Relevance Vector Machine and Support Vector Machine Classifier Analysis of Scanning Laser Polarimetry Retinal Nerve Fiber Layer Measurements

Christopher Bowd,[1] *Felipe A. Medeiros,*[1] *Zuohua Zhang,*[2] *Linda M. Zangwill,*[1] *Jiucang Hao,*[2,3] *Te-Won Lee,*[2,3] *Terrence J. Sejnowski,*[2,3] *Robert N. Weinreb,*[1] *and Michael H. Goldbaum*[1,4]

Purpose. To classify healthy and glaucomatous eyes using relevance vector machine (RVM) and support vector machine (SVM) learning classifiers trained on retinal nerve fiber layer (RNFL) thickness measurements obtained by scanning laser polarimetry (SLP).

Methods. Seventy-two eyes of 72 healthy control subjects (average age = 64.3 ± 8.8 years, visual field mean deviation = −0.71 ± 1.2 dB) and 92 eyes of 92 patients with glaucoma (average age = 66.9 ± 8.9 years, visual field mean deviation = −5.32 ± 4.0 dB) were imaged with SLP with variable corneal compensation (GDx VCC; Laser Diagnostic Technologies, San Diego, CA). RVM and SVM learning classifiers were trained and tested on SLP-determined RNFL thickness measurements from 14 standard parameters and 64 sectors (approximately 5.6° each) obtained in the circumpapillary area under the instrument-defined measurement ellipse (total 78 parameters). Ten-fold cross-validation was used to train and test RVM and SVM classifiers on unique subsets of the full 164-eye data set and areas under the receiver operating characteristic (AUROC) curve for the classification of eyes in the test set were generated. AUROC curve results from RVM and SVM were compared to those for 14 SLP software-generated global and regional RNFL thickness parameters. Also reported was the AUROC curve for the GDx VCC software-generated nerve fiber indicator (NFI).

Results. The AUROC curves for RVM and SVM were 0.90 and 0.91, respectively, and increased to 0.93 and 0.94 when the training sets were optimized with sequential forward and backward selection (resulting in reduced dimensional data sets). AUROC curves for optimized RVM and SVM were significantly larger than those for all individual SLP parameters. The AUROC curve for the NFI was 0.87.

Conclusions. Results from RVM and SVM trained on SLP RNFL thickness measurements are similar and provide accurate classification of glaucomatous and healthy eyes. RVM may be preferable to SVM, because it provides a Bayesian-derived probability of glaucoma as an output. These results suggest that these machine learning classifiers show good potential for glaucoma diagnosis. (*Invest Ophthalmol Vis Sci.* 2005;46:1322–1329) DOI:10.1167/iovs.04-1122

Scanning laser polarimetry (SLP) is an optical imaging technique used to measure retinal nerve fiber layer (RNFL) thickness that has been successfully applied to glaucoma diagnosis.[1–9] Like other optical imaging technologies, such as confocal scanning laser ophthalmoscopy (CSLO) and optical coherence tomography, SLP imaging results in a large number of measurement parameters that can be difficult for the clinician to interpret. To aid in clinical interpretation, the newest generation SLP includes a summary parameter called the Nerve Fiber Indicator (NFI), developed using a machine learning classifier, the support vector machine (SVM).[10,11] Several studies have demonstrated superior diagnostic performance using the NFI, compared with individual standard parameters.[4,5] In addition, studies have demonstrated superior diagnostic performance of SVM classification, compared with individual parameters and statistical classifiers, with other glaucoma testing modalities.[12,13]

A related machine learning classifier, the relevance vector machine (RVM), recently has been introduced,[14,15] which, unlike SVM, incorporates probabilistic output (probability of class membership, e.g., probability of glaucoma) through Bayesian inference. Its decision function depends on fewer input variables than SVM, possibly allowing better classification estimates for small data sets with high dimensionality (i.e., a large number of input variables).[15] In the present study we compared the performance of RVM and SVM for classifying eyes as healthy or glaucomatous using SLP data.

Methods

Subjects

This observational cross-sectional study included one randomly selected eye from each of 164 study participants (92 patients with glaucoma and 72 healthy controls) enrolled in the University of California, San Diego, Diagnostic Innovations in Glaucoma Study (DIGS).

Each study participant underwent a comprehensive ophthalmic evaluation, including review of medical history, best-corrected visual acuity testing, slit-lamp biomicroscopy, intraocular pressure measurement with Goldmann applanation tonometry, gonioscopy, dilated fundus examination with a 78-D lens, simultaneous stereoscopic optic disc photography (TRC-SS; Topcon Instruments Corp. of America, Paramus, NJ), and standard automated perimetry (SAP) with the 24-2 Swedish Interactive Threshold Algorithm (SITA; Humphrey Field Analyzer II; Carl Zeiss Meditec, Dublin, CA). To be included in the study, partici-

From the [1]Hamilton Glaucoma Center and the [2]Institute for Neural Computation, University of California, San Diego, California; the [3]Computational Neurobiology Laboratories, The Salk Institute, La Jolla, California; and the [4]VA San Diego Healthcare System, San Diego, California.

Supported by Glaucoma Research Foundation (CB), the Foundation for Eye Research (FAM), and National Eye Institute Grants EY11008 (LMZ) and EY13928 (MHG).

Submitted for publication September 21, 2004; revised November 22, 2004; accepted December 18, 2004.

Disclosure: **C. Bowd**, None; **F.A. Medeiros**, None; **Z. Zhang**, None; **L.M. Zangwill**, None; **J. Hao**, None; **T.-W. Lee**, None; **T.J. Sejnowski**, None; **R.N. Weinreb**, Laser Diagnostic Technologies (F); **M.H. Goldbaum**, None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Christopher Bowd, Hamilton Glaucoma Center, University of California, San Diego, La Jolla, CA 92037-0946; cbowd@eyecenter.ucsd.edu.

