Published in final edited form as:

Invest Ophthalmol Vis Sci. 2007 December ; 48(12): 5582–5590. doi:10.1167/iows.06-0897.

Assessing Visual Field Clustering Schemes Using Machine Learning Classifiers in Standard Perimetry

Catherine Boden¹, Kwokleung Chan^{2,3}, Pamela A. Sample¹, Jiucang Hao², Te-Wan Lee^{2,3}, Linda M. Zangwill¹, Robert N. Weinreb¹, and Michael H. Goldbaum¹

¹Hamilton Glaucoma Center and Visual Function Laboratory, Department of Ophthalmology, University of California at San Diego, La Jolla, California

²Institute for Neural Computation, University of California at San Diego, La Jolla, California

³Computational Neurobiology Laboratories, Salk Institute, La Jolla, California.

Abstract

Purpose—To compare machine learning classifiers trained on three clustering schemes to determine whether distinguishing healthy eyes from those with glaucomatous optic neuropathy (GON) can be optimized by training with clustered data.

Methods—Two machine learning classifiers—quadratic discriminant analysis (QDA) and support vector machines with Gaussian kernel (SVMg)—were trained separately using standard perimetry data from the Diagnostic Innovations in Glaucoma Study (DIGS), clustered using three clustering schemes on a training data set (123 eyes/123 glaucoma patients with GON; 135 eyes/135 normal control subjects). Trained classifiers were then applied to an independent data set containing 69 eyes of 69 glaucoma patients with early visual field loss and 83 eyes of 83 normal control subjects. Two control conditions were included: unclustered data and a random assignment of locations to clusters.

Results—Areas under the receiver operating characteristic (ROC) curve ranged from 0.85 (SVMg, thresholds clustered by Glaucoma Hemifield Test sectors) to 0.92 (QDA, thresholds clustered by Garway-Heath mapping) for the training data set. Use of clustered data showed no significant optimization of sensitivity over use of unclustered data, and no single clustering method resulted in significantly higher performance in the independent data set. Sensitivities tended to be higher with QDA than with SVMg, regardless of specificity cutoff and clustering method.

Conclusions—QDA performed better with the early glaucoma data set than did the SVMg. Clustering may be advantageous when data–dimension reduction is needed—for example, when combining field results with other high-dimensional data (e.g., structural imaging data)—but it is not necessary for visual field data alone.

Machine learning classifiers (MLCs) are computational methods that enable machines to learn from experience. MLCs are effective in a variety of applications and classification problems in glaucoma, including visual field interpretation,^{1–7} diagnosis of glaucoma through structural measures,^{8–14} identification of patterns of visual field loss,^{15,16} and detection of glaucomatous progression.^{17–20} MLCs trained to identify patients with glaucomatous optic neuropathy (GON) from standard automated perimetry (SAP) of visual fields have been shown

Corresponding author: Pamela A. Sample, Department of Ophthalmology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0946; psample@glaucoma.ucsd.edu..

Disclosure: **C. Boden**, None; **K. Chan**, None; **P.A. Sample**, Carl Zeiss Meditec, Inc. (F), Haag Streit (F); **J. Hao**, None; **T.-W. Lee**, None; **L.M. Zangwill**, Carl Zeiss Meditec, Inc. (F), Heidelberg Engineering (F, R); **R.N. Weinreb**, Carl Zeiss Meditec, Inc. (F), Heidelberg Engineering (F, R); **M.H. Goldbaum**, None

to be as sensitive as traditional Statpac-like analysis at specificities of 90% and greater.² As with all new methods, however, further optimization of the performance of the MLCs would be advantageous.

One approach to optimizing the analysis of visual fields has been to group visual field locations. Mapping the visual field into clusters of related locations has been used to clarify the structure–function relationship^{21–24} and to aid in the detection of glaucomatous progression by reducing the effect of long-term variability.^{25–27} Clustering may also serve as a dimension–reduction tool to optimize MLCs and to increase our understanding of the visual field regional relationships. However, it is not clear whether there is an advantage to selecting one map over the others or whether dimension reduction with structure-derived clusters compares favorably with mathematical dimension-reduction.

In the present study we compared the sensitivity and specificity of two MLCs trained separately on three clustering schemes to determine (1) whether MLC ability to categorize healthy and GON eyes can be optimized by training with clustered data; (2) which MLC, visual field mapping scheme or MLC/map combination achieves the highest performance; and (3) how structure-derived schemes compare to the mathematical dimension-reducing scheme.

Methods

Visual field data came from participants who were part of the prospective, longitudinal Diagnostic Innovations in Glaucoma Study (DIGS). One eye was randomly selected as the study eye, except in participants in whom only one eye was eligible. All participants provided written informed consent to participate in the study, and all methods were approved by the University of California, San Diego, Human Subjects Committee. The study adhered to the Declaration of Helsinki for research involving human subjects.

