Heidelberg Retina Tomograph Measurements of the Optic Disc and Parapapillary Retina for Detecting Glaucoma Analyzed by Machine Learning Classifiers

Linda M. Zangwill,[1] *Kwokleung Chan,*[2,3] *Christopher Bowd,*[1] *Jicuang Hao,*[2,3] *Te-Won Lee,*[2,3] *Robert N. Weinreb,*[1] *Terrence J. Sejnowski,*[2,3] *and Michael H. Goldbaum*[1]

Purpose. To determine whether topographical measurements of the parapapillary region analyzed by machine learning classifiers can detect early to moderate glaucoma better than similarly processed measurements obtained within the disc margin and to improve methods for optimization of machine learning classifier feature selection.

Methods. One eye of each of 95 patients with early to moderate glaucomatous visual field damage and of each of 135 normal subjects older than 40 years participating in the longitudinal Diagnostic Innovations in Glaucoma Study (DIGS) were included. Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Dossenheim, Germany) mean height contour was measured in 36 equal sectors, both along the disc margin and in the parapapillary region (at a mean contour line radius of 1.7 mm). Each sector was evaluated individually and in combination with other sectors. Gaussian support vector machine (SVM) learning classifiers were used to interpret HRT sector measurements along the disc margin and in the parapapillary region, to differentiate between eyes with normal and glaucomatous visual fields and to compare the results with global and regional HRT parameter measurements. The area under the receiver operating characteristic (ROC) curve was used to measure diagnostic performance of the HRT parameters and to evaluate the cross-validation strategies and forward selection and backward elimination optimization techniques that were used to generate the reduced feature sets.

Results. The area under the ROC curve for mean height contour of the 36 sectors along the disc margin was larger than that for the mean height contour in the parapapillary region (0.97 and 0.85, respectively). Of the 36 individual sectors along the disc margin, those in the inferior region between 240° and 300°, had the largest area under the ROC curve (0.85–0.91). With SVM Gaussian techniques, the regional parameters showed the best ability to discriminate between normal eyes and eyes with glaucomatous visual field damage, followed by the global parameters, mean height contour measures along the disc margin, and mean height contour measures in the parapapillary region. The area under the ROC curve was 0.98, 0.94, 0.93, and 0.85, respectively. Cross-validation and optimization techniques demonstrated that good discrimination (99% of peak area under the ROC curve) can be obtained with a reduced number of HRT parameters.

Conclusions. Mean height contour measurements along the disc margin discriminated between normal and glaucomatous eyes better than measurements obtained in the parapapillary region. (*Invest Ophthalmol Vis Sci.* 2004;45:3144–3151) DOI: 10.1167/iovs.04-0202

Various imaging technologies are being used to assist the clinician in the assessment of the optic disc and RNFL. Each of these technologies has the advantage of providing independent quantitative, objective, and reproducible measurements of the optic disc and retinal nerve fiber layer (RNFL) without relying on the subjective clinical examination or qualitative review of photographs. One such technique, confocal scanning laser ophthalmoscopy (CSLO), provides reproducible measurements of the optic disc and parapapillary retina.[1–3] Although studies have assessed the ability of topographic optic disc measurements obtained with the Heidelberg Retina Tomograph (HRT, Heidelberg Engineering, Dossenheim, Germany), a confocal scanning laser ophthalmoscope, to detect early to moderate glaucoma,[4–6] few studies with the HRT have focused on RNFL layer measurements implied by retinal thickness in the parapapillary area.[7]

There are several possible advantages to measuring the overall RNFL and particularly to measuring the parapapillary RNFL. First, there is evidence that subjective assessment of change in the RNFL often precedes change in the cup-disc ratio in eyes in which glaucomatous visual field damage subsequently develops.[8] Therefore, quantitative objective measurements of the RNFL may be particularly useful in detecting early glaucoma. Second, HRT RNFL thickness measurements have been shown to correlate with histomorphometric axon counts in monkey eyes.[9] Third, HRT measurements in the flat parapapillary retina have been shown to be more reproducible than measurements along steeper areas of the cup within the disc margin.[10]

In an effort to summarize the large amount of data provided by the HRT, several different analysis strategies, including linear discriminant functions[11–14] and machine learning classifiers (including neural networks),[14] have been used to improve the discriminating ability of HRT parameters. Recently, machine classifiers, such as support vector machines (SVMs) trained on HRT global and regional topographic optic disc parameters, have been shown to improve on previously used linear discriminant functions for detecting early to moderate glaucoma.[15] However, the HRT images can provide more information that has not been previously included in analyses using machine learning classifiers, including sectoral RNFL measurements.

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

Supported in part by National Eye Institute Grants EY11008 (LMZ) and EY13235 (MHG) and the Glaucoma Research Foundation (CB).

Submitted for publication February 24, 2004; revised March 29, 2004; accepted April 7, 2004.

Disclosure: **L.M. Zangwill**, None; **K. Chan**, None; **C. Bowd**, None; **J. Hao**, None; **M.H. Goldbaum**, None; **T.-W. Lee**, None; **T.J. Sejnowski**, None; **R.N. Weinreb**, Heidelberg Engineering (F)

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: Linda M. Zangwill, Diagnostic Imaging Laboratory, Hamilton Glaucoma Center and Department of Ophthalmology, University of California at San Diego, La Jolla, CA 92093-0946; zangwill@eyecenter.ucsd.edu.

