



Heidelberg Retina Tomograph 3 machine learning classifiers for glaucoma detection

K A Townsend, G Wollstein, D Danks, K R Sung, H Ishikawa, L Kagemann, M L Gabriele and J S Schuman

Br. J. Ophthalmol. 2008;92;814-818
doi:10.1136/bjo.2007.133074

Updated information and services can be found at:
<http://bjo.bmj.com/cgi/content/full/92/6/814>

These include:

References

This article cites 20 articles, 12 of which can be accessed free at:
<http://bjo.bmj.com/cgi/content/full/92/6/814#BIBL>

Rapid responses

You can respond to this article at:
<http://bjo.bmj.com/cgi/eletter-submit/92/6/814>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *British Journal of Ophthalmology* go to:
<http://journals.bmj.com/subscriptions/>

Heidelberg Retina Tomograph 3 machine learning classifiers for glaucoma detection

K A Townsend,¹ G Wollstein,¹ D Danks,^{2,3} K R Sung,¹ H Ishikawa,¹ L Kagemann,¹ M L Gabriele,¹ J S Schuman¹

¹UPMC Eye Center, Eye and Ear Institute, Ophthalmology and Visual Science Research Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; ²Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, USA; ³Institute for Human and Machine Cognition, Pensacola, FL, USA

Correspondence to: Dr J S Schuman, UPMC Eye Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, 203 Lothrop Street, Eye and Ear Institute, Suite 816, Pittsburgh, PA 15213, USA; schumanjs@upmc.edu

KAT and GW contributed equally to this study.

Meeting Presentation: Association for Research in Vision and Ophthalmology Annual Meeting, Ft Lauderdale, FL, May 2007.

Accepted 19 March 2008

ABSTRACT

Aims: To assess performance of classifiers trained on Heidelberg Retina Tomograph 3 (HRT3) parameters for discriminating between healthy and glaucomatous eyes.

Methods: Classifiers were trained using HRT3 parameters from 60 healthy subjects and 140 glaucomatous subjects. The classifiers were trained on all 95 variables and smaller sets created with backward elimination. Seven types of classifiers, including Support Vector Machines with radial basis (SVM-radial), and Recursive Partitioning and Regression Trees (RPART), were trained on the parameters. The area under the ROC curve (AUC) was calculated for classifiers, individual parameters and HRT3 glaucoma probability scores (GPS). Classifier AUCs and leave-one-out accuracy were compared with the highest individual parameter and GPS AUCs and accuracies.

Results: The highest AUC and accuracy for an individual parameter were 0.848 and 0.79, for vertical cup/disc ratio (vC/D). For GPS, global GPS performed best with AUC 0.829 and accuracy 0.78. SVM-radial with all parameters showed significant improvement over global GPS and vC/D with AUC 0.916 and accuracy 0.85. RPART with all parameters provided significant improvement over global GPS with AUC 0.899 and significant improvement over global GPS and vC/D with accuracy 0.875.

Conclusions: Machine learning classifiers of HRT3 data provide significant enhancement over current methods for detection of glaucoma.

Glaucoma is an optic neuropathy characterised by visual field loss with gradual thinning of the retinal nerve fibre layer (RNFL) and cupping of the optic nerve head (ONH).¹ The diagnosis of glaucoma currently requires assessment of both the functional visual abilities and structural measurements through visual field (VF) testing and imaging methods such as confocal scanning laser ophthalmoscopy (CSLO).² However, the detection of VF defects typically occurs after there has been substantial structural damage.³ Subjective assessment through clinical examination has inherent limitations due to inter-observer variability,^{4,5} so an objective method of examining the ONH with an imaging device such as CSLO may have significant diagnostic value.

The Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany) is a CSLO instrument that acquires three-dimensional maps of the ONH and peripapillary retina. Numerous quantitative parameters are calculated both globally and in segmented areas of the ONH. These parameters can then be used as input for machine learning classifiers. Machine learning

classifiers are systems for determining the relationship between their input parameters and desired output classification, based on a training set whose classification is known a priori. The optimised use of all available HRT structural information might improve the discrimination ability more than an individual parameter.⁶⁻⁸ The purpose of this study was to assess the performance of classifiers trained on HRT3 ONH parameters for discriminating between healthy and glaucomatous eyes.