Inclusion Criteria for DIGS

All subjects underwent complete ophthalmic examination including slit lamp biomicroscopy, intraocular pressure measurement, dilated stereoscopic fundus examination, and stereophotography of the optic nerve heads. Simultaneous stereoscopic photographs were obtained in all subjects and were of adequate quality for the subject to be included. All subjects had open angles, best corrected acuity of 20/40 or better, spherical refraction within ± 5.0 D, and cylinder correction within ± 3.0 D. A family history of glaucoma was allowed.

Exclusion Criteria for DIGS

Subjects were excluded if they had a history of intraocular surgery (except for uncomplicated cataract or glaucoma surgery). We also excluded all subjects with nonglaucomatous secondary causes of elevated IOP (e.g., iridocyclitis, trauma), other intraocular eye disease, other diseases affecting visual field (e.g., pituitary lesions, demyelinating diseases, HIV⁺ or AIDS, or diabetic retinopathy), with medications known to affect visual field sensitivity, or with problems other than glaucoma affecting color vision.

Inclusion Criteria for This Report

Participants were required to have at least one reliable SAP field test result in their study eye. Healthy eyes were characterized by healthy optic discs and had no history of elevated IOP (IOP > 22 mm Hg). Glaucomatous eyes had GON on masked, independent stereophotograph review. Other than reliability, visual field results were not used for inclusion or diagnostic purposes.

Participants were allocated to one of two data sets: the initial evaluation or the independent data set. The initial evaluation data set included 156 healthy eyes and 189 glaucomatous eyes

before age matching. These eyes were used in our previous report of the performance of MLCs trained with SAP raw thresholds and age.² The normal and glaucomatous groups were then age matched by randomly removing participants within 10-year age bins, to equalize the number of healthy and GON eyes in each age bin, yielding 135 healthy eyes and 123 glaucomatous eyes. This data set was used for training and initial evaluation of the MLCs and included eyes with glaucoma of various severities.

A separate group of 95 healthy eyes and 73 glaucomatous eyes met our inclusion criteria for the independent data set. To determine whether the performance of the machine classifiers trained with clustered data could match or even exceed the performance of existing clinical methods in a more difficult classification decision, glaucomatous eyes in the independent data set had to meet the additional criterion of having early visual field loss. Early visual field loss was defined by using clinically available analyses as follows: (1) the field had to be abnormal (i.e., the pattern SD [PSD] had to be $P < 5\%$ or the Glaucoma Hemifield Test (GHT) result had to be outside normal limits), and (2) the visual loss had to be relatively mild (i.e., the mean deviation [MD] had to be better than or equal to -6 dB and the PSD no worse than the 1% probability level). A significant difference in age between GON and healthy eyes was noted (unpaired t -test; $P < 5\%$), and so the groups were age matched using the same procedure as just described. An age-matched group of 83 healthy eyes and 69 glaucomatous eyes were used for all analyses.

Optic Disc Stereoscopic Photographs

Simultaneous stereoscopic photographs (Simultaneous Stereo Camera TRC SS; Topcon Instrument Corp. of America, Paramus, NJ) were obtained in all patients. Those stereophotographs closest in time to the visual field date were examined for GON. Each masked stereophotograph was graded independently by two experienced reviewers. GON was diagnosed if there was evidence of excavation, focal or diffuse rim thinning, or nerve fiber layer defects by both reviewers. In cases of disagreement, a third reviewer adjudicated. Healthy eyes had no evidence of GON.

Visual Fields

SAP visual fields were performed with the 24-2 program on the Humphrey Visual Field Analyzer (Carl Zeiss Meditec, Inc., Dublin, CA), with a white Goldmann size III (0.43°) stimulus on a 31.5-apostilb background. The full-threshold algorithm was used for all fields. Only reliable visual fields were included (i.e., false-positive and false-negative responses and fixation losses were all less than or equal to 25%).

Visual field locations have been grouped anatomically^{21,28,29} or by statistical clustering.²⁵⁻²⁷ Anatomically derived clusters of field locations are based on localized nerve fiber bundle or wedge defects^{21,29} and/or prominent nerve fiber bundles.^{21,28} Garway-Heath et al.²¹ developed their map from 69 eyes with normal-tension glaucoma that had discrete retinal nerve fiber layer defects and/or prominent nerve fiber bundles (Fig. 1A). Visual field locations with defects that tended to correspond with a particular nerve fiber layer bundle were clustered. The GHT sectors were derived by superimposing the 30-2 test point pattern of the Humphrey Visual Field Analyzer on photographs of the retinal nerve fiber layer of normal subjects, to identify the normal arrangement of retinal nerve fibers (Fig. 1B).³⁰ Test points arranged along nerve fibers were clustered together to derive 10 sectors.