Investigative Ophthalmology & Visual Science, April 2005, Vol. 46, No. 4

TABLE 1. RNFL Thickness Measurements for Healthy and Glaucomatous Eyes for All SLP Parameters

SLP Parameter	Healthy	Glaucoma	*P*
NFI	20.6 (7.8)	47.0 (22.3)	< 0.0001
Normalized Inferior area (μm^2)	0.132 (0.024)	0.089 (0.030)	< 0.0001
Inferior RNFL thickness (μm)	60.6 (8.0)	48.0 (10.0)	< 0.0001
RNFL thickness standard deviation	21.1 (4.0)	15.2 (4.9)	< 0.0001
Average RNFL thickness (μm)	52.9 (4.5)	45.0 (7.5)	< 0.0001
Inferior ratio	3.29 (1.16)	2.22 (0.92)	< 0.0001
Superior ratio	3.17 (1.00)	2.20 (0.92)	< 0.0001
Normalized superior area (μm^2)	0.127 (0.020)	0.095 (0.036)	< 0.0001
Superior RNFL thickness (μm)	64.6 (6.5)	53.4 (12.1)	< 0.0001
Maximum modulation	2.50 (1.06)	1.65 (0.93)	< 0.0001
Superior maximum (μm)	74.0 (7.8)	63.0 (13.7)	< 0.0001
Inferior maximum (μm)	76.8 (11.2)	63.7 (14.2)	< 0.0001
Ellipse modulation	3.85 (1.42)	2.70 (1.28)	< 0.0001
Superior/nasal ratio	2.33 (0.50)	2.06 (0.68)	0.01
Symmetry	0.98 (0.12)	1.01 (0.24)	0.219

Data are the mean (±SD).

pants had to have a best-corrected acuity better than or equal to 20/40, spherical refraction within ±5.0 D, cylinder correction within ±3.0 D, and open angles on gonioscopy. Eyes with coexisting retinal disease, uveitis, or nonglaucomatous optic neuropathy were excluded.

For labeling eyes during classifier training, glaucomatous eyes were defined as those with repeatable (two consecutive) SAP results outside normal limits by pattern standard deviation (PSD; $P < 5\%$) or Glaucoma Hemifield Test (GHT). The first abnormal SAP was on or before the imaging date. Neither optic disc appearance nor intraocular pressure was part of the inclusion criteria for the glaucoma group. Average SAP mean deviation (MD) of the glaucomatous eyes was −5.32 ± 4.0 dB (range, −20.14 to −0.26 dB). According to the scale of Hodapp et al.[16] for glaucoma severity, 54 (59%) patients had early, 24 (26%) had moderate, and 14 (15%) had severe visual field defects. The mean age of the patients with glaucoma was 66.9 ± 8.9 years (range, 48.7–85.9).

Healthy eyes were defined as those with healthy-appearing optic discs on clinical examination, SAP results (MD, PSD, GHT) within normal limits, and no history of intraocular pressure > 22 mm Hg. Average SAP MD of the healthy eyes was −0.71 ± 1.22 dB (range, −3.92–1.87 dB) and was significantly different from that of healthy eyes (*t*-test, $P < 0.001$). The mean age of the healthy participants was 64.3 ± 8.8 years (range, 48.2–86.8) and was similar to that of the patients with glaucoma (*t*-test, $P = 0.07$).

This research adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from each participant, and the University of California, San Diego, Human Research Protection Program approved all methodology.

Scanning Laser Polarimetry

Study participants underwent ocular imaging with the commercially available SLP with variable corneal compensation, the GDx VCC (software version 5.01; Laser Diagnostic Technologies, San Diego, CA). Scanning laser polarimetry measures the retardation of light reflected from the birefringent RNFL fibers and provides an estimated RNFL thickness based on the linear relationship between observed retardation, measured using a prototype instrument, and RNFL thickness, determined histologically.[17] Details of this technique have been described previously.[18,19] Because corneal polarization axis and magnitude effect SLP measurements and are not similar across eyes,[20,21] the GDx VCC employs a variable corneal polarization compensator that allows eye-specific compensation. After determining the axis and magnitude of corneal polarization in each eye by macular scanning,[22] three appropriately compensated retinal polarization images per eye were automatically obtained and combined, to form each mean image used for analysis. Only well-focused, evenly illuminated, and centered scans with residual anterior segment retardation ≤15.0 nm and atypical scan scores <25, determined by GDx VCC software, were included (cutoffs suggested by written communication, Michael Sinai, PhD, Laser Diagnostic Technologies, June 2004). The atypical scan score indicates the presence of atypical patterns of retardation that can generate spurious RNFL thickness measurements.

We trained the machine learning classifiers on RNFL thickness measurements from 14 standard RNFL measurements (described in detail elsewhere[5]; Table 1) in addition to RNFL thickness measurements from 64 sectors (approximately 5.6° each) obtained in the circumpapillary area under the instrument-defined measurement ellipse (total, 78 parameters). Sector 1 was located temporally, with sectors 16 and 48 located superiorly and inferiorly, respectively (i.e., results were normalized to a right eye). These measurements were determined automatically by GDx VCC software ver. 5.01. We also conducted a subanalysis in which only the 64 sectoral RNFL thickness measurements were included in the machine learning classifier training set to determine which RNFL sectors were most important for classifying eyes as healthy or glaucomatous in our sample.

Machine Classifiers

SVM[10,11] is a machine classification method that directly minimizes the classification error without requiring a statistical data model. This method is popular because of its simple implementation and consistently high classification accuracy when applied to many real-world classification situations. The SVM algorithm can be applied to both classification and regression (model fitting) problems. In classification, an SVM classifier can separate data (e.g., SLP results from healthy and glaucomatous eyes) that are not easily separable in the original data space (i.e., two-dimensional x, y) by mapping data into a higher dimensional (transformed) space. SVM uses a kernel function to find a hyperplane that maximizes the distance (margin) between the two classes (e.g., healthy versus glaucomatous eyes), while minimizing training error.[21] The resultant model is sparse, depending on only a few training samples (the "support vectors"). The number of support vectors increases linearly with the available training data,[15] requiring much higher computational complexity when classifying very large data sets (e.g., tens or hundreds of thousands of input variables). SVMs have been used by us and others for various clinical medicine classification applications including (but not limited to) detecting[12,13,23,24] and predicting[25] glaucoma, detecting central auditory processing disorder,[26] detecting seizure onset,[27] and detecting[28] and characterizing[29] breast lesions.

