science. 2012;53(4):2314-2320.

47. Lu AT, Wang M, Varma R, et al. Combining nerve fiber layer parameters to optimize glaucoma diagnosis with optical coherence tomography. Ophthalmology. 2008;115(8):1352-1357, 1357 e1351-1352.

48. Chen HY, Huang ML, Hung PT. Logistic regression analysis for glaucoma diagnosis using Stratus Optical Coherence Tomography. Optom Vis Sci. 2006;83(7):527-534.

49. AR. B. Statistical Learning from a Regression Perspective. NY: Springer Science + Business Media, LLC 2008.

50. Lisboa R, Leite MT, Zangwill LM, et al. Diagnosing Preperimetric Glaucoma with Spectral Domain Optical Coherence Tomography. Ophthalmology. 2012.

51. 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.

52. Hart WM, Jr., Becker B. The onset and evolution of glaucomatous visual field defects. Ophthalmology. 1982;89(3):268-279.

53. Heijl A, Lundqvist L. The frequency distribution of earliest glaucomatous visual field defects documented by automatic perimetry. Acta Ophthalmol (Copenh). 1984;62(4):658-664.
54. Garway-Heath DF, Poinoosawmy D, Fitzke FW, et al. Mapping the visual field to the optic disc in normal tension glaucoma eyes. Ophthalmology. 2000;107(10):1809-1815.
55. Lee JY, Hwang YH, Lee SM, et al. Age and retinal nerve fiber layer thickness measured by spectral domain optical coherence tomography. Korean J Ophthalmol. 2012;26(3):163-168.
56. Parikh RS, Parikh SR, Sekhar GC, et al. Normal age-related decay of retinal nerve fiber layer thickness. Ophthalmology. 2007;114(5):921-926.
57. Feuer WJ, Budenz DL, Anderson DR, et al. Topographic differences in the age-related changes in the retinal nerve fiber layer of normal eyes measured by Stratus optical coherence tomography. J Glaucoma. 2011;20(3):133-138.
58. Knight OJ, Girkin CA, Budenz DL, et al. Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by Cirrus HD-OCT. Arch Ophthalmol. 2012;130(3):312-318.
59. Ooto S, Hangai M, Tomidokoro A, et al. Effects of age, sex, and axial length on the three-dimensional profile of normal macular layer structures. Investigative ophthalmology & visual science. 2011;52(12):8769-8779.
60. Kang SH, Hong SW, Im SK, et al. Effect of myopia on the thickness of the retinal nerve fiber layer measured by Cirrus HD optical coherence tomography. Investigative ophthalmology & visual science. 2010;51(8):4075-4083.
61. Zou H, Zhang X, Xu X, et al. Quantitative in vivo retinal thickness measurement in chinese healthy subjects with retinal thickness analyzer. Investigative ophthalmology & visual science. 2006;47(1):341-347.
62. Chan CM, Yu JH, Chen LJ, et al. Posterior pole retinal thickness measurements by the retinal thickness analyzer in healthy Chinese subjects. Retina. 2006;26(2):176-181.
63. Hoh ST, Lim MC, Seah SK, et al. Peripapillary retinal nerve fiber layer thickness variations with myopia. Ophthalmology. 2006;113(5):773-777.
64. 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 research notes. 2011;4:299.
65. 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.

	'Glaucomatous' VF group			'Normal' VF group			
	mean	sd	range	mean	sd	range	p value
Age (years)	53.6	13.2	17 - 85	48.5	12.7	17 - 48	< 0.01
MD (dB)	-6.2	5.2	-30	-0.5	1.2	-4.9	<0.01
AL (mm)	25.1	1.7	22.2 - 29.3	26	1	22.8 - 29.5	0.11
m-RNFL (µm)	25.5	7.9	1.0 - 46.6	35.6	5.4	27.5 - 63.1	<0.01
cp-RNFL (µm)	88.3	15.1	49.0 - 123.4	104	15	66.9 - 150.9	<0.01
GCC (µm)	68.8	15.3	43.7 - 106.5	89.3	19.7	55.7 - 127.3	<0.01
Rim area (mm^2)	1.1	0.5	0.3 - 3.8	1.6	0.6	0.6 - 3.7	<0.01
Eye (right / left)	116 / 108			35 / 34			
Gender (male / female)	108 / 114			40 /31			

Table 1

	Measurement
cp-RNFL	total, 4 sectors (superior, temporal, nasal, inferior), 12 sectors
m-RNFL	total, 2 sectors (superior, inferior), 100 sectors
GCL + IPL	total, 2 sectors (superior, inferior), 100 sectors
Optic disc	Disc area, Cup area, Rim area, Cup volume, Rim volume, C/D area ratio, Linear C/D ratio, Vertical C/D ratio, Disc diameter (vertical), Disc diameter (horizontal)
m-RNFL	Significant according to normative database (P < 5%)
m-RNFL	Significant according to normative database (P < 1%)
GCL + IPL	Significant according to normative database (P < 5%)
GCL + IPL	Significant according to normative database (P < 1%)
Age	
Gender	
AL	
Eye (right/left)	

Table2

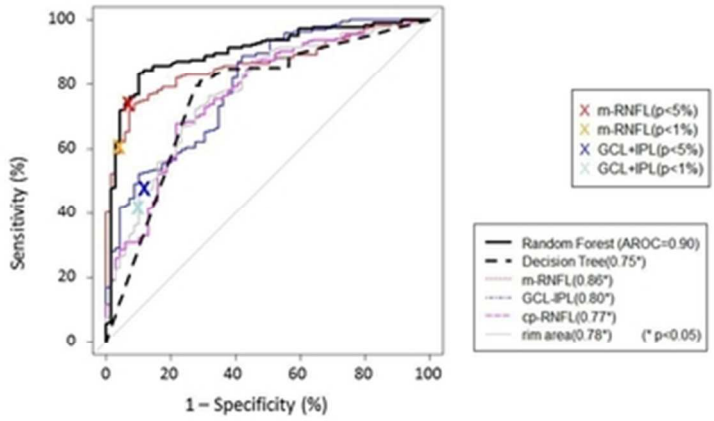


Figure 1

30x22mm (300 x 300 DPI)

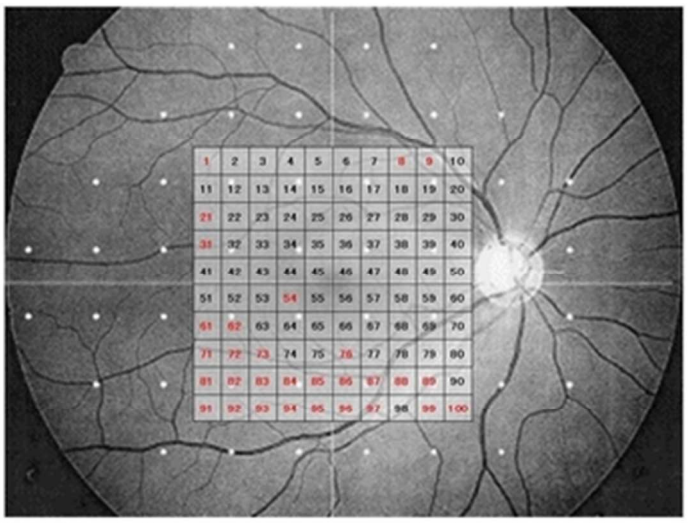


Figure 2a

30x22mm (300 x 300 DPI)

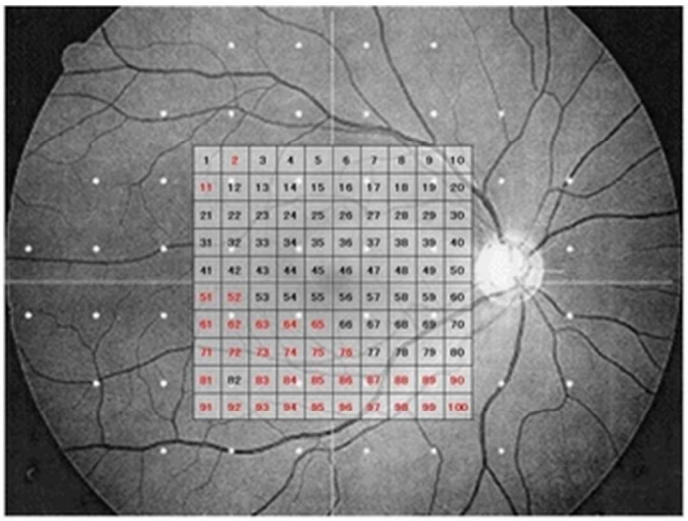


Figure 2b

30x22mm (300 x 300 DPI)

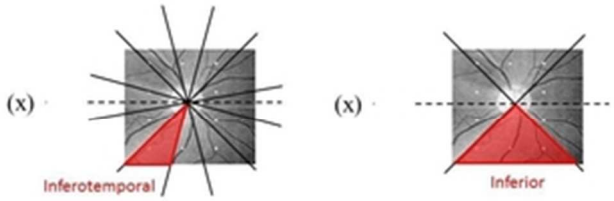


Figure 2c

30x22mm (300 x 300 DPI)

STARD checklist for reporting of studies of diagnostic accuracy

(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	2
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	3
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	3
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	3
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	3
Test methods	7	The reference standard and its rationale.	4
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	3
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	4
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	N/A
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	N/A
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	4
	13	Methods for calculating test reproducibility, if done.	N/A
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	3
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	6
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	3
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	N/A
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	Table 2
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	Figure 1
	20	Any adverse events from performing the index tests or the reference standard.	N/A
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	6
	22	How indeterminate results, missing data and outliers of the index tests were handled.	N/A
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	N/A
	24	Estimates of test reproducibility, if done.	N/A
DISCUSSION	25	Discuss the clinical applicability of the study findings.	7-9

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?

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003114.R1
Article Type:	Research
Date Submitted by the Author:	20-Jun-2013
Complete List of Authors:	Sugimoto, Koichiro; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Murata, Hiroshi; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Aihara, Makoto; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Mayama, Chihiro; University of Tokyo Graduate School of Medicine, Department of Ophthalmology Asaoka, Ryo; University of Tokyo Graduate School of Medicine, Department of Ophthalmology
Primary Subject Heading:	Ophthalmology
Secondary Subject Heading:	Ophthalmology
Keywords:	Glaucoma < OPHTHALMOLOGY, Optical Coherence Tomography, Visual Field, Random Forest

Title: 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?