Statistical clustering groups locations that are statistically associated, rather than related on anatomic grounds. Variational Bayesian independent components analysis with a mixture model (vb-ICA) is a sophisticated clustering analysis, related to principle components analysis. It is an unsupervised machine learning classifier, in that the classifier is not provided with

feedback concerning the diagnostic status of the cases. The details of vb-ICA are described in.³¹ vb-ICA was selected for the present study, as we had found it to be effective for describing patterns of glaucomatous field loss³¹ and identifying progression.³² When applied to a data set of visual field data from healthy and GON eyes,³¹ vb-ICA generated two clusters (one predominantly containing visual fields from eyes with GON and the other predominantly from healthy eye) and described six maximally independent axes within the GON cluster. For the present study, a visual field map was derived from vb-ICA by assigning visual field locations to groupings based on the relevance of each location to these six axes. Used in this manner, vb-ICA represents a mathematical dimension-reducing scheme (Fig. 1C). The resultant map consists of six clusters (instead of the original 52 dimensions representing the 52 separate visual field locations) that respect the horizontal meridian and show similarity to the arcuate pattern expected in glaucoma (Fig. 1C). SAP locations were clustered according to (1) the nerve fiber bundle map of Garway-Heath et al.,²¹ (2) the GHT sectors,^{30,33,34} (3) vb-ICA-derived clusters,³¹ and (4) random assignment to clusters (Fig. 1). The first 2 methods were derived from topographic relationship to glaucomatous optic disc damage,^{21-24,29,35} whereas the third method was derived from unsupervised learning by using a variational Bayesian independent component analysis. The GHT analysis of the Statpac 2 program is used clinically also. We used the GHT in two ways: (1) GHT sectors were used to cluster visual field locations for training MLCs for the purpose of optimizing MLC classification, and (2) GHT results, which represent traditional statistical methods of defining visual field abnormality, from Statpac 2 (e.g., outside normal limits) are reported. Schematics of the allocation of visual field test locations to clusters in each of the maps are presented in Figure 1. Cluster scores were calculated by taking the average of the raw thresholds of the individual locations forming the cluster.

Random assignment of the visual field locations to 10 clusters served as one control condition (Random map). A second control condition, unclustered control, was based on unclustered visual field data (i.e., MLCs were trained and tested with all 52 raw thresholds and age).

Machine Learning Classifiers

We chose to employ two MLCs with different properties, which were used in our previous studies^{1,2,9,36-39}—a quadratic discriminant analysis (QDA) and a support vector machine with a Gaussian kernel (SVMg). The MLCs were trained with average cluster thresholds and age from each of the four maps separately using visual fields from the initial evaluation data set. Thus, there were eight classifier/map combinations and two classifier/unclustered control combinations.

For the initial evaluation data set, separation of teaching and test examples was accomplished with 10-fold cross-validation, to minimize bias in testing the sensitivity and specificity of the machine classifiers. Data from normal and glaucomatous eyes were randomly divided into 10 partitions. One partition of the data set was retained as a test set, and nine partitions were combined to act as the training set. The training-test process was repeated until each partition had the opportunity to be in the test set.

Machine classifiers were generated with custom computer programs (written in MatLab; The Mathworks Inc., Natick, MA) developed by one of the authors (KC).² Detailed descriptions of the classifiers can be found in Goldbaum et al.² However, a brief description of the classifiers is provided in the following sections.

Quadratic Discriminant Analysis

In a discriminant analysis, two classes (GON and healthy eyes) of multivariate observations (e.g., 52 visual field locations plus age) together form a training set. A discriminant rule is

learned from these data and can then be used to classify new observations into one of the two classes. Quadratic discriminant analysis is closely related to linear discriminant analysis but differs in some of the underlying assumptions about the characteristics of the data and by assuming that the discriminant rule is a quadratic function. QDA fits one normal distribution to each class, and the output is the combination of normal distribution for each class weighted by class probability (percent of normal and glaucoma). QDA estimates the covariance of each class, whereas LDA assumes that the two classes have the same covariance. This results in a linear discriminant rule for LDA, and a quadratic rule for QDA. This type of classification tool has been shown previously to be one of the more effective MLCs with standard perimetry data.² QDA is disadvantaged, however, in high-dimensional space, as it requires the estimation of many parameters. Reducing the dimensionality of the data from 53 dimensions to 11 and fewer might improve the performance of QDA in classifying eyes as healthy or GON based on visual fields.

Support Vector Machine with a Gaussian Kernel

Support vector machines are a class of supervised learning algorithms.^{40,41} Unlike QDA, the data are mapped to higher-dimensional space by nonlinear mapping through Gaussian kernel with the expectation that it will provide better separation instead of by adapting the separating surface. A multivariate Gaussian distribution kernel (SVMg) analysis is designed to cope with data that cannot be easily separated in the original data space.⁴² The data are mapped to higher-dimensional space by nonlinear mapping with the expectation that the data will be more separable in higher-dimensional space with a hyperplane. Support vectors are input data closest to the decision boundary in a projected higher-dimensional space. A linear separation hyperplane is learned based on maximizing the distance of support vectors from the separation hyperplane. The linear boundary in projected space results in nonlinear separation in data space. SVMs have been used for a variety of classification problems, with success.⁴³⁻⁴⁶ They are able to deal better with high-dimensional data than QDA and therefore do not require dimension reduction for maximal performance. However, SVMs are still sensitive with low-dimensional data (e.g., the clustered data). If a particular mapping scheme clusters the data more optimally for our data set, it is assumed that the performance of SVM at separating GON from healthy eyes will be better for some mapping schemes than others.