The SVM was implemented by using Platt's sequential minimal optimization algorithm in commercial software (MatLab, ver. 5.0; The MathWorks, Natick, MA). For classification of the SLP data, Gaussian (nonlinear) kernels of various widths were tested, and a Gaussian kernel with width = $\sqrt{(2 \times \text{number of input variables})}$ was chosen that gave the highest area under the receiver operating characteristic

TABLE 2. AUROC Curves and Sensitivities at Fixed Specificities for Classifying Eyes as Healthy or Glaucomatous for All Techniques and Parameters

Technique	AUROC Curve ± SE	Sensitivity at 75% Specificity (%)	Sensitivity at 90% Specificity (%)
Optimized RVM (forward)	0.92 ± 0.02	87	77
Optimized RVM (backward)	0.93 ± 0.02	92	84
Optimized SVM (forward)	0.94 ± 0.02	92	77
Optimized SVM (backward)	0.94 ± 0.02	90	76
Optimized 64 sector RVM (forward)	0.92 ± 0.02	91	75
Optimized 64 sector RVM (backward)	0.94 ± 0.02	92	84
Optimized 64 sector SVM (forward)	0.94 ± 0.02	92	84
Optimized 64 sector SVM (backward)	0.94 ± 0.02	90	81
RVM	0.90 ± 0.02	83	79
SVM	0.91 ± 0.02	85	76
RVM 64 sector	0.91 ± 0.02	84	79
SVM 64 sector	0.91 ± 0.02	86	74
NFI	0.87 ± 0.03	85	66
Normalized inferior area	0.87 ± 0.03	79	56
Inferior RNFL thickness	0.84 ± 0.03	79	52
RNFL thickness standard deviation	0.83 ± 0.03	81	44
Average RNFL thickness	0.81 ± 0.03	70	61
Inferior ratio	0.80 ± 0.03	69	50
Superior ratio	0.79 ± 0.04	74	40
Normalized superior area	0.79 ± 0.04	71	22
Superior RNFL thickness	0.78 ± 0.04	71	56
Maximum modulation	0.76 ± 0.04	68	33
Superior maximum	0.75 ± 0.04	65	21
Inferior maximum	0.75 ± 0.04	61	29
Ellipse modulation	0.74 ± 0.04	64	33
Superior/nasal ratio	0.65 ± 0.05	42	15
Symmetry	0.51 ± 0.05	36	22

(AUROC) curve using 10-fold cross-validation. The penalty for error/margin tradeoff C was 1.0.

Because SVM does not model the data distribution, but instead directly minimizes the classification error, the resultant output is a binary decision. Although a binary decision is sufficient for many applications, it is difficult to arrive at a meaningful disease-versus-no disease cutoff for glaucoma. This concern can be alleviated with a new machine learning classifier, the RVM.[14,15] The RVM has the same functional form as the SVM within a Bayesian framework. This classifier is a sparse Bayesian model that provides probabilistic predictions (e.g., probability of glaucoma based on the training examples) through Bayesian inference.[15] Its decision function depends on fewer input data (i.e., more sparse) than a comparable SVM, because SVM minimizes the training error under the constraint of maximum smoothness, requiring more decision points.[14] The benefit of a sparser classifier is that its results are more generalizable (i.e., it decreases overfitting). RVM predications are more reliable than SVM predictions because they are directly generated through Bayesian inference, whereas SVM can provide pseudoprobabilistic outputs (i.e., between 0 and 1.0) only through postprocessing. In classification, RVM outputs probabilities of class membership rather than point estimates like SVM. This provides a conditional distribution that allows the expression of uncertainty in the prediction.[30] A Medline search indicates that the RVM has not been applied to clinical data.

The RVM was implemented using a commercially available algorithm (SparseBayes ver. 1.0; Microsoft Research, Cambridge, UK, for MatLab, The MathWorks). For classification of the SLP data, a Gaussian kernel with width = $\sqrt{}$(2 × number of input variables) was chosen because it gave the highest AUROC curve with 10-fold cross-validation.

Analyses

AUROC curves for classifying eyes as healthy or glaucomatous were determined for each machine learning classification technique and each individual parameter automatically provided by the GDx VCC software (Table 1). Significant differences in AUROC curves among RVM, SVM, and all individual parameters were determined by using the method of DeLong et al.[31] We also reported classification results of the GDx VCC NFI, but did not compare these results to the other machine learning classifier results, because we thought such a comparison was somewhat biased (see the Discussion section).

Training and Testing Machine Learning Classifiers

Ten-fold cross-validation was used to train and test RVM and SVM classifiers to avoid training and testing on the same data. First, glaucomatous and healthy eyes were randomly divided into 10 approximately equal, exhaustive, and mutually exclusive subsets. Next, classifiers were trained on 9 subsets and subsequently tested on the 10th subset. This sequence was repeated 10 times, with each subset serving as the test set one time, so that each tested eye was never part of its training set and was tested only once. The test results from 164 eyes were then used to plot the bias-corrected ROC curve. Sensitivities at 75% and 90% specificities, arbitrarily chosen to represent moderate and high specificity, respectively, also were reported, although these values can be estimated from visual inspection of ROC curves.

As the dimensionality of the data sets (number of parameters) is relatively large but the size of the data sets (number of observations) is relatively small, we used sequential forward selection and backward elimination to reduce the data dimension to alleviate the "curse of dimensionality" (reduced classifier performance caused by the forced inclusion of irrelevant parameters in the solution set).[30] For the sake of simplicity, for RVM these techniques were performed using RVM and for SVM these techniques were performed using SVM, although RVM can be optimized using SVM and vice versa. For forward selection, we started with an empty feature set and sequentially added parameters that improved the performance of the feature set the most, until peak performance (e.g., highest AUROC curve after which inclusion of additional parameters decreased the AUROC curve) was reached. For backward elimination, we started with the full-dimensional feature set