Investigative Ophthalmology & Visual Science, September 2004, Vol. 45, No. 9

TABLE 1. Demographic Characteristics of the Glaucoma and Normal Study Groups

	Glaucomatous Eyes (n = 95)	Normal Eyes (n = 135)	P
Age (y, mean ± SD, range)	65.9 ± 12.2 (40.0–93.0)	59.5 ± 12.0 (40.0–80.0)	<0.0001
Gender (% male)	54	67	0.046
Race (% white)	80	90	0.045
Optic disc area (mm^2, mean ± SD, range)	1.93 ± 0.51 (0.95–3.63)	1.83 ± 0.38 (0.96–2.90)	0.085
Visual field mean deviation (dB, mean ± SD, range)	−5.73 ± 5.32 (−21.86–0.67)	0.06 ± 1.32 (−3.66–3.51)	<0.0001
Visual field pattern standard deviation (dB, mean ± SD, range)	6.2 ± 3.5 (5.70–6.70)	1.74 ± .34 (1.30–2.10)	<0.0001

We therefore focused this investigation on evaluating HRT parapapillary RNFL measurements using machine learning classifier techniques. The purpose of this study was to use machine learning classifiers to determine whether assessment of HRT sectoral measurements in the parapapillary retina, where measurements are often less variable,[10] can improve differentiation between normal eyes and eyes with early to moderate glaucoma compared with sectoral measurements along the disc margin. A secondary objective of the study was to use these data to improve previously described methods[15] for optimization of machine learning classifier feature selection to develop robust methods for identifying a reduced feature set of HRT parameters that discriminate as well as the full set of features.

METHODS

Subjects

One randomly selected eye from each of 95 patients with glaucoma and from each of 135 normal subjects participating in the longitudinal Diagnostic Innovations in Glaucoma Study (DIGS) was included in the study. All subjects completed an ophthalmic examination, including slit lamp biomicroscopy, intraocular pressure (IOP) measurement, stereoscopic fundus examination, stereoscopic photography of the optic disc, and full-threshold standard automated perimetry (SAP; Humphrey Field Analyzer, Carl Zeiss Meditec Inc., Dublin, CA). Best corrected visual acuity at the time of testing was 20/40 or better.

For purposes of this study, patients were classified as having glaucoma if they had at least two consecutive standard automated perimetry examinations with either a corrected pattern standard deviation outside the 95% normal limits or a glaucoma hemifield test result outside the 99% normal limits. At least one of the abnormal fields was obtained on or before the date of CSLO imaging. The appearance of the optic disc was not used as a criterion for designation of glaucoma. The average age (95% confidence interval) of patients with glaucoma was 65.9 (63.4–68.3) years. Average mean deviation and pattern standard deviation (95% confidence interval) of the SAP closest to the CSLO imaging date was −5.74 (−4.64 to −6.83) dB and 6.2 (5.7–6.7) dB, respectively, indicating early to moderate glaucomatous visual field damage.

Healthy eyes had intact rims, with no evidence of notching, glaucomatous excavation, or RNFL defect, and had symmetrical optic discs (asymmetry of vertical cup-disc ratio <0.2) based on a dilated clinical examination. IOP was ≤22 mm Hg with no history of elevated IOP. SAP results were within normal limits. Healthy patients had no history of diabetes and no ophthalmic or neurologic surgery or disease. The average age (±SD) of healthy subjects was 59.5 (57.4–61.5) years. Healthy subjects were significantly younger than patients with glaucoma (*t*-test, $P < 0.001$; Table 1).

The research adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all participants and the University of California, San Diego, Human Subjects Committee approved all methodology.

Confocal Scanning Laser Ophthalmoscopy

The HRT provides topographical measures of the optic disc and parapapillary retina and has been described in detail elsewhere.[3,16]

Three 15° field of view scans centered on the optic were obtained in each test eye. A mean topography image of these three scans judged to be of acceptable quality was created with the HRT (software version 2.01). The optic disc margin was outlined on the mean topography image by trained technicians while viewing simultaneous stereoscopic photographs of the optic disc. All mean topography images had an SD of ≤50 μm.

We included sectoral measurements along the operator-drawn contour line outlining the margin of the optic disc and in the parapapillary retina along a contour of 1.7 mm radius concentric with the contour line (Fig. 1). Specifically, 36 ten-degree sector measurements of the mean height contour along the operator-drawn contour line surrounding the optic disc margin (called mean height contour at disc margin, or MHCDM) and 36 ten-degree sector measurements of the mean retinal height or mean height contour around a contour of 1.7-mm radius (MHC1.7), with its center at the center of the optic disc were included in the analysis. The contour of 1.7-mm radius was established by automatically changing the radius of the contour line outlining the disc margin to 1.7 mm. As illustrated in Figure 1, the shape of the contour in the parapapillary retina therefore matches the shape of the contour line outlining the disc margin.

The sectoral mean height contour at the disc margin measurements describe the height of the neuroretinal rim at the optic disc margin relative to the standard reference ring. In addition, HRT RNFL measurements along the contour line outlining the disc margin (RNFLDM) and also along the contour of the 1.7-mm radius (RNFL1.7) concentric with the contour line were included in the analysis. HRT RNFL measurements are an indirect measure of RNFL thickness. The RNFL measurements are calculated by subtracting the mean height contour from the reference plane height. The reference plane height is defined as the mean thickness 50 μm posterior to mean retinal heights between 350° and 356° temporal along the contour line.

