Comparing Neural Networks and Linear Discriminant Functions for Glaucoma Detection Using Confocal Scanning Laser Ophthalmoscopy of the Optic Disc

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

Purpose. To determine whether neural network techniques can improve differentiation between glaucomatous and nonglaucomatous eyes, using the optic disc topography parameters of the Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany) .

Methods. With the HRT, one eye was imaged from each of 108 patients with glaucoma (defined as having repeatable visual field defects with standard automated perimetry) and 189 subjects without glaucoma (no visual field defects with healthy-appearing optic disc and retinal nerve fiber layer on clinical examination) and the optic nerve topography was defined by 17 global and 66 regional HRT parameters. With all the HRT parameters used as input, receiver operating characteristic (ROC) curves were generated for the classification of eyes, by three neural network techniques: linear and Gaussian support vector machines (SVM linear and SVM Gaussian, respectively) and a multilayer perceptron (MLP), as well as four previously proposed linear discriminant functions (LDFs) and one LDF developed on the current data with all HRT parameters used as input.

Results. The areas under the ROC curves for SVM linear and SVM Gaussian were 0.938 and 0.945, respectively; for MLP, 0.941; for the current LDF, 0.906; and for the best previously proposed LDF, 0.890. With the use of forward selection and backward elimination optimization techniques, the areas under the ROC curves for SVM Gaussian and the current LDF were increased to approximately 0.96.

Conclusions. Trained neural networks, with global and regional HRT parameters used as input, improve on previously proposed HRT parameter-based LDFs for discriminating between glaucomatous and nonglaucomatous eyes. The performance of both neural networks and LDFs can be improved with optimization of the features in the input. Neural network analyses show promise for increasing diagnostic accuracy of tests for glaucoma. (*Invest Ophthalmol Vis Sci.* 2002;43: 3444–3454)

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

Supported by The Glaucoma Research Foundation (CB), and by National Eye Institute Grants EY13235 (MHG) and EY11008 (LMZ).

Submitted for publication February 4, 2002; revised June 5, 2002; accepted June 17, 2002.

Commercial relationships policy: N.

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, Department of Ophthalmology, University of California at San Diego, La Jolla, CA 92093-0946; cbowd@eyecenter.ucsd.edu.

Methods of early detection of glaucoma often focus on the assessment of optic disc topography and retinal nerve fiber layer (RNFL) thickness in an attempt to identify patients at risk for development of visual field defects. Because clinical examination and fundus photography are subjective and qualitative, optical imaging techniques that provide objective and quantitative measures for evaluating the optic disc and RNFL may be advantageous. For example, confocal scanning laser ophthalmoscopy (CSLO) provides quantitative measures that are reproducible and correlate with histomorphometric measurements in monkey eyes.[1–5] CSLO shows promise for discriminating between eyes with characteristic glaucomatous damage and healthy eyes, although the reported success for classifying these eyes varies.[6–9]

In an attempt to classify eyes effectively as glaucomatous or healthy, analysis strategies have been developed that use as input different CSLO optic disc topography measurement parameters, by using statistical methods such as linear discriminant function (LDF) analyses.[8,10–14] LDF analysis assumes that data representing different groups are linearly separable. If this assumption is not well met, the classifier's performance is degraded. Other investigators have used artificial neural networks (specifically, multilayer perceptrons [MLPs] with back-propagated learning) trained on CSLO parameters to classify eyes as glaucomatous or healthy.[8,15] Using this method, the neural network classifier is trained to detect a relationship between input (CSLO parameters) and a predefined gold-standard diagnosis by comparing its prediction with the labeled diagnosis and by learning from its mistakes. In general, neural network techniques differ from basic statistical techniques such as LDFs, because they can adapt to the distribution of the data rather than assume a predefined distribution. The success of statistical or neural network classification methods is most often measured by reporting areas under the receiver operating characteristic (ROC) curve or by reporting sensitivity at different specificities.

The purpose of the current study was to compare the performance of previously proposed HRT parameter-based LDFs with three artificial neural network methods in a single sample. Comparing different classification methods in a single sample reduces the effects of confounding variables, such as subject demographics and severity of glaucoma. Because of their adaptability, we hypothesized that neural network techniques would perform as well as or better than LDF classifiers in discriminating between glaucomatous and healthy eyes.

Methods

Subjects

One randomly selected eye from each of 108 patients with glaucoma and 189 normal subjects was included in the study. All subjects underwent a complete ophthalmic examination, including slit lamp biomicroscopy, measurement of intraocular pressure (IOP), stereoscopic fundus examination, stereoscopic photography of the optic disc, and

TABLE 1. HRT Parameters Included in the Full Dimensional Input Set

Parameter	Location
Disc area (mm^2)	(Global, T, TS, TI, N, NS, NI)
Area below reference (Cup Area) (mm^2)	(Global, T, TS, TI, N, NS, NI)
Mean height contour (mm)	(Global, T, TS, TI, N, NS, NI)
Peak height contour (mm)	(Global, T, TS, TI, N, NS, NI)
Height variation contour (mm)	(Global)
Volume below surface (mm^3)	(Global, T, TS, TI, N, NS, NI)
Volume above surface (mm^3)	(Global, T, TS, TI, N, NS, NI)
Volume below reference (mm^3)	(Global, T, TS, TI, N, NS, NI)
Volume above reference (mm^3)	(Global, T, TS, TI, N, NS, NI)
Maximum cup depth (mm)	(Global, T, TS, TI, N, NS, NI)
Cup shape	(Global, T, TS, TI, N, NS, NI)
Mean cup depth (mm)	(Global)
RNFL thickness (mm^2)	(Global)
RNFL cross-sectional area (mm^3)	(Global)
Reference height (mm)	(Global)
Area above reference (Rim Area) (mm^2)	(Global)
Cup-to-Disk area ratio	(Global, T, TS, TI, N, NS, NI)

T, temporal; TS, temporal superior; TI, temporal inferior; N, nasal; NS, nasal superior; NI, nasal inferior.

standard full-threshold automated perimetry (SAP; Humphrey Field Analyzer, Humphrey Instruments, Dublin, CA). Informed consent was obtained from all participants and the University of California, San Diego Human Subjects Committee approved all methodology. All methods adhered to the provisions of the Declaration of Helsinki guidelines for research in human subjects.

Because CSLO-measured optic disc topography was being evaluated, we chose the best indicator of glaucoma that is not dependent on optic disc appearance for training and evaluating the neural network techniques. Patients with open-angle glaucoma were defined as those with at least two consecutive SAP fields with either a corrected pattern standard deviation (CPSD) outside the 95% normal limits or a glaucoma hemifield test (GHT) result outside the 99% normal limits. At least one of the abnormal fields was obtained on or before the date of CSLO imaging. Mean deviation (±SD) of the SAP closest to the CSLO imaging date was −6.08 ± 5.77 dB, indicating mild to moderate visual field damage. Patients with glaucoma had no history of diabetes and no apparent cataracts and were not using medication known to affect visual sensitivity at the time of visual field testing. Best corrected visual acuity at the time of SAP and CSLO testing was 20/40 or better. The average age (±SD) of patients with glaucoma was 65.2 ± 13.6 years.

Healthy eyes had a measured IOP of 22 mm Hg or more with no history of elevated IOP. These eyes had intact rims, no evidence of hemorrhage, notching, glaucomatous excavation, or RNFL defect and had symmetrical optic discs (asymmetry of vertical cup-disc ratio < 0.2) based on clinical examination. SAP results were within normal limits. Healthy patients had no history of diabetes or other systemic disease and no ophthalmic or neurologic surgery or disease. Best corrected visual acuity at the time of testing was 20/40 or better. The average age (±SD) of healthy subjects was 54.2 ± 16.3 years, significantly younger than patients with glaucoma (*t*-test; $P < 0.05$).