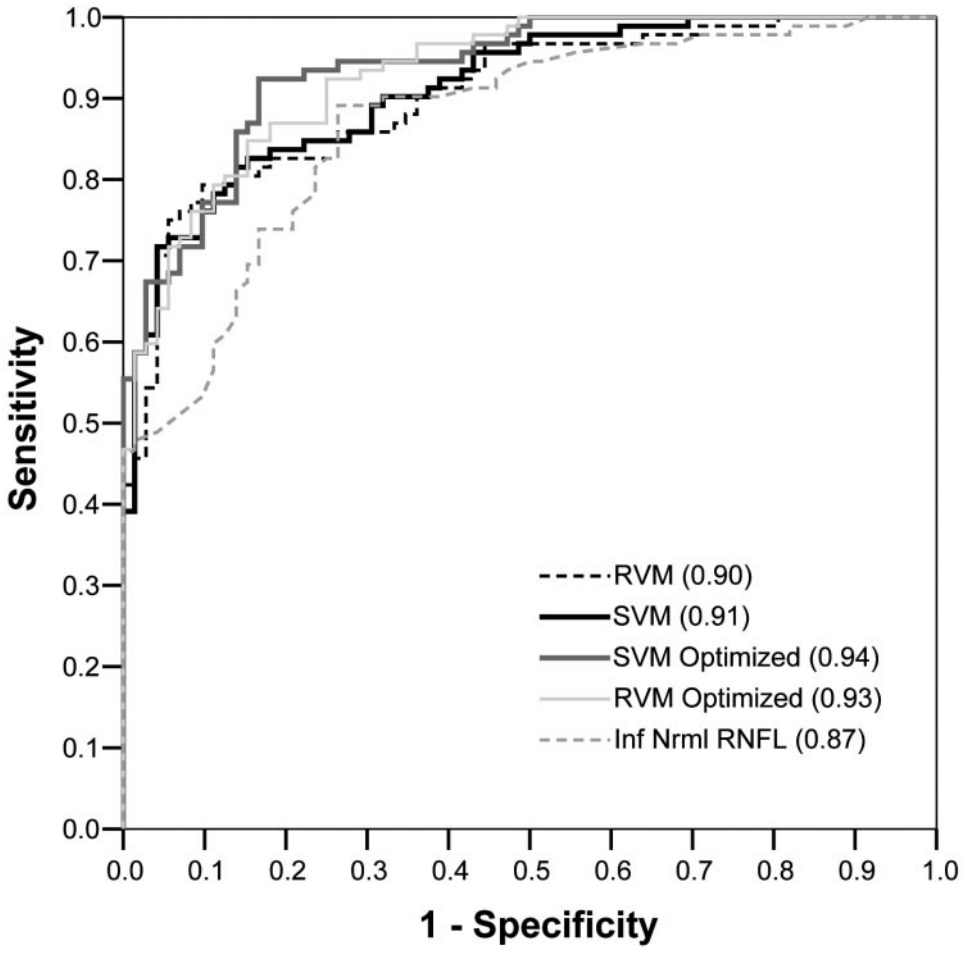

FIGURE 1. ROC curves for "best" optimized RVM, "best" optimized SVM, full-dimensional RVM, full-dimensional SVM, and normalized inferior area. AUROC curves are shown in parentheses.

(e.g., 78 RNFL thickness measurements) and sequentially deleted the parameters that improved the performance of the feature set the most, until performance began to decline.

Similar to using 10-fold cross-validation to minimize bias in the testing and training of the full-dimension RVM and SVM, we used cross-validation to minimize bias in the test sets used during the process of optimizing the feature sets. The data were randomly divided into five approximately equal-sized subsets. Four of the five subsets were used as feature-selection sets to determine the optimized feature set, selected based on the maximum AUROC curve when using 10-fold cross-validation. The optimized feature set was then trained on the initial four subsets, resulting in a single classifier. Next, this classifier was tested on the remaining single subset, and a bias-corrected AUROC curve was generated. This sequence was repeated five times, with each partition serving as the test set one time, resulting in five unbiased estimates of the AUROC curve. This technique for optimization is discussed in greater detail elsewhere.[24]

Several optimized feature sets were investigated for RVM and SVM using both the full data set (i.e., 78 RNFL thickness measurements) and the sector-only data set (i.e., 64 sectoral RNFL thickness measurements). However, we report only results from the optimized feature set for each technique that resulted in the largest AUROC curve.

RESULTS

Table 1 compares the mean values for each individual SLP parameter between healthy and glaucomatous eyes. Statistically significant differences were found for all parameters except symmetry (*t*-test, $P = 0.219$).

ROC Curve Areas and Sensitivities

AUROC curves for nonoptimized RVM and SVM trained using the full (78 RNFL thickness measurements) data set were 0.90 and 0.91, respectively. AUROC curves for nonoptimized RVM and SVM, trained using the sector-only (64-dimensional) data set, were both 0.91. AUROC curves for individual SLP parameters ranged from 0.51 (for symmetry) to 0.87 (for normalized inferior area). The AUROC curves for all individual SLP parameters, except for normalized inferior area, were significantly lower than AUROC curves for RVM and SVM analysis of RNFL thickness (method of DeLong et al.,[31] $P \leq 0.01$; Table 2). The AUROC curve for the NFI was 0.87 and was significantly higher than the AUROC curves for all other individual RNFL measurements (method of DeLong et al., $P \leq 0.01$) except for normalized inferior area (0.87) and inferior RNFL thickness (0.84).

When RVM and SVM trained on the full data set were optimized using forward and backward selection, AUROC curves ranged from 0.92 (RVM forward selection) to 0.94 (SVM forward selection and SVM backward selection). AUROC curves for the best optimized techniques were significantly larger than those for the nonoptimized techniques (method of DeLong et al., all $P < 0.03$) ROC curves for the "best" optimized RVM and SVM, nonoptimized RVM and SVM, and normalized inferior area are shown in Figure 1.

Sensitivities at 75% specificity were 83% for RVM and 85% for SVM (Table 2). These values improved to 92% for both optimized RVM and SVM. Sensitivity at 75% specificity for global RNFL thickness measurements ranged from 36% (sym-

TABLE 3. RNFL Thickness Sectors Included in Optimized Training Sets for RVM and SVM Machine Learning Classifiers

RVM with Forward Selection	RVM with Backward Elimination	SVM with Forward Selection	SVM with Backward Elimination
Sector 51 (IT)	Sector 52 (IT)	Sector 52 (IT)	Sector 53 (IT)
Sector 62 (IT)	Sector 53 (IT)	Sector 51 (IT)	Sector 52 (IT)
Sector 53 (IT)	Sector 62 (IT)	Sector 60 (IT)	Sector 60 (IT)
Sector 52 (IT)	Sector 44 (IN)	Sector 14 (ST)	Sector 13 (ST)
Sector 7 (ST)	Sector 59 (IT)	Sector 27 (SN)	Sector 62 (IT)
Sector 31 (SN)	Sector 28 (SN)	Sector 26 (SN)	Sector 44 (IN)
Sector 13 (ST)	Sector 6 (ST)	Sector 28 (SN)	Sector 22 (SN)
Sector 59 (IT)	Sector 61 (IT)	Sector 15 (ST)	Sector 33 (IN)
	Sector 60 (IT)	Sector 5 (ST)	Sector 51 (IT)
	Sector 27 (SN)	Sector 25 (SN)	Sector 19 (SN)
			Sector 57 (IT)
			Sector 26 (SN)

Sector 1 is temporal, sector 16 is superior, and sector 46 is inferior. RNFL quadrants corresponding to each sector also are shown. IT, inferior temporal quadrant; IT, inferior temporal quadrant; SN, superior nasal quadrant; ST, superior temporal quadrant.

metry) to 81% (RNFL thickness SD). Sensitivity at 75% specificity of the NFI was 85%.