In addition to the parameters just described, we included the following global topographic measures for evaluation: disc area, area below reference, mean height of contour, peak height of contour, height variability of contour, volume below surface, volume above surface, volume below reference, volume above reference, maximum cup depth, cup shape, mean cup depth, RNFL thickness, RNFL cross-sectional area, reference plane height, rim area, cup-disc ratio, and rim-disc ratio. Six regional topographic measures of each parameter except mean cup depth, RNFL cross-sectional area, and reference plane height were also included. Regions were defined as temporal superior (46°–90° unit circle), nasal superior (91°–135°), nasal (136°–225°), nasal inferior (226°–270°), temporal inferior (271°–315°), and

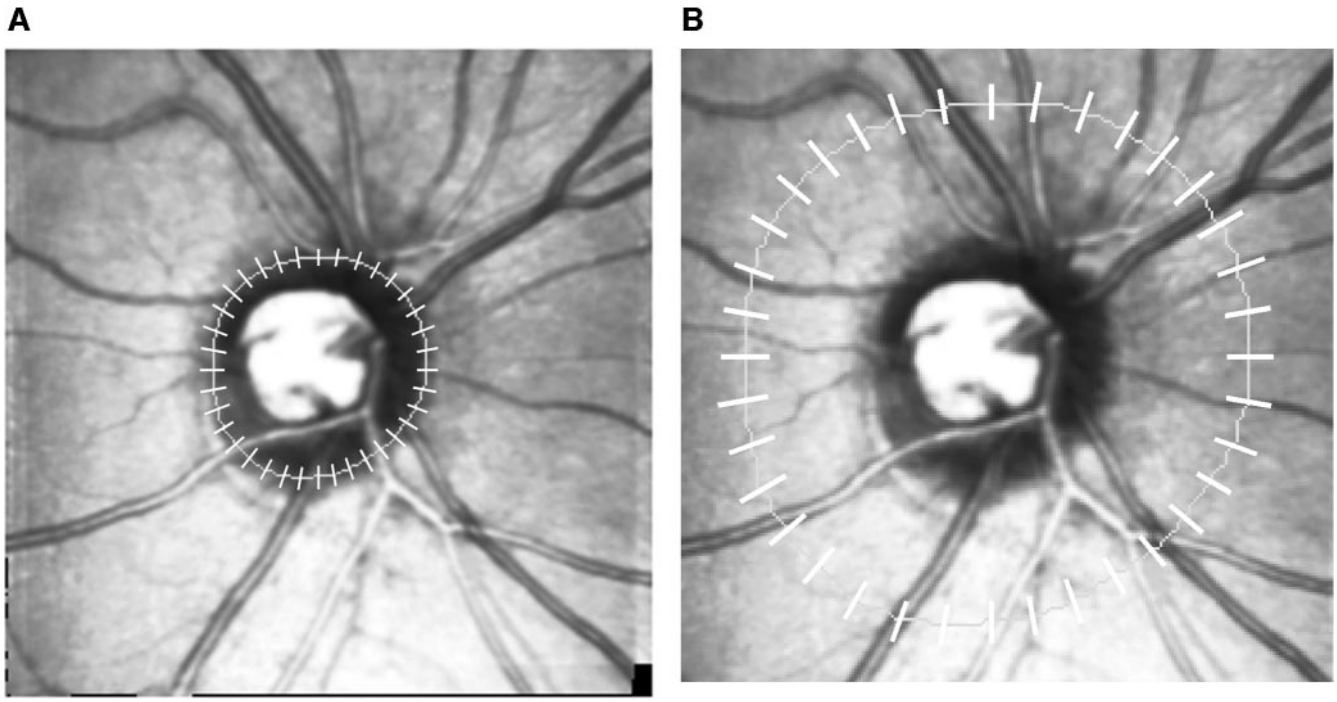

FIGURE 1. (**A**) Thirty-six sectors along the contour line outlining the disc margin. (**B**) Thirty-six sectors along a contour concentric with the disc margin contour line at a radius of 1.7 mm.

temporal (316°–45). These parameters have been discussed in more detail elsewhere.[3,16–18] A total of 22 global and 86 regional parameters were included in these analyses.

Machine Learning Classifier Techniques

There are many techniques for classifying data. Machine learning classifiers have the advantage that they can adapt their internal parameters to account for structure in data, thus often achieving performance superior to that of nonadaptive classifiers. Among machine learning classifiers SVMs[19,20] have recently gained popularity because of their consistent and high classification performance.[21] They are especially suitable for binary classification (two-class problem) and for data of relatively small size and high dimensions.[19] These properties match well with our task.

We evaluated the performance of two machine learning classifier techniques, linear SVM (SVM linear) and Gaussian SVM (SVM Gaussian) for classifying eyes as glaucomatous or healthy. Both SVM linear and SVM Gaussian have been described elsewhere and have been used to classify eyes as glaucomatous or nonglaucomatous, based on HRT[15] and visual field data.[22,23] SVM classifiers were chosen because they yielded the highest area under the ROC curves for discriminating between healthy and glaucomatous eyes in another study using HRT parameters.[15] In brief, SVMs are techniques used for solving classification and regression problems. During training, the SVM with a nonlinear kernel maps the training data to a high dimensional space where a hyperplane is fitted that maximizes the margin of separation between classes while minimizing the generalization error (ability to generalize results from finite training set to previously unseen data set) using statistical learning theory. Constraints imposed on the construction of the separating surface result in a subset of training data that is involved in the decision function (called support vectors). The hyperplane that splits the positive and negative vectors is oriented to maximize the distance between itself and the nearest positive and negative examples. SVM linear and SVM Gaussian differ because they assume different distributions of input data. SVM linear uses linear mapping, resulting in a dot product kernel and SVM Gaussian uses nonlinear mapping resulting in a Gaussian kernel.[19,20]

ROC curves for classifying eyes as glaucomatous or healthy were determined for the linear and Gaussian SVMs applied to the sectoral mean height contour at the disc margin and to RNFL at the disc margin measurements only, to the sectoral parapapillary mean height contour and parapapillary RNFL measurements only, to the global HRT parameters only, to the regional HRT parameters only, and to all these measurements combined. For each area under the ROC curve reported, 10-fold cross-validation techniques were completed **to** keep training and testing data separate by dividing the glaucomatous and healthy eyes randomly into 10 mutually exclusive subsets each. The 10 sets were composed of an approximately equal ratio of glaucomatous-to-healthy eyes. One subset of data was used as the test set, and the remaining nine composed the training set. This procedure was repeated, with each subset serving once as the test set. The results obtained for the 10 test sets were combined to generate a single ROC curve for each classification method. We also reported sensitivities at specificities of 75% (representing moderate specificity) and 90% (representing high specificity).

