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 Hiroyo Hirasawa,1 Makoto Aihara,1,2 Chihiro Mayama,1 Ryo Asaoka1

ABSTRACT

Objectives: 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.

Design: Case–control study.

Participants: 293 eyes of 179 participants with open angle glaucoma (OAG) or suspected OAG.

Interventions: Spectral domain OCT (Topcon 3D OCT-2000) and perimetry (Humphrey Field Analyser, 24-2 or 30-2 SITA standard) measurements were conducted in all of the participants. VF damage (Ocular Hypertension Treatment Study criteria (2002)) was used as a ‘gold-standard’ to classify glaucomatous eyes. The ‘Random Forest’ method was then used to analyse 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.

Main outcome measures: 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: (1) macular retinal nerve fibre layer (m-RNFL) alone; (2) circumpapillary (cp-RNFL) alone; (3) ganglion cell layer and inner plexiform layer (GCL+IPL) alone; (4) rim area alone and (5) 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.

SUMMARY

Strengths and limitations of this study
• Combining optical coherence tomography measurements.
• Accurate prediction of the presence of perimetric deterioration in glaucoma suspects.
• Lack of a normative population to act as a reference.

INTRODUCTION

Glaucoma is the second most common cause of blindness. As glaucomatous visual field (VF) damage is irreversible, the early diagnosis of glaucoma is essential. Structural changes at the optic nerve head1 and retinal nerve fibre layer (RNFL) around the optic disc2 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 OCT (SD-OCT) has greatly improved the imaging speed and resolution of the device,4 and has enabled imaging scans of the macular RNFL (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 process5 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
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.\textsuperscript{5} reported that the diagnostic performance of circumpapillary RNFL (cp-RNFL) thickness measurements tended to be better in patients with a small optic disc, and an inverse effect was observed using the macular ganglion cell complex (GCC) measurement. Conversely, GCC may be preferential to detect glaucomatous change in patients with high myopia.\textsuperscript{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\textsuperscript{19}; however, decision trees suffer from the problem of ‘over-fitting’, which influences the diagnostic accuracy.\textsuperscript{20} On the contrary, the Random Forest classifier overcomes this problem by summarising 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.\textsuperscript{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 patients with glaucoma without VF defects (preperimetric glaucoma).\textsuperscript{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 preperimetric glaucoma.\textsuperscript{22} Predicting the presence of VF damage from structural measurements is clinically very important, especially in patients who cannot reliably perform VF test, for example, due to inability to concentrate, mental disorders, locomotor disabilities, etc. The purpose of this study was to improve the prediction of the presence of VF damage in glaucoma suspects by analysing multiple OCT measurements concurrently using the Random Forest algorithm.

The patients underwent complete ophthalmic examinations, including slit lamp biomicroscopy, gonioscopy, intraocular pressure measurement and funduscopy. If glaucomatous structural changes were confirmed from these tests, axial length (AL; IOL Master, Carl Zeiss Meditec, Dublin, California, USA), imaging with SD-OCT and VF testing were performed. The 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 Swedish interactive threshold algorithm (SITA) Standard strategy, with the Goldmann size III target. Near refractive correction was used as necessary, calculated according to the patient’s age by the HFA software. Unreliable VFs were excluded according to the HFA criteria (fixation losses greater than 25%, false-negative responses greater than 15%). A false negative rate was not used as an indicator of test reliability following a previous report.\textsuperscript{24} A glaucomatous VF was defined as a pattern SD value beyond the normal limit (p<0.05), or a Glaucoma Hemifield Test result outside normal limits following the criteria.\textsuperscript{25} All patients with glaucoma 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 three-dimensional (3D) disc scan and 3D Macula scan (128 horizontal scan lines comprised of 512 A scans for an image area of 6x6 mm). SD-OCT uses a superluminescent diode laser with a centre wavelength of 840 nm and a bandwidth of 50 nm as the light source. The transverse and axial resolutions are less than 20 and 5 µm, respectively. The acquisition speed is 50 000 A scans/s. In the selected eye, the macula was imaged by six radial lines centred 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.\textsuperscript{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.\textsuperscript{28} Indeed there are many reports that suggest that the Random Forest gives the best prediction accuracy among various machine learning methods and it has been used in many research fields, including gene selection and cancer classification.\textsuperscript{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.23

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 analysed), age, gender, AL and right/left eye (see table 1). In this procedure, 10 000 trees were grown and 5 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 participants (n=178) was used (leave-one-out cross validation) so that right and left eyes of a participant were not used for 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 (V.2.14.2, The R Foundation for Statistical Computing, Vienna, Austria) and MedCalc V.11.4.2.0; MedCalc statistical software, Mariakerke, Belgium). The R package ‘randomForest’ and ‘rpart’ was used to carry out the analysis of the Random Forest method and decision tree method, respectively.