MATERIALS AND METHODS

Subjects

Healthy subjects and glaucoma patients from glaucoma clinics meeting eligibility criteria were enrolled in this cross-sectional study. All subjects received a comprehensive ophthalmic evaluation, with all tests completed within 6 months. The evaluation included medical history, best-corrected visual acuity, manifest refraction, intraocular pressure (IOP) measurements by Goldmann applanation, gonioscopy, slit-lamp examination before and after pupil dilation, VF testing and HRT2 scanning of the disc. If necessary to obtain HRT scans, subjects underwent pupillary dilation with tropicamide and phenylephrine. Diagnosis was determined clinically by glaucoma experts using the criteria below.

All subjects had best-corrected visual acuity of 20/40 or better and refractive error -6.00 to $+6.00$ dioptres (spherical equivalent). Subjects were excluded if they exhibited signs of ocular pathologies other than glaucoma, if media opacity or poorly dilating pupils interfered with clinical viewing or fundus imaging, or if they chronically used medications known to affect retinal thickness. Patients were also excluded if they had systemic diseases that may affect retinal thickness or visual field, or if they had previous ocular operations other than uneventful cataract extraction.

Glaucomatous eyes

Eyes were defined as glaucomatous if they displayed glaucomatous optic neuropathy and glaucomatous VF loss. Glaucomatous optic neuropathy was defined as inter-eye cup-disc ratio asymmetry >0.2 , accounting for disc size; general rim thinning or focal notching; peripapillary haemorrhages; or cup-disc ratio ≥ 0.6 . Glaucomatous VF loss is diagnosed if the glaucoma hemifield test was outside normal limits, pattern standard deviation (PSD) was $<5\%$, or a cluster of three or more non-edge points were depressed on the pattern

deviation plot at a level of $p < 0.05$, with one point depressed at a level of $p < 0.01$, in two consecutive VF tests.

Healthy eyes

Eyes were defined as healthy if there was no history or evidence of glaucoma, IOP was less than 21 mm Hg, the ONH did not display signs of optic neuropathy, and the Humphrey 24-2 pattern VF appeared without reproducible points outside normal limits.

VF testing

All subjects had Humphrey Swedish interactive thresholding algorithm (SITA) standard 24-2 perimetry (Carl Zeiss Meditec, Dublin, CA). A reliable VF test was defined as one with fewer than 30% fixation losses, false-positive or false-negative responses. The VF results were considered reproducible if the same type, location and index of abnormality were evident in two consecutive VF tests.

HRT scanning

HRT scans were performed using an HRT2 device. The files were then transferred to HRT3 software to be processed. Eligible images had a pixel SD $< 50 \mu\text{m}$ with even illumination, acceptable centration and focus. Images were also assessed for misalignment or incorrect contour line placement. Ninety-five global and sectoral parameters available through the HRT3 Stereometric Parameters export function (table 1) were used as input for the machine classifiers.

Among the parameters included in the analysis was the Glaucoma Probability Score (GPS), a discriminatory parameter that is defined without the need of subjective determination of the disc margin.⁹ Another common HRT classifier, the Moorfields regression analysis,¹⁰ was not used in this study because its clinical output is categorical, as opposed to the other

analyses, which provide continuous numerical values, leading to an unfair comparison.

Machine classifiers

Seven types of machine learning classifiers were trained: Linear Discriminant Analysis (LDA), Support Vector Machine with linear kernel (SVM-linear), Support Vector Machine with radial kernel (SVM-radial), Generalised Additive Model (GAM), Generalised Linear Model with Gaussian error (GLM-Gauss), Generalised Linear Model with binomial error (GLM-bin), and Recursive Partitioning and Regression Trees (RPART). Classifiers were selected either because of success using the method in previously published work,^{8 11 12} or to ensure an appropriate breadth of model-type (eg, linear versus non-linear). All classifiers were implemented in R statistical software (R version 2.2; R-Project, available at <http://cran.r-project.org>).

In addition to using these models with all global and sectoral input parameters ($n = 95$), specific models were created using only the 10 parameters with the highest unconditional Pearson correlation to diagnosis. Backward selection in this 10-predictor set was performed using Akaike information criteria (AIC). AIC was used to remove redundancy in the dataset and avoid overfitting in addition to the use of the machine classifiers that by themselves extract relevant information. For those methods for which AIC was defined, we used the full dataset to generate models for all possible subsets of the 10 parameters, and then used the predictor set whose model had the best AIC; depending on the method, these sets had seven, eight or nine parameters.