Sensitivities at 90% specificity were 79% for RVM and 76% for SVM (Table 2). These values improved to 84% and 77% for optimized RVM and SVM, respectively. Sensitivity at 90% specificity for global RNFL thickness measurements ranged from 15% (superior-to-nasal ratio) to 61% (average RNFL thickness). Sensitivity at 90% specificity of the NFI was 66%.

Best RNFL Sectors Determined by Optimization

To determine what individual RNFL sectors are most informative for classifying healthy and glaucomatous eyes in our sample, we determined the "best" reduced-dimensional data sets (defined as those with the largest AUROC curves) by optimizing the sector-only data set (i.e., 64 sector RNFL thickness measurements) using RVM and SVM with forward and backward selection. Results are shown in Table 3. In all cases the highest ranked sectors were those in the inferior temporal circumpapillary quadrant. A significant number of sectors in the superior temporal and superior nasal quadrants also were identified. However, only two sectors in the inferior nasal quadrant were identified.

Relevance Vector Machine Probabilistic Results

RVM output is in the form of a probability of belonging to the glaucoma group of the training set. Average mean RVM output was 0.22 (22%) for healthy eyes and 0.81 (81%) for glaucomatous eyes (*t*-test, $P < 0.001$). Figure 2 shows the number of healthy and glaucomatous eyes that fell into each 10% probability bin based on RVM output. Of the healthy eyes, 84.6% were assigned a probability of glaucoma under 51%, and of the glaucomatous eyes, 85.8% were assigned a probability of glaucoma over 50%, although there was significant overlap between the groups. Figures 3 and 4 show clinical examples of this overlap.

Figure 3 shows GDx VCC retardation images from three healthy eyes assigned the probabilities of 0.01 (A), 0.48 (B), and 0.92 (C), respectively, by RVM. (Eyes were selected as closest to 0.0, 0.5, and 1.0 probabilities, respectively). These eyes all had normal SAP results by definition. Eye (A) had a SAP MD of −0.13 dB and a PSD of 1.70 dB. GDx VCC NFI output was 9, suggesting a very low likelihood of glaucomatous damage. Eye (B) had a SAP MD of 0.10 dB and a PSD of 1.56 dB. GDx VCC NFI output was 14, also suggesting a very low likelihood of glaucomatous damage. Eye (C) had a SAP MD of −1.58 dB and a PSD of 1.49 dB. GDx VCC NFI output was 36, suggesting possible glaucomatous damage.

Figure 4 shows GDx VCC retardation images from three glaucomatous eyes assigned the probabilities of 0.14 (A), 0.49 (B), and 0.99 (C). (Again, eyes were selected as closest to 0.0, 0.5, and 1.0 probabilities, respectively.) These eyes all had abnormal SAP results by definition. Eye (A) had a SAP MD of −4.41 dB and a PSD of 7.93 dB. GDx VCC NFI output was 26, suggesting a low likelihood of glaucomatous damage. Eye (B) had a SAP MD of −3.51 dB and a PSD of 3.09 dB. GDx VCC NFI output was 22, also suggesting a low likelihood of glaucomatous damage. Eye (C) had a SAP MD of −8.53 dB and a PSD of

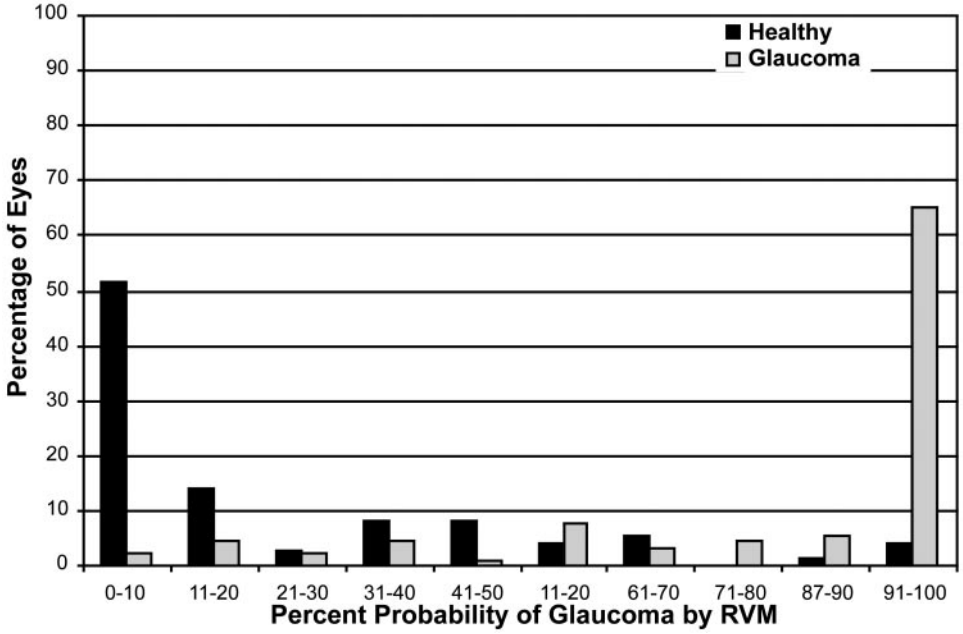

FIGURE 2. Percentage of healthy or glaucomatous eyes assigned by RVM to each 10% probability bin. Of the healthy eyes, 84.6% were assigned a probability under 51%, and of glaucomatous eyes, 85.8% were assigned a probability over 50%.