Feature Selection by Forward Selection and Backward Elimination

We used feature-selection techniques to reduce the feature set to the most important features. We determined how the number of HRT parameters on which the SVM model was trained influenced its discriminating ability. Feature selection was accomplished with forward selection and backward elimination. The forward selection strategies started with an empty feature set for input. In each round we added to previously selected features the one parameter that improved most the performance of the classifier. The process was repeated until all parameters were included in the feature set. For backward elimination we started with a full feature set. In each round, we deleted from the input feature set the one parameter that decreased least the performance of the classifier. The process was repeated until the feature set was empty.

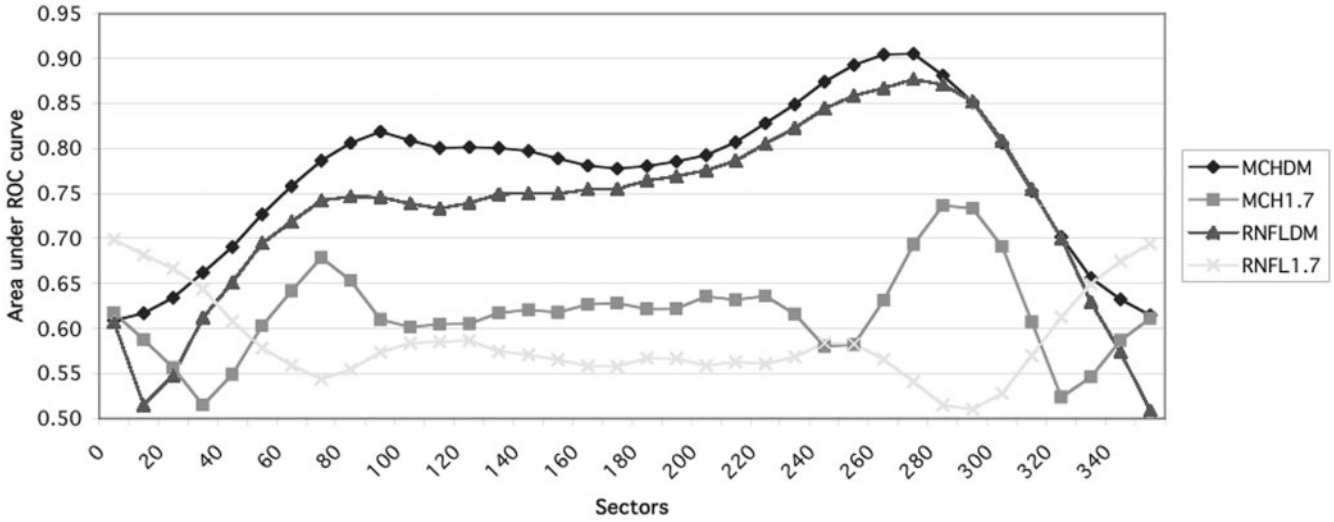

FIGURE 2. Area under the ROC curve by mean height contour and RNFL sectors along the disc margin and in the parapapillary region (at a radius of 1.7 mm from the center of the disc). Mean height contour (♦) and RNFL (▲) sectors along the disc margin have a larger area under ROC curve than mean height contour (■) and RNFL (**X**) in the parapapillary region.

Cross-Validation Techniques

To minimize bias in the testing sets used during the process of optimizing the feature sets, we used "external cross-validation." The whole data set was divided into a "selection set" and an "evaluation set." Eighty percent of the eyes were included in the selection set and 20% in the independent evaluation set. Feature-selection methods were applied only on the selection set, whereas the independent evaluation set was used solely for assessing the performance of the reduced feature sets obtained by the forward selection and backward elimination feature-selection methods.

With the 80% selection data set, internal cross-validation was used during the feature selection to select the next HRT parameters, which contributed most to the correct classification of eyes as healthy or glaucomatous. The 80% of glaucomatous and healthy eyes were divided again randomly into 10 mutually exclusive subsets. The 10 sets were composed of an approximately equal ratio of glaucoma to healthy eyes. One subset of data was used as the test set (8%), and the remaining nine composed the training set (72%). This was repeated with each subset serving once as the test set. The results obtained for the 10 test sets were combined to generate a single ROC curve for each classification method. The ROC area served as the performance measure to decide which HRT parameter to delete or add. After the reduced feature set for any size (n) was established by either forward selection or backward elimination, an SVM using that feature subset was tested with the external evaluation data set. This was repeated for various n. Since the evaluation data set is not used in the feature-selection process, this gives a less biased assessment of each feature-selection method.

The whole process of an 80-to-20 split, forward selection–backward elimination, and assessment using evaluation sets was repeated five times, each time with a totally different 20% of the original data as the evaluation set. This method provided five rankings of the parameters from backward elimination and another five rankings from forward selection.

The various resulting rankings were unified by a simple voting scheme. The parameter that appeared first the most often across the different rankings was considered the most important and was removed. The remaining parameters were shifted up accordingly. The parameter that now appeared first the most often was considered the next most important and was removed. The remaining parameters are shifted up again. The process of voting and removal was repeated until no parameters were left.

We also evaluated the ability of using a limited number of HRT parameters to discriminate between glaucomatous and healthy eyes. With the selection data set, the number of parameters included when the area under the ROC curve reached its "peak" or the highest value was determined. In addition, the number of parameters included in the SVM Gaussian model that corresponded to 97.5%, 99.0%, and 99.9% of peak area under the ROC curve was also identified. The area under the ROC curve and sensitivities at fixed specificities were estimated for these models with a reduced number of HRT parameters. This procedure was completed to identify a smaller set of parameters that can discriminate as well as the full set of 178 parameters.