LDA uses a linear combination of the parameters to separate subjects into glaucoma and healthy.¹³ It assumes the data form a Gaussian distribution, and separates the data with linear discrimination boundaries that maximise the variance between the two classes while minimising the variance within classes. Each new data point is classified based on the likelihood it was generated by each of the categories, glaucoma or healthy.

SVM maps the multidimensional parameters into a feature space and creates a hyperplane to separate glaucomatous and healthy eyes with maximal distance between all cases and the hyperplane.¹⁴ In this study, both linear and radial kernels were used. SVMs tend to be better than other types of classifiers at identifying more important parameters and ignoring those that are less relevant.

GAM assumes the expectation of glaucoma severity can be expressed as a sum of univariate smooth functions of the parameters.¹⁵ SVMs tend to be better than other types of classifiers at identifying more important parameters and ignoring those that are less relevant. GAM cannot be trained on fewer datapoints than variables, so since cross-validation has only 25 data points in each fold, the GAM was only trained on the 10 parameter set, and the smaller eight-parameter set from backward selection.

GLM is a generalised form of least-squares regression.¹⁶ It assumes that the log of the odds ratio of a subject having glaucoma over being healthy can be expressed as a linear function of the provided parameters. The boundary between glaucoma and healthy is defined as the hyperplane where the odds of a subject having glaucoma is equal to the odds that the same patient is healthy. We generated GLMs with both Gaussian and binomial error models.

RPART is a decision-tree partitioning algorithm.¹⁷ It recursively partitions the parameter space along individual parameters. The parameters that are chosen to split and the points at which they are split are chosen in order to maximise categorisation accuracy, resulting in partitioned regions called

Table 1 Heidelberg Retina Tomograph 3 output parameters.

Parameter	G	T	TS	TI	N	NS	NI
Disc area	*	*	*	*	*	*	*
Cup area	*	*	*	*	*	*	*
Rim area	*	*	*	*	*	*	*
Cup/disc area ratio	*	*	*	*	*	*	*
Rim/disc area ratio	*	*	*	*	*	*	*
Cup volume	*	*	*	*	*	*	*
Rim volume	*	*	*	*	*	*	*
Mean cup depth	*	*	*	*	*	*	*
Maximum cup depth	*	*	*	*	*	*	*
Height variation contour	*	*	*	*	*	*	*
Cup shape measure	*	*	*	*	*	*	*
Mean RNFL thickness	*	*	*	*	*	*	*
RNFL cross-sectional area	*	*	*	*	*	*	*
Horizontal cup/disc ratio	*						
Vertical cup/disc ratio	*						
Maximum contour elevation	*						
Maximum contour depression	*						
CLM temporal-superior	*						
CLM temporal-inferior	*						
Average variability (SD)	*						
Reference height	*						
FSM discriminant function value	*						
RB discriminant function value	*						

Parameters in bold were not used as input to machine classifiers. CLM, contour line modulation; G, global; N, nasal; NI, nasal inferior; NS, nasal superior; RNFL, retinal nerve fibre layer; T, temporal; TI, temporal inferior; TS, temporal superior.

leaves. Each new case is classified by majority of the training cases belonging to the same leaf as the new case.

Data analysis and statistics

Classifiers were assessed using eightfold cross-validation and leave-one-out (LOO) analyses. For eightfold cross-validation, the data set was split into eight folds with 25 data points each. For each classifier, eight different models were generated, using seven folds to train the classifiers and the eighth to test the classifier. This was repeated so all eight folds are used as the testing set once. Classifiers that overfit the data do poorly in cross-validation, as they perform poorly on all eight test sets. The average area under the receiver operating characteristic curve (AUC) for each classifier was computed by pointwise averaging across folds. The AUCs were compared using the DeLong method.¹⁸

LOO accuracy was also calculated to assess the discrimination abilities of each classifier, GPS, and the individual parameters. GPS and individual parameter accuracies were based on optimised cut-offs, rather than the manufacturer provided cut-offs, to test our classifiers against the best performance possible for this data set. LOO accuracy was calculated by training each classifier on the entire data set except one eye, and then testing the remaining eye. This was repeated so all eyes were chosen as the test eye. Accuracy was then calculated as number of true predictions divided by total number of observations. The alpha level of significance was set to 0.05.