Analysis

Student's *t*-tests were used to compare clinical characteristics between glaucomatous and healthy eyes for the initial evaluation and the independent data sets (JMP software; SAS Institute Inc., Cary, NC). $P < 0.05$ was considered significant.

Sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve were calculated for classifier/map combinations to evaluate performance of (1) clustered versus the two control conditions (random map and unclustered visual field data) for each classifier, (2) clustering methods (i.e., maps) across classifiers, (3) classifiers across clustering methods, and (4) classifier/map combinations versus clinically available analyses. Areas under the ROC curve do not fully account for the shape of the curve. Thus, areas under the ROC curve are reported but the primary analyses were conducted on sensitivity at the high specificities desired in glaucoma diagnostic tests. The specificity cutoffs were derived from the initial evaluation data set and then applied to the independent data set. The specificity was recalculated in the independent data set to determine how much the specificity changed, if at all, when the cutoff was applied to the new set of data. Statistical comparison of sensitivity and specificity were conducted using the McNemar test for nonparametric, paired data (SPSS software; SPSS Inc., Chicago, IL).

For comparison with a clinical standard, the sensitivity and specificity of MD and PSD were calculated using a cutoff of $P < 5\%$ provided by Statpac 2 to define abnormality. The GHT result was considered abnormal if it was outside normal limits.

Results

Descriptive statistics for both the initial evaluation and independent data sets are presented in Table 1. Specificity and sensitivity in the independent data set were determined using criteria that achieved 96% and 90% specificity in the initial evaluation data set. ROC curves and a scatterplot of sensitivities for the various MLC-map combinations at two specificities are presented in Figure 2.

Figure 3 depicts the comparisons that were made. Table 2 shows the probabilities from comparison of pairs of classifiers, and Figure 4 shows the sensitivity of each of the classifiers and the scatterplots of sensitivity versus specificity. Statistically significant differences in sensitivity due to the type of clustering (Table 2, gray-shaded within-MLC comparisons; Fig. 4) were noted for the 96% specificity cutoff only: Sensitivity of GH-QDA (51%) was lower than all the other clustering-classifier combinations with QDA (un-clustered-QDA, 80%; GHT-QDA, 76%; vb-ICA-QDA, 69%; random-QDA, 67%); GHT-SVMg (44%) and random-SVMg showed significantly lower sensitivity than unclustered-SVMg (60%). No within-MLC differences were noted across cluster methods for the 90% specificity cutoff. The sensitivity of MLCs was not significantly higher for mathematical (vb-ICA) or anatomic (GHT, GH) clustering over random assignment (random map) of test locations to clusters, though a trend in that direction was present. Overall, the sensitivity of QDA tended to be higher than SVMg (Table 2, Fig. 4).

Specificity is often lower when a cutoff from one data set is applied to a new data set. Therefore, we evaluated the actual specificity in the independent data set, as well as analyzing sensitivity. The plots of sensitivity versus specificity (Fig. 4) show the tradeoff of sensitivity and specificity for the various classifiers when applied to the independent data set. The classifiers with the highest sensitivity did not necessarily retain the highest specificity and vice versa. For instance, with the 96% specificity cutoff, the unclustered-QDA showed high sensitivity but the actual specificity when applied to the independent data set was only 82% to 14% lower than desired. In contrast, the sensitivity of GHT-QDA was only 4% lower than unclustered-QDA but specificity was 88%, only 7% lower than desired. This trend is only partially supported statistically, in that many of the differences in specificity were not statistically significant (Table 2).

Classifier/Map Combinations and Clinical Analyses

Individual Statpac criteria did not achieve a specificity of 96% when clinically available criteria were used. The GHT Statpac 2 results had a specificity of 89% and sensitivity of 71%. No significant difference in specificity between the GHT Statpac outcome and any of the classifiers was found for 96% specificity cutoff (selected to more closely approximate the theoretical specificity of the GHT Statpac analysis). The unclustered-QDA (sensitivity, 80%) showed significantly higher sensitivity than the GHT Statpac result, whereas GHT-SVMg (sensitivity, 44%), random-SVMg (sensitivity, 47%) and vb-ICA-SVMg (sensitivity, 51%) showed significantly lower sensitivity than the GHT Statpac result.

PSD yielded a specificity of 89% and sensitivity of 61%. The unclustered-QDA showed significantly higher sensitivity than the PSD, whereas the GHT-SVMg showed significantly lower sensitivity. No other significant differences were found.