Confocal Scanning Laser Ophthalmoscope

The Heidelberg Retina Tomograph (HRT-1, Heidelberg Engineering, Heidelberg, Germany) provides topographical measures of the optic disc and parapapillary retina, with confocal scanning laser technology. The topographical image is derived from 32 optical sections at consecutive focal depth planes. Each image consists of 256 × 256 pixels with each pixel corresponding to retinal height at its location. This instrument has been discussed in detail elsewhere.[1,16]

Procedure. Three 15° field-of-view scans centered on the optic disc and judged to be of acceptable quality were obtained for each test eye. A mean topography image of these three scans was created with the HRT. The optic disc margin was outlined on the mean topography image by a trained technician using information obtained by viewing stereoscopic photographs of the optic disc.

HRT Parameters. Eighty-three topographic parameters (automatically provided by HRT software, ver. 2.01) were used in this study (Table 1). We used global (360-degree) measures for each parameter and for some parameters, also used regional measures. 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 temporal (316–45°). Regional parameters were not evaluated for height variation of contour, mean cup depth, RNFL thickness, RNFL cross-sectional area, reference height, or rim area. All these parameters have been discussed in more detail elsewhere.[1,16–18]

Linear Discriminant Functions

We evaluated the performance of four published linear discriminant analysis formulas developed by Mikelberg et al.,[10] Bathija et al.,[12]

TABLE 2. Area under the ROC Curve and Sensitivities

Technique	Area Under ROC Curve: Total ± SE	Area Under ROC Curve: Specificity 0.9–1.0	Sensitivity (%): Specificity of 75%	Sensitivity (%): Specificity of 90%
LDFs				
Current LDF	0.906 ± 0.02	0.184	88	81
LDF Bathija et al.	0.890 ± 0.02	0.134	83	67
LDF Mardin et al.	0.873 ± 0.02	0.143	81	70
LDF Iester et al.	0.860 ± 0.02	0.138	80	69
LDF Mikelberg et al.	0.848 ± 0.02	0.116	81	64
Neural networks				
SVM Gaussian	0.945 ± 0.01	0.203	92	83
SVM linear	0.938 ± 0.01	0.182	91	78
MLP	0.941 ± 0.01	0.178	95	78
Optimized techniques				
SVM Gaussian (forward selection)	0.966 ± 0.01	0.236	97	91
SVM Gaussian (backward elimination)	0.965 ± 0.01	0.213	98	85
Current LDF (forward selection)	0.960 ± 0.01	0.213	95	86
Current LDF (backward elimination)	0.961 ± 0.01	0.223	95	88

Techniques are ordered based on largest area under the ROC curve in each analysis class (LDF, neural network, or optimized technique).

* Data adapted, with permission, from journals listed in Figure 1.

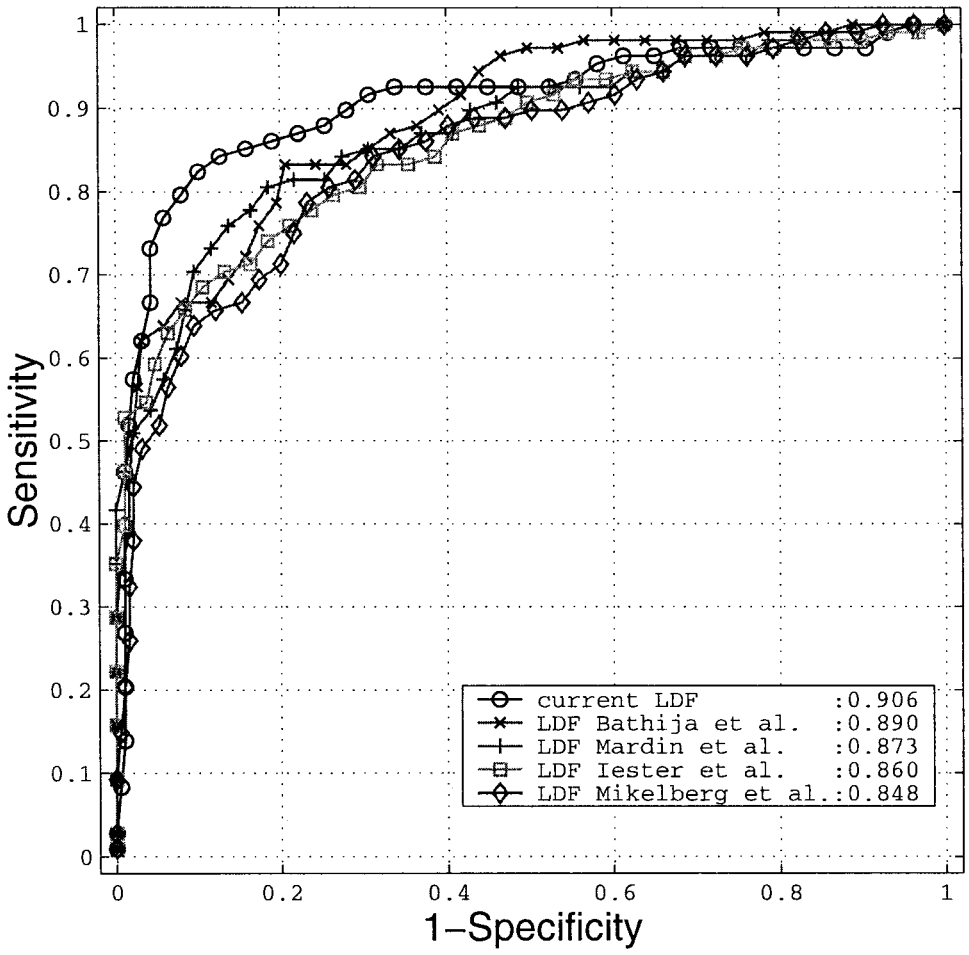

FIGURE 1. ROC curves (and areas under the curves) for the four previously proposed LDFs and current LDF. Areas for current LDF, LDF Bathija et al., and LDF Mardin et al. were significantly greater than the area for LDF Mikelberg et al.

Mardin et al.,[11] and Iester et al.[14] (and Iester M, personal communication, June 2001) for classifying eyes as glaucomatous or healthy. These formulas were developed with the available optic disc topography parameters provided by the HRT software. The Mikelberg et al.[10] formula is available in HRT software version 2.01 as "glaucoma classification."

- LDF Mikelberg et al.[10]: (rim volume × 1.95) + (height variation contour × 30.12) − (corrected cup shape × 28.52) − 10.08, where corrected cup shape is cup shape + [0.0019 × (50 − age)]
- LDF Bathija et al.[12]: −3.72 − (5.57 × height variation contour) + (11.78 × RNFL thickness) − (4.37 × cup shape) + (1.85 × rim area)
- LDF Mardin et al.[11]: −2.77 + (0.3 × rim area) + (3.70 × rim volume) + (4.30 × RNFL thickness) − (3.70 × cup shape)− (3.10 × cup volume) − (0.90 × cup area)
- LDF Iester et al.[14]: (10.07 × cup area temporal inferior sector) − (7.02 × effective area temporal inferior sector) + (4.18 × mean height contour nasal sector) + (3.10 × mean height contour temporal sector) − (2.08 × peak height contour nasal superior sector) + (6.09 × cup shape) − (11.09 × rim volume temporal superior sector) − (8.05 × volume below surface temporal sector) + 1.83

We also developed and evaluated an LDF (called "current" LDF) that used all 83 parameters as input. This LDF was developed and tested, with 10-fold cross-validation used to reduce bias in developing and testing on the same samples (described later).