Koichiro Sugimoto[1], Hiroshi Murata[1], Makoto Aihara[1, 2], Chihiro Mayama[1], Ryo Asaoka[1]

Institutions: [1] Department of Ophthalmology, University of Tokyo Graduate School of Medicine, Tokyo , Japan; [2] Shirato Eye Clinic, Tokyo, Japan

Key words: Visual Field, Glaucoma, Optical Coherence Tomography, Random Forest

Word count: 3032

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

Summary

1) Article Focus

- Cross sectional study to investigate the usefulness of interpreting optical coherence tomography parameters simultaneously when predicting visual field damage in glaucoma suspects.

2) Key Messages

- Combining optical coherence tomography parameters with the Random Forest machine learning method improves the prediction of visual field damage in glaucoma suspects.

3) Strengths and Limitations

- Strengths: This method could be used to improve the clinical management of glaucoma suspects and glaucoma patients.
- Limitations: No inclusion of normal subjects.

Abstract

Purpose

To develop a classifier to predict the presence of visual field (VF) deterioration in glaucoma suspects based on optical coherence tomography (OCT) measurements using the machine learning method known as the 'Random Forest' algorithm.

Methods

Spectral domain OCT (Topcon 3D OCT-2000) and perimetry (Humphrey Field Analyzer, 24-2 or 30-2 SITA standard) measurements were conducted in 293 eyes of 179 subjects with open angle glaucoma (OAG) or suspected OAG. VF damage (OHTS criteria (2002)) was used as a 'gold-standard' to classify glaucomatous eyes. The 'Random Forest' method was then used to analyze the relationship between the presence/absence of glaucomatous VF damage and the following variables: age, gender, right or left eye, axial length plus 237 different OCT measurements. The area under the receiver operating characteristic curve (AROC) was then derived using the probability of glaucoma as suggested by the proportion of votes in the Random Forest classifier. For comparison, five AROCs were derived based on: (i) macular retinal nerve fiber layer (m-RNFL) alone, (ii) circumpapillary (cp-RNFL) alone, (iii) ganglion cell layer and inner plexiform layer (GCL + IPL) alone, (iv) rim area alone, and (v) a decision tree method using the same variables as the Random Forest algorithm.

Results

The AROC from the combined Random Forest classifier (0.90) was significantly larger than the AROCs based on individual measurements of m-RNFL (0.86), cp-RNFL (0.77), GCL + IPL (0.80), rim area (0.78), and the decision tree method (0.75) ($p < 0.05$).

Conclusions

Evaluating OCT measurements using the Random Forest method provides an accurate prediction of the presence of perimetric

deterioration in glaucoma suspects.

Introduction

Glaucoma is the second most common cause of blindness. As glaucomatous visual field (VF) damage is irreversible, early diagnosis of glaucoma is essential. Structural changes at the optic nerve head[1] and retinal nerve fiber layer (RNFL) around the optic disc[2] can also indicate glaucomatous damage, and may precede measurable VF loss.

Optical coherence tomography (OCT) is an imaging technology widely used in the diagnosis of glaucoma, enabling high-resolution measurements of the retina. [3] The recent advancement of OCT from the time domain to the spectral domain (SD-OCT) has greatly improved the imaging speed and resolution of the device, [4] and has enabled imaging scans of the macular retinal nerve fiber layer (m-RNFL) and the macular ganglion cell layer and inner plexiform layer (GCL + IPL). It has been reported that these retinal layers are damaged early in the glaucoma disease process[5 6] and many studies have investigated the diagnostic performance of thickness measurements of these structures to discriminate between healthy and glaucomatous eyes[7-14]. However, in these previous studies, the different measurements were interpreted independently, yet damage to these structures does not necessarily occur in parallel [15 16] and thus there is no consensus on which structure is optimum for diagnosing glaucoma. Indeed, specific structures may be preferentially damaged in any given patient. For example, Cordeiro et al. reported that the diagnostic performance of cp-RNFL thickness measurements tended to be better in patients with a small optic disc, and an inverse effect was observed using the GCC measurement. [17] Conversely, GCC may be preferential to detect glaucomatous change in high myopic patients.[18] Thus, it appears that no single structural measurement is best for diagnosing glaucoma.

The 'Random Forest' method is a decision support tool which consists of many decision trees. Decision trees have previously been used to diagnose glaucoma[19]; however, decision trees suffer from the problem of 'over-fitting', which influences the diagnostic accuracy. [20] In

contrary, the Random Forest classifier overcomes this problem by summarizing the results of many decision trees. Another noteworthy advantage of the Random Forest algorithm over traditional methods, such as logistic regression, is that any interaction or correlation between variables does not adversely affect the classification since it is capable of representing high order interactions. [21] Furthermore, predictors that might otherwise be masked by their correlation with other variables, using other classification methods, can contribute to the Random Forest classifier.

Glaucomatous structural change is often apparent in glaucomatous patients without VF defects (pre-perimetric glaucoma)[22] [23]. Therefore, it may be possible to predict the presence of the VF deterioration from structural measurements, as it has been reported that there is a significant difference in structural measurements between patients with perimetric and pre-perimetric glaucoma. [22] Predicting the presence of VF damage from structural measurements is clinically very important, especially in patients who cannot reliably perform VF test, due to, for example, inability to concentrate, mental disorders, locomotor disabilities, and so on. The purpose of this study was to improve the prediction of the presence of VF damage in glaucoma suspects by analyzing multiple OCT measurements concurrently using the Random Forest algorithm.

Materials and Methods

The study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at the University of Tokyo. 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.

This retrospective study comprised 293 eyes of 179 consecutive patients referred to the University of Tokyo Hospital for glaucoma or suspected glaucoma between August 2010 and July 2012. Patients were referred based on optic disc damage: focal or diffuse neuroretinal rim thinning, localized notching or nerve fiber layer defects. Subjects underwent complete ophthalmic examinations, including slit lamp biomicroscopy, gonioscopy, intraocular pressure (IOP) measurement and funduscopy. If glaucomatous structural changes were confirmed from these tests, axial length (AL; IOL Master, Carl Zeiss Meditec, Dublin, CA), imaging with SD-OCT and VF testing were performed. Criteria for inclusion were visual acuity better than 6/12; no previous ocular surgery, except cataract extraction and intraocular lens implantation; open anterior chamber angle (patients with angle closure glaucoma and secondary open angle glaucoma were excluded); no other anterior and posterior segment eye disease. AL was not used for the inclusion / exclusion criteria.

VF testing was performed using the Humphrey Field Analyzer (HFA, Carl Zeiss Meditec); 24-2 or 30-2 test pattern and the SITA Standard strategy, with the Goldmann size III target. Near refractive correction was used as necessary, calculated according to the subject's age by the HFA software. Unreliable VFs were excluded according to HFA criteria (fixation losses greater than 25%, or false-positive responses greater than 15%). False negative rate was not used as an indicator of test reliability following a previous report [24]. A glaucomatous VF was defined as a pattern standard deviation (PSD) value beyond the normal limit ($P < 0.05$), or a Glaucoma Hemifield Test

(GHT) result outside normal limits following the criteria in [25]. All glaucoma patients had previous experience in visual field testing.

SD-OCT (3D OCT-2000; Topcon Corp., Tokyo, Japan) was used to obtain tomographic images of the parapapillary fundus with the 3D Disc scan and 3D Macula scan (128 horizontal scan lines comprised of 512 A-scans for an image area of 6×6 mm). SD-OCT uses a superluminescent diode laser with a center wavelength of 840nm and a bandwidth of 50nm as the light source. The transverse and axial resolutions are less than 20μm and 5μm, respectively. The acquisition speed is 50,000 A scans per second. In the selected eye, the macula was imaged by 6 radial lines centered at the fovea spaced 30° apart. All of the measurements were performed after pupil dilation with 1% Tropicamide and all of the images had signal strength of at least 60, as recommended by the manufacturer.

The 'Random Forest' algorithm is an ensemble machine learning classifier proposed by Breiman in 2001. [26 27] The Random Forest consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. Thus the Random Forest is an ensemble classifier, which has been reported to improve the prediction accuracy of decision tree.[28] Indeed there are many reports which suggested the Random Forest gives best prediction accuracy among various machine learning methods and this method has been used in many research fields, including gene selection and cancer classification.[29-32] In the random Forest method, when classifying a new object from an input vector, the input vector is classified by each of the trees in the forest, and the tree "votes" for that class. The forest then chooses the classification having the most votes over all the trees in the forest. Each tree is constructed using a different bootstrap sample from the original data. Thus, cross-validation is performed internally and there is no need for a separate cross-validation data set to obtain an unbiased estimate of the test set error. For classification, node impurity was measured using the Gini index [33].

The Random Forest method was used to classify the presence or

absence of glaucomatous VF damage using: OCT measurements (237 different measurements in total were analyzed), age, gender, AL and right/left eye (see **Table 2)**. In this procedure, 10,000 trees were grown and five among the 241 parameters were used at each node. The area under the receiver operating characteristic curve (AROC) was derived from the probability of glaucoma (the proportion of votes) as suggested by the method; for each individual, only the data from all other subjects (n=178) was used (leave-one-out cross validation) so that right and left eyes of a subject are not used for both training and testing simultaneously. For comparison, the AROCs were also derived using only individual raw thickness measurements of: m-RNFL, or cp-RNFL, or GCL + IPL, or rim area and the prediction with the decision tree method. The diagnostic sensitivity and specificity was also calculated for the age-matched normative limits of the different measurements (P ≤ %5, or, P ≤ 1%): m-RNFL, and GCL + IPL, as shown on the instrument's print out.

Finally, variable importance was calculated by randomly permuting a variable at each decision tree and observing whether the number of correct decisions decreased[27].

All statistical analyses were carried out using the statistical programming language R (ver. 2.14.2, The R Foundation for Statistical Computing, Vienna, Austria) and Medcalc version 11.4.2.0; MedCalc statistical software, Mariakerke, Belgium). The R package "randomForset'' and "rpart" was used to carry out the analysis of the Random Forest method and decision tree method, respectively.

Results

Subject characteristics are given in **Table 1**. VFs of 224 eyes in 150 patients were diagnosed as glaucomatous while the remaining 69 eyes of 57 patients were judged as normal. The average total m-RNFL thickness, cp-RNFL thickness, GCL + IPL thickness and rim area were significantly smaller in the glaucomatous group compared with the normal group ($P < 0.05$, non-paired t-test).

