

Using Machine Learning Classifiers to Identify Glaucomatous Change Earlier in Standard Visual Fields

Pamela A. Sample,¹ Michael H. Goldbaum,¹ Kwokleung Chan,² Catherine Boden,¹ Te-Won Lee,² Christiana Vasile,¹ Andreas G. Boehm,^{1,3} Terrence Sejnowski,⁴ Chris A. Johnson,⁵ and Robert N. Weinreb¹

PURPOSE. To compare the ability of several machine learning classifiers to predict development of abnormal fields at follow-up in ocular hypertensive (OHT) eyes that had normal visual fields in baseline examination.

METHODS. The visual fields of 114 eyes of 114 patients with OHT with four or more visual field tests with standard automated perimetry over three or more years and for whom stereophotographs were available were assessed. The mean (\pm SD) number of visual field tests was 7.89 ± 3.04 . The mean number of years covered (\pm SD) was 5.92 ± 2.34 (range, 2.81–11.77). Fields were classified as normal or abnormal based on Statpac-like methods (Humphrey Instruments, Dublin, CA) and by several machine learning classifiers. The machine learning classifiers were two types of support vector machine (SVM), a mixture of Gaussian (MoG) classifier, a constrained MoG, and a mixture of generalized Gaussian (MGG). Specificity was set to 96% for all classifiers, using data from 94 normal eyes evaluated longitudinally. Specificity cutoffs required confirmation of abnormality.

RESULTS. Thirty-two percent (36/114) of the eyes converted to abnormal fields during follow-up based on the Statpac-like methods. All 36 were identified by at least one machine classifier. In nearly all cases, the machine learning classifiers predicted the confirmed abnormality, on average, 3.92 ± 0.55 years earlier than traditional Statpac-like methods.

CONCLUSIONS. Machine learning classifiers can learn complex patterns and trends in data and adapt to create a decision surface without the constraints imposed by statistical classifiers. This adaptation allowed the machine learning classifiers to identify abnormality in visual field converts much earlier than the traditional methods. (*Invest Ophthalmol Vis Sci.* 2002;43:2660–2665)

In this study, we investigated whether machine learning classifiers, including neural networks, are useful for identifying which individuals with initially normal visual fields will

have development of abnormal visual fields later, due to glaucoma. Neural networks are a subset of machine learning classifiers. The terminology has been changed in the artificial intelligence community to the latter term to include classifiers that “learn”, but do not necessarily mimic, a simple neural pathway of the brain, as neural networks do. Machine learning classifiers usually use a form of supervised learning. Supervised learning refers to systems that are trained, instead of programmed, by a set of examples that are input-output pairs.¹ The input is the data and the output is the classification made by the machine learning method. During training, the classifier is told whether it is correct or incorrect based on a gold standard, and, after each run-through, it adjusts its internal parameters to arrive at more correct responses. This process is repeated until the classification performance does not improve. After training, the goal is that the machine classifier has learned and can correctly classify new input data that were not part of the original training sets. An attractive aspect of these classifiers is their ability to learn complex patterns and trends in data and to create decision rules adaptively, without the constraints imposed by statistical classifiers.^{2,3}

In a previous study, we compared the ability of several classifiers to detect early field loss.⁴ The inputs to the classifiers in that study were threshold values from standard visual fields plus the age from either healthy eyes or from eyes with glaucomatous optic neuropathy (GON). Because there is no absolute agreed-on gold standard for the presence of early glaucoma, the surrogate gold standard in this previous study, used to train the classifiers, was the absence or presence of GON. Visual field results were not used to select subjects or as a gold standard to train the output. The output from each classifier was a designation of either “normal field” or “glaucomatous field”. The classifiers’ results were also compared with those of two glaucoma experts and the Statpac 2 indices^{5,6} (Humphrey Instruments, Dublin, CA) that are typically used to identify field abnormalities. We found that several machine learning classifiers representing different methods of learning and reasoning performed well in comparison with both Statpac 2 and the glaucoma experts when classifying the visual fields.

The purpose of the present study was to apply the best candidate machine learning classifiers from our previous study, along with more Statpac-like traditional classifiers, to a new set of longitudinal standard automated perimetry (SAP) data from 114 ocular hypertensive eyes. If the classifiers could identify visual field converts from this group, they might have great utility in situations in which experts in glaucoma are not available and for standardization of methods in clinical trials.