RESULTS

Subject characteristics

Two hundred and sixty-three patients were retrospectively evaluated for this cross-sectional study. Twenty-three were excluded due to non-reliable VFs or non-reproducible defects. Eleven were excluded due to HRT3 SD >50 µm. Sixteen were excluded due to ocular pathologies other than glaucoma. Thirteen subjects had bilateral findings that did not conform to the inclusion criteria of either group, and were excluded as glaucoma suspects. If both eyes qualified, one was randomly selected. A total of 200 eyes (200 subjects) were included in the study, and their characteristics are summarised in table 2.

There was a significant difference in age, gender, MD and PSD between healthy and glaucoma subjects. The mean MD reflects a moderate level of damage in the glaucoma group.

Table 2 Subject characteristics

Characteristic	Healthy eyes (n = 60)	Glaucomatous eyes (n = 140)	p
Age (years)	43.8 (16.1) (18 to 81)	62.4 (13.7) (20 to 94)	<0.0001*
Male/female	19/41	66/74	0.042†
Visual field mean deviation (dB)	-0.89 (1.62) (-7.02 to 1.69)	-6.67 (7.09) (-27.30 to 0.69)	<0.0001*
Visual field pattern standard deviation (dB)	1.66 (0.94) (0.97 to 6.94)	5.96 (4.23) (1.02 to 14.79)	<0.0001*
Ethnicity			0.17†
Caucasian	51	109	
Others	9	31	
Disc size (mm ²)	1.76 (0.42) (0.89 to 3.25)	1.93 (0.55) (0.75 to 3.65)	0.029*

*Two-tailed t test.

†Chi-squared test.

Selected parameters for smaller data sets

Machine classifier analysis was conducted on the complete 95-parameter set, and smaller sets of 10, nine, eight and seven parameters. The 10-parameter set was selected as the 10 single parameters with the best correlation to diagnosis. These parameters are listed in table 3.

From the 10-parameter set, backward selection using AIC was performed to find smaller predictor sets for the GAM, GLM-Gaussian and GLM-Binomial. Because AIC is not well defined for RPART, backward selection could not be performed, so it was tested on (1) a nine-parameter set excluding cup area in the temporal inferior sector, the only parameter excluded from all parameter data sets created by backward selection; (2) the narrowed sets created by backward selection for GAM and the GLMs; and (3) the 10-parameter data set.

Discriminating between healthy and glaucomatous eyes

Among individual HRT3 parameters, the global vertical cup/disc ratio (vC/D) had the greatest AUC of 0.848 (table 4). Among the seven GPS sectoral measurements, the global GPS had the greatest AUC (0.829; data shown for global only). RPART AUC with all 95 parameters showed a significant improvement over global GPS, though not over global vC/D. SVM-radial with all 95 parameters displayed a significant improvement over global vC/D, as well as over global GPS. All machine classifiers except GLM-binomial with all parameters and all RPART with narrowed data sets had an AUC greater than or equal to global GPS, though other than those previously mentioned, the differences were non-significant. Both RPART with all parameters and SVM-radial had a significant improvement in accuracy over global GPS and vC/D with accuracy improved from 6 to 9.5%. All AUC and accuracy results can be seen in table 4, and significant AUC curves can be seen in fig 1.

DISCUSSION

In this study, we demonstrated that the use of certain machine classifiers improves the discrimination ability of HRT as compared with current methods. Validation of the classifiers' performance was done through eightfold cross-validation to generate AUC, and LOO accuracy. These methods provide an ample training set for the classifiers, while avoiding training the classifiers on the data set used to test its performance.

Out of the seven types of machine classifiers we trained, RPART with all 95 of the parameters and SVM-radial were superior over the

Table 3 Selected parameters for the backwards selection narrowed data sets.

Parameter	Correlation to diagnosis	GAM	GLM-Gaussian	GLM-binomial	RPART
Global vertical C/D ratio	0.55	*		*	*
TI C/D area ratio	0.57	*	*	*	*
Global C/D area ratio	0.54	*	*	*	*
Global cup shape	0.51	*	*		*
Linear C/D ratio	0.53		*		*
TS C/D area ratio	0.52	*	*	*	*
NI C/D area ratio	0.53	*	*	*	*
TI cup area	0.50				
NS C/D area ratio	0.48	*	*	*	*
TS cup shape	0.48	*	*	*	*

*Parameters used in the dataset.