As shown in the **Figure 1**, The AROC of the Random Forest method utilizing all measurements (0.90) was significantly larger than that with m-RNFL alone (0.86), cp-RNFL alone (0.77), GCL-IPL (0.80) and rim area alone (0.78) ($p < 0.05$). Furthermore, the diagnostic performance (sensitivity and specificity) of the age-matched normative database (as shown on the OCT printout) were also plotted in **Figure 1**. The sensitivity and specificity for thickness values outside normal limits were: m-RNFL ($P < 5\%$): 0.74 and 0.93; m-RNFL ($P < 1\%$): 0.61 and 0.96; GCL-IPL ($P < 5\%$): 0.48 and 0.88; GCL + IPL ($P < 1\%$): 0.42 and 0.90 (sensitivity and specificity, respectively).

Figure 2 illustrates the OCT measurements analyzed. Among 237 measurements, 76 had a significant variable importance measure including: total and inferior m-RNFL thickness, total and inferior GCL + IPL thickness, an m-RNFL thickness value outside normal limits ($P < 5\%$), various sectorial m-RNFL thickness values (**Figure 2a**), various GCL + IPL thickness values (**Figure 2b**), and two cp-RNFL thickness values (**Figure 2c**). Age, AL, gender, and right or left eye were not significant.

Discussion

In the current study, the 'Random Forest' decision tree classifier was used to predict the presence of VF damage in glaucoma suspects. As a result, it was shown that the AROC given by the Random Forest method was significantly larger than those derived from any single OCT parameter and the simple decision tree method.

Previous attempts have been made to interpret multiple structural parameters in order to aid the diagnosis of glaucoma. Chen et al. used a logistical diagnostic model to diagnose glaucoma; the model analyzed a patient's optic cup:optic disc vertical ratio, cp-RNFL thickness and rim area simultaneously, but the authors found that diagnostic performance was not significantly improved compared with using individual measurements. [34] On the other hand, Burgansky-Eliash et al. used a support vector machine classifier of multiple Stratus OCT parameters to diagnose glaucoma, and showed that the AROC was significantly larger. [35] Other studies also support combining multiple structural measurements to diagnose glaucoma. [36 37]In addition, a recent study suggested the decision tree method is useful to discriminate between glaucoma patients and normal subjects.[19] However, in the current study, the decision tree method, which often suffers from the problem of over-fitting[38], failed to show benefit in discriminating glaucoma. On the other hand, it was beneficial to use the Random Forest method, which is an ensemble classifier of decision trees. Recent reports have revealed that distinguishing between perimetric glaucoma and pre-perimetric glaucoma is more difficult than differentiating normal subjects from glaucoma patients[39] with early VF damage[22]. A noteworthy advantage of the current study is that it is the first of its kind to analyze m-RNFL and GCL + IPL layers simultaneously with cp-RNFL, optic disc shape parameters as well age and AL.

It must be noted that a clear caveat of the current study is the lack of a normative population to act as a reference. Therefore, AROCs derived in the current study are not directly relevant to distinguishing between healthy subjects and glaucomatous patients. A further study

should be carried out with normative and glaucomatous populations (particularly patients with early stage glaucoma) in order to further investigate the merits of the Random Forest classifier. Nonetheless, the method's ability to accurately differentiate glaucoma suspects from glaucoma patients suggests that the classifier may be even more useful in this context.

The variable importance measure from the Random Forest method suggested that total m-RNFL thickness, total GCL + IPL thickness, and m-RNFL thickness outside normal limits (P < 5%) significantly contributed to the diagnosis of glaucoma. In contrast, age, AL, gender, eye (right / left), and optic disc measurements such as rim area, were not significant. Reports have suggested that optic disc shape parameters are useful for classifying glaucomatous eyes, but are less useful compared to RNFL parameters[16 40]. However, previous results have been based on HRT measurements of the optic disc, and there are notable differences between the corresponding measurements in SD-OCT. For instance, the margin of the optic disc and cup is automatically identified in SD-OCT, whereas it is manually drawn by the examiner in HRT. Furthermore, it has been reported that HRT measurements of optic disc shape detect a different population of glaucoma patients to OCT measurements of the RNFL[16]. Accordingly, the diagnostic performance of the Random Forest classifier may be further improved by also including various optic disc shape parameters derived from HRT. We intend to investigate this hypothesis in a future study.

Interestingly, our results question the validity of SD-OCT's normal limits to discriminate glaucoma. For example, the blue cross in **Figure 1** indicates that GCL + IPL measurements outside normal limits at the P < 1% level have a specificity of 90%. The normal limits of the SD-OCT are derived by testing 'normal' subjects without ocular disease; Rao et al. have reported that cp-RNFL thickness measurements from normal subjects and patients with glaucoma overlap considerably [41]. A significant advantage of the Random Forest classifier is that normal limits could be established based on results from normal subjects and

glaucoma patients; these would be expected to better reflect the 'true' specificity of the test result. Another merit of the Random Forest method, in comparison to the current standard, is that the method gives an exact probability of glaucoma, rather than a binary classification (glaucoma or not at P < 1%, or P < 5%); such a value could be interpreted in a manner similar to that of the 'Nerve Fiber Index' (NFI) score in the nerve fibre analyzer imaging instrument (GDx, Carl Zeiss Meditec), which is a continuous numeric score from 0 to 99.

In our Random Forest classifier, many sectorial thickness measurements of the m-RNFL, GCL + IPL and cp-RNFL layers were deemed significant for the prediction of glaucomatous VF damage. Significant sectors were generally located in the inferior hemi-retina, although a few sectors were also situated in the superior hemi-retina (see **Figure 2**). Previous studies have suggested that glaucomatous VF damage preferably affects the superior hemifield[42 43]. Interestingly, the significant m-RNFL, GCL + IPL and cp-RNFL sectors in our classifier were principally distributed along the inferotemporal RNFL bundle, which likely corresponds to an arcuate defect in the superior VF[44]. Thus, these results also suggest that glaucomatous RNFL / GCL + IPL damage tends to occur in the inferior hemi-retina.

OCT structural measurements are influenced by ageing; cp-RNFL [45-47], rim area [48], m-RNFL, and GCL + IPL all become thinner with age [49]. In addition, studies suggest that AL may have an effect on measurements of the cp-RNFL[48 50], rim area [48 50], m-RNFL [49], and GCL + IPL [49]; however any such effects remain contentious[51-53]. In our study, removing age and AL factors did not affect the AROC of the Random Forest classifier.

Other machine learning methods, such as support vector machines, boosting and bagging classifiers could also be used to diagnose glaucoma. Previous reports suggest that the Random Forest method outperforms most other methods [31 54 55]; hence the Random Forest algorithm was used in the current study. Nevertheless, in a future study, we intend to investigate the performance of machine

learning methods for discriminating perimetric and preperimetric glaucoma.

In conclusion, we have shown that combining SD-OCT measurements of the m-RNFL, cp-RNFL, GCL + IPL layers, using the Random Forest method, is beneficial for predicting the presence of glaucomatous VF damage in glaucoma suspects, especially when compared with the current OCT reference-standard of comparing these measurements to an age-matched normative database.

Figure and Table Legends

Table 1

Characteristics of the study participants.
MD: mean deviation, m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Table 2

The variables used in the analysis, including 237 optical coherence tomography parameters.
m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Figure 1

Receiver operating characteristic (ROC) curves with the probability of glaucoma suggested by the Random Forest classifier and raw thickness measurements of: m-RNFL alone, cp-RNFL alone, and GCL + IPL alone, and decision tree method.
The area under the ROC with the Random Forest method was significantly larger than those of individual measurements and decision tree method ($P < 0.05$). The colored "X" represent the sensitivity and specificity of the SD-OCT normative database (red: m-RNFL ($P < 5\%$), orange: m-RNFL ($P < 1\%$), green: GCC ($P< 5\%$), blue: GCC ($P < 1\%$)).
m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Figure 2

Variables in the Random Forest classifier having a significant effect on the presence of glaucomatous visual field damage.
Sectors of the cp-RNFL, m-RNFL, and GCL + IPL were superimposed onto a fundus photograph[44]; significant sectors are highlighted in red. If a subject's left eye was tested, the recorded data were mapped to a

right eye format for analysis. Figure 2a: cp-RNFL, Figure 2b: m-RNFL, Figure 2c: GCL-IPL

m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Funding

This research received no specific funding.

Competing interests

There are no competing interests.

Contributorship statement

Conceived and designed the experiments: KS, HM, RA

Performed the experiments: KS, HM, RA

Analyzed the data: KS, HM, RA

Contributed reagents/materials/analysis tools: KS, HM, RA

Wrote the manuscript: KS, HM, RA

Gave advice from the viewpoint of glaucom specialist: MA, CM

Data sharing

There are no additional data available.

References

1. 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.
2. 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.
3. Chang R, Budenz DL. New developments in optical coherence tomography for glaucoma. *Current opinion in ophthalmology* 2008;19(2):127-35.
4. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science* 1991;254(5035):1178-81.
5. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *American journal of ophthalmology* 1989;107(5):453-64.
6. Nakano N, Ikeda HO, Hangai M, et al. Longitudinal and simultaneous imaging of retinal ganglion cells and inner retinal layers in a mouse model of glaucoma induced by N-methyl-D-aspartate. *Investigative ophthalmology & visual science* 2011;52(12):8754-62.
7. 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. *Investigative ophthalmology & visual science* 2010;51(12):6401-7.
8. 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* 2011;25(1):57-65.
9. 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. *Investigative ophthalmology & visual science* 2010;51(9):4646-51.
10. 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. *Canadian journal of ophthalmology. Journal canadien d'ophtalmologie* 2011;46(6):543-7.
11. 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(1):133-9.