Neural Network Techniques

We evaluated the performance of three artificial neural network techniques for classifying eyes as glaucomatous or healthy. For all neural network techniques, all HRT parameters described earlier were included initially in the training set. Details and mathematical descriptions of the neural network techniques used have been described elsewhere by us and by others.[19–24]

Multilayer Perceptron. The MLP, a feed-forward back-propagation network, is the most frequently used neural network technique in glaucoma research. Researchers have used this method to assess optic disc topography,[8,15] to interpret and classify visual fields[19,25–28] and to detect visual field progression.[29] Briefly, MLPs are supervised learning classifiers that consist of an input layer (multiple HRT parameters, in our case), an output layer (glaucoma or not glaucoma, in our case), and one or more hidden layers that extract useful information during learning and assign modifiable weighting coefficients to components of the input layer. In the first (forward) pass, weights assigned to the input units and the nodes in the hidden layers and between the nodes in the hidden layer and the output, determine the output. The output is compared with the target output (binary glaucoma or no glaucoma). An error signal is then back propagated and the connection weights are adjusted correspondingly. During training, MLPs construct a multidimensional space, defined by activation of the hidden nodes, so that the two classes (glaucoma, not glaucoma) are as separable as possible. The separating surface adapts to the data.

We used a 10-unit MLP in the present study constructed in a commercial software program (Neural Network toolbox ver. 3.0 of Matlab; The MathWorks, Inc., Natick, MA). Input nodes fed into a 10-node hidden layer activated by hyperbolic tangent functions. Output was a single node with a logistic function for glaucoma (1) and healthy (0) eyes. Training was accomplished with the Levenberg-Marquant enhancement of back propagation. We evaluated MLPs with different numbers of units and found the 10-unit MLP performed best, as measured by performance of cross-validation.

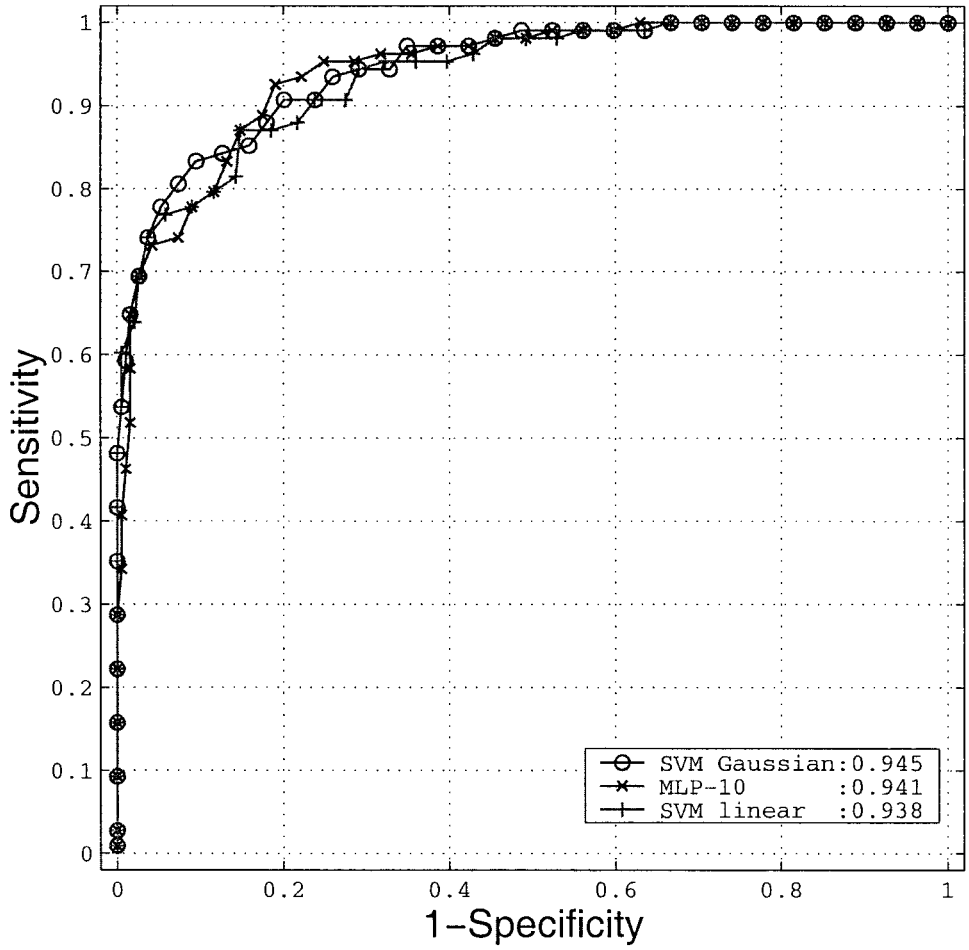

FIGURE 2. ROC curves (and areas) for the three neural network techniques investigated. No statistically significant differences between neural network areas under the ROC curve were observed.

Linear Kernel (SVM linear) and Gaussian Kernel Support Vector Machine (SVM Gaussian). SVMs are newly developed techniques used for solving classification and regression problems. SVM architecture resembles the architecture of MLPs (input layer, hidden layer, output layer). During training, the SVM nonlinearly maps the training data to a high dimensional space where a hyper plane is fit that maximizes the margin of separation between classes while minimizing the generalization error (ability to generalize results from finite training set to data set), with the use of 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 SVM attempts to split the positive and negative vectors to optimize the distance between the hyperplane and the nearest of the 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 unknown nonlinear mapping, resulting in a Gaussian kernel. Both SVM linear and SVM Gaussian have been used to classify eyes as glaucomatous or nonglaucomatous, based on visual field data.[19]

The SVM was programmed using the software program (Matlab, ver. 5.0; the MathWorks) and trained using Platt's sequential minimal optimization algorithm. The programmer chose the parameters for penalty and the kernel by trial and error. The penalty used was $C = 1.0$.

Analysis

ROC curves for classifying eyes as glaucomatous or healthy were determined for all techniques. These curves describe the continuous relationship between sensitivity and specificity at specificities ranging from 0% to 100% and quantify the diagnostic accuracy of a test in a single number. An area under the ROC curve of 0.50 is equivalent to chance discrimination, and an area of 1.00 is equivalent to perfect discrimination. For SVMs and current LDF, 10-fold cross-validation was used to evaluate the classifiers. The glaucomatous and healthy eyes were each divided randomly into 10 approximately equal subsets. Ten mutually exclusive partitions were formed for cross validation (to measure the true rather than the estimated error rate) by combining one of the 10 healthy subsets with one of the 10 glaucoma subsets. One partition was used as the test set and the remaining nine partitions were combined to form the training set. The process was iterated, with each partition 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. For MLP, cross-validation was similar, except eight partitions were used for training, one was used as a test set, and one was used as a stopping set to avoid overtraining. We provided sensitivities at specificities of 75% (representing moderate specificity) and 90% (representing high specificity), although this information is available in the graphic representations of ROC curves also presented. Finally, we reported the area under the ROC curve when specificity was 90% or more for the different techniques. These areas are bound by the ROC curve, the point at 100% specificity, and the line that passes through the point at 90% specificity and is perpendicular to the diagonal that represents chance discrimination. This information was provided to examine differences between techniques when specificity was high. The 90% specificity level was chosen because it theoretically forces the cases presumed to be the most difficult into the disease group by allowing only 10% of these cases into the healthy group.

We used the method of DeLong et al.[30] to determine statistically significant differences in overall areas under the ROC curves.