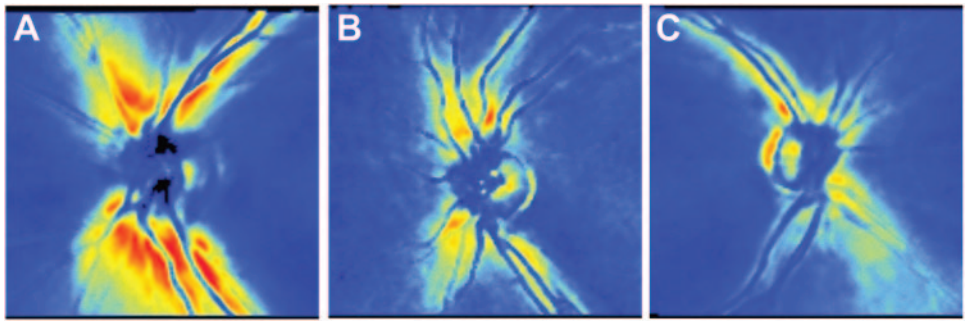

FIGURE 3. Scanning laser polarimetry RNFL thickness maps from three eyes classified as normal based on visual field results. *Brighter colors* represent a thicker RNFL. Eye (**A**) was assigned a 0.1% probability of belonging to the glaucoma group by RVM analysis, with a SAP MD of −0.13 dB and a PSD of 1.70 dB. GDx VCC NFI output was 9. Eye (**B**) was assigned a 48% probability of belonging to the glaucoma group by RVM analysis, with a SAP MD of 0.10 dB and a PSD of 1.56 dB. GDx VCC NFI output was 14. Eye (**C**) was assigned a 92% probability of belonging to the glaucoma group by RVM analysis, with a SAP MD of −1.58 dB and a PSD of 1.49 dB. GDx VCC NFI output was 36.

7.61 dB. GDx VCC NFI output was 98, indicating a very high likelihood of glaucomatous damage.

DISCUSSION

Results in this study indicate that RVM and SVM classifiers, trained on SLP RNFL thickness measurements, can improve on standard software-provided SLP RNFL thickness measurements for classifying eyes as healthy or glaucomatous. This improvement is increased when optimization is used to select features (i.e., a subset of RNFL parameters) that are likely to be the most relevant for this classification. RVM performed as well as SVM at this classification task (maximum AUROC curves, 0.93 and 0.94, respectively). The fact that RVM provides equal-interval, probabilistic output based on Bayesian inference makes results using this technique more informative than those with SVM, which requires a nonlinear transformation to output values between 0.0 and 1.0. These general results are likely to be applicable to other glaucoma imaging technologies such as CSLO and optical coherence tomography, although the likelihood has not been thoroughly investigated (Goldbaum MH, et al. *IOVS* 2002;43:ARVO E-Abstract 2137).

Probabilistic results from the RVM indicated that most of the glaucomatous eyes fell within the 51% to 100% probability range and most healthy eyes fell within the 0% to 50% range, although some overlap was observed. Figures 3B and 3C may suggest that SLP-trained RVM can detect eyes with normal SAP results that have glaucomatous abnormalities detectable by other diagnostic tests, although this suggestion is not conclusive. The eye shown in Figure 3B, with a SAP MD of 0.10 dB and PSD of 1.56 dB (i.e., normal result), was assigned a 48% probability of belonging to the glaucoma group. At the time of SLP imaging, this eye had an abnormal short-wavelength automated perimetry result with MD of −5.99 dB ($P < 5\%$), PSD of 4.40 dB (within normal limits), and GHT result within normal limits. However, Heidelberg Retina Tomograph (HRT II; Heidelberg Engineering, Dossenheim, Germany) Moorfields Regression Analysis results were all within normal limits. The eye shown in Figure 3C, with a SAP MD of −1.58 dB and PSD of 1.49 dB (normal result), was assigned a 92% probability of belonging to the glaucoma group. At the time of SLP imaging, this eye had a normal short-wavelength automated perimetry result with MD of −2.60 dB (within normal limits), PSD of 4.40 dB (within normal limits), and GHT result within normal limits. The Heidelberg Retina Tomograph Moorfields Regression Analysis of the superior nasal sector was assigned a borderline result.

In the above cases, RVM probabilistic output can be used clinically to help determine the posttest probability of disease. If the output is greater than 50% (0.50), the probability of

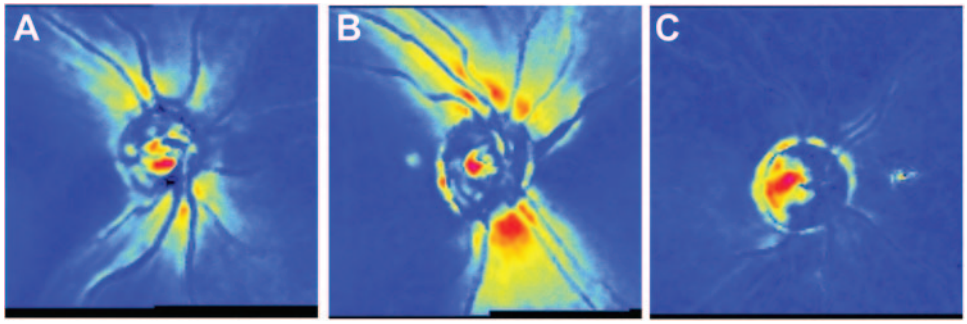

FIGURE 4. Scanning laser polarimetry RNFL thickness maps from three eyes classified as glaucomatous based on visual field results (repeatable SAP Glaucoma Hemifield Test results outside normal limits and/or PSD < 5%). Eye (**A**) was assigned a 14% probability of belonging to the glaucoma group by RVM analysis, with a SAP MD of −4.41 dB and a PSD of 7.93 dB. GDx VCC NFI output was 26. Eye (**B**) was assigned a 49% probability of belonging to the glaucoma group by RVM analysis, with a SAP MD of −3.51 dB and a PSD of 3.09 dB. GDx VCC NFI output was 22. Eye (**C**) was assigned a 99% probability of belonging to the glaucoma group by RVM analysis with a SAP MD of −8.53 dB and a PSD of 7.61 dB. GDx VCC NFI output was 98.

having glaucoma is increased compared with the pretest probability. If the output is <50%, the probability of having glaucoma is decreased.

Machine learning classifier analyses, using forward and backward selection on sector-only RNFL thickness measurements, were used to determine the RNFL sectors most essential for classifying healthy and glaucomatous eyes in our sample. We determined that sectors in the inferior temporal quadrant were most important, followed by sectors in the superior temporal and the superior nasal quadrants. In general, inferior nasal sectors were not included. These results are in agreement with results from other studies in which SLP, CSLO, or optical coherence tomography were used, indicating that the inferior temporal RNFL and neuroretinal rim are the most important regions for discriminating between healthy and early-stage glaucomatous eyes.[24,32,33] The difference in the relative importance of superior nasal and inferior nasal quadrants for classifying eyes as healthy or glaucomatous may be due, in part, to the presence of split superior RNFL bundles in many eyes.[34] Based on viewing many SLP images, we suspect that the effect of split superior RNFL bundles is that the superior nerve fiber bundle is displaced nasally in many eyes, thus increasing this region's importance in glaucoma discrimination.