12. Rao HL, Kumbar T, Addepalli UK, et al. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. *Investigative ophthalmology & visual science* 2012;53(2):1058-65.
13. 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. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 2011;249(7):1039-45.
14. 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):2305-14 e1-2.
15. Tuulonen A, Lehtola J, Airaksinen PJ. Nerve fiber layer defects with normal visual fields. Do normal optic disc and normal visual field indicate absence of glaucomatous abnormality? *Ophthalmology* 1993;100(5):587-97; discussion 97-8.
16. Leung CK, Choi N, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: pattern of RNFL defects in glaucoma. *Ophthalmology* 2010;117(12):2337-44.
17. Cordeiro DV, Lima VC, Castro DP, et al. Influence of optic disc size on the diagnostic performance of macular ganglion cell complex and peripapillary retinal nerve fiber layer analyses in glaucoma. *Clinical ophthalmology* 2011;5:1333-7.
18. Shoji T, Nagaoka Y, Sato H, et al. Impact of high myopia on the performance of SD-OCT parameters to detect glaucoma. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 2012.
19. Baskaran M, Ong EL, Li JL, et al. Classification algorithms enhance the discrimination of glaucoma from normal eyes using high-definition optical coherence tomography. *Investigative ophthalmology & visual science* 2012;53(4):2314-20.
20. Mitchell T. *MACHINE LEARNING*. NY: McGraw-Hill Higher Education, 1997.
21. Strobl C, Boulesteix AL, Kneib T, et al. Conditional variable importance for random forests. *BMC bioinformatics* 2008;9:307.
22. Cvenkel B, Kontestabile AS. Correlation between nerve fibre layer thickness

measured with spectral domain OCT and visual field in patients with different stages of glaucoma. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 2011;249(4):575-84.

23. Jeoung JW, Park KH. Comparison of Cirrus OCT and Stratus OCT on the ability to detect localized retinal nerve fiber layer defects in preperimetric glaucoma. *Investigative ophthalmology & visual science* 2010;51(2):938-45.
24. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Investigative ophthalmology & visual science* 2000;41(8):2201-4.
25. Gordon MO, Beiser JA, Brandt JD, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;120(6):714-20; discussion 829-30.
26. Breiman L. Random Forests. *Machine Learning* 2001;45:5-32.
27. Breiman L, Cutler A. Random Forests, 2004.
28. Dietterich TG. *Ensemble learning. In The Handbook of Brain Theory and Neural Networks, 2nd ed*. Cambridge: The MIT Press, 2002.
29. Palmer DS, O'Boyle NM, Glen RC, et al. Random forest models to predict aqueous solubility. *Journal of chemical information and modeling* 2007;47(1):150-8.
30. Wu B, Abbott T, Fishman D, et al. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics* 2003;19(13):1636-43.
31. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. *BMC bioinformatics* 2006;7:3.
32. Svetnik V, Liaw A, Tong C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences* 2003;43(6):1947-58.
33. Gini C. 1909 Concentration and dependency ratios (in Italian). *English translation in Rivista di Politica Economica* 1997;87:769-89.
34. Fang Y, Pan YZ, Li M, et al. Diagnostic capability of Fourier-Domain optical coherence tomography in early primary open angle glaucoma. *Chinese medical journal* 2010;123(15):2045-50.
35. Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Investigative ophthalmology & visual science* 2005;46(11):4147-52.
36. Lu AT, Wang M, Varma R, et al. Combining nerve fiber layer parameters to optimize

glaucoma diagnosis with optical coherence tomography. *Ophthalmology* 2008;115(8):1352-7, 57 e1-2.

37. Chen HY, Huang ML, Hung PT. Logistic regression analysis for glaucoma diagnosis using Stratus Optical Coherence Tomography. *Optometry and vision science : official publication of the American Academy of Optometry* 2006;83(7):527-34.
38. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer, 2001.
39. Morooka S, Hangai M, Nukada M, et al. Wide 3-dimensional macular ganglion cell complex imaging with spectral-domain optical coherence tomography in glaucoma. *Investigative ophthalmology & visual science* 2012;53(8):4805-12.
40. Lisboa R, Leite MT, Zangwill LM, et al. Diagnosing Preperimetric Glaucoma with Spectral Domain Optical Coherence Tomography. *Ophthalmology* 2012.
41. 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):1692-9, 99 e1.
42. Hart WM, Jr., Becker B. The onset and evolution of glaucomatous visual field defects. *Ophthalmology* 1982;89(3):268-79.
43. Heijl A, Lundqvist L. The frequency distribution of earliest glaucomatous visual field defects documented by automatic perimetry. *Acta ophthalmologica* 1984;62(4):658-64.
44. Garway-Heath DF, Poinoosawmy D, Fitzke FW, et al. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000;107(10):1809-15.
45. Lee JY, Hwang YH, Lee SM, et al. Age and retinal nerve fiber layer thickness measured by spectral domain optical coherence tomography. *Korean journal of ophthalmology : KJO* 2012;26(3):163-8.
46. Parikh RS, Parikh SR, Sekhar GC, et al. Normal age-related decay of retinal nerve fiber layer thickness. *Ophthalmology* 2007;114(5):921-6.
47. Feuer WJ, Budenz DL, Anderson DR, et al. Topographic differences in the age-related changes in the retinal nerve fiber layer of normal eyes measured by Stratus optical coherence tomography. *Journal of glaucoma* 2011;20(3):133-8.
48. Knight OJ, Girkin CA, Budenz DL, et al. Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by Cirrus HD-OCT. *Arch Ophthalmol* 2012;130(3):312-8.

49. Ooto S, Hangai M, Tomidokoro A, et al. Effects of age, sex, and axial length on the three-dimensional profile of normal macular layer structures. *Investigative ophthalmology & visual science* 2011;52(12):8769-79.
50. Kang SH, Hong SW, Im SK, et al. Effect of myopia on the thickness of the retinal nerve fiber layer measured by Cirrus HD optical coherence tomography. *Investigative ophthalmology & visual science* 2010;51(8):4075-83.
51. Zou H, Zhang X, Xu X, et al. Quantitative in vivo retinal thickness measurement in chinese healthy subjects with retinal thickness analyzer. *Investigative ophthalmology & visual science* 2006;47(1):341-7.
52. Chan CM, Yu JH, Chen LJ, et al. Posterior pole retinal thickness measurements by the retinal thickness analyzer in healthy Chinese subjects. *Retina* 2006;26(2):176-81.
53. Hoh ST, Lim MC, Seah SK, et al. Peripapillary retinal nerve fiber layer thickness variations with myopia. *Ophthalmology* 2006;113(5):773-7.
54. 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 research notes* 2011;4:299.
55. 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):544-53.

	'Glaucomatous' VF group			'Normal' VF group			
	mean	sd	range	mean	sd	range	p value
Age (years)	53.6	13.2	17 - 85	48.5	12.7	17 - 48	< 0.01
MD (dB)	-6.2	5.2	-28.2 - 1.8	-0.5	1.2	-3.6 - 1.3	<0.01
AL (mm)	25.1	1.7	22.2 - 29.3	26.0	1.0	22.8 - 29.5	0.11
m-RNFL (µm)	25.5	7.9	1.0 - 46.6	35.6	5.4	27.5 - 63.1	<0.01
cp-RNFL (µm)	88.3	15.1	49.0 - 123.4	104.0	15.0	66.9 - 150.9	<0.01
GCC (µm)	68.8	15.3	43.7 - 106.5	89.3	19.7	55.7 - 127.3	<0.01
Rim area (mm^2)	1.1	0.5	0.3 - 3.8	1.6	0.6	0.6 - 3.7	<0.01
Eye (right / left)		116 / 108			35 / 34		
Gender (male / female)		108 / 114			40 /31		

Table1

	Measurement
cp-RNFL	total, 4 sectors (superior, temporal, nasal, inferior), 12 sectors
m-RNFL	total, 2 sectors (superior, inferior), 100 sectors
GCL + IPL	total, 2 sectors (superior, inferior), 100 sectors
Optic disc	Disc area, Cup area, Rim area, Cup volume, Rim volume, C/D area ratio, Linear C/D ratio, Vertical C/D ratio, Disc diameter (vertical), Disc diameter (horizontal)
m-RNFL	Significant according to normative database (P < 5%)
m-RNFL	Significant according to normative database (P < 1%)
GCL + IPL	Significant according to normative database (P < 5%)
GCL + IPL	Significant according to normative database (P < 1%)
Age	
Gender	
AL	
Eye (right/left)	

Table2

Title: Cross sectional study: does ~~C~~combining Optical Coherence Tomography measurements using the 'Random Forest' decision tree classifier improve~~s~~ the ~~diagnosis~~prediction of the presence of perimetric deterioration in ~~of~~ glaucoma suspects?

Koichiro Sugimoto[1], Hiroshi Murata[1], Makoto Aihara[1, 2], Chihiro Mayama[1], Ryo Asaoka[1]

Institutions: [1] Department of Ophthalmology, University of Tokyo Graduate School of Medicine, Tokyo , Japan; [2] Shirato Eye Clinic, Tokyo, Japan

Key words: Visual Field, Glaucoma, Optical Coherence Tomography, Random Forest

Word count: 3032~~2987~~

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

Summary

1) Article Focus

- Cross sectional study to investigate the usefulness of ~~I~~interpreting optical coherence tomography parameters simultaneously when predicti~~o~~ng ~~of~~visual field damage in~~discrimination between glaucoma and~~ glaucoma suspects.
- ~~It is beneficial to use a machine learning algorithm of Random Forest to combine the optical coherence tomography parameters.~~

~~-~~2) Key Messages

- ~~It was beneficial to c~~Combining~~e the~~ optical coherence tomography parameters with the Random Forest machine learning method ~~to combine the optical coherence tomography parameters for~~ improves the prediction of visual field damage in~~discriminating between glaucoma and~~ glaucoma suspects.

3) Strengths and Limitations

- Strengths: This method could be used to improve the clinical management of glaucoma suspects and glaucoma patients.
- Limitations: No inclusion of normal subjects.

Abstract

Purpose

To develop a classifier to ~~diagnose~~ predict the presence of visual field (VF) deterioration in glaucoma suspects ~~glaucoma~~ based on optical coherence tomography (OCT) measurements ~~of these structures~~ using the machine learning method known as the 'Random Forest' algorithm.

Methods

Spectral domain OCT~~optical coherence tomography~~ (Topcon 3D OCT-2000) and perimetry (Humphrey Field Analyzer, 24-2 or 30-2 SITA standard) measurements were conducted in 293 eyes of 179 subjects with open angle glaucoma (OAG) or suspected OAG. ~~Visual field (~~VF~~)~~ damage (OHTS criteria (2002)) was used as a 'gold-standard' to classify glaucomatous eyes. The 'Random Forest' method was then used to analyze the relationship between the presence/absence of glaucomatous VF damage and the following variables: age, gender, right or left eye, axial length plus 237 different OCT measurements. The area under the receiver operating characteristic curve (AROC) was then derived using the probability of glaucoma as suggested by the proportion of votes in the Random Forest classifier. For comparison, five AROCs were derived based on: (i) macular retinal nerve fiber layer (m-RNFL) alone, (ii) circumpapillary (cp-RNFL) alone, (iii) ganglion cell layer and inner plexiform layer (GCL + IPL) alone, (iv) rim area alone, and (v) a decision tree method using the same variables as the Random Forest algorithm.