METHODS

Subjects

One hundred fourteen eyes of 114 patients with ocular hypertension (OHT) followed up longitudinally at the University of California, San Diego (UCSD) were included. All patients had a minimum of four field tests with standard automated perimetry (SAP) over three or more

From the ¹Glaucoma Center, Department of Ophthalmology, and the ²Institute for Neural Computation, University of California at San Diego, La Jolla, California; the ⁴Computational Neurobiology Laboratory, Salk Institute, La Jolla, California; the ³Department of Ophthalmology, University of Dresden, Germany; and ⁵Discoveries in Sight, Devers Eye Institute, Portland, Oregon.

Supported by National Eye Institute Grants EY08208 (PAS), EY13235 (MHG), and EY03424 (CAJ); Research to Prevent Blindness (PAS); the Foundation for Eye Research (CV); and the Howard Hughes Medical Institute (TS).

Submitted for publication September 25, 2001; revised March 20, 2002; accepted April 1, 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: Pamela A. Sample, Department of Ophthalmology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0946; psample@eyecenter.ucsd.edu.

TABLE 1. Final Criteria for Abnormal Visual Field

SAGE Criterion	Specificity (%)	Confidence Limits (%)	Abnormal (%)	n
PSD worse than 1%, confirmed on the next test	100.0	96-100	18	21
GHT outside normal limits, confirmed on the next test	98.9	94-100	21	24
One hemifield cluster worse than the 1% level, confirmed on the next test	100.0	96-100	23	26
Two hemifield clusters worse than the 5% level, confirmed on the next test	97.9	93-100	15	17
Four points worse than the 5% level on pattern deviation, confirmed on next test	98.9	94-100	24	27
Eyes conversion confirmed by any of the above criteria			32	36

The specificity for each criterion is shown, along with the confidence limits and the percentage of the 114 eyes showing confirmed change from normal to abnormal fields based on the traditional methods. Overall specificity was 96%.

years and stereo optic disc photographs. The Human Subjects Committee of the University of California, San Diego, approved this study and its protocol adhered to the Declaration of Helsinki, with informed written consent obtained from all participants.

Each subject underwent a complete ophthalmic examination, which included review of relevant medical history, best corrected visual acuity, slit lamp biomicroscopy (including gonioscopy), applanation tonometry, dilated funduscopy, and fundus photography.

All patients had OHT with intraocular pressures more than 23 mm Hg when measured on at least two separate occasions and normal SAP visual field test results (defined later) on baseline examination. All had best corrected acuity of 20/40 or better, spherical refraction within +5.0 D, and cylinder correction within +3.0 D. Patients with significant lens opacity at the baseline clinical examination or on subsequent ophthalmic examinations were included. Patients with other disorders known to affect visual fields were excluded. Neither visual field nor optic nerve status was used to select these subjects.

Visual Fields

Visual field testing consisted of SAP, with the Full-Threshold test strategy and the 24-2 stimulus presentation pattern of the Humphrey Visual Field Analyzer (Humphrey Instruments) with 31.5 apostilbs (10 candelas/m²) white background and a Goldmann size III stimulus. Patients had to have normal visual fields (see definition of abnormality below) at baseline with at least three additional follow-up fields over a 3-year period. The mean (\pm SD) number of fields was 7.89 ± 3.04 . The mean (\pm SD) number of years covered was 5.92 ± 2.34 (range, 2.81-11.77 years). All visual fields from all eyes were evaluated by all machine learning and statistical classification methods.

Optic Disc

Included eyes also had serial simultaneous stereophotographs evaluated for evidence of glaucomatous optic neuropathy determined independently in a masked review by two glaucoma specialists at the Optic Disc Reading Center at USC. Photographs were masked for temporal order. Disagreements were resolved by consensus or adjudication. Optic discs were considered abnormal when one or more of the following was present: excavation or undermining of the cup, nerve fiber layer defects, notching or rim thinning, or cup-to-disc asymmetry between eyes of more than 0.2. Normal optic discs showed no evidence of these abnormal findings. The findings were not part of the inclusion-exclusion criteria for the study, but the presence and timing of detectable GON is reported in those eyes showing conversion from normal to abnormal visual fields by the various classifiers.

Normative Data

There were two sets of normative data used in this study. The first was used to develop a Statpac-like analysis package for determining single field abnormality. The second was a longitudinal data set used to set specificity cutoffs for conversion from normal to abnormal fields. Each is described in the following sections.