The SVM was programmed on computer (MatLab, ver. 5.0; The MathWorks, Natick, MA) and trained using the Platt sequential minimal optimization algorithm. The programer chose the parameters for penalty and the kernel by trial and error. The penalty used was $C = 1.0$. Details and mathematical descriptions of the SVM techniques used have been described elsewhere.[20,23–27] The method of DeLong et al.[28] was used to determine statistically significant differences in overall area under the ROC curves between classifiers trained on each set of data (e.g., mean height contour along the disc margin, global parameters). The area under the ROC curve for the 36 sectors in Figure 2 was determined on computer by using nominal logistic regression (JMP software; SAS, Cary, NC).

RESULTS

Our results indicate that HRT mean height contour and RNFL measurements along the disc margin better discriminated between normal and glaucomatous eyes than measurements obtained in the parapapillary retina. We determined the area under the ROC curve for each of the 36 mean height contour and RNFL sectors individually to compare the discriminating ability of measurements obtained along the disc margin and in the parapapillary retina and to determine which locations were most informative for classifying eyes as healthy or glaucomatous. Figure 2 shows that measurements of RNFL and mean height contour along the disc margin had larger areas under the ROC curve than measurements in the parapapillary region. In addition, the figure shows a double-hump–like pattern with an apparent peak located inferiorly (approximately 240°–280°) and superiorly (approximately 80°–120°), indicating that these sectors have the largest area under the ROC curve and there-

TABLE 2. Area under the ROC Curves and Sensitivities at Set Specificities from SVM Gaussian for Different Data Sets

Training Set	Classifier	ROC Area (SE)	Sensitivity at 75% Specificity (%)	Sensitivity at 90% Specificity (%)
All parameters combined*	SVM-G	0.964 (0.010)	97	85
Regional parameters*	SVM-G	0.959 (0.011)	94	86
Global parameters†	SVM-G	0.935 (0.016)	89	81
Sectoral mean height contour along the disc margin (MHCDM)†	SVM-G	0.914 (0.018)	87	77
Sectoral parapapillary mean height contour (MCH1.7)	SVM-G	0.808 (0.027)	70	46
Sectoral RNFL thickness along the disc margin (RNFLDM)‡	SVM-G	0.863 (.026)	81	68
Sectoral Parapapillary RNFL thickness (RNFL1.7)	SVM-G	0754 (.032)	59	38

* Area under the ROC curve significantly larger than area under the ROC curve for global parameters, sectoral mean height contour and RNFL along the disc margin (MCHDM and RNFLDM) and sectoral parapapillary mean retinal height contour and RNFL measurements (MCH1.7 and RNFL1.7).

† Area under the ROC curve significantly larger than area under the ROC curve for sectoral parapapillary mean retinal height contour and RNFL (MCH1.7 and RNFL1.7).

‡ AUROC significantly larger than parapapillary RNFL (RNFL1.7).

fore the greatest ability to discriminate between normal and glaucomatous eyes compared with other sectors.

The next objective was to use machine learning classifiers to compare the area under the ROC curve of mean height contour and RNFL measurements along the disc margin with measurements obtained in the parapapillary retina (Table 2). With training sets using SVM Gaussian techniques, the area under the ROC curve (±SE) was significantly greater when using the 36 sectoral mean height contour measurements along the disc margin (0.914 ± 0.018) than when using the 36 sectoral parapapillary mean height contour measurements (0.808 ± 0.027). Sensitivities at 75% and 90% specificity were higher with the 36 mean height contour along the disc margin sectors (87% and 77%, respectively) than with the 36 parapapillary mean height contour sectors (70% and 46%, respectively). Similarly, the area under the ROC curve when using the 36 RNFL thickness sectors along the disc margin (0.863 ± 0.026) was higher than when using measurements in the parapapillary retina (0.754 ± 0.032).

We also compared the sectoral measurements along the disc margin and parapapillary retina to other HRT regional and global parameters. The area under the ROC curve and sensitivities at specificities of 90% and 75% for these training sets and for the SVM Gaussian for all parameters combined are presented in Table 2. When SVM Gaussian results for the different training sets (global, regional, mean height contour at the disc margin, parapapillary mean height contour, and all combined) were compared, two training sets had the largest area under the ROC curve (±SE): the set containing all parameters combined (0.964 ± 0.010) and the set that included regional parameters only (ROC area, 0.959 ± 0.011). The area under the ROC curve (±SE) of these two training sets was significantly larger than that of the training sets including global parameters only (0.935± 0.016), sectoral mean height contour measurements along the disc margin (0.911± 0.020), sectoral parapapillary mean height contour measurements (0.796± 0.030), sectoral RNFL measurements along the disc margin (0.863 ± 0.020), and sectoral parapapillary RNFL measurements (0.754 ± 0.032) (all comparisons, $P \leq 0.014$; Table 2). Further, two training sets, the set that included global parameters only and the set containing sectoral mean height contour measurements along the disc margin had significantly larger areas under the ROC curves than training sets that included sectoral RNFL measurements along the disc margin, and parapapillary mean height contour and RNFL measurements (all comparisons, $P \leq 0.03$). Finally, the area under the ROC curve for sectoral RNFL measurements along the disc margin was significantly greater than that for sectoral parapapillary RNFL measurements ($P = 0.007$).

For each training set in Table 2, the area under the ROC curve was somewhat larger for SVM Gaussian than for SVM linear, but these differences did not reach statistical significance except when global parameters were used in the model ($P = 0.023$; data not shown). We therefore limited the reporting to results using SVM Gaussian techniques.

Optimizing Machine Learning Classifiers by Using Feature Selection

Additional analyses were completed to determine whether better performance of the model, as measured by the area under the ROC curve, could be achieved by including only the most effective features (HRT parameters). Forward selection and backward elimination procedures were completed as outlined in the Methods section. Figure 3 illustrates the area under the ROC curve (*y*-axis) by the number of features included in the model (*x*-axis). The top curve in Figures 3A (forward selection) and 3B (backward elimination) indicate the estimate of the area under the ROC curve after the selection process, without independent verification. As expected, when the selection of features is verified in an independent data set, the estimate for the area under the ROC curve is lower (Fig. 3A, 3B, bottom curves).

Figures 3A and 3B also illustrate that good performance (99% of peak area under the ROC curve) can be attained when less than 10 of the features are included in the model. With forward-selection techniques, the number of parameters necessary to obtain 97.5%, 99.0%, and 99.9% of peak area under the ROC curve for SVM Gaussian including all 178 parameters was 4, 8, and 26, respectively. With backward-elimination optimization techniques, the number of parameters necessary to obtain 97.5%, 99.0%, and 99.9% of peak area under the ROC curve for SVM Gaussian was 4, 7, and 14, respectively.

DISCUSSION

HRT sectoral mean height contour and RNFL measurements obtained along the disc margin were found to discriminate between healthy eyes and eyes with early to moderate glaucoma better than measurements obtained in the parapapillary retina. In addition, global and regional parameters also had better discriminating ability (larger area under the ROC curve)

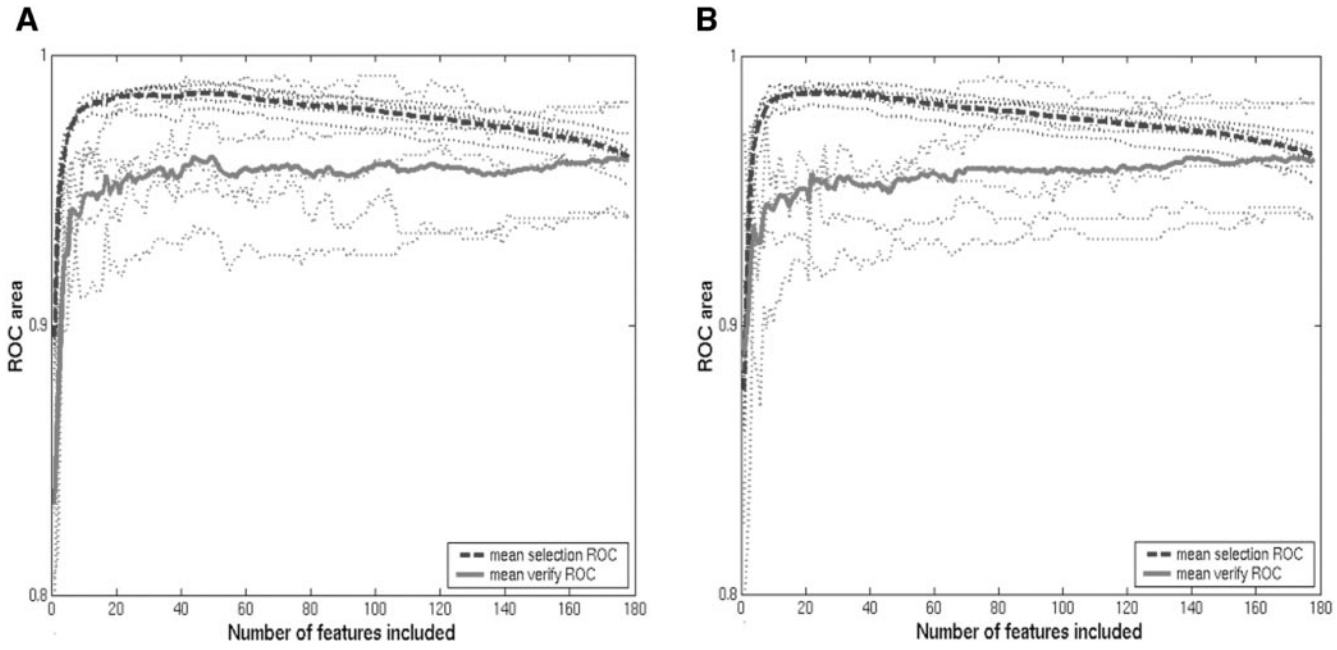

FIGURE 3. SVM Gaussian optimization using a combination of parameters: global, regional, and mean height contour along disc margin. Optimization using (**A**) forward selection and (**B**) backward elimination show that the area under ROC curve is higher for the selection (*thick dashed line*) ROC curve than the verification (*thick solid line*) ROC curve. Five cross-validation replications (represented by the five *thin dashed lines*) surrounding the selection ROC curve and five replications (represented by the five *thin dotted lines*) were used to create the selection and verification ROC curve, respectively. In addition, forward selection and backward elimination show that less than 10 HRT parameters are required for a model to discriminate at 99% of the peak area under ROC curve (*arrows*).

than sectoral measurements in the parapapillary region. Furthermore, it was found that machine learning classifiers can summarize large amounts of data and that optimization techniques can be used to identify a reduced set of features that provide good discrimination.

Among the sectoral measurements along the disc margin, a relatively small number of sectors in the inferior and superior region discriminated between healthy and glaucomatous eyes better than other sectors. The graph of the area under the ROC curve for the disc margin measurements in Figure 2 resembles the anatomic double-hump pattern of RNFL thickness distribution around the optic disc, with the largest area under the ROC curve in the inferior and superior regions, corresponding to the known pattern of thicker RNFL in these regions. The use of sectoral data to discriminate between healthy eyes and eyes with early to moderate glaucoma confirms previous reports using other imaging instruments that also found that inferior or superior RNFL parameters often provide better discrimination than measurements in other areas.[5,29,30]

To our knowledge, this is the first report comparing the discriminating ability of HRT measurements along the disc margin to measurements in the parapapillary area. HRT RNFL thickness along the disc margin has been reported to be among the best parameters for discriminating between normal and healthy eyes.[6,12] Furthermore, HRT RNFL thickness along the disc margin was the HRT parameter most strongly correlated with histomorphometric axon counts in monkey eyes.[9] As RNFL damage, as assessed by review of red-free RNFL photographs of the parapapillary region, has been shown to precede visual field damage by as many as 6 years, it is of interest to determine whether parapapillary measurements available with each HRT image can discriminate as well as measurements along the disc margin. In addition, it is of interest to evaluate whether the parapapillary measurements in combination with traditional measurements along and within the optic disc provide an improvement in discriminating ability above that of optic nerve head parameters alone. Our results indicate that HRT measurements in the parapapillary area did not discriminate better than measurements obtained along the disc margin and did not provide additional discriminating power when evaluated in combination with other optic nerve head parameters.

There are several possible explanations of why HRT mean height contour and RNFL measurements along the disc margin discriminate better than measurements obtained in the parapapillary region. The HRT measures retinal surface topography and only indirectly measures RNFL thickness. The sectoral measurements along the disc margin are thicker and may better reflect the tissue of interest, the RNFL. Compared with measurements along the disc margin, the thinner measurements in the parapapillary region may include in its measurement a greater proportion of the other layers of the retina in addition to RNFL tissue. A standard reference plane was used to calculate RNFL thickness measurements both along the disc margin and in the parapapillary retina. Because the discriminating ability of mean height contour measurements, which are not based on a reference plane, performed as well or better than RNFL measurements, it is unlikely that the reference plane per se is the reason for the better discriminating ability of measurements taken at the disc margin.

The better discriminating ability of HRT measurements along the disc margin compared with the parapapillary should not be generalized to RNFL measurements obtained by other imaging methods, such as scanning laser polarimetry and optical coherence tomography, that are designed to measure directly the RNFL in the parapapillary region. For the present study, we defined the parapapillary measurements at a fixed distance (1.7-mm radius) from the center of the optic cup rather than at a distance relative to the disc margin, so that data from eye to eye would be more comparable. Other imaging instruments that measure features of the parapapillary RNFL, including the GDx variable corneal and lens compensator

(VCC; Laser Diagnostic Technologies, Inc., San Diego, CA) and the Optical Coherence Tomograph (Carl Zeiss Meditec Inc.), obtain measurements at fixed distances from the center of the optic cup. Both instruments include a default measurement along a circle of 1. 7-mm radius, as it has been shown to be among the most reproducible locations.[31] More important, perhaps, if measurements are obtained relative to the disc margin, then they will be located at different locations relative to the center of the disc, limiting the comparability of the measurements across eyes. For example, if two subjects have the same number of nerve fibers and axons, but one eye has a larger optic disc diameter than another, the RNFL thickness should be similar in both subjects if measured at a fixed radius, as long as it is sufficiently far from the disc margin. However, if measurements are obtained at a parapapillary location relative to the disc margin, such as 1.5 disc diameters from the disc margin, then thinner measurements would likely be obtained in the larger disc (compared to smaller disc), because the location of the measurement would be more peripheral.

Investigators have applied several analysis strategies in an effort to summarize the large amount of data provided by the HRT and other diagnostic instruments, including standard automated perimetry. Discriminant functions, Fourier analysis,[32,33] and machine learning classifier techniques have been shown to improve the ability of these instruments to discriminate between normal and glaucomatous eyes.[15,22,23] More recently, machine classifiers and neural network techniques including SVM have been shown to improve the ability of HRT optic disc parameters,[15] and visual field results[22,23] to differentiate between normal and glaucomatous eyes. Specifically, using HRT regional parameters, SVM techniques improved the ability to detect early glaucoma (area under the ROC curve, 0.97) compared with commercially available and previously published HRT-based linear discriminant functions (area under the ROC curves all <0.91).[15] For these reasons, SVM techniques were included in the study as the method to use to integrate the large quantity of data automatically provided by the HRT.

The area under the ROC curves in the present study are not as large as those reported by Bowd et al.,[15] probably because one of the objectives of this study was to refine cross-validation and optimization techniques to provide better estimates (reduce bias) of the area under the ROC curve. Extension of cross-validation techniques with a separate evaluation test set is a more conservative statistical approach. It is preferred because it avoids overstating the discriminating ability (area under ROC curve) of the models that might occur from using the selection set as the evaluation set. In contrast to our previous study,[15] the present study used an evaluation set separate from the selection set. As illustrated in Figure 3, it is likely that the verification ROC curve (solid line), based on separate selection and evaluation sets, provides a better estimate of the discriminating ability than the selection ROC curve (dashed line), which used the selection set as the evaluation set, too.

Another objective of the present study was to identify a smaller set of HRT parameters that would discriminate as well as the full dimensional data. It was found that using fewer than 10 parameters can result in discrimination at 99% of the peak area under the ROC curve using the full data set. Identifying a smaller set of parameters will not reduce patient testing time (all data are available from a single scan). However, using a reduced number of variables may facilitate postprocessing of these data. It can also indicate which features are likely to be in a well-performing reduced feature set.

A machine learning classifier can combine features into its input from several tests, such as automated perimetry, optic nerve topography, and various risk factors. Eliminating less useful features from each of the tests may ease the integration of these data for analysis by one classifier. Our data from this study and others[15,22,23] suggest that machine learning classifiers in general, and SVMs in particular, may be well suited to integrate data on both structural and functional tests to provide the clinician with a more comprehensive diagnostic tool for detecting glaucoma.

In the present study, our inclusion criteria for normal subjects required a normal visual field and normal optic nerve appearance at the clinical examination, whereas the criteria for our glaucoma group required a glaucomatous visual field. The requirement for a normal clinical examination was necessary to avoid the inclusion of subjects with glaucomatous optic neuropathy but normal visual fields in the control group. Although disc appearance was not used explicitly as one of the inclusion criteria in our glaucoma group, it can be assumed that nearly all eyes with visual field damage also had optic disc damage. It can be argued that these inclusion criteria lead to an overestimate of the area under the ROC curve. Therefore, our results may not be generalizable to other studies where the diagnosis is not known and borderline cases are present. It should be noted that this is a methodological issue common to glaucoma diagnostic studies and no practical solution to this problem is available at this time.

In conclusion, mean height contour measurements along the disc margin discriminated better between healthy eyes and eyes with visual field damage than parapapillary measurements away from the disc margin. Optimization of machine learning classifier techniques can be used to identify a reduced set of features that provide good discrimination. Extension of cross-validation techniques with a separate evaluation test set avoids overstating the discriminating ability of the models. This study provides further evidence that the use of machine learning classifiers, trained with adequate cross-validation methods, can assist in identifying which combination of HRT parameters can best detect glaucoma. The application of these results in clinical practice could result in a more accurate diagnosis of glaucoma than possible with any single optic disc parameter such as cup-disc ratio or rim area.

References

1. Chauhan BC, LeBlanc RP, McCormick TA, Rogers JB. Test-retest variability of topographic measurements with confocal scanning laser tomography in patients with glaucoma and control subjects. *Am J Ophthalmol*. 1994;118:9–15.
2. Rohrschneider K, Burk ROW, Kruse FE, Volcker HE. Reproducibility of the optic nerve head topography with a new laser tomographic scanning device. *Ophthalmology*. 1994;101:1044–1049.
3. Weinreb RN, Lusky M, Bartsch DU, Morsman D. Effect of repetitive imaging on topographic measurements of the optic nerve head. *Arch Ophthalmol*. 1993;111:636–638.
4. Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology*. 1998;105:1557–1563.
5. Zangwill LM, Bowd C, Berry CC, et al. Discriminating between normal and glaucomatous eyes using the heidelberg retina tomograph, GDx nerve fiber analyzer, and optical coherence tomograph. *Arch Ophthalmol*. 2001;119:985–993.
6. Mardin CY, Horn FK, Jonas JB, Budde WM. Preperimetric glaucoma diagnosis by confocal scanning laser tomography of the optic disc. *Br J Ophthalmol*. 1999;83:299–304.
7. Burk RO, Tuulonen A, Airaksinen PJ. Laser scanning tomography of localised nerve fibre layer defects. *Br J Ophthalmol*. 1998;82:1112–1117.
8. Sommer A, Katz J, Quigley HA et al. Clinically detectable nerve

fiber atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol*. 1991;109:77–83.
9. Yucel YH, Gupta N, Kalichman MW et al. Relationship of optic disc topography to optic nerve fiber number in glaucoma. *Arch Ophthalmol*. 1998;116:493–497.
10. Brigatti L, Weitzman M, Caprioli J. Regional test-retest variability of confocal scanning laser tomography. *Am J Ophthalmol*. 1995;120: 433–440.
11. Mikelberg FS, Parfitt CM, Swindale NV, Graham SL, Drance SM, Gosine R. Ability of the Heidelberg Retina Tomograph to detect early glaucomatous visual field loss. *J Glaucoma*. 1995;4:242–247.
12. Bathija R, Zangwill L, Berry CC, Sample PA, Weinreb RN. Detection of early glaucomatous structural damage with confocal scanning laser tomography. *J Glaucoma*. 1998;7:121–127.
13. Iester M, Jonas JB, Mardin CY, Budde WM. Discriminant analysis models for early detection of glaucomatous optic disc changes. *Br J Ophthalmol*. 2000;84:464–468.
14. 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.
15. 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.
16. Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the Heidelberg Retina Tomograph. *J Glaucoma*. 1993;2:101–103.
17. Zangwill L, Bowd C, Weinreb RN. Evaluating the optic disc and retinal nerve fiber layer in glaucoma II: optical image analysis. *Semin Ophthalmol*. 2000;15:206–220.
18. Bowd C, Zangwill LM, Blumenthal EZ, et al. Imaging of the optic disc and retinal nerve fiber layer: the effects of age, optic disc area, refractive error, and gender. *J Opt Soc Am A Opt Image Sci Vis*. 2002;19:197–207.
19. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining Knowledge Discovery*. 1998;2:121–167.
20. Vapnik V. *Statistical Learning Theory.* New York: John Wiley & Sons; 1998.
21. Duda RO, Hart PE, Stork DG. *Pattern Classification.* 2nd ed. New York: John Wiley & Sons; 2001.
22. 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.
23. 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.
24. Rumelhart DE, Hinton G, Williams R. Learning representations of back-propagation errors. *Nature*. 1986;323:533–536.
25. Broomhead DS, Lowe D. Multivariable functional interpolation and adaptive networks. *Complex Systems*. 1988;2:321–355.
26. Bishop CM. *Neural Networks for Pattern Recognition.* Oxford, UK: Clarendon Press; 1995.
27. Vapnik V. *The Nature of Statistical Learning Theory.* 2nd ed. New York: Springer; 2000.
28. 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.
29. Greaney MJ, Hoffman DC, Garway-Heath DF, Nakla M, Coleman AL, Caprioli J. Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma. *Invest Ophthalmol Vis Sci*. 2002;43:140–145.
30. 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.
31. Schuman JS, Pedut-Kloizman T, Hertzmark E, et al. Reproducibility of nerve fiber layer thickness measurements using optical coherence tomography (see comments). *Ophthalmology*. 1996;103: 1889–1898.
32. Essock EA, Sinai MJ, Zangwill LM, Bowd C, Weinreb RN. Fourier analysis of optical coherence tomography and scanning laser polarimetry retinal nerve fiber layer measurements in the diagnosis of glaucoma. *Arch Ophthalmol*. 2003;121:1238–1245.
33. Medeiros FA, Zangwill LM, Bowd C, Bernd AS, Weinreb RN. Fourier analysis of scanning laser polarimetry measurements with variable corneal compensation in glaucoma. *Invest Ophthalmol Vis Sci*. 2003;44:2606–2612.