Results

The AROC from the combined Random Forest classifier (0.90) was significantly larger than the AROCs based on individual measurements of m-RNFL (0.86), cp-RNFL (0.77), GCL + IPL (0.80), rim area (0.78), and the decision tree method (0.75) ($p < 0.05$).

Conclusions

Evaluating OCT measurements using the Random Forest method provides an accurate ~~diagnosis~~ prediction of the presence of perimetric deterioration in glaucoma suspects~~of glaucoma~~.

Introduction

Glaucoma is the second most common cause of blindness. As glaucomatous visual field (VF) damage is irreversible, early diagnosis of glaucoma is essential. Structural changes at the optic nerve head[1] and retinal nerve fiber layer (RNFL) around the optic disc[2] can also indicate glaucomatous damage, and may precede measurable VF loss.

Optical coherence tomography (OCT) is an imaging technology widely used in the diagnosis of glaucoma, enabling high-resolution measurements of the retina. [3] The recent advancement of OCT from the time domain to the spectral domain (SD-OCT) has greatly improved the imaging speed and resolution of the device, [4] and has enabled imaging scans of the macular retinal nerve fiber layer (m-RNFL) and the macular ganglion cell layer and inner plexiform layer (GCL + IPL). It has been reported that these retinal layers are damaged early in the glaucoma disease process[5 6] and many studies have investigated the diagnostic performance of thickness measurements of these structures to discriminate between healthy and glaucomatous eyes[7-14]. However, in these previous studies, the different measurements were interpreted independently, yet damage to these structures does not necessarily occur in parallel [15 16] and thus there is no consensus on which structure is optimum for diagnosing glaucoma. Indeed, specific structures may be preferentially damaged in any given patient. For example, Cordeiro et al. reported that the diagnostic performance of cp-RNFL thickness measurements tended to be better in patients with a small optic disc, and an inverse effect was observed using the GCC measurement. [17] Conversely, GCC may be preferential to detect glaucomatous change in high myopic patients.[18] Thus, it appears that no single structural measurement is best for diagnosing glaucoma.

~~The purpose of this study was to improve the structural diagnosis of glaucoma (using VF damage as a gold-standard classifier) by analyzing multiple OCT measurements concurrently using the machine learning method known as the 'Random Forest' algorithm.~~

The 'Random Forest' method is a decision support tool which

consists of many decision trees. Decision trees have previously been used to diagnose glaucoma[19]; however, decision trees suffer from the problem of 'over-fitting', which influences the diagnostic accuracy. [20] In contrary, the Random Forest classifier overcomes this problem by summarizing the results of many decision trees. Another noteworthy advantage of the Random Forest algorithm over traditional methods, such as logistic regression, is that any interaction or correlation between variables does not adversely affect the classification since it is capable of representing high order interactions. [21] Furthermore, predictors that might otherwise be masked by their correlation with other variables, using other classification methods, can contribute to the Random Forest classifier. ~~In this study we have employed the Random Forest algorithm to explore multiple OCT parameters concurrently in order to build an unbiased glaucoma classifier.~~

Glaucomatous structural change is often apparent in glaucomatous patients without VF defects (pre-perimetric glaucoma)[22] [23]. Therefore, it may be possible to predict the presence of the VF deterioration from structural measurements, as it has been reported that there is a significant difference in structural measurements between patients with perimetric and pre-perimetric glaucoma. [22] Predicting the presence of VF damage from structural measurements is clinically very important, especially in patients who cannot reliably perform VF test, due to, for example, inability to concentrate, mental disorders, locomotor disabilities, and so on. The purpose of this study was to improve the prediction of the presence of VF damage in glaucoma suspects by analyzing multiple OCT measurements concurrently using the Random Forest algorithm.

Materials and Methods

The study was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine at the University of Tokyo. 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.

This retrospective study comprised 293 eyes of 179 consecutive patients referred to the University of Tokyo Hospital for glaucoma or suspected glaucoma between August 2010 and July 2012. Patients were referred based on optic disc damage: focal or diffuse neuroretinal rim thinning, localized notching or nerve fiber layer defects. Subjects underwent complete ophthalmic examinations, including slit lamp biomicroscopy, gonioscopy, intraocular pressure (IOP) measurement and funduscopy. If glaucomatous structural changes were confirmed from these tests, axial length (AL~~) measurement (;~~ IOL Master, Carl Zeiss Meditec, Dublin, CA), ~~as well~~ imaging with SD-OCT and VF testing were performed. Criteria for inclusion were visual acuity better than 6/12; no previous ocular surgery, except cataract extraction and intraocular lens implantation; open anterior chamber angle (patients with angle closure glaucoma and secondary open angle glaucoma were excluded); no other anterior and posterior segment eye disease. AL was not used for the inclusion / exclusion criteria.

VF testing was performed using the Humphrey Field Analyzer (HFA, Carl Zeiss Meditec); 24-2 or 30-2 test pattern and the SITA Standard strategy, with the Goldmann size III target. Near refractive correction was used as necessary, calculated according to the subject's age by the HFA software. Unreliable VFs were excluded according to HFA criteria (fixation losses greater than 25%, or false-positive responses greater than 15%). False negative rate was not used as an indicator of test reliability following a previous report [24]. A glaucomatous VF was defined as a pattern standard deviation (PSD) value beyond the normal limit ($P < 0.05$), or a Glaucoma Hemifield Test (GHT) result outside normal limits following the criteria in [25]. All

glaucoma patients had previous experience in visual field testing.

SD-OCT (3D OCT-2000; Topcon Corp., Tokyo, Japan) was used to obtain tomographic images of the parapapillary fundus with the 3D Disc scan and 3D Macula scan (128 horizontal scan lines comprised of 512 A-scans for an image area of 6×6 mm). SD-OCT uses a superluminescent diode laser with a center wavelength of 840nm and a bandwidth of 50nm as the light source. The transverse and axial resolutions are less than 20μm and 5μm, respectively. The acquisition speed is 50,000 A scans per second. In the selected eye, the macula was imaged by 6 radial lines centered at the fovea spaced 30° apart. All of the measurements were performed after pupil dilation with 1% Tropicamide and all of the images had signal strength of at least 60, as recommended by the manufacturer.

The 'Random Forest' algorithm is an ensemble machine learning classifier proposed by Breiman in 2001. [26 27] The Random Forest consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. Thus the Random Forest is an ensemble classifier, which has been reported to improve the prediction accuracy of decision tree.[28] Indeed there are many reports which suggested the Random Forest gives best prediction accuracy among various machine learning methods and this method has been used in many research fields, including gene selection and cancer classification.[29-32] In the random Forest method, when classifying a new object from an input vector, the input vector is classified by each of the trees in the forest, and the tree "votes" for that class. The forest then chooses the classification having the most votes over all the trees in the forest. Each tree is constructed using a different bootstrap sample from the original data. Thus, cross-validation is performed internally and there is no need for a separate cross-validation data set to obtain an unbiased estimate of the test set error. For classification, node impurity was measured using the Gini index [33].

The Random Forest method was used to classify the presence or absence of glaucomatous VF damage using: OCT measurements (237

different measurements in total were analyzed), age, gender, AL and right/left eye (see **Table 2)**. In this procedure, 10,000 trees were grown and five among the 241 parameters were used at each node. The area under the receiver operating characteristic curve (AROC) was derived from the probability of glaucoma (the proportion of votes) as suggested by the method; for each individual, only the data from all other subjects (n=178) was used (leave-one-out cross validation) so that right and left eyes of a subject are not used for both training and testing simultaneously. For comparison, the AROCs were also derived using only individual raw thickness measurements of: m-RNFL, or cp-RNFL, or GCL + IPL, or rim area and the prediction with the decision tree method. The diagnostic sensitivity and specificity was also calculated for the age-matched normative limits of the different measurements (P ≤ %5, or, P ≤ 1%): m-RNFL, and GCL + IPL, as shown on the instrument's print out.

Finally, variable importance was calculated by randomly permuting a variable at each decision tree and observing whether the number of correct decisions decreased[27].

All statistical analyses were carried out using the statistical programming language R (ver. 2.14.2, The R Foundation for Statistical Computing, Vienna, Austria) and Medcalc version 11.4.2.0; MedCalc statistical software, Mariakerke, Belgium). The R package "randomForset" and "rpart" was used to carry out the analysis of the Random Forest method and decision tree method, respectively.

Results

Subject characteristics are given in **Table 1**. VFs of 224 eyes in 150 patients were diagnosed as glaucomatous while the remaining 69 eyes of 57 patients were judged as normal. The average total m-RNFL thickness, cp-RNFL thickness, GCL + IPL thickness and rim area were significantly smaller in the glaucomatous group compared with the normal group ($P < 0.05$, non-paired t-test).

As shown in the **Figure 1**, The AROC of the Random Forest method utilizing all measurements (0.90) was significantly larger than that with m-RNFL alone (0.86), cp-RNFL alone (0.77), GCL-IPL (0.80) and rim area alone (0.78) ($p < 0.05$). Furthermore, the diagnostic performance (sensitivity and specificity) of the age-matched normative database (as shown on the OCT printout) were also plotted in **Figure 1**. The sensitivity and specificity for thickness values outside normal limits were: m-RNFL ($P < 5\%$): 0.74 and 0.93; m-RNFL ($P < 1\%$): 0.61 and 0.96; GCL-IPL ($P < 5\%$): 0.48 and 0.88; GCL + IPL ($P < 1\%$): 0.42 and 0.90 (sensitivity and specificity, respectively).

Figure 2 illustrates the OCT measurements analyzed. Among 237 measurements, 76 had a significant variable importance measure including: total and inferior m-RNFL thickness, total and inferior GCL + IPL thickness, an m-RNFL thickness value outside normal limits ($P < 5\%$), various sectorial m-RNFL thickness values (**Figure 2a**), various GCL + IPL thickness values (**Figure 2b**), and two cp-RNFL thickness values (**Figure 2c**). Age, AL, gender, and right or left eye were not significant.