C/D, cup/disc; GAM, Generalised Additive Model; GLM, Generalised Linear Model; NI, nasal inferior; NS, nasal superior; TI, temporal inferior; TS, temporal superior.

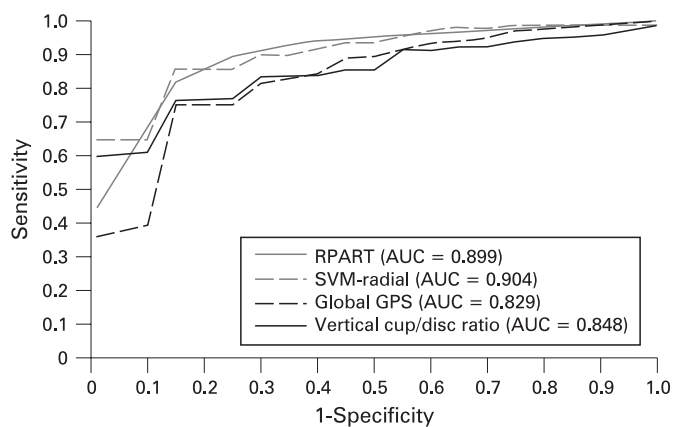


Figure 1 Selected ROC curves for best predictive single variable (vertical cup/disc ratio), best predictive GPS (global), and the two classifiers that displayed a significant improvement over them.

other classifiers, all sectors of the GPS and all individual parameters. They had the highest AUC and accuracy of all classifiers and displayed a significant improvement in both AUC and accuracy over the best GPS and individual parameters. An exception was RPART's AUC, whose improvement was just under the significant level for vC/D. GPS AUC results in our study were similar to those reported in previous studies ranging from 0.70 to 0.88.^{19–22}

Minimising the data set improved the performance of the linear classifiers (GAM, GLM and LDA) and decreased the performance of the non-linear classifiers (RPART, SVM). Most classifiers with smaller data sets performed better than those with all parameters except for the non-linear classifiers RPART and

both SVMs (linear and radial), which performed worse with fewer parameters. This reflects similar results in other studies examining machine classifiers for glaucoma discrimination.^{11 12} The smaller data sets may perform better for the linear classifiers because they remove confounding “noise” parameters that have minimal correspondence to glaucoma diagnosis, and help prevent overfitting. RPART and SVM, as non-linear methods, are better at weeding out insignificant information than linear methods, which may explain why mechanically providing them with less information may have been detrimental.

It was interesting to observe that the list of the 10 best correlating parameters (table 3) included nine parameters that describe the shape of the disc without size consideration or area ratios that cancel out effects of disc size. When the 10 parameter sets were narrowed further to improve the performance, the only parameter of the ten that was eliminated by all data sets through backward selection was cup area in the temporal inferior sector, the one parameter of the 10 that is affected by disc size. This provides evidence that dependence on disc size can be a confounding factor when using individual parameters for glaucoma discrimination. Other studies have found similar results, though they do not emphasise as completely the parameters not affected by disc size.^{10 11 23 24}

Some of our models found a significant improvement over current differentiation techniques; however, this improvement may have been limited by our sample size and sampled population. Our validation methods do not eliminate confounders that may be present in our selected study group. It would be beneficial to test our models on an independent data set; however this would require an extremely large data set, which was not feasible at the current date. The cross-validation