Statpac-like Visual Field Analysis. To compare the machine learning classifiers to more traditional statistical approaches for analyzing

visual fields, we used a Statpac-like analysis developed for the short wavelength automated perimetry (SWAP) ancillary arm of the Ocular Hypertension Treatment Study (OHTS).⁷ This analysis with its own normative database was developed to allow extraction of data that are not on the field analyzer printout (e.g., numerical glaucoma hemifield sector values) and to provide export of all data to spreadsheets. Although some of the information could be extracted from the field analyzer printout, we thought it important that all analyses be based on the same normative data set, to allow us to make comparisons between SAP and SWAP in future studies on machine learning classification of SWAP, using the same normative data for both tests.

The normative database consisted of one eye from each of the same 348 normal subjects tested on both SAP and SWAP between the ages of 20 and 85. The data were collected at five different centers by a standardized test protocol identical with that used to establish the field analyzer's internal normative database. To be included in the normative database, all subjects had to have a normal findings in an eye examination, 20/30 or better visual acuity, normal color vision, no history of ocular or neurologic disease or surgery, refractive error of less than 5 D spherical equivalent and 3 D cylinder, no diabetes, and normal optic nerve appearance. They could not be taking any medications known to affect visual fields or color vision.

After age correction for each of the 52 test locations of program 24-2 (the two blind spot locations are not included), the total deviation plot, pattern deviation plot, their associated probability cutoffs, and probability plots were computed. The package then computed the global indices and the cutoff values at specific probabilities for mean deviation (MD) and pattern standard deviation (PSD) along with an asymmetry analysis patterned after the glaucoma hemifield test (GHT) analysis.⁶

Setting Specificities. When developing an algorithm for conversion from normal to abnormal fields it is important to identify a meaningful specificity for field conversion in longitudinal data sets. To determine the parameters from the analysis that would provide a high specificity for visual field conversion, 94 normal eyes from a longitudinal study that had been performed at the University of California, Davis, were used. Each had been followed-up annually and had had four visual field tests. They ranged in age from 21 to 85. The inclusion-exclusion criteria for these normal subjects was identical with those for the 114 eyes with OHT used in the present study, with the exception that they had intraocular pressure less than 20 mm Hg and a family history of glaucoma. None was part of the normative database sample of 348 eyes used to develop the Statpac-like analysis package described in the previous section. Fifty candidate criteria for change were evaluated, and the specificity for each in the 94 longitudinally followed-up normal eyes was determined. This analysis resulted in five best criteria for change from a normal to an abnormal field.⁷ The specificities and confidence limits for the five are summarized in Table 1. The resultant algorithm for abnormality based on any one of the five criteria with confirmation is called Statpac-like Analysis for Glaucoma Evaluation (SAGE).

Confirmation of a suspected change greatly improved specificity for all criteria. We used these criteria in combination in the present

study to classify fields from the 114 eyes with OHT into normal and abnormal categories. A field conversion with SAGE was defined a normal visual field at baseline in which abnormality developed later, based on at least one of the criteria and confirmed by that same criterion on the subsequent visual field test. Because any one criterion was sufficient, the specificity overall was approximately 96% (0% false-positive results for PSD worse than 1%, 0% for the 1 hemifield cluster at 1%, 1.1% for GHT result of "outside normal limits," 1.1% for 4 points on the PSD plot at 5%, plus 2.1% for two hemifield clusters at 5%; all confirmed). A specificity of 96% was also found for our best glaucoma expert in our previous study.⁴ These best-candidate criteria of SAGE are all elements of the statistical analysis package (Statpac II) used on the Humphrey Visual Field Analyzer and can be considered among the current standards for classification of visual fields.