RESULTS

Participant’s characteristics are given in table 2. 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 utilising 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 analysed. 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 2A). 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 analysed a patient’s optic cup:optic disc vertical ratio, cp-RNFL thickness and rim area simultaneously, but the authors

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The variables used in the analysis, including 237 optical coherence tomography parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement</td>
<td>cp-RNFL</td>
</tr>
<tr>
<td></td>
<td>Total, 4 sectors (superior, temporal, nasal, inferior), 12 sectors</td>
</tr>
<tr>
<td>AL, axial length; cp-RNFL, circumpapillary retinal nerve fibre layer; GCL+IPL, ganglion cell layer and inner plexiform layer; m-RNFL, macular RNFL.</td>
<td></td>
</tr>
</tbody>
</table>
found that diagnostic performance was not significantly improved compared with using individual measurements. 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. Other studies also support combining multiple structural measurements to diagnose glaucoma. In addition, a recent study suggested the decision tree method is useful to discriminate between patients with glaucoma and normal participants. However, in the current study, the decision tree method, which often suffers from the problem of over-fitting, 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 preperimetric glaucoma is more difficult than differentiating normal participants from patients with glaucoma with early VF damage. A noteworthy advantage of the current study is that it is the first of its kind to analyse 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 participants and patients with glaucoma. 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

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Characteristics of the study participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Glaucamous’ VF group</td>
<td>‘Normal’ VF group</td>
</tr>
<tr>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Age (years)</td>
<td>53.6</td>
</tr>
<tr>
<td>MD (dB)</td>
<td>−6.2</td>
</tr>
<tr>
<td>AL (mm)</td>
<td>25.1</td>
</tr>
<tr>
<td>m-RNFL (μm)</td>
<td>25.5</td>
</tr>
<tr>
<td>cp-RNFL (μm)</td>
<td>88.3</td>
</tr>
<tr>
<td>GCL+IPL (μm)</td>
<td>68.8</td>
</tr>
<tr>
<td>Rim area (mm²)</td>
<td>1.1</td>
</tr>
<tr>
<td>Eye (right/left)</td>
<td>116/108</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>108/116</td>
</tr>
</tbody>
</table>

AL, axial length; cp-RNFL, circumpapillary retinal nerve fibre layer; GCL+IPL, ganglion cell layer and inner plexiform layer; MD, mean deviation; m-RNFL, macular; VF, visual field.

Figure 1 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 coloured ‘X’ represent the sensitivity and specificity of the SD-OCT normative database (red: m-RNFL (p<5%), orange: m-RNFL (p<1%), green: GCL+IPL (p<5%), blue: GCL+IPL (p<1%)). AL, axial length; cp-RNFL, circumpapillary retinal nerve fibre layer; GCL+IPL, ganglion cell layer and inner plexiform layer; m-RNFL, macular RNFL; ROC, receiver operating characteristic.
suspects from patients with glaucoma 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. However, previous results have been based on Heidelberg retina tomography (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 patients with glaucoma to OCT measurements of the RNFL. 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’ participants without ocular disease; Rao et al. have reported that cp-RNFL thickness measurements from normal participants and patients with glaucoma overlap considerably. A significant advantage of the Random Forest classifier is that normal limits could be established based on results from normal participants and patients with glaucoma; 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’ score in the nerve fibre analyser 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 hemiretina, although a few sectors were also situated in the superior hemiretina (see figure 2). Previous studies have suggested that glaucomatous VF damage preferably affects the superior hemifield.\(^42\)\(^45\) 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 hemiretina.

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,\(^50\) however any such effects remain contentious.\(^51\)

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.

Acknowledgements The authors express huge thanks to Hiyori Hirasawa for her invaluable help with manuscript preparation and publication.

Contributors MA and CM gave advice from the viewpoint of a glaucoma specialist. KS, HM and RA conceived and designed the experiments, performed the experiments, analysed the data, contributed in arranging reagents/materials/analysis tools and wrote the manuscript.

Funding This research was supported in part by grants 25861618 (HM), 60645000 (HH), and 50570701 (CM) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests None.

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

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES