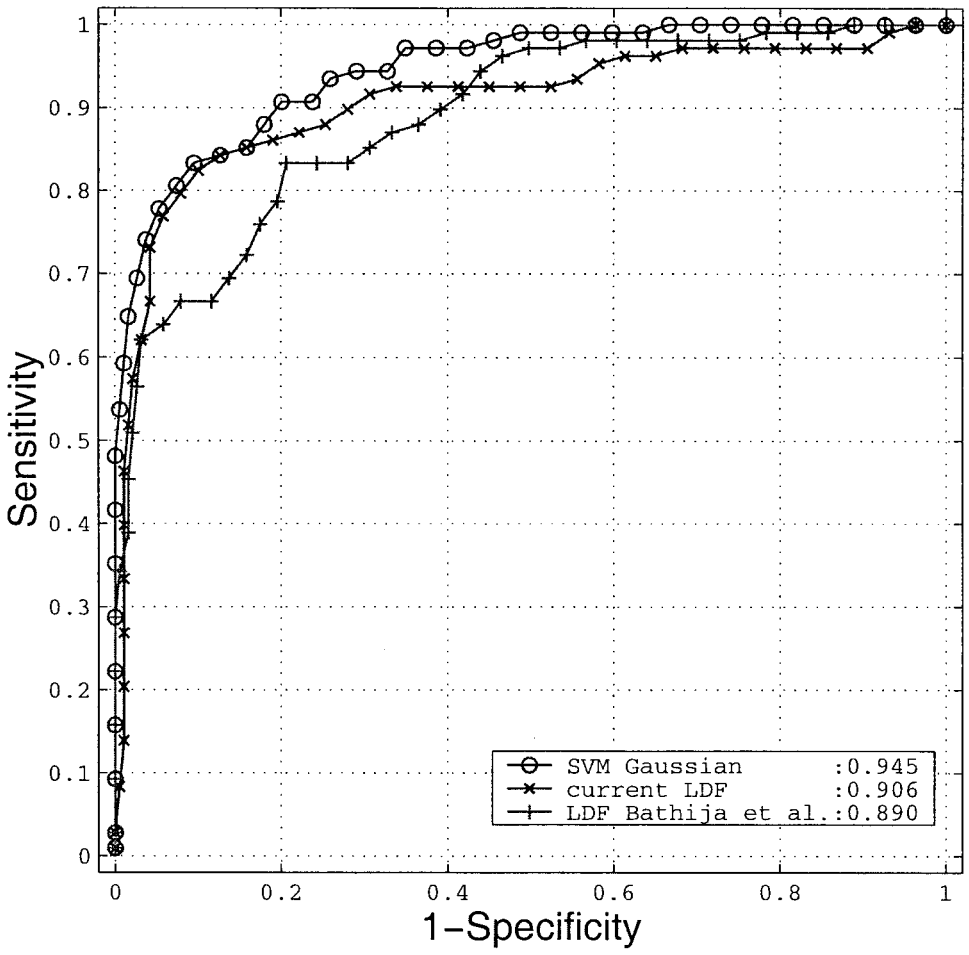

FIGURE 3. ROC curves (and areas under the curves) for the best LDF (Bathija et al., attribution in Figure 1), best neural network technique (SVM Gaussian), and the current LDF. The area for SVM Gaussian is significantly greater than that for both LDFs.

RESULTS

Comparing Linear Discriminant Functions

Areas under the ROC curves (with sensitivities at 75% and 90% specificities) for all classification techniques evaluated are shown in Table 2. The area under the ROC curve (± SE) for the best-performing LDF was for the current LDF (0.906 ± 0.02). The area for the best previously proposed LDF was for LDF Bathija et al.[12] (0.890 ± 0.02), followed by LDF Mardin et al.[11] (0.873 ± 0.02), LDF Iester et al.[14] (0.860 ± 0.02), and the HRT classification by Mikelberg et al.[10] (0.848 ± 0.02). Areas under the ROC curves of the current LDF, LDF Bathija et al., and LDF Mardin et al. were significantly greater than that of LDF Mikelberg et al. (all $P = 0.02$). No other statistically significant differences between areas under the ROC curves of proposed LDF were observed. ROC curves for the five LDFs are shown in Figure 1. Areas under the curves when specificity was constrained from 90% to 100% were 0.184, 0.134, 0.143, 0.138, and 0.116, for current LDF, LDF Bathija et al., LDF Mardin et al., LDF Iester et al., and LDF Mikelberg et al., respectively.

Sensitivities at 75% specificity for current LDF, LDF Bathija et al., LDF Mardin et al., LDF Iester et al., and LDF Mikelberg et al., were 88%, 83%, 81%, 80%, and 81%, respectively. Sensitivities at 90% specificity were 81%, 67%, 70%, 69%, and 64%.

Comparing Neural Network Techniques

For MLP, the area under the ROC curve was 0.941 ± 0.01; for SVM linear, 0.938 ± 0.01; and for SVM Gaussian, 0.945 ± 0.01. No statistically significant differences between areas under the curves of neural network techniques were observed. ROC curves for the three neural network techniques are shown in Figure 2. Areas under the curves when specificity was constrained from 90% to 100% were 0.178, 0.182, and 0.203, for MLP, SVM linear, and SVM Gaussian, respectively.

Sensitivities at 75% specificity for MLP, SVM linear, and SVM Gaussian were 95%, 91%, and 92%, respectively; sensitivities at 90% specificity were 78%, 78%, and 83%, respectively.

Comparing Linear Discriminant Functions with Neural Network Techniques

Areas under the ROC curves were significantly higher for MLP, SVM linear, and SVM Gaussian than with all previously proposed LDFs (all $P < 0.01$), and with the current LDF (all $P < 0.03$). ROC curves for the best neural network (SVM Gaussian), the current LDF, and the best previously proposed LDF (LDF Bathija et al.) are shown in Figure 3.

Optimizing Neural Network and LDF Results

The neural network technique that provided the largest area under the ROC curve when all HRT parameters were included as input to the training set was the Gaussian SVM. We performed feature selection both by sequential forward selection and sequential backward elimination of features[31] to determine whether relying on more effective features and removing less effective features would improve the performance of a classifier as measured by area under the ROC curve. During forward selection, an optimum training (input) set was determined by starting with an empty subset and adding one input parameter

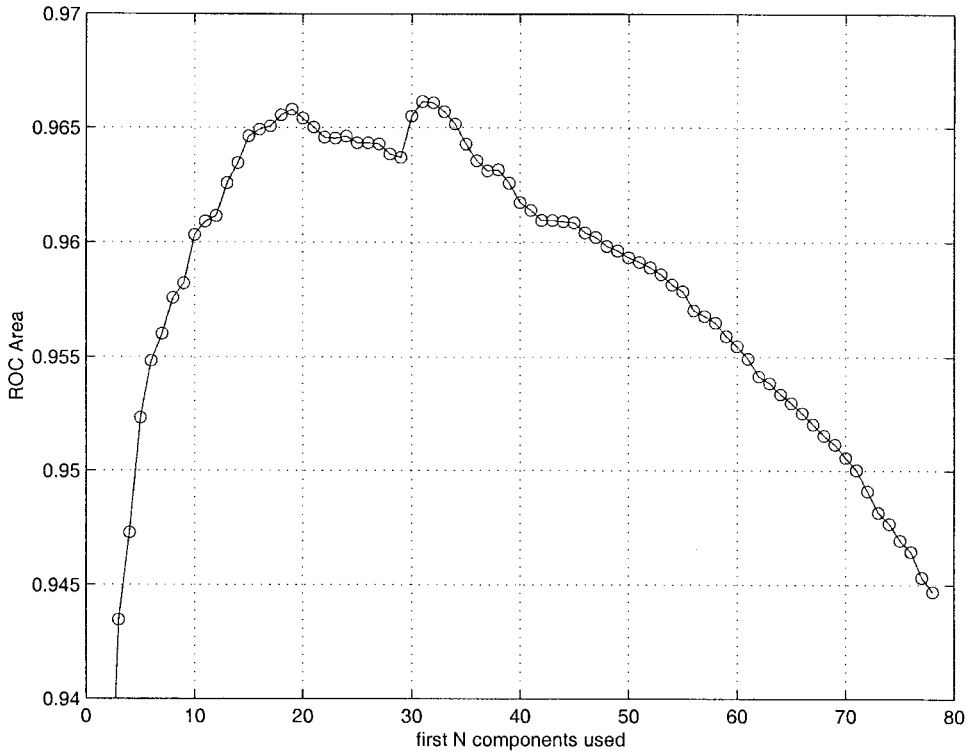

FIGURE 4. Use of forward selection to determine an optimum training set for SVM Gaussian. The area under the ROC curve (*y*-axis) is shown as a function of the number of HRT parameters in the training set (*x*-axis). The training set is optimized at maximum area under the curve (0.976, $n = 31$ parameters).

at a time (e.g., the one that most increased the area under the curve in combination with the previously selected parameters) to the previously selected features until the area reached a maximum. During backward elimination, an optimal training set was found by starting with the full dimensional set from which the least effective input parameter was removed, one input parameter at a time (e.g., the one that resulted in the smallest increase in area under the ROC curve) until the maximum area was reached.

Figures 4 and 5 show that we achieved the optimal area under the ROC curve with either forward selection or backward elimination when we were using approximately 40% of the input parameters. These figures show areas under the ROC curve (*y*-axis) as a function of the number of HRT parameters in the training set (*x*-axis). The areas were maximized with a reduced dimension data set (subset of available input parameters) that contained an optimal combination of features determined by each optimization method, compared with using the

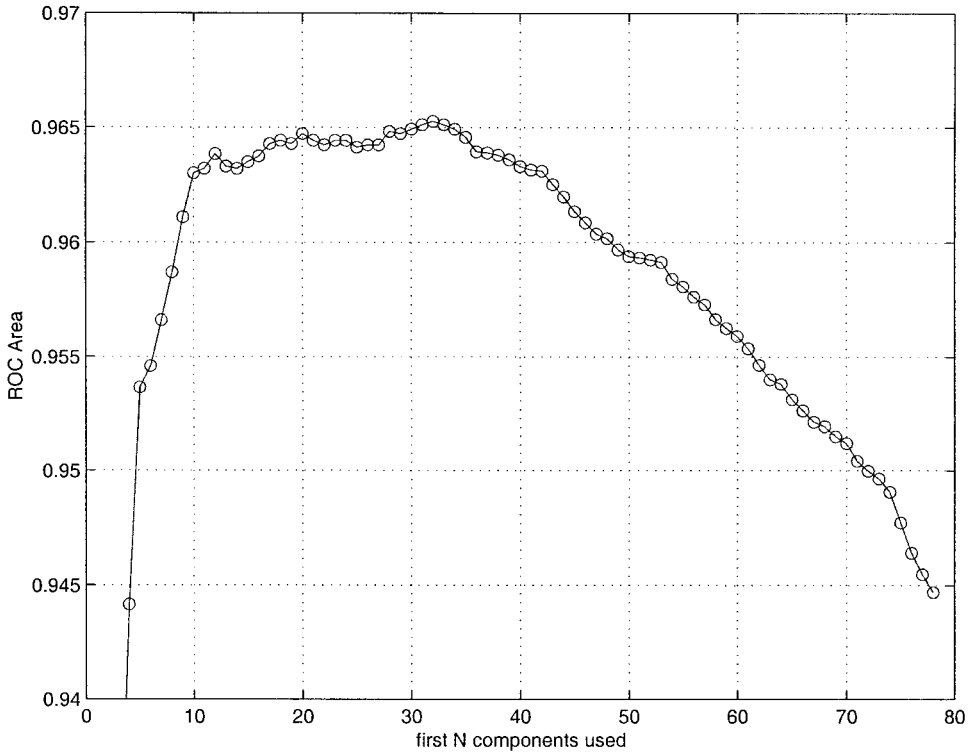

FIGURE 5. Use of backward elimination to determine an optimum training set for SVM Gaussian. The area under the ROC curve (*y*-axis) is shown as a function of number of HRT parameters in the training set (*x*-axis). The training set is optimized at maximum area under the curve (0.965, $n = 32$ parameters).

TABLE 3. HRT Parameters Included in Optimized Training Sets for SVM Gaussian and Current LDF

SVM Gaussian (Forward Selection)	SVM Gaussian (Backward Elimination)	Current LDF (Forward Selection)	Current LDF (Backward Elimination)
Peak height contour TI	Peak height contour TI	Peak height contour TI	Mean height contour TI
Cup shape G	Cup disk ratio TS	Cup shape G	Cup shape G
Volume below surf G	Disc area T	Volume above surf NS	Reference height
Volume above ref TS	Volume above surf NS	Volume above ref G	Rim area
Disc area N	Area below ref NS	Volume below surf NS	Disc area TI
Volume below ref G	Area below ref TI	Volume above surf NI	Area below ref G
Area below ref TS	Volume above surf NI	Volume above surf NI	Area below ref NS
Volume below ref NI	Cup disk ratio NS	Cup shape N	Cup disk ratio NS
Volume below surf T	Cup shape G	Area below ref NS	Cup disk ratio TS
Cup disk ratio NS	Volume below surf NS	Volume below ref NI	Mean height contour NI
Max cup depth NS	Volume above surf TI	Cup shape TS	Volume above ref G
Volume above surf N	Max cup depth N	Area below ref NI	Volume below surf G
Disc area T	Disc area N	Cup disk ratio N	Disc area G
Area below ref TI	Height var contour G	Area below ref TS	Mean height contour NS
Area below ref NS	Mean height contour TI	Peak height contour T	Volume below surf T
Cup shape TI	Cup disk ratio TI	Disc area T	Volume above ref T
Volume below surf NS	Area below ref TS	Disc area TI	RNFL thickness G
Volume above ref NS	Volume below ref T	Volume below surf N	Mean height contour N
Volume below surf TS	Cup shape TI	Mean height contour T	Disc area T
Area below ref NI	Max cup depth T	Cup disk ratio TS	Mean height contour TS
Volume below surf N	Cup disk ratio NI	Reference height	Volume below surf NS
Area below ref G	Mean height contour NI	Volume above ref TS	Volume above surf T
Cup disk ratio TS	Disc area TI	RNFL cross section G	Max cup depth G
Volume below ref T	Cup disk ratio G	Volume below surf TI	Mean cup depth G
Cup disk ratio NI	Max cup depth NS	Max cup depth G	Max cup depth NI
Cup shape T	Volume above ref NS	Max cup depth NS	Peak height contour TS
Volume below surf NI	Disc area NS	Mean cup depth G	Cup shape TI
Volume below surf TI	Volume below ref G		Cup disk ratio TI
Volume above surf NS	Volume below surf NI		Peak height contour TI
Cup shape NS	Area below ref NI		
Cup disk ratio G	Volume below surf G		
Area below ref N			

Parameters are ranked as having the most to least effect on increasing (forward selection) or decreasing (backward selection) ROC curve area. Regions of parameters are defined in Table 1. Max, maximum; ref, reference; surf, surface; var, variation.

full-dimensional feature set (all available input parameters). Using forward selection, the area under the ROC curve (± SE) increased from 0.945 (± 0.01) with all input parameters, to a maximum of 0.967 (± 0.01) with 31 input parameters. When the optimal feature set was analyzed at specificities constrained from 90% to 100%, the area under the ROC curve increased from 0.203 to 0.236. Sensitivity at 75% specificity increased from 92% to 97%, and sensitivity at 90% increased from 83% to 91%. When backward elimination was used, the area under the ROC curve increased to 0.965 ± 0.01 and reached its maximum with 32 input parameters. When specificity was constrained from 90% to 100%, the area was 0.213. Sensitivity at 75% specificity was 98% and sensitivity at 90% specificity was 85%. HRT parameters included in the optimized SVM Gaussian training set with both methods are shown in Table 3.

In an attempt to maximize the performance of the current LDF, we optimized the training set by using the same methods described herein. When forward selection was used, the area under the ROC curve (±SE) increased from 0.906 ± 0.02, with all input parameters, to 0.960 ± 0.01, with 29 input parameters. The areas when specificity was constrained from 90% to 100% increased from 0.184, with all input parameters, to 0.213, with 29 input parameters. Sensitivity at 75% specificity increased from 88% to 95%, and sensitivity at 90% increased from 81% to 86%. When backward elimination was used, the area under the ROC curve increased to 0.961 ± 0.01, with 27 input parameters. When specificity was constrained from 90% to 100%, the area was 0.223. Sensitivity at 75% specificity was 95% and sensitivity at 90% specificity was 88%. HRT parameters included in the optimized current LDF training set with both methods are shown in Table 3. The glaucomatous–healthy classification performance of the optimized current LDF was similar to that of the optimized SVM Gaussian, indicating that, with an optimal feature set, the data are linearly separable, and adaptive classifiers may not be necessary. Areas under the ROC curves for the optimized and full-dimensional current LDF and SVM Gaussian are shown in Figure 6.

Optimal Parameters within the Full-Dimensional Input

To determine some of the most informative HRT parameters, we identified a subset of input parameters from the full dimensional data set that most affected the area under the ROC by using forward selection and backward elimination.[32] With each optimization method, input parameters were ranked from having the most (rank 1) to the least (rank 78) effect on the area under the curve when combined with other effective parameters. These ranks were plotted on a two-dimensional graph (forward selection rank on the *y*-axis, backward elimination rank on the *x*-axis). Those parameters closest to the origin were considered the most informative ones, because they presumably had the greatest influence on the area under the ROC curve with both optimization methods. This method was applied to both SVM Gaussian and current LDF (Figs. 7, 8). For SVM Gaussian, the three most informative parameters were peak height contour in the temporal inferior region, global cup shape, and disc area in the nasal region. For current LDF, the three most informative parameters were global cup shape,

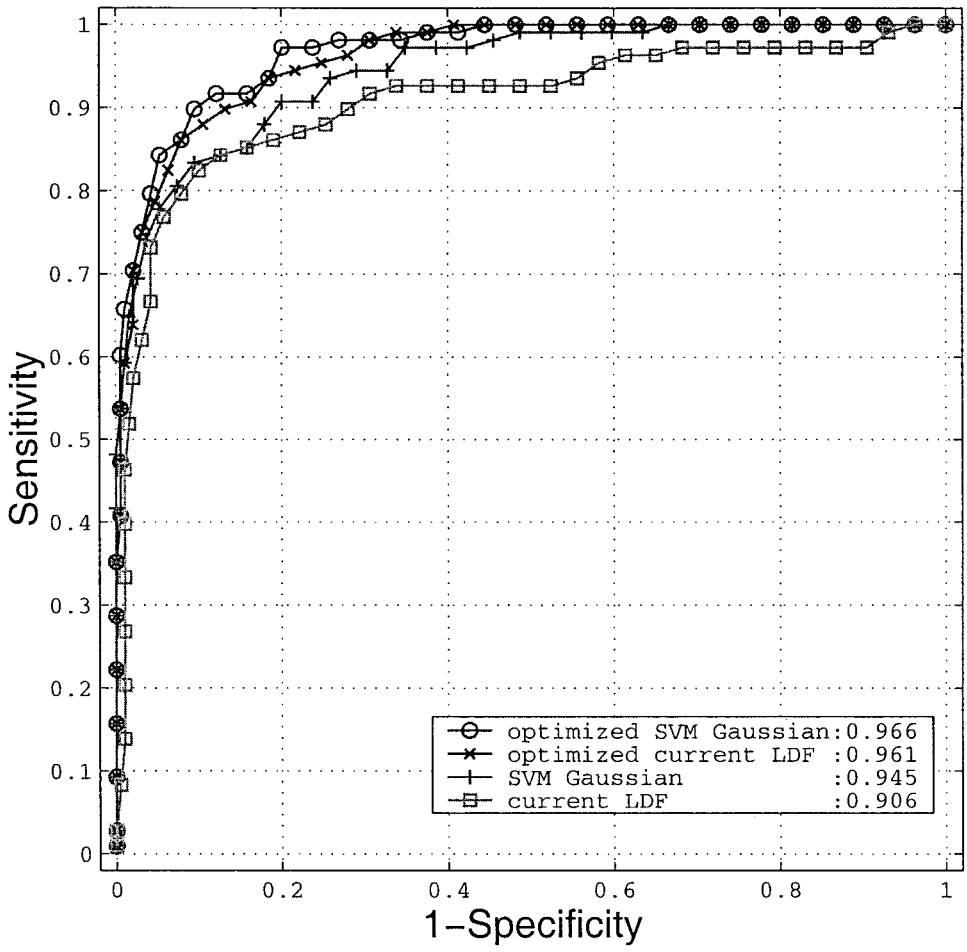

FIGURE 6. ROC curves (and areas) for optimized and nonoptimized SVM Gaussian and current LDF. The areas under the ROC curves for optimized SVM Gaussian and optimized current LDF were significantly greater than the areas for nonoptimized SVM Gaussian and nonoptimized current LDF.

global rim volume, and cup area (area below reference) in the nasal superior region.

DISCUSSION

In our sample, all investigated HRT-based neural network techniques performed as well as or better than the HRT-based linear discriminant functions. ROC curves for nonoptimized neural network techniques ranged from 0.938 to 0.945, compared with 0.848 to 0.906 for LDF methods. Further, optimization of the feature set significantly increased discrimination ability, probably because of the removal of parameters that add information that has less value than the cost of including them in the training process. These results suggest that neural network classification techniques trained on HRT parameters are promising for discriminating between healthy eyes and those with mild to moderate glaucomatous visual field defects.

In the present study, the nonoptimized technique that resulted in the largest area under the ROC curve for discriminating between glaucomatous and healthy eyes (area under ROC curve = 0.945 for SVM Gaussian) yielded a sensitivity of 83% at 90% specificity. In previous work, Uchida et al.[8] reported an area under the ROC curve of 0.94 and sensitivity and specificity of 92% and 91%, respectively, when using a back-propagation multilayer perceptron trained with nine global HRT parameters. These results are similar to the best optimized results from the present study (optimized SVM Gaussian: 91% sensitivity at 90% specificity). Severity of glaucoma in the patients in the present study was slightly higher than that of Uchida et al. (SAP mean deviation of −6.1 and −4.8 dB, respectively). Our study is the first to investigate the performance of SVMs trained on optical imaging data for discriminating between glaucomatous and healthy eyes.

Other studies have examined the success of individual HRT parameters, linear discriminant analyses, and specially developed parameters for classifying eyes as glaucomatous or healthy. For example, Iester et al.[6] and Uchida et al.[8] found the HRT cup shape measure to be the best individual parameter for identifying glaucoma. These authors reported areas under the ROC curve for glaucoma detection of 0.81 and 0.84, respectively, using this parameter. These areas are slightly higher than that reported by Zangwill et al.[9] (0.78) for the same parameter.

The most frequently investigated HRT parameter-based LDF is that developed by Mikelberg et al.[10] Using this model, reported sensitivity for detecting glaucomatous eyes (defined by abnormal visual fields and/or abnormal appearing optic discs) ranges from 42% to 92%, and reported specificity for detecting healthy eyes ranges from 84% to 96%.[9,10,12,33] In the current study, sensitivity was 81% and 64% at 75% and 90% specificity, respectively. Using other HRT parameter-based LDFs for discriminating between glaucomatous and healthy eyes, Mardin et al.[11] and Iester et al.,[14] reported sensitivities of 84% and 70%, respectively, and specificities of 95% and 92%, respectively. In the present study, at a set specificity of 90%, we found a sensitivity of 70% with the Mardin et al. LDF and a sensitivity of 69%, with the Iester et al. LDFs. Finally, Bathija et al.[12] reported a sensitivity and specificity of 62% and 94%, respectively, within a demographically similar sample with similar inclusion criteria and severity of glaucoma as in the current study.

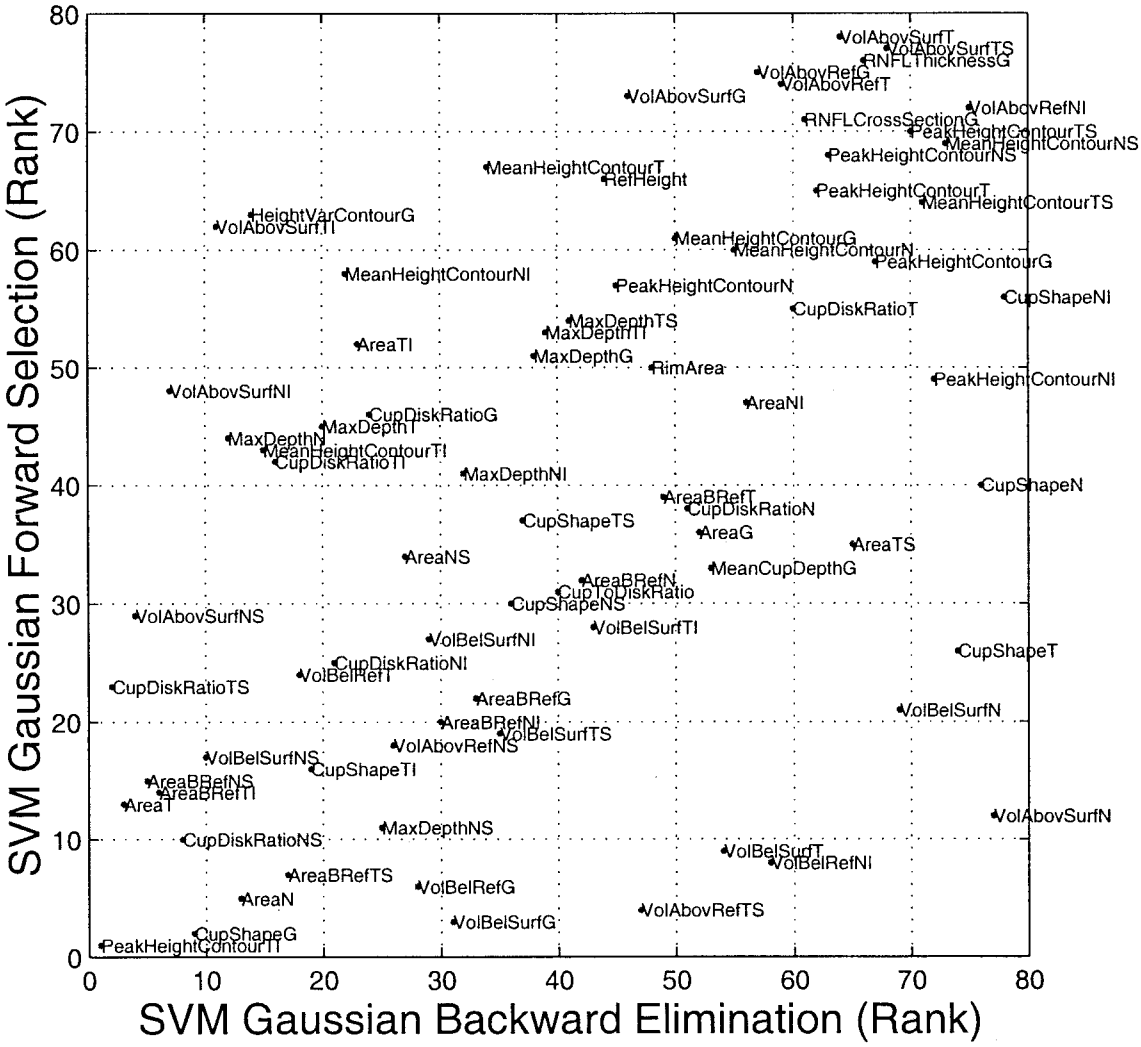

FIGURE 7. Identifying optimal parameters for full dimensional input for SVM Gaussian. Input parameters (see Table 1) are ranked from having the most (rank 1) to the least (rank 78) effect on area under the ROC curve with forward selection and backward elimination and are plotted on a two-dimensional graph. Parameters closest to the origin are considered the most informative ones because they presumably have the greatest influence on area under the ROC curve with both optimization methods.

Using non-standard HRT parameters to discriminate between glaucomatous and healthy eyes, Caprioli et al.[34] reported a sensitivity of 83% and specificity of 85% (parapapillary slope derived from radial height measures around the disc), Iester et al.[14] reported a sensitivity of 65% and specificity of 100% (measurement of retinal height differential), and Wollstein et al.[35] reported a sensitivity of 84% and specificity of 96% (rim area adjusted for disc area). The maximum sensitivity and specificity reported in the present study were 91% and 90%, respectively (for optimized SVM Gaussian). Comparisons across studies are difficult, however, because of differences in population demographics, definition and severity of glaucoma, and differences in sensitivity and specificity at chosen cutoff values. We include this information to provide a context for our results.

In the current study, we identified HRT parameters that most affected the area under the ROC curve for discriminating between glaucomatous and healthy eyes with forward selection and backward elimination used in the SVM Gaussian and current LDF techniques. For both SVM and LDF techniques, global cup shape was an important parameter, identified by both forward selection and backward elimination. This finding is interesting, because other research has shown that cup shape is among the best individual parameters for discriminating between glaucomatous and healthy eyes.[6,8,9] Other parameters identified by our techniques are novel. For instance, with the SVM Gaussian technique, peak height contour in the temporal inferior region and disc area in the nasal region were identified. It is possible that techniques without clinical bias may identify unexpected parameters that are important for discriminating glaucomatous from healthy eyes. The discordance in the most effective parameters with different classifiers may be explained by differences in the classifier reasoning process.

Two possible limitations of the current study were the lack of independent samples on which to develop and test current LDF and neural network techniques and the significant difference in age between healthy subjects and patients. Although we used cross-validation to train the neural network classifiers and the current LDF, part of the reason for the improved performance of these techniques compared with previously proposed LDFs might be that our techniques were trained and tested on groups with similar demographics and severity of glaucoma. We plan to test these techniques on outside populations. Because of the age difference between healthy subjects and patients, we did not include age as input when training the