Discussion

In the current study, the 'Random Forest' decision tree classifier was used to predict the presence of VF damage in glaucoma suspects. As a result, it was shown that the AROC given by the Random Forest method was significantly larger than those derived from any single OCT parameter and the simple decision tree method.

~~Glaucoma can be diagnosed and monitored using structural measurements from OCT; to date, cp-RNFL thickness measurements have generally been used to quantify glaucomatous damage using time-domain OCT [27-31]. Another "traditional" way to measure structural glaucomatous damage is to use scanning laser tomography (HRT; Heidelberg retina tomography, Heidelberg engineering, Heidelberg, Germany) to measure characteristics of the optic disc such as size and shape. HRT works on the principle of confocal scanning laser ophthalmoscopy, and is long established as a diagnostic tool for glaucoma. HRT measurements of rim area, among the various optic disc parameters, have been reported as most clinically meaningful, repeatable and reliable[32-34].~~

~~The development of SD-OCT has improved the scan speed, resolution[35] and repeatability[36] of captured images. These improvements may, in turn, strengthen the association between structural SD-OCT measurements and functional VF measurements in glaucoma patients, which is referred to as the 'structure-function' relationship [37, 38]. The advent of SD-OCT has also enabled the measurement of the m-RNFL and GCL + IPL layers. Recent studies have investigated the structure-function relationship in glaucoma patients using measurements of the macular ganglion cell complex (GCC), which is the total thickness of the GCL + IPL and m-RNFL layers[7, 11, 13, 14, 39-42]. This research suggests that structural measurements of the cp-RNFL layer, or the GCC, give rise to an analogous structure-function relationship and diagnostic ability to detect glaucoma, and there is no consensus on which structure is optimum for diagnosing glaucoma. Indeed, specific structures may be preferentially damaged in any given patient. For example, Cordeiro et al. reported that the diagnostic~~

~~performance of cp-RNFL thickness measurements tended to be better in patients with a small optic disc, and an inverse effect was observed using the GCC measurement.~~ [43] ~~Conversely, GCC may be preferential to detect glaucomatous change in high myopic patients.~~ [44] ~~Thus, it appears that no single structural measurement is best for diagnosing glaucoma.~~

Previous attempts have been made to interpret multiple structural parameters in order to aid the diagnosis of glaucoma. Chen et al. used a logistical diagnostic model to diagnose glaucoma; the model analyzed a patient's optic cup:optic disc vertical ratio, cp-RNFL thickness and rim area simultaneously, but the authors found that diagnostic performance was not significantly improved compared with using individual measurements. [34] On the other hand, Burgansky-Eliash et al. used a support vector machine classifier of multiple Stratus OCT parameters to diagnose glaucoma, and showed that the AROC was significantly larger. [35] Other studies also support combining multiple structural measurements to diagnose glaucoma. [36 37] ~~Also~~ In addition, a recent study suggested the decision tree method is useful to discriminate between glaucoma patients and normal subjects.[19] However, in the current study, the decision tree method, which often suffers from the problem of over-fitting[38], failed to show benefit in discriminating glaucoma~~and glaucoma subject~~. On the other hand, ~~However~~ it was beneficial to use the Random Forest method, which is an ensemble classifier of decision trees. Recent reports have revealed that distinguishing between perimetric glaucoma and pre-perimetric glaucoma is more difficult than differentiating normal subjects from glaucoma patients[39] with early VF damage[22]. ~~studies which suggested support the merit of combining multiple structural measurements to diagnose glaucoma,~~ [38 39] ~~yet~~ A noteworthy advantage of the current study is that it is the first of its kind to analyze~~d the~~ m-RNFL and GCL + IPL layers simultaneously with cp-RNFL, optic disc shape parameters as well age and AL.

It must be noted that a clear caveat of the current study is the lack of a normative population to act as a reference. Therefore, AROCs

derived in the current study are not directly relevant to distinguishing between healthy subjects and glaucomatous patients. A further study should be carried out with normative and glaucomatous populations (particularly patients with early stage glaucoma) in order to further investigate the merits of the Random Forest classifier. Nonetheless, the method's ability to accurately differentiate glaucoma suspects from glaucoma patients suggests that the classifier may be even more useful in this context.

~~A merit of the Random Forest method is that the importance of each parameter for its classification can be tested.~~ The variable importance measure from the Random Forest method suggested that total m-RNFL thickness, total GCL + IPL thickness, and m-RNFL thickness outside normal limits (P < 5%) significantly contributed to the diagnosis of glaucoma. In contrast, age, AL, gender, eye (right / left), and optic disc measurements such as rim area, were not significant. Reports have suggested that optic disc shape parameters are useful for classifying glaucomatous eyes, but are less useful compared to RNFL parameters[16 40]. However, previous results have been based on HRT measurements of the optic disc, and there are notable differences between the corresponding measurements in SD-OCT. For instance, the margin of the optic disc and cup is automatically identified in SD-OCT, whereas it is manually drawn by the examiner in HRT. Furthermore, it has been reported that HRT measurements of optic disc shape detect a different population of glaucoma patients to OCT measurements of the RNFL[16]. Accordingly, the diagnostic performance of the Random Forest classifier may be further improved by also including various optic disc shape parameters derived from HRT. We intend to investigate this hypothesis in a future study.

Interestingly, our results question the validity of SD-OCT's normal limits to discriminate glaucoma. For example, the blue cross in **Figure 1** indicates that GCL + IPL measurements outside normal limits at the P < 1% level have a specificity of 90%. The normal limits of the SD-OCT are derived by testing 'normal' subjects without ocular disease; Rao et al. have reported that cp-RNFL thickness measurements from normal

subjects and patients with glaucoma overlap considerably [41]. A significant advantage of the Random Forest classifier is that normal limits could be established based on results from normal subjects and glaucoma patients; these would be expected to better reflect the 'true' specificity of the test result. Another merit of the Random Forest method, in comparison to the current standard, is that the method gives an exact probability of glaucoma, rather than a binary classification (glaucoma or not at P < 1%, or P < 5%); such a value could be interpreted in a manner similar to that of the 'Nerve Fiber Index' (NFI) score in the nerve fibre analyzer imaging instrument (GDx, Carl Zeiss Meditec), which is a continuous numeric score from 0 to 99.

In our Random Forest classifier, many sectorial thickness measurements of the m-RNFL, GCL + IPL and cp-RNFL layers were deemed significant for the ~~diagnosis~~ prediction of glaucomatous VF damage. Significant sectors were generally located in the inferior hemi-retina, although a few sectors were also situated in the superior hemi-retina (see **Figure 2**). Previous studies have suggested that glaucomatous VF damage preferably affects the superior hemifield[42 43]. Interestingly, the significant m-RNFL, GCL + IPL and cp-RNFL sectors in our classifier were principally distributed along the inferotemporal RNFL bundle, which likely corresponds to an arcuate defect in the superior VF[44]. Thus, these results also suggest that glaucomatous RNFL / GCL + IPL damage tends to occur in the inferior hemi-retina.

OCT structural measurements are influenced by ageing; cp-RNFL [45-47], rim area [48], m-RNFL, and GCL + IPL all become thinner with age [49]. In addition, studies suggest that AL may have an effect on measurements of the cp-RNFL[48 50], rim area [48 50], m-RNFL [49], and GCL + IPL [49]; however any such effects remain contentious[51-53]. In our study, removing age and AL factors did not affect the AROC of the Random Forest classifier.

Other machine learning methods, such as support vector machines, boosting and bagging classifiers could also be used to diagnose glaucoma. Previous reports suggest that the Random Forest

method outperforms most other methods [31 54 55]; hence the Random Forest algorithm was used in the current study. Nevertheless, in a future study, we intend to investigate the performance of machine learning methods for discriminating perimetric and preperimetric glaucoma.

In conclusion, we have shown that combining SD-OCT measurements of the m-RNFL, cp-RNFL, GCL + IPL layers, using the Random Forest method, is beneficial for ~~diagnosing~~ predicting the presence of glaucomatous VF damage in glaucoma suspects, especially when compared with the current OCT reference-standard of comparing these measurements to an age-matched normative database.

Figure and Table Legends

Table 1

Characteristics of the study participants.

MD: mean deviation, m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Table 2

The variables used in the analysis, including 237 optical coherence tomography parameters.

m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Figure 1

Receiver operating characteristic (ROC) curves with the probability of glaucoma suggested by the Random Forest classifier and raw thickness measurements of: m-RNFL alone, cp-RNFL alone, and GCL + IPL alone, and decision tree method.

The area under the ROC with the Random Forest method was significantly larger than those of individual measurements and decision tree method ($P < 0.05$). The colored "X" represent the sensitivity and specificity of the SD-OCT normative database (red: m-RNFL ($P < 5\%$), orange: m-RNFL ($P < 1\%$), green: GCC ($P< 5\%$), blue: GCC ($P < 1\%$)).

m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

Figure 2

Variables in the Random Forest classifier having a significant effect on the ~~diagnosis~~ presence of glaucomatous visual field damage.

Sectors of the cp-RNFL, m-RNFL, and GCL + IPL were superimposed onto a fundus photograph[44]; significant sectors are highlighted in red. If a subject's left eye was tested, the recorded data were mapped to a right eye format for analysis. Figure 2a: cp-RNFL, Figure 2b: m-RNFL,

Figure 2c: GCL-IPL
m-RNFL: macular retinal nerve fiber layer (RNFL), cp-RNFL: circumpapillary RNFL, GCL + IPL: ganglion cell layer and inner plexiform layer, and AL: axial length

References

1. Quigley HA, Katz J, Derick RJ, Gilbert D, Sommer A. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. *Ophthalmology* 1992;99(1):19-28.
2. Sommer A, Katz J, Quigley HA, Miller NR, Robin AL, Richter RC, et al. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol* 1991;109(1):77-83.
3. Chang R, Budenz DL. New developments in optical coherence tomography for glaucoma. *Current opinion in ophthalmology* 2008;19(2):127-35.
4. Huang D, Swanson EA, Lin CP, Schuman JS, Stinson WG, Chang W, et al. Optical coherence tomography. *Science* 1991;254(5035):1178-81.
5. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *American journal of ophthalmology* 1989;107(5):453-64.
6. Nakano N, Ikeda HO, Hangai M, Muraoka Y, Toda Y, Kakizuka A, et al. Longitudinal and simultaneous imaging of retinal ganglion cells and inner retinal layers in a mouse model of glaucoma induced by N-methyl-D-aspartate. *Investigative ophthalmology & visual science* 2011;52(12):8754-62.
7. Cho JW, Sung KR, Lee S, Yun SC, Kang SY, Choi J, et al. Relationship between visual field sensitivity and macular ganglion cell complex thickness as measured by spectral-domain optical coherence tomography. *Investigative ophthalmology & visual science* 2010;51(12):6401-7.
8. 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* 2011;25(1):57-65.
9. Kim NR, Lee ES, Seong GJ, Kim JH, An HG, Kim CY. Structure-function relationship and diagnostic value of macular ganglion cell complex measurement using Fourier-domain OCT in glaucoma. *Investigative ophthalmology & visual science* 2010;51(9):4646-51.
10. Moreno PA, Konno B, Lima VC, Castro DP, Castro LC, Leite MT, et al. Spectral-domain optical coherence tomography for early glaucoma assessment: analysis of macular ganglion cell complex versus peripapillary retinal nerve fiber layer. *Canadian journal of ophthalmology. Journal canadien d'ophtalmologie* 2011;46(6):543-7.
11. Rao HL, Babu JG, Addepalli UK, Senthil S, Garudadri CS. Retinal nerve fiber layer

and macular inner retina measurements by spectral domain optical coherence tomograph in Indian eyes with early glaucoma. *Eye* 2012;26(1):133-9.

12. Rao HL, Kumbar T, Addepalli UK, Bharti N, Senthil S, Choudhari NS, et al. Effect of spectrum bias on the diagnostic accuracy of spectral-domain optical coherence tomography in glaucoma. *Investigative ophthalmology & visual science* 2012;53(2):1058-65.
13. Schulze A, Lamparter J, Pfeiffer N, Berisha F, Schmidtmann I, Hoffmann EM. Diagnostic ability of retinal ganglion cell complex, retinal nerve fiber layer, and optic nerve head measurements by Fourier-domain optical coherence tomography. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 2011;249(7):1039-45.
14. Tan O, Chopra V, Lu AT, Schuman JS, Ishikawa H, Wollstein G, et al. Detection of macular ganglion cell loss in glaucoma by Fourier-domain optical coherence tomography. *Ophthalmology* 2009;116(12):2305-14 e1-2.
15. Tuulonen A, Lehtola J, Airaksinen PJ. Nerve fiber layer defects with normal visual fields. Do normal optic disc and normal visual field indicate absence of glaucomatous abnormality? *Ophthalmology* 1993;100(5):587-97; discussion 97-8.
16. Leung CK, Choi N, Weinreb RN, Liu S, Ye C, Liu L, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: pattern of RNFL defects in glaucoma. *Ophthalmology* 2010;117(12):2337-44.
17. Cordeiro DV, Lima VC, Castro DP, Castro LC, Pacheco MA, Lee JM, et al. Influence of optic disc size on the diagnostic performance of macular ganglion cell complex and peripapillary retinal nerve fiber layer analyses in glaucoma. *Clinical ophthalmology* 2011;5:1333-7.
18. Shoji T, Nagaoka Y, Sato H, Chihara E. Impact of high myopia on the performance of SD-OCT parameters to detect glaucoma. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 2012.
19. Baskaran M, Ong EL, Li JL, Cheung CY, Chen D, Perera SA, et al. Classification algorithms enhance the discrimination of glaucoma from normal eyes using high-definition optical coherence tomography. *Investigative ophthalmology & visual science* 2012;53(4):2314-20.
20. Mitchell T. *MACHINE LEARNING*. NY: McGraw-Hill Higher Education, 1997.
21. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable

importance for random forests. *BMC bioinformatics* 2008;9:307.

22. Cvenkel B, Kontestabile AS. Correlation between nerve fibre layer thickness measured with spectral domain OCT and visual field in patients with different stages of glaucoma. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* 2011;249(4):575-84.
23. Jeoung JW, Park KH. Comparison of Cirrus OCT and Stratus OCT on the ability to detect localized retinal nerve fiber layer defects in preperimetric glaucoma. *Investigative ophthalmology & visual science* 2010;51(2):938-45.
24. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Investigative ophthalmology & visual science* 2000;41(8):2201-4.
25. Gordon MO, Beiser JA, Brandt JD, Heuer DK, Higginbotham EJ, Johnson CA, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* 2002;120(6):714-20; discussion 829-30.
26. Breiman L. Random Forests. *Machine Learning* 2001;45:5-32.
27. Breiman L, Cutler A. Random Forests, 2004.
28. Dietterich TG. *Ensemble learning. In The Handbook of Brain Theory and Neural Networks, 2nd ed*. Cambridge: The MIT Press, 2002.
29. Palmer DS, O'Boyle NM, Glen RC, Mitchell JB. Random forest models to predict aqueous solubility. *Journal of chemical information and modeling* 2007;47(1):150-8.
30. Wu B, Abbott T, Fishman D, McMurray W, Mor G, Stone K, et al. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics* 2003;19(13):1636-43.
31. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. *BMC bioinformatics* 2006;7:3.
32. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences* 2003;43(6):1947-58.
33. Gini C. 1909 Concentration and dependency ratios (in Italian). *English translation in Rivista di Politica Economica* 1997;87:769-89.
34. Fang Y, Pan YZ, Li M, Qiao RH, Cai Y. Diagnostic capability of Fourier-Domain optical coherence tomography in early primary open angle glaucoma. *Chinese*

medical journal 2010;123(15):2045-50.

35. Burgansky-Eliash Z, Wollstein G, Chu T, Ramsey JD, Glymour C, Noecker RJ, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Investigative ophthalmology & visual science* 2005;46(11):4147-52.
36. Lu AT, Wang M, Varma R, Schuman JS, Greenfield DS, Smith SD, et al. Combining nerve fiber layer parameters to optimize glaucoma diagnosis with optical coherence tomography. *Ophthalmology* 2008;115(8):1352-7, 57 e1-2.
37. Chen HY, Huang ML, Hung PT. Logistic regression analysis for glaucoma diagnosis using Stratus Optical Coherence Tomography. *Optometry and vision science : official publication of the American Academy of Optometry* 2006;83(7):527-34.
38. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer, 2001.
39. Morooka S, Hangai M, Nukada M, Nakano N, Takayama K, Kimura Y, et al. Wide 3-dimensional macular ganglion cell complex imaging with spectral-domain optical coherence tomography in glaucoma. *Investigative ophthalmology & visual science* 2012;53(8):4805-12.
40. Lisboa R, Leite MT, Zangwill LM, Tafreshi A, Weinreb RN, Medeiros FA. Diagnosing Preperimetric Glaucoma with Spectral Domain Optical Coherence Tomography. *Ophthalmology* 2012.
41. Rao HL, Zangwill LM, Weinreb RN, Sample PA, Alencar LM, Medeiros FA. Comparison of different spectral domain optical coherence tomography scanning areas for glaucoma diagnosis. *Ophthalmology* 2010;117(9):1692-9, 99 e1.
42. Hart WM, Jr., Becker B. The onset and evolution of glaucomatous visual field defects. *Ophthalmology* 1982;89(3):268-79.
43. Heijl A, Lundqvist L. The frequency distribution of earliest glaucomatous visual field defects documented by automatic perimetry. *Acta ophthalmologica* 1984;62(4):658-64.
44. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000;107(10):1809-15.
45. Lee JY, Hwang YH, Lee SM, Kim YY. Age and retinal nerve fiber layer thickness measured by spectral domain optical coherence tomography. *Korean journal of ophthalmology : KJO* 2012;26(3):163-8.

46. Parikh RS, Parikh SR, Sekhar GC, Prabakaran S, Babu JG, Thomas R. Normal age-related decay of retinal nerve fiber layer thickness. *Ophthalmology* 2007;114(5):921-6.
47. Feuer WJ, Budenz DL, Anderson DR, Cantor L, Greenfield DS, Savell J, et al. Topographic differences in the age-related changes in the retinal nerve fiber layer of normal eyes measured by Stratus optical coherence tomography. *Journal of glaucoma* 2011;20(3):133-8.
48. Knight OJ, Girkin CA, Budenz DL, Durbin MK, Feuer WJ, Cirrus OCTNDSG. Effect of race, age, and axial length on optic nerve head parameters and retinal nerve fiber layer thickness measured by Cirrus HD-OCT. *Arch Ophthalmol* 2012;130(3):312-8.
49. Ooto S, Hangai M, Tomidokoro A, Saito H, Araie M, Otani T, et al. Effects of age, sex, and axial length on the three-dimensional profile of normal macular layer structures. *Investigative ophthalmology & visual science* 2011;52(12):8769-79.
50. Kang SH, Hong SW, Im SK, Lee SH, Ahn MD. Effect of myopia on the thickness of the retinal nerve fiber layer measured by Cirrus HD optical coherence tomography. *Investigative ophthalmology & visual science* 2010;51(8):4075-83.
51. Zou H, Zhang X, Xu X, Yu S. Quantitative in vivo retinal thickness measurement in chinese healthy subjects with retinal thickness analyzer. *Investigative ophthalmology & visual science* 2006;47(1):341-7.
52. Chan CM, Yu JH, Chen LJ, Huang CH, Lee CT, Lin TC, et al. Posterior pole retinal thickness measurements by the retinal thickness analyzer in healthy Chinese subjects. *Retina* 2006;26(2):176-81.
53. Hoh ST, Lim MC, Seah SK, Lim AT, Chew SJ, Foster PJ, et al. Peripapillary retinal nerve fiber layer thickness variations with myopia. *Ophthalmology* 2006;113(5):773-7.
54. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, de Mendonca A. 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 research notes* 2011;4:299.
55. Douglas PK, Harris S, Yuille A, Cohen MS. Performance comparison of machine learning algorithms and number of independent components used in fMRI decoding of belief vs. disbelief. *NeuroImage* 2011;56(2):544-53.