The GDx VCC currently includes an SVM-based classification parameter, the NFI. We did not compare our RVM or SVM results directly with results from the NFI because we suspected that the outcome of such a comparison would be biased in favor of our RVM/SVM, because we trained and tested our classifiers on very similar data sets. Both the training and testing data sets were constrained by the selection criteria for study inclusion, and each data set reflected the age, race, and glaucoma severity characteristics of our clinic population, which were not necessarily similar to that used to develop the NFI. For instance, the mean age of our patients with glaucoma was 66.9 ± 8.9 years (range, 48–89). The mean age of the patients with glaucoma included in the development the NFI was similar (65.4 ± 13.33 years), but the NFI data set included 10% of patients under the age of 46 years (Sinai M, Laser Diagnostics Technologies, personal communication, June 2004). The mea SAP MD of patients in the present study was −5.3 ± 4.0 dB (range, −20.14 to +0.26). Although the mea SAP MD of patients included in the development of the NFI was similar (−5.42 ± 6.11 dB), the range was greater than that in our study (−31.57 to +2.81 dB; Sinai M, Laser Diagnostics Technologies, personal communication, June 2004). These differences in study population characteristics of the machine classifier training sets would be likely to confound any direct comparison between our techniques and the NFI.

Results for the NFI and other standard RNFL thickness measurements are similar to those in other work reported from our laboratory (using GDx VCC and a prototype instrument, with similar inclusion criteria)[4,5,8] and others. For instance, Tannenbaum et al.[35] reported AUROC curves for average, superior, and inferior RNFL thicknesses of 0.81, 0.87, and 0.85, respectively. The AUROC curve for superior thickness in their study was higher than that reported in our study (0.78). This may be a function of the inclusion in their study of more advanced cases, suggested by the range of SAP MD to approximately −30 dB, or by the presence of a larger number of subjects with inferior visual field defects.

The current results, demonstrating that machine learning classification techniques are superior to single parameters for discriminating between healthy and glaucomatous eyes using optical imaging technologies, support previously published results on this topic.[12,24] However, in prior publications on this topic, CSLO data were used for machine learning classifier training (with the exception of studies reporting results from the standard software-provided machine learning classifiers available with the current and previous versions of the commercially available SLP).

Previous studies using back-propagated, multi-layer perceptrons trained using global CSLO optic disc topography parameters found good discrimination between normal and glaucomatous eyes with a diagnostic precision of 80%[36] and 92% (with an AUROC curve of 0.94).[37] More recently, "bagging" (boot-strap aggregating) classification trees, trained using a large number of global and regional HRT parameters, showed decreases in the normal versus glaucoma misclassification error, compared with linear discriminant functions.[38,39] For instance, reported misclassification was 15% using the bagged classification tree, compared with approximately 20% for previously published linear discriminant analyses.[38,39] In a study similar to the current one, SVM trained on CSLO global and regional topographic optic disc parameters were shown to improve on multi-layer perceptron techniques and previously published linear discriminant functions for differentiating between mild and moderate glaucoma.[12] AUROC curves for nonoptimized neural network techniques ranged from 0.94 to 0.95, compared with 0.85 to 0.91 for statistic-based methods. After forward selection and backward elimination were applied, the discriminating ability of the SVM increased significantly to 0.97. These findings are similar to those from the present study.

AUROC curve results from our machine learning classifier techniques may be somewhat overestimated, because we used cross-validation instead of truly independent training and test sets. Although RVM and SVM were trained and tested on different data, each data set was generated from the same rather homogeneous pool. This fact may exaggerate somewhat the differences in classification ability between RVM and SVM and standard SLP RNFL thickness parameters, although we were careful to employ cross-validation to separate training and test sets at all steps in training, testing, and optimizing our machine learning classifiers.

In addition, when using optimized techniques on sector-only data to identify RNFL sectors that were most important for discriminating between healthy and glaucomatous eyes, the resultant measurement regions (i.e., RNFL sectors) may be specific to the constraints of the early-to-moderate glaucoma damage inclusion criterion of the present study. It is possible that in more advanced glaucoma, or with a larger training set, different sectors would be identified as most important for the classification task. Our training set may have included more examples of inferior temporal RNFL defects, than other defects, because of the glaucoma severity investigated or because of the size of the training set. This possibility suggests that the development (training and testing) of different classifiers for different degrees of glaucoma severity may be necessary. In addition, a larger training set including more examples of glaucomatous eyes is desirable.

Overall, our results showed that optimized RVM and SVM, trained on SLP RNFL thickness measurements, classify glaucomatous and healthy eyes more accurately than current software-provided RNFL thickness measurements. These results suggest that these machine learning classifiers show good potential for glaucoma diagnosis. Moreover, results from relevance vector machine analyses showed that most glaucomatous eyes were assigned a high probability of being glaucomatous, based on labeled training examples, and that most healthy eyes were assigned a low probability of being glaucomatous. RVM and SVM performed similarly at classifying eyes as healthy or glaucomatous. Because RVM output provides a Bayesian-derived probability of glaucoma and SVM output does not, RVM classifiers are likely to provide more information than SVM classifiers for the diagnosis of glaucoma.