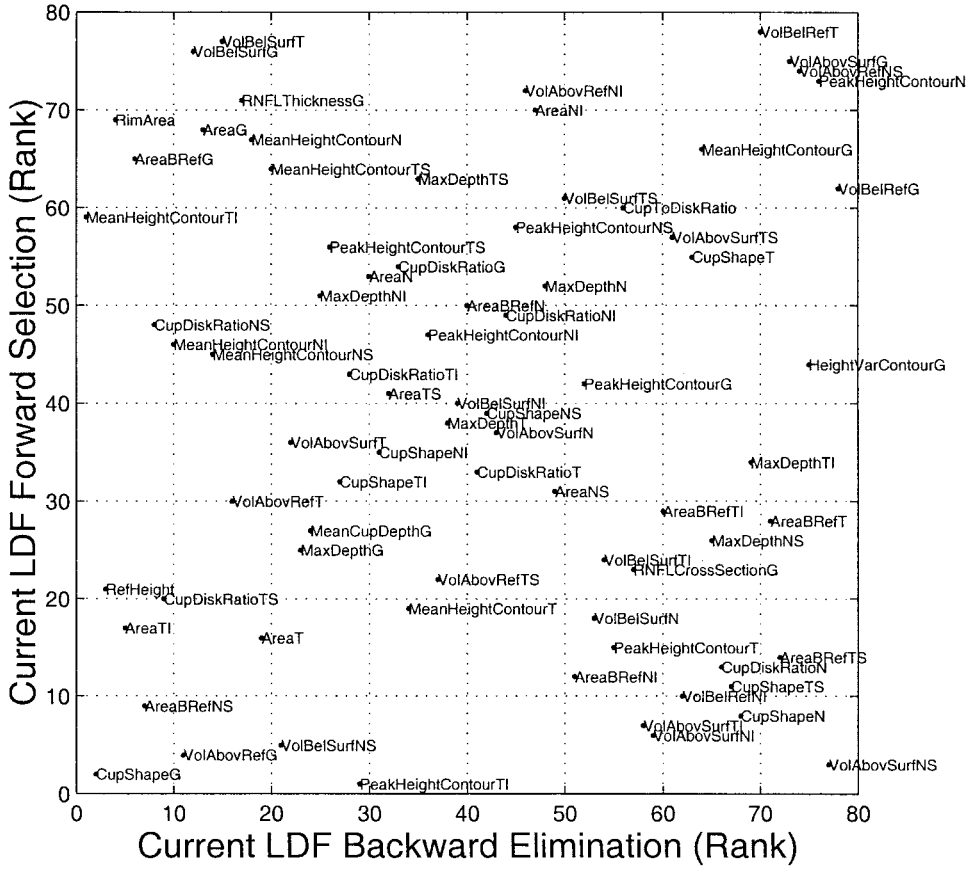

FIGURE 8. Identifying optimal parameters for full dimensional input for current LDF with forward selection and backward elimination.

neural networks or developing the current LDF. The inclusion of age might allow the neural networks to classify eyes as glaucomatous or healthy based on age alone. This classification is clearly not practical from a diagnostic standpoint. When age was included in the training set, areas under the ROC curves for all neural network techniques and current LDF increased by 0.01 or less. These increases were not statistically significant. Sensitivities at the chosen specificity cutoffs increased by less than 5%. We also trained and tested the neural network techniques on a subset of our data in which age in both the healthy subjects and patients was constrained to between 40 years and 81 years. This subset was composed of 133 healthy subjects (mean age, 65.10 years) and 90 patients with glaucoma (mean age, 66.78 years). Age was not significantly different between groups (*t*-test, $P > 0.10$). Areas under the ROC curve for all neural network techniques and current LDF changed by 0.01 or less and no changes were statistically significant. Sensitivities at the chosen specificity cutoffs changed by less than 5%.

Although neural networks successfully discriminated between healthy and glaucomatous eyes in this study, these techniques incur one general criticism. Due to the complexity of the classifiers, they do not allow the interaction of important variables to be identified and measured. Other classification techniques, such as Bayesian networks, allow better assessment of the relative contribution of features.

In summary, neural network techniques were more successful at discriminating between glaucomatous and healthy eyes than previously proposed LDFs. This improvement suggests that neural network techniques show at least as much potential for use in diagnosis of glaucoma as linear discriminant techniques. In addition, support vector machines demonstrated better generalization performance, and therefore better classification performance than MLPs (see also Refs. 19,34,35). This result, coupled with the fact that SVMs are faster to train than MLPs, suggests that SVMs show superior potential for use in diagnosis of glaucoma when compared with MLPs.

References

1. 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.
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. Janknecht P, Funk J. Optic nerve head analyser and Heidelberg retina tomograph: accuracy and reproducibility of topographic measurements in a model eye and in volunteers. *Br J Ophthalmol.* 1994;78:760–768.
4. 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.
5. 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.
6. Iester M, Mikelberg FS, Swindale NV, Drance SM. ROC analysis of Heidelberg Retina Tomograph optic disc shape measures in glaucoma. *Can J Ophthalmol.* 1997;32:382–388.
7. Nakla M, Nduaguba C, Rozier M, Joudeh M, Hoffman D, Caprioli J. Comparison of imaging techniques to detect glaucomatous optic nerve damage [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 1999; 40(4):**S**397. Abstract nr 2089.
8. 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.
9. 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.

10. 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.
11. 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.
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. Iester M, Parfitt CM, Swindale NV, Mikelberg FS. Sector-based analysis of Heidelberg Retina Tomograph (HRT) parameters in normal and glaucomatous eyes [ARVO abstract]. *Invest Ophthalmol Vis Sci.* 1997;38(4):**S**835. Abstract nr 3892.
15. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol.* 1996;121:511–521.
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.* 2002;19:197–207.
19. 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.
20. Rumelhart DE, Hinton G, Williams R. Learning representations of back-propagation errors. *Nature.* 1986;323:533–536.
21. Broomhead DS, Lowe D. Multivariable functional interpolation and adaptive networks. *Complex Syst.* 1988;2:321–355.
22. Bishop CM. *Neural Networks for Pattern Recognition.* Oxford, UK: Clarendon Press; 1995.
23. Vapnik V. *Statistical Learning Theory.* New York: Wiley; 1998.
24. Vapnik V. *The Nature of Statistical Learning Theory.* 2nd ed. New York: Springer; 2000.
25. Spenceley SE, Henson DB, Bull DR. Visual field analysis using artificial neural networks. *Ophthalmic Physiol Opt.* 1994;14:239–248.
26. Lietman T, Eng J, Katz J, Quigley HA. Neural networks for visual field analysis: how do they compare with other algorithms? *J Glaucoma.* 1999;8:77–80.
27. Goldbaum MH, Sample PA, White H, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci.* 1994;35:3362–3373.
28. Mutlukan E, Keating D. Visual field interpretation with a personal computer based neural network. *Eye.* 1994;8:321–323.
29. Brigatti L, Nouri-Mahdavi K, Weitzman M, Caprioli J. Automatic detection of glaucomatous visual field progression with neural networks. *Arch Ophthalmol.* 1997;115:725–758.
30. 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.
31. Ripley BD. *Pattern Recognition and Neural Networks.* Cambridge, UK: Cambridge University Press; 1996.
32. Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Scholkopf B, Smola A, eds. *Advances in Kernel Methods: Support Vector Machines.* Cambridge, MA: MIT Press; 1998:185–208.
33. Broadway DC, Drance SM, Parfitt CM, Mikelberg FS. The ability of scanning laser ophthalmoscopy to identify various glaucomatous optic disk appearances. *Am J Ophthalmol.* 1998;125:593–604.
34. Caprioli J, Park HJ, Ugurlu S, Hoffman D. Slope of the peripapillary nerve fiber layer surface in glaucoma. *Invest Ophthalmol Vis Sci.* 1998;39:2321–2328.
35. Wollstein G, Garway-Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology.* 1998;105:1557–1563.
36. Platt J, Cristianini N, Shawe-Taylor J. Large margin DAGs for multiclass classification. In: Solla SA, Leen TK, Müller K-R (eds.) *Advances in Neural Information Processing.* Vol 12. Cambridge, MA: MIT Press; 2000:547–553.
37. Roth V, Steinhag V. Nonlinear discriminant analysis using kernel functions. In: Solla SA, Leen TK, Müller K-R (eds.) *Advances in Neural Information Processing.* Vol 12. Cambridge, MA: MIT Press; 2000:568–574.