Table 4 Machine classifier glaucoma discrimination results

	AUC				p Value		Accuracy		
	AUC	AUC SE	Sensitivity at 80% specificity	Sensitivity at 95% specificity	Classifier vs GPS	Classifier vs global vertical C/D ratio	Accuracy	Classifier vs GPS	Classifier vs global vertical C/D ratio
Global vertical C/D ratio	0.848	0.027	0.766	0.603	0.544		0.790	0.728	
GPS	0.829	0.028	0.749	0.374		0.544	0.780		0.728
RPART (GLM-B7)	0.756	0.034	0.648	0.310	0.065	<0.001	0.685	0.002	<0.001
RPART (GLM-G8)	0.783	0.032	0.613	0.243	0.199	0.013	0.780	1.000	0.728
GLM-Binomial	0.783	0.032	0.630	0.474	0.223	0.097	0.795	0.670	0.931
RPART (GAM-8)	0.789	0.032	0.688	0.310	0.279	0.027	0.770	0.733	0.488
RPART (9)	0.798	0.031	0.679	0.316	0.362	0.052	0.745	0.232	0.119
RPART (10)	0.805	0.031	0.681	0.316	0.475	0.089	0.740	0.172	0.098
GLM-Gaussian	0.829	0.029	0.764	0.479	0.984	0.568	0.785	0.932	0.862
LDA	0.829	0.029	0.764	0.479	0.984	0.568	0.770	0.733	0.488
GAM (10)	0.837	0.028	0.734	0.524	0.803	0.556	0.785	0.932	0.862
SVM-radial (10)	0.849	0.027	0.797	0.444	0.491	0.960	0.780	1.000	0.728
SVM-linear (10)	0.852	0.026	0.755	0.631	0.475	0.752	0.780	1.000	0.728
SVM-linear	0.859	0.026	0.761	0.620	0.363	0.718	0.825	0.146	0.259
GAM (8)	0.861	0.025	0.789	0.574	0.280	0.492	0.790	0.798	1.000
GLM-Gaussian (10)	0.874	0.024	0.811	0.630	0.135	0.079	0.790	0.798	1.000
LDA (10)	0.874	0.024	0.811	0.630	0.135	0.079	0.790	0.798	1.000
GLM-Gaussian (8)	0.875	0.024	0.797	0.644	0.131	0.087	0.765	0.609	0.386
GLM-Binomial (10)	0.876	0.024	0.805	0.610	0.112	0.162	0.780	1.000	0.728
GLM-Binomial (7)	0.879	0.023	0.797	0.637	0.104	0.088	0.740	0.172	0.098
RPART	0.899	0.021	0.857	0.546	0.014	0.074	0.875	0.001	0.002
SVM-radial	0.904	0.021	0.857	0.648	0.006	0.018	0.850	0.017	0.037

AUC, area under the receiver operating characteristics curve; C/D, cup/disc; GAM, Generalised Additive Model; GLM, Generalised Linear Model; GPS, Global Positioning System; RPART, Recursive Partitioning and Regression Trees; SE, standard error; SVM, Support Vector Machines.

Bold cells indicate a significant improvement. See Materials and Methods section for explanation of classifier abbreviations. Classifiers with values in parentheses were trained on narrowed data sets, either selected for that classifier with backward selection or, in the case of RPART, trained on backwards selection data sets from the other classifiers.

techniques in the current study are the best way to utilise all data available, while still avoiding training and testing on the same data at all times. Investigation with a larger data set would be beneficial, as it may result in further tuning of the methods and better performance of the machine classifiers. Another limitation to this study was the comparison of our classifiers to the GPS classifier, which is trained on separate data from the HRT3 output parameters. While this is a confounding factor, the aim was to compare the clinical classifier currently available with our non-clinically available classifiers.

Our glaucoma group primarily included people with moderate glaucomatous damage, which might improve the overall AUC over a less advanced group. The discriminating ability afforded by the machine classifiers in the earliest stages of glaucoma remains to be determined. Also, the glaucoma group was significantly older than the healthy. As people age, they lose neural tissue in the optic nerve, so this age difference may result in the AUC being overestimated, even though age is not an input parameter. However, since we compared the AUC of the machine classifiers, the individual parameters and GPS in the same data set, the age effect and moderate level of glaucoma damage can be expected to affect all results equally.

There have been a few instances of using machine learning classifiers on ocular imaging data for glaucoma discrimination.^{6-8 11 12} However, they differ from the present study through the use of different input data and having no clinically available classifier for comparison. Huang and Chen⁶ and Burgansky-Eliash *et al*¹² both used optical coherence tomography data as input, and no clinically available machine classifier currently exists to be compared with their machine classifiers. Mardin *et al*⁷ created classifiers incorporating both visual field and HRT2 exported parameters, but did not compare these to the HRT3 GPS clinical output. Zangwill *et al*⁸ developed their own HRT2 mean height contour measurements in 36 sectors along the disc margin and in the parapapillary region as input for their classifiers, and Bowd *et al*¹¹ used the HRT2 output parameters, which are similar to the HRT3 parameters used in this study, but created two neural network techniques as their classifiers and compared them with previously published linear discrimination functions that are not available in direct clinical outputs. This publication is able to uniquely compare clinically available classifiers and our classifiers developed on the most up-to-date HRT data.

In conclusion, machine classifiers can provide a significant improvement in HRT3 diagnostic power over the current methods of discrimination of single parameters and GPS. In particular, SVM-radial and RPART with all parameters displayed the greatest improvement in glaucoma discrimination.