References

1. Bagga H, Greenfield DS. Quantitative assessment of structural damage in eyes with localized visual field abnormalities. *Am J Ophthalmol*. 2004;137:797–805.
2. Bagga H, Greenfield DS, Feuer W, Knighton RW. Scanning laser polarimetry with variable corneal compensation and optical coherence tomography in normal and glaucomatous eyes. *Am J Ophthalmol*. 2003;135:521–529.
3. Bowd C, Zangwill LM, Weinreb RN. Association between scanning laser polarimetry measurements using variable corneal polarization compensation and visual field sensitivity in glaucomatous eyes. *Arch Ophthalmol*. 2003;121:961–966.
4. Medeiros FA, Zangwill LM, Bowd C, Mohammadi K, Weinreb RN. Comparison of scanning laser polarimetry using variable corneal compensation and retinal nerve fiber layer photography for detection of glaucoma. *Arch Ophthalmol*. 2004;122:698–704.
5. Medeiros FA, Zangwill LM, Bowd C, Weinreb RN. Comparison of the GDx VCC scanning laser polarimeter, HRT II confocal scanning laser ophthalmoscope, and Stratus OCT optical coherence tomograph for the detection of glaucoma. *Arch Ophthalmol*. 2004;122:827–837.
6. Reus NJ, Lemij HG. The relationship between standard automated perimetry and GDx VCC measurements. *Invest Ophthalmol Vis Sci*. 2004;45:840–845.
7. Weinreb RN. Evaluating the retinal nerve fiber layer in glaucoma with scanning laser polarimetry. *Arch Ophthalmol*. 1999;117:1403–1406.
8. Weinreb RN, Bowd C, Zangwill LM. Glaucoma detection using scanning laser polarimetry with variable corneal polarization compensation. *Arch Ophthalmol*. 2003;121:218–224.
9. Garway-Heath DF, Greaney MJ, Caprioli J. Correction for the erroneous compensation of anterior segment birefringence with the scanning laser polarimeter for glaucoma diagnosis. *Invest Ophthalmol Vis Sci*. 2002;43:1465–1474.
10. Vapnik V. *Statistical Learning Theory.* New York: Wiley; 1998.
11. Vapnik V. *The Nature of Statistical Learning Theory.* 2nd ed. New York: Springer; 2000.
12. Bowd C, Chan K, Zangwill LM, et al. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci*. 2002;43:3444–3454.
13. Goldbaum MH, Sample PA, Chan K, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2002;43:162–169.
14. Bishop CM, Tipping ME. Variational relevance vector machines. In: Boutilier C, Goldszmidt M, eds. *Uncertainty in Artificial Intelligence.* Cambridge, UK: Microsoft Research; 2000:45–63.
15. Tipping ME. Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res*. 2001;1:211–244.
16. Hodapp E, Parrish RK, Anderson DR. *Clinical Decisions in Glaucoma.* St. Louis, MO: Mosby; 1993.
17. Weinreb RN, Dreher AW, Coleman A, Quigley H, Shaw B, Reiter K. Histopathologic validation of Fourier-ellipsometry measurements of retinal nerve fiber layer thickness. *Arch Ophthalmol*. 1990;108:557–560.
18. Bagga H, Greenfield DS. Retinal nerve fiber layer assessment using scanning laser polarimetry. *Int Ophthalmol Clin*. 2004;44:29–42.
19. Zangwill LM, Medeiros FA, Bowd C, Weinreb RN. Optic nerve imaging: recent advances. In: Grehn F, Stamper R, eds. *Glaucoma.* Berlin: Springer-Verlag; 2004:63–91.
20. Greenfield DS, Knighton RW, Huang XR. Effect of corneal polarization axis on assessment of retinal nerve fiber layer thickness by scanning laser polarimetry. *Am J Ophthalmol*. 2000;129:715–722.
21. Weinreb RN, Bowd C, Greenfield DS, Zangwill LM. Measurement of the magnitude and axis of corneal polarization with scanning laser polarimetry. *Arch Ophthalmol*. 2002;120:901–906.
22. Zhou Q, Weinreb RN. Individualized compensation of anterior segment birefringence during scanning laser polarimetry. *Invest Ophthalmol Vis Sci*. 2002;43:2221–2228.
23. Sample PA, Goldbaum MH, Chan K, et al. Using machine learning classifiers to identify glaucomatous change earlier in standard visual fields. *Invest Ophthalmol Vis Sci*. 2002;43:2660–2665.
24. Zangwill LM, Chan K, Bowd C, et al. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci*. 2004;45:3144–3151.
25. Bowd C, Zangwill LM, Medeiros FA, et al. Confocal scanning laser ophthalmoscopy classifiers and stereophotograph evaluation for prediction of visual field abnormalities in glaucoma-suspect eyes. *Invest Ophthalmol Vis Sci*. 2004;45:2255–2262.
26. Strauss DJ, Delb W, Plinkert PK. Objective detection of the central auditory processing disorder: a new machine learning approach. *IEEE Trans Biomed Eng*. 2004;51:1147–1155.
27. Shoeb A, Edwards H, Connolly J, Bourgeois B, Treves T, Guttag J. Patient-specific seizure onset detection. *Epilepsy Behav*. 2004;5:483–498.
28. Campanini R, Dongiovanni D, Iampieri E, et al. A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Phys Med Biol*. 2004;49:961–975.
29. Piliouras N, Kalatzis I, Dimitropoulos N, Cavouras D. Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound. *Comput Med Imaging Graph*. 2004;28:247–255.
30. Bishop CM, Tipping ME. Bayesian regression and classification. In: Suykens J, Horvath G, Basu S, Micchelli C, Vandewalle J, eds. *Advances in Learning Theory: Methods, Models and Applications* (NATO Science Series. Series III, Computer and Systems Sciences. Vol. 190). Amsterdam: IOS Press; 2003:267–285.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
32. Bowd C, Zangwill LM, Berry CC, et al. Detecting early glaucoma by assessment of retinal nerve fiber layer thickness and visual function. *Invest Ophthalmol Vis Sci*. 2001;42:1993–2003.
33. Medeiros FA, Susanna R Jr. Comparison of algorithms for detection of localised nerve fibre layer defects using scanning laser polarimetry. *Br J Ophthalmol*. 2003;87:413–419.
34. Colen TP, Lemij HG. Prevalence of split nerve fiber layer bundles in healthy eyes imaged with scanning laser polarimetry. *Ophthalmology*. 2001;108:151–156.
35. Tannenbaum DP, Hoffman D, Lemij HG, Garway-Heath DF, Greenfield DS, Caprioli J. Variable corneal compensation improves discrimination between normal and glaucomatous eyes with the scanning laser polarimeter. *Ophthalmology*. 2004;111:259–264.
36. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol*. 1996;121:511–521.
37. Uchida H, Brigatti L, Caprioli J. Detection of structural damage from glaucoma with confocal laser image analysis. *Invest Ophthalmol Vis Sci*. 1996;37:2393–2401.
38. Hothorn T, Lausen B. Bagging tree classifiers for laser scanning images: a data- and simulation-based strategy. *Artif Intell Med*. 2003;27:65–79.
39. Mardin CY, Hothorn T, Peters A, Junemann AG, Nguyen NX, Lausen B. New glaucoma classification method based on standard Heidelberg Retina Tomograph parameters by bagging classification trees. *J Glaucoma*. 2003;12:340–346.