	'Glaucomatous' VF group			'Normal' VF group			
	mean	sd	range	mean	sd	range	p value
Age (years)	53.6	13.2	17 - 85	48.5	12.7	17 - 48	< 0.01
MD (dB)	-6.2	5.2	-28.2 - 1.8	-0.5	1.2	-3.6 - 1.3	<0.01
AL (mm)	25.1	1.7	22.2 - 29.3	26.0	1.0	22.8 - 29.5	0.11
m-RNFL (µm)	25.5	7.9	1.0 - 46.6	35.6	5.4	27.5 - 63.1	<0.01
cp-RNFL (µm)	88.3	15.1	49.0 - 123.4	104.0	15.0	66.9 - 150.9	<0.01
GCC (µm)	68.8	15.3	43.7 - 106.5	89.3	19.7	55.7 - 127.3	<0.01
Rim area (mm^2)	1.1	0.5	0.3 - 3.8	1.6	0.6	0.6 - 3.7	<0.01
Eye (right / left)		116 / 108			35 / 34		
Gender (male / female)		108 / 114			40 /31		

Table1

	Measurement
cp-RNFL	total, 4 sectors (superior, temporal, nasal, inferior), 12 sectors
m-RNFL	total, 2 sectors (superior, inferior), 100 sectors
GCL + IPL	total, 2 sectors (superior, inferior), 100 sectors
Optic disc	Disc area, Cup area, Rim area, Cup volume, Rim volume, C/D area ratio, Linear C/D ratio, Vertical C/D ratio, Disc diameter (vertical), Disc diameter (horizontal)
m-RNFL	Significant according to normative database (P < 5%)
m-RNFL	Significant according to normative database (P < 1%)
GCL + IPL	Significant according to normative database (P < 5%)
GCL + IPL	Significant according to normative database (P < 1%)
Age	
Gender	
AL	
Eye (right/left)	

Table2

Formatted: Font: (Default) Verdana, 10 pt

Formatted: Font: (Default) Verdana, 10 pt

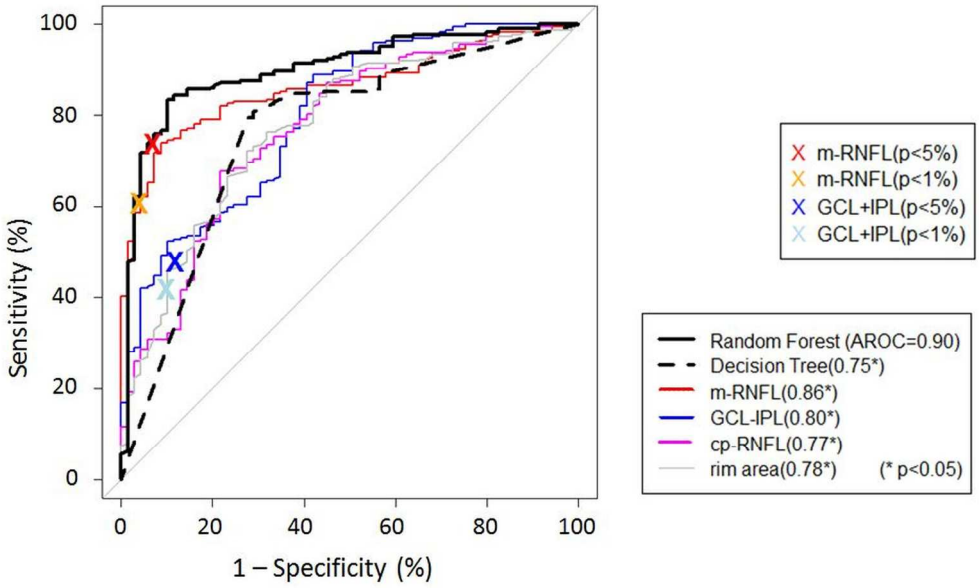
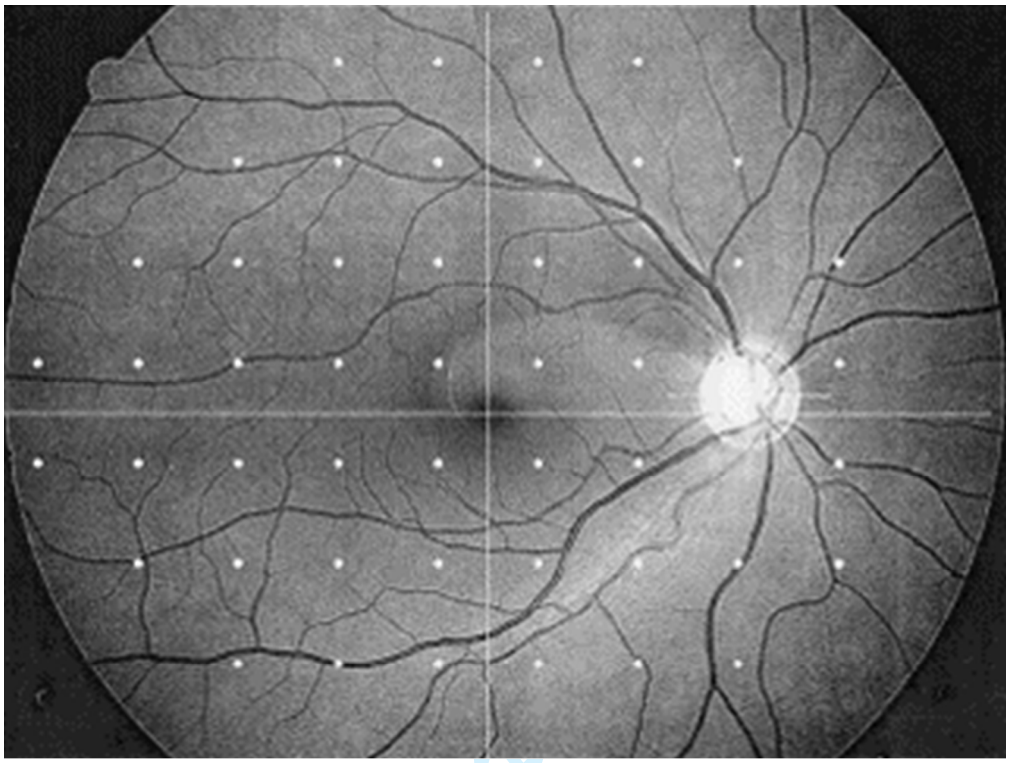
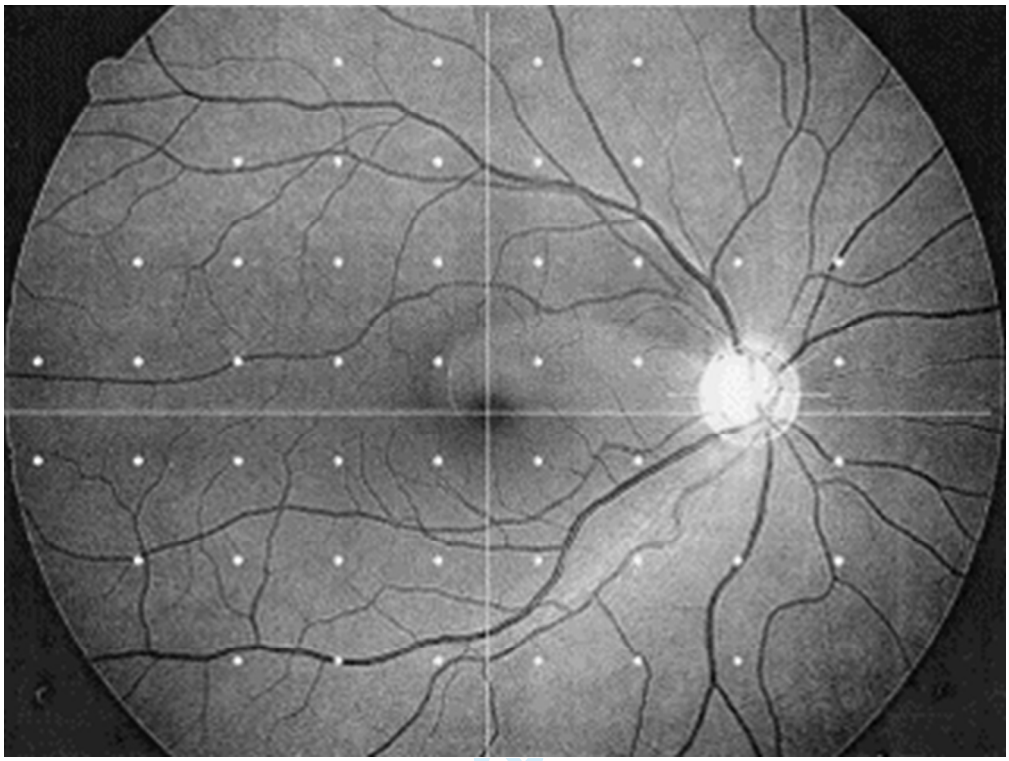
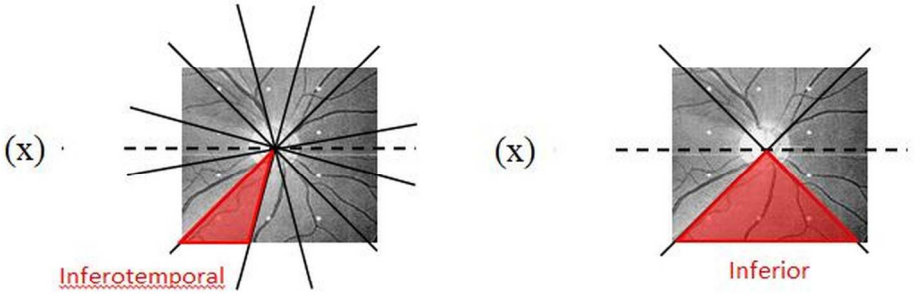


Figure 1

252x167mm (300 x 300 DPI)







210x90mm (300 x 300 DPI)

STARD checklist for reporting of studies of diagnostic accuracy

(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	1
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	4
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	6
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	6
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	6
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	6
Test methods	7	The reference standard and its rationale.	7, 8
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	7, 8
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	~~4~~9
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	N/A
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	N/A
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	7, 8
	13	Methods for calculating test reproducibility, if done.	N/A
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	6
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	6
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	~~3~~6
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	N/A
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	Table 1
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	Figure 1
	20	Any adverse events from performing the index tests or the reference standard.	N/A
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	9
	22	How indeterminate results, missing data and outliers of the index tests were handled.	~~N/A~~7
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	N/A
	24	Estimates of test reproducibility, if done.	N/A
DISCUSSION	25	Discuss the clinical applicability of the study findings.	10~~7~~-13~~9~~