Discussion

In the present study, we examined the potential for dimension reduction, by grouping visual field locations mathematically or anatomically, to optimize MLC classification of visual fields as glaucomatous or normal. An additional goal was to determine whether a particular MLC and/or a particular clustering method was most efficacious. The results demonstrated a marked advantage for QDA for the classification of SAP data. Previous work has also shown the QDA to be one of the more sensitive and specific classifiers.^{2,38} The data do not strongly support an advantage for one particular clustering method, except there was a benefit for the combination of QDA trained with data clustered according to the GHT sectors. Although we do not know why this is the case, it should be noted that the GHT clustering differed from the others by excluding some visual field locations near the blind spot which could theoretically “confuse” the classifier in some cases. SVMg copes better with high-dimensional data than QDA, perhaps explaining the lack of marked benefit of using the clustered data for this classifier. Previous structure–function comparisons have shown topographic relationships between visual field and optic disc imaging for various groupings of visual field locations, 21-24,47,48, which supports the finding that there is not just one effective way of grouping the visual field locations.

Turpin et al.⁴⁹ report comparable results in their analysis of visual field progression using a support vector machine with a linear kernel (SVMI). SVMI was able to classify eyes as progressing or stable with similar sensitivity as point-wise linear regression when the age-corrected thresholds from the 76 locations of the 30-2 pattern of SAP were used as input into the classifiers. This outcome was not improved in any significant way by clustering the data by using an anatomically derived map developed by Weber and Ulrich.⁵⁰

There was not a strong advantage of clustering data with either statistical or anatomic clustering methods. Clustering visual field locations to reduce data dimensionality, however, did not adversely affect MLC performance in general. Thus, in circumstances in which reducing data dimensionality may be helpful, such as when visual field data are combined with high-dimensional data from another diagnostic instrument, clustering visual field maps is a viable option. Moreover, there appeared to be a trend in the data plotting sensitivity versus specificity, indicating the potential for a tradeoff. Some classifiers maintained sensitivity while losing specificity when applied to a new data set—for instance, the unclustered-QDA classifier. The GHT-QDA classifier appeared to have good sensitivity while approaching the desired specificity. Future studies examining the diagnostic potential of new classifiers should address specifically the possibility of this tradeoff. A good diagnostic analysis package should be both sensitive and maintain good specificity when applied to the population for which it has been developed.

Testing of the initial evaluation group confirmed that sensitivities at high specificities using the two MLCs trained with clustered visual field input were similar to those found in a previous study on the same population for classifiers trained with SAP raw thresholds and age.² The present study differed from our previous studies in that the classifiers were trained on one data set with GON eyes of mixed disease severity (initial evaluation group), but tested on a separate data set containing eyes with early glaucoma (defined by visual fields)—a more stringent evaluation of the classifiers. In this case, the analysis of primary interest was the discrimination of healthy eyes from GON eyes with early field loss. As expected, sensitivities were lower for all classifier/map combinations. However, we believe this data set provides better separation of classifier/map combinations, because it excludes the most obviously defective cases.

The random control condition was used to determine whether data reduction per se is sufficient to optimize MLC performance. Although sensitivity was not always significantly lower for

classifiers trained with the random map, the trend was present. When data reduction is desired, the other maps developed specifically to group related visual field locations would be most useful.

With respect to current clinical indices, a slight, but not statistically significant advantage for the QDA/GHT combination over the PSD was noted and matched the performance of the GHT. Because of the selection criteria for the patients in the independent data set, we do not know if the performance of the GHT can be improved on. QDA trained with data clustered by either one of these two maps may increase the ability to detect early visual field loss while maintaining high sensitivity. The GHT Statpac analysis is well documented, is commonly used in clinical practice, and has been valuable in identifying glaucoma.⁵¹ A future study comparing the classification by QDA combined with GHT clustering and GHT Statpac analysis in longitudinal data of ocular hypertensive and eyes with suspected glaucoma could help determine whether a QDA/map combination is able to detect field loss earlier than any of the current commercially available Statpac analyses.

A limitation of the present study is the lack of data from an independent source of participants who might be of a different socioeconomic and/or racial make-up than the participants in the DIGS. This limitation is true of most MLC studies in this field because of the difficulty of acquiring a large enough data set. However, since our purpose was to compare the relative performance of various combinations of classifiers and mapping schemes, rather than estimate diagnostic performance, the lack of data acquired independently from the test set should not have a large impact on our conclusions.

In summary, the results of the present study do not support the use of one method of clustering visual field locations over the others. There was a marked advantage for one MLC (QDA). In addition, one combination of classifier and data reduction, the QDA with data clustered by the GHT sectors, worked better than other such combinations. Future work is needed to examine the potential for using visual field clustering to optimize MLC performance when field data are combined with other high-dimensional test results, such as imaging data.

Acknowledgements

Supported by National Eye Institute Grants EY08208 (PAS), EY13235 (MHG), EY11008 (LMZ). Participant retention incentive grants in the form of glaucoma medication at no cost were provided by Alcon Laboratories Inc, Allergan, Pfizer, Inc., and Santen Inc.

References

1. 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. [PubMed: 8056511]
2. 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. [PubMed: 11773027]
3. Mutlukan E, Keating D. Visual field interpretation with a personal computer based neural network. *Eye* 1994;8:321–323. [PubMed: 7958038]
4. Bengtsson B, Bizios D, Heijl A. Effects of input data on the performance of a neural network in distinguishing normal and glaucomatous visual fields. *Invest Ophthalmol Vis Sci* 2005;46:3730–3736. [PubMed: 16186356]
5. 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. [PubMed: 10084278]
6. Henson DB, Spenceley SE, Bull DR. Artificial neural network analysis of noisy visual field data in glaucoma. *Artif Intell Med* 1997;10:99–113. [PubMed: 9201381]

7. Spenceley SE, Henson DB, Bull DR. Visual field analysis using artificial neural networks. *Ophthalmic Physiol Opt* 1994;14:239–248. [PubMed: 7970738]
8. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol* 1996;121:511–521. [PubMed: 8610794]
9. 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. [PubMed: 12407155]
10. Huang ML, Chen HY. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci* 2005;46:4121–4129. [PubMed: 16249489]
11. Yu J, Abidi SS, Artes P, McIntyre A, Heywood M. Diagnostic support for glaucoma using retinal images: a hybrid image analysis and data mining approach. *Stud Health Technol Inform* 2005;116:187–192. [PubMed: 16160257]
12. Poinoosawmy D, Tan JC, Bunce C, Hitchings RA. The ability of the GDx nerve fibre analyser neural network to diagnose glaucoma. *Graefes Arch Clin Exp Ophthalmol* 2001;239:122–127. [PubMed: 11372541]
13. 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;46:4147–4152. [PubMed: 16249492]
14. Bowd C, Medeiros FA, Zhang Z, et al. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. *Invest Ophthalmol Vis Sci* 2005;46:1322–1329. [PubMed: 15790898]
15. Sample PA, Chan K, Boden C, et al. Using unsupervised learning with variational bayesian mixture of factor analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci* 2004;45:2596–2605. [PubMed: 15277482]
16. Henson DB, Spenceley SE, Bull DR. Spatial classification of glaucomatous visual field loss. *Br J Ophthalmol* 1996;80:526–531. [PubMed: 8759263]
17. Brigatti L, Nouri-Mahdavi K, Weitzman M, Caprioli J. Automatic detection of glaucomatous visual field progression with neural networks. *Arch Ophthalmol* 1997;115:725–728. [PubMed: 9194722]
18. Lin A, Hoffman D, Gaasterland DE, Caprioli J. Neural networks to identify glaucomatous visual field progression. *Am J Ophthalmol* 2003;135:49–54. [PubMed: 12504697]
19. Tucker A, Vinciotti V, Liu X, Garway-Heath D. A spatio-temporal Bayesian network classifier for understanding visual field deterioration. *Artif Intell Med* 2005;34:163–177. [PubMed: 15894180]
20. Madsen EM, Yoltan RL. Demonstration of a neural network expert system for recognition of glaucomatous visual field changes. *Mil Med* 1994;159:553–557. [PubMed: 7824147]
21. Garway-Heath DF, Poinoosawmy D, Fitzke FW, Hitchings RA. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 2000;107:1809–1815. [PubMed: 11013178]
22. Anton A, Yamagishi N, Zangwill L, Sample P, Weinreb R. Mapping structural to functional damage in glaucoma with standard automated perimetry and confocal laser ophthalmoscopy. *Am J Ophthalmol* 1998;125:436–446. [PubMed: 9559728]
23. Bosworth CF, Sample P, Williams JM, Zangwill L, Lee B, Weinreb RN. Spatial relationship of motion automated perimetry and optic disc topography in patients with glaucomatous optic neuropathy. *J Glaucoma* 1999;8:281–289. [PubMed: 10529926]
24. Yamagishi N, Anton A, Sample PA, Zangwill L, Lopez A, Weinreb RN. Mapping structural damage of the optic disk to visual field defect in glaucoma. *Am J Ophthalmol* 1997;123:667–676. [PubMed: 9152072]
25. Mandava S, Zulauf M, Zeyen T, Caprioli J. An evaluation of clusters in the glaucomatous visual field. *Am J Ophthalmol* 1993;116:684–691. [PubMed: 8250069]
26. Suzuki Y, Araie M, Ohashi Y. Sectorization of the central 30 degrees visual field in glaucoma. *Ophthalmology* 1993;100:69–75. [PubMed: 8433830]
27. Suzuki Y, Kitazawa Y, Araie M, et al. Mathematical and optimal clustering of test points of the central 30-degree visual field of glaucoma. *J Glaucoma* 2001;10:121–128. [PubMed: 11316094]

28. Wirtschafter JD, Becker WL, Howe JB, Younge BR. Glaucoma visual field analysis by computed profile of nerve fiber function in optic disc sectors. *Ophthalmology* 1982;89:255–267. [PubMed: 7088509]
29. Weber J, Dannheim F, Dannheim D. The topographical relationship between optic disc and visual field in glaucoma. *Acta Ophthalmol* 1990;68:568–574. [PubMed: 2275353]
30. Asman P, Heijl A. Glaucoma Hemifield Test: automated visual field evaluation. *Arch Ophthalmol* 1992;110:812–819. [PubMed: 1596230]
31. Goldbaum MH, Sample PA, Zhang Z, et al. Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci* 2005;46:3676–3683. [PubMed: 16186349]
32. Sample PA, Boden C, Zhang Z, et al. Unsupervised machine learning with independent component analysis to identify areas of progression in glaucomatous visual fields. *Invest Ophthalmol Vis Sci* 2005;46:3684–3692. [PubMed: 16186350]
33. Asman P, Heijl A. Evaluation of methods for automated Hemifield analysis in perimetry. *Arch Ophthalmol* 1992;110:820–826. [PubMed: 1596231]
34. Heijl, A.; Lindgren, G.; Lindgren, A., et al. Extended empirical statistical package for evaluation of single and multiple fields in glaucoma: Statpac 2. In: Mills, RP.; Heijl, A., editors. *Perimetry Update*. Vol. 91. Kugler Publications; Amsterdam: 1990. p. 303-315.
35. Garway-Heath DF, Holder GE, Fitzke FW, Hitchings RA. Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma. *Invest Ophthalmol Vis Sci* 2002;43:2213–2220. [PubMed: 12091419]
36. Bowd C, Zangwill LM, Medeiros FA, et al. Confocal scanning laser ophthalmoscopy classifiers and stereophotograph evaluation for prediction of visual field abnormalities in glaucoma-suspect eyes. *Invest Ophthalmol Vis Sci* 2004;45:2255–2262. [PubMed: 15223803]
37. Chan K, Lee TW, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Trans Biomed Eng* 2002;49:963–974. [PubMed: 12214886]
38. 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. [PubMed: 12147600]
39. Zangwill LM, Chan K, Bowd C, et al. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci* 2004;45:3144–3151. [PubMed: 15326133]
40. Vapnik V, Chapelle O. Bounds on error expectation for support vector machines. *Neural Comput* 2000;12:2013–2036. [PubMed: 10976137]
41. Boser, BE.; Guyon, IM.; Vapnik, VN. 5th Annual ACM Workshop on Computational Learning Theory. ACM Press; Pittsburgh, PA: 1992. A training algorithm for optimal margin classifiers; p. 144-152.
42. Haykin, S. *Neural Networks: A Comprehensive Foundation*. Vol. 2nd ed. Prentice Hall; Upper Saddle River, NJ: 1999.
43. LeCun, Y.; Jackel, LD.; Bottou, L., et al. *Neural Networks: Proceedings of the CTP-PBSRI Joint Workshop on Theoretical Physics*. World scientific; Singapore: 1995. Learning algorithms for classification: a comparison on handwritten digit recognition; p. 261-276.
44. Dumais, ST.; Platt, JC.; Heckerman, D.; Sahami, M. Inductive learning algorithms and representations for text categorization. In: Gardarin, G.; French, J.; Pissinou, N.; Makki, K.; Bouganim, L., editors. *Proceedings of CIKM '98 7th International Conference on Information and Knowledge Management*. New York: 1998. p. 148-155.
45. Mangasarian OL, Street WN, Wolberg WH. Breast cancer diagnosis and prognosis via linear programming. *Oper Res* 1995;43:570–577.
46. Burbidge R, Trotter M, Buxton B, Holden S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput Chem* 2001;26:5–14. [PubMed: 11765851]
47. El Beltagi TA, Bowd C, Boden C, et al. Retinal nerve fiber layer thickness measured with optical coherence tomography is related to visual function in glaucomatous eyes. *Ophthalmology* 2003;110:2185–2191. [PubMed: 14597528]

48. Weinreb RN, Shakiba S, Sample PA, et al. Association between quantitative nerve fiber layer measurement and visual field loss in glaucoma. *Am J Ophthalmol* 1995;120:732–738. [PubMed: 8540546]
49. Turpin, A.; Frank, E.; Hall, M.; Witten, IH.; Johnson, CA. The 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD). Hong Kong: 2001. Detecting progression in glaucoma using data mining techniques; p. 136-147.
50. Weber J, Ulrich H. A perimetric nerve fiber bundle map. *Int Ophthalmol* 1991;15:193–200. [PubMed: 2050474]
51. Anderson, DR.; Patella, VM. Automated Static Perimetry. Vol. 2nd ed.. Mosby; St. Louis: 1999.

Figure 1. Schematics of (A) Garway-Heath map, (B) GHT sector map, (C) vb-ICA map, and (D) random map. *White circles:* the 24-2 test locations. The number within each circle indicates the cluster to which that location belongs. *White lines:* cluster boundaries. *Light gray sections* on the GHT sector map indicate the eight locations that are not included in the map.

Figure 2.

Areas under the ROC curve for each combination of classifier and map combination (*top*) and a scatterplot of sensitivities for map-MLC combinations at two specificities (*bottom*) for the initial evaluation group. A value of 1 indicates perfect discrimination, whereas 0.5 indicates chance discrimination.

* In this case, "GHT" refers to the Glaucoma Hemifield Test output from Statpac, "Outside Normal Limits" = abnormal

Figure 3.

The four components of the analysis. Each box represents a unique combination of map and MLC. *Ovals*: clinical indices from Statpac 2 (Carl Zeiss Meditec, Inc., Dublin, CA). *Gray*: output from QDA; *white*: output from SVMg. *Lines*: comparisons between pairs of conditions. Analyses were conducted on both specificity and sensitivity in the independent data set.

^Statistically significant difference in sensitivity when using the 96% specificity criterion;

**statistically significant difference in sensitivity when using the 90% specificity criterion.

Figure 4.

Sensitivity for each combination of classifier and map for the independent data set for specificities of 96% and 90% (*top*), scatterplot of sensitivity versus actual specificity for 90% (*bottom left*) and 96% (*bottom right*) specificity cutoffs.

Table 1

Participant Descriptive Statistics

	Healthy Eyes		GON Eyes	
	Independent	Initial Evaluation	Independent	Initial Evaluation
Age (y) *	65.1 (1.2)	50.0 (1.1)	67.3 (1.2)	62.1 (0.9)
Gender (% female)	68		63	
Race	1 African descent, 1 Asian, 74 white, 7 unknown		4 African descent, 1 Asian, 63 white, 3 unknown	
MD (dB)	-0.6 (1.6) [†]	-0.11 (0.1) [†]	-2.0 (1.6) [‡]	-3.9 (0.3) [‡]
PSD (dB)	1.9 (0.6)	1.9(0.1)	2.9 (0.6) [§]	5.2 (0.2) [§]

* There was no significant difference in age between healthy and GON eyes in either data set (t -test, $P > 5\%$).

[†] Healthy eyes from independent and initial evaluation data sets were significantly different ($P < 0.05$).

[‡] GON eyes from independent and initial evaluation data sets were significantly different ($P < 0.05$).

[§] GON eyes from independent and initial evaluation data sets were significantly different ($P < 0.05$).

Table 2
Pair-wise Comparison of Map/MLC Sensitivity with Probabilities from the McNemar Test

	QDA					SVMg				
	GH	GHT	vbICA	Unclustered	Random	GH	GHT	vbICA	Unclustered	Random
96% Specificity Cutoffs										
QDA										
GH	—	0.002	0.02	<0.001*	0.04	0.82	0.38	1.0	0.30	0.66
GHT		—	0.36	0.61	0.31	0.003	<0.001	<0.001	0.02	<0.001
vbICA			—	0.13	1.0	0.08	0.12	0.03	0.31	0.009
Unclustered				—	0.06	<0.001	<0.001	<0.001	0.004	<0.001
Random					—	0.11	0.003	0.04	0.30	0.007
SVMg										
GH						—	0.06	0.73	0.34	0.18
GHT							—	0.12	0.001	0.69
vbICA								—	0.18	0.45
Unclustered									—	0.01
Random										—
90% Specificity Cutoffs										
QDA										
GH	—	0.66*	0.82	0.80	0.24	0.007	0.01*	0.02*	0.02	0.002
GHT		—	0.30*	0.27	0.65	1.0	0.05	0.09	0.08	0.01
vbICA			—	1.0	0.12	0.003	0.007*	0.007*	0.01	0.001
Unclustered				—	0.08	<0.001	<0.001	0.001	0.001	<0.001
Random					—	0.13	0.19	0.26	0.26	0.05
SVMg										
GH						—	1.0	0.69	0.62	0.73
GHT							—	1.0	1.0	0.38
vbICA								—	1.0	0.12
Unclustered									—	0.29
Random										—

Sensitivities were calculated when the 96% and 90% cutoffs from the training set were applied to the independent data set. The shaded quadrants of the table represent within-MLC comparisons. QDA, quadratic discriminant function; SVMg, support vector machine, Gaussian kernel; GH, Garway-Heath map; GHT, Glaucoma Hemifield Test sector map; vbICA, variational Bayesian

independent component analysis map. Unclustered, 52 locations, no clustering; Random, random map (i.e., locations randomly assigned to cluster). Bold type indicates comparisons that are significant at $P < 0.05$.

* Significant difference in specificity, $P < 0.05$.