Funding: Supported in part by NIH grants R01-EY013178, P30-EY008098 (Bethesda, MD), The Eye and Ear Foundation (Pittsburgh) and an unrestricted grant from Research to Prevent Blindness (New York).

Competing interests: Dr Schuman receives royalties for intellectual property licensed by Massachusetts Institute of Technology to Carl Zeiss Meditec. Dr Wollstein received research funding from Carl Zeiss Meditec.

Ethics approval: The study was approved by the Institutional Review Board/Ethics Committee, and adhered to the Declaration of Helsinki and Health Insurance Portability and Accountability Act regulations.

Patient consent: Patient consent was obtained.

REFERENCES

1. Sommer A, Miller NR, Pollack I, *et al*. The nerve fiber layer in the diagnosis of glaucoma. *Arch Ophthalmol* 1977;**95**:2149–56.
2. Quigley HA, Katz J, Derick RJ, *et al*. An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. *Ophthalmology* 1992;**99**:19–28.
3. 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.
4. Garway-Heath DF, Poinosawmy D, Wollstein G, *et al*. Inter- and intraobserver variation in the analysis of optic disc images: comparison of the Heidelberg retina tomograph and computer assisted planimetry. *Br J Ophthalmol* 1999;**83**:664–9.
5. Iester M, Mikelberg FS, Courtright P, *et al*. Interobserver variability of optic disk variables measured by confocal scanning laser tomography. *Am J Ophthalmol* 2001;**132**:57–62.
6. Huang ML, Chen HY. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci* 2005;**46**:4121–9.
7. Mardin CY, Peters A, Horn F, *et al*. Improving glaucoma diagnosis by the combination of perimetry and HRT measurements. *J Glaucoma* 2006;**15**:299–305.
8. Zangwill LM, Chan K, Bowd C, *et al*. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci* 2004;**45**:3144–51.
9. Swindale NV, Stjepanovic G, Chin A, *et al*. Automated analysis of normal and glaucomatous optic nerve head topography images. *Invest Ophthalmol Vis Sci* 2000;**41**:1730–42.
10. Wollstein G, Garway-Heath TF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology* 1998;**105**:1557–63.
11. 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**:3444–54.
12. Burgansky-Eliash Z, Wollstein G, Chu T, *et al*. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci* 2005;**46**:4147–52.
13. Fisher R. The use of multiple measurements in taxonomic problems. *Ann Eugen* 1936;**7**:179–88.
14. Vapnick V. *Statistical learning theory*. New York: Wiley-Interscience, 1998:401–41.
15. Hastie T, Tibshirani RJ. *Generalized additive models*. Vol. 43: Chapman & Hall/CRC, 1990:352.
16. McCullagh P, Nelder JA. *Generalized linear models*. London: Chapman & Hall, 1989.
17. Breiman L FJ, Olshen R, Stone C. Classification and regression trees. In: *Classification and regression trees*. Belmont, CA: Wadsworth, 1984.
18. DeLong E, DeLong D, Clarke-Pearson D. Comparing the area under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics* 1988;**44**:837–45.
19. Burgansky-Eliash Z, Wollstein G, Bilonick RA, *et al*. Glaucoma detection with the Heidelberg retina tomograph 3. *Ophthalmology* 2007;**114**:466–71.
20. Coops A, Henson DB, Kwartz AJ, *et al*. Automated analysis of Heidelberg retinal tomograph optic disc images by glaucoma probability score. *Invest Ophthalmol Vis Sci* 2006;**47**:5348–55.
21. Ferreras A, Pajarin AB, Polo V, *et al*. Diagnostic ability of Heidelberg Retina Tomograph 3 classifications glaucoma probability score versus Moorfields regression analysis. *Ophthalmology* 2007;**114**:1981–7.
22. Zangwill LM, Jain S, Racette L, *et al*. The effect of disc size and severity of disease on the diagnostic accuracy of the Heidelberg retina tomograph glaucoma probability score. *Invest Ophthalmol Vis Sci* 2007;**48**:2653–60.
23. DeLeon-Ortega JE, Arthur SN Jr., GM, *et al*. Discrimination between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective optic nerve head assessment. *Invest Ophthalmol Vis Sci* 2006;**47**:3374–80.
24. 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**:985–93.