Machine Learning Classifiers

Several machine learning classifiers were compared with the results obtained with SAGE to assess their ability to classify the fields of the 114 ocular hypertensive eyes to determine which fields would be classified as abnormal. The machine learning classifiers were chosen based on the results of our previous study, in which we trained a set of classifiers to categorize SAP visual fields as normal or abnormal.⁴ In that study, the surrogate gold standard for glaucoma was the presence of glaucomatous optic neuropathy. Visual fields were not used to classify the subjects. The study indicated several classifiers that could separate fields from normal eyes and eyes with GON with a high specificity and sensitivity. These already trained classifiers were used in the present study. The sensitivity (the proportion of fields from eyes with GON classified as abnormal) and the specificity (the proportion of fields from normal subjects classified as normal) depended on the selection of a threshold cutoff value along the range of outputs for each classifier. We set the cutoff values for the present study to obtain a specificity of 96% for each of the classifiers, using the same 94 normal eyes from Dr. Johnson's longitudinal study that determined the 96% overall specificity for SAGE.⁷ As with SAGE, the cutoff needed for that specificity was based on two confirmed abnormal visual fields.

Consistent with the previous study, the input to each of the classifiers listed in the following sections was the absolute threshold at each of the 52 locations of the visual field and age. Training and testing were performed in our previous study, using cross-validation to classify eyes with known GON versus healthy eyes. In the present study, we asked these same already trained classifiers to classify visual fields from an independent group of 114 eyes selected based on IOP and normal baseline visual fields and not on optic nerve status.

SAGE is a type of classifier that uses statistically determined cutoffs to distinguish between classes. The attractive aspect of machine learning classifiers is their ability to learn complex patterns and adapt to the data. They are not constrained to linear analysis, which would, for example, result in a decision surface that is a line in two dimensions between the data of patients with abnormal fields and those with normal fields, or a plane in three dimensions. Instead, the decision surface can be any shape in the dimension that provides the best separation between the groups. Abnormal fields fall on one side of the surface and normal on the other. The decision surface itself is a boundary between the two clusters of data. A brief description of the classifiers used follows. More detail on each can be found in the detailed Appendix to our previous study.⁴

Support Vector Machines. These are a new class of supervised machine learning algorithms or neural networks that are able to solve a variety of classification and regression (model-fitting) problems.^{8,9} A support vector machine (SVM) can separate data that are not easily separable in the original data space, by mapping the data of interest into a much higher dimensional space until a decision surface is identified that allows the separation of the input data—in our case into two groups of visual fields: normal and abnormal. The name of this classifier refers to support vectors, which are those data points that lie closest to the decision surface and therefore are the most difficult to

classify. As such, they have a direct bearing on the optimum location of the decision surface.² Training maximizes the margin of separation between the normal and abnormal vectors while minimizing the estimated generalization errors in classification.¹⁰⁻¹² The architecture (structure of the network) is similar to that of a multilayer perceptron (MLP), a basic form of neural network. It is a feed-forward network with an input layer, a hidden layer, and an output layer.

We used two types of SVM. For linearly separable data, the parameters used in the SVM-linear (SVMI) analysis are chosen so that the margin between the decision plane and the training examples is at maximum. To avoid the assumption of linear separability, we also used a multivariate Gaussian distribution (SVMg) analysis.² Both SVMs significantly outperformed the MLP in our previous study. They have also shown good generalization of performance in face recognition,¹³ text categorization,¹⁴ recognition of handwritten digits,¹⁵ and breast cancer diagnosis and prognosis.¹⁶

Mixture of Gaussian. The MoG is a special case of a committee machine.² Committee machines use a set of hidden analyses to divide a computationally complex task into a number of computationally simple tasks, performed by "committee members". Each member does well at modeling its own simplified data set. In the associative MoG model, the members perform self-organized learning (unsupervised learning) on the input data to achieve good partitioning. The fusion of all the members' outputs is combined with supervised learning to model the desired response. In our case, the desired response is "normal visual field" or "abnormal visual field." We made two adjustments to facilitate the computation of MoG. To help the MoG manage the high dimensionality of 53 inputs, we constrained the MoG analyses to one Gaussian cluster for each class. This constraint results in a quadratic discrimination function (QDF). In our previous work we found that this improved performance relative to Statpac indices. Also, these classifiers sometimes have difficulty with high-dimension input, and therefore we also projected the data by using principal component analysis (PCA) from the original 53 dimensions to a space of eight dimensions.¹⁷ PCA is a way of reducing the dimensionality of the data space by retaining most of the information in terms of its variance. Our previous work showed that for visual field data, QDF on the full dimension data worked comparably to the MoG on principle component analyzed data. We used both the QDF and MoG with PCA for the present study.

Mixture of Generalized Gaussians. The mixture of generalized Gaussians (MGG) uses the same architecture as MoG, except it is designed for situations in which the underlying distributions of the data for the classification problem are not necessarily Gaussian. For instance, the data may distribute with heavier tails or may even be bimodal. It would degrade performance of the classifier to model these problems with Gaussian distributions. With the development of a generalized Gaussian mixture model,^{17,18} we are able to model the class conditional density distributions with higher flexibility, while preserving a comprehension of the statistical properties of the data in terms of, for example, means, variances, and kurtosis. It has been demonstrated in real-data experiments that this model generally improves classification performance over the standard MoG in those cases in which the assumption of a Gaussian distribution of data is incorrect.^{17,18}

In summary, for our initial study, we chose the classifiers that have recently become popular due to their excellent classification performance and robustness in analysis of many data sets in different applications.^{9,19-21} The best among these for separating fields from eyes with GON and fields from normal eyes were used in the present study.⁴ The best were two types of discriminative classifiers, SVMI and SVMg, as well as, three generative classifiers (QDF, MoG, and MGG). Discriminative classifiers, such as SVM, minimize error by finding optimal boundaries between classes, whereas generative classifiers try to estimate the probability density of each class. These two principles are currently the state of the art in classifiers applied to pattern-recognition tasks.^{9,19-21}

TABLE 2. Machine Learning Classifier Results

	SVMI	SVMg	MoG	QDF	MGG	Mean Gap (y)
Abnormal at baseline	9	9	2	8	1	
Additional confirmed prior to traditional method	17	18	22	22	22	
Confirmed same time as traditional method	1	1	1	0	1	
Confirmed after traditional method	1	1	1	1	1	
Confirmed abnormal by traditional and classifier	28	29	26	31	25	
Average gap (y)*	4.393	4.43324	3.387	4.122	3.275	3.922
Standard deviation	2.919	2.747	1.546	1.780	1.590	0.554

SAGE identified 36 eyes converting to confirmed abnormal visual fields. Shown are classifier results including the mean (\pm standard deviation) number of years earlier that repeatable abnormal fields were identified by a classifier before SAGE identified the abnormality in the same eyes. Classifier data are the number of eyes.

*Gap between the traditional method date and classifier date when both determined eye to be abnormal.

RESULTS

The last two columns in Table 1 present the percentage and number of eyes with fields that converted from normal SAP visual fields to a field with a confirmed visual field defect, for each of the five criteria of SAGE. Some fields met more than one criterion. SAGE identified 36 (32%) of the 114 eyes as having fields that converted.

Table 2 shows the results for the various classifiers. Thirty-eight percent (43/114) of the eyes converted by one or more methods. One or more of the classifiers identified all 36 eyes identified by SAGE. QDF identified 31 plus 1 additional eye; SVMI, 28 plus 6 additional eyes; SVMg, 29 plus 7 additional eyes; MoG, 26 plus 1 additional eye; and MGG, 25 plus 1 additional eye.

The agreement among SAGE and all classifiers for categorization of eyes as having fields that would convert to abnormal versus remaining normal was good to excellent ($\kappa = 0.63$ – 0.91 ; Table 3).²² For the most part, the same eyes were identified as converting to abnormal fields by SAGE and by the classifiers under test. The two versions of SVM showed agreement of 96%, as did the MoG and MGG. This can be attributed to the similarities in their architecture. The QDF agreed most closely with the SAGE method (95% agreement).

In the 36 eyes with confirmed field abnormality based on SAGE and by at least one of the classifiers, the machine classifier predicted the abnormality several years before SAGE (Table 2). The average gap in years (\pm SD) with the use of SVMI was 4.39 ± 2.92 ; with SVMg, 4.43 ± 2.75 ; with MoG, 3.39 ± 1.55 ; with QDF, 4.12 ± 1.78 ; and with MGG, 3.28 ± 1.59 . These classifiers were all significantly earlier than the model in determining repeatable abnormality ($P < 0.0001$). Constraining MoG to QDF significantly improved the timing of detection among the generative classifiers. It was significantly earlier than both the unconstrained MoG ($P < 0.024$) and MGG ($P < 0.011$), but not the two SVMs. The number of eyes each classifier called abnormal on the baseline visits is shown in Table 2. Six of the seven eyes that were identified as abnormal by a classifier that always remained normal with SAGE are still under observation.

The classifiers used in this study were chosen because we had previously shown they could successfully separate the visual fields of normal eyes from those of eyes with GON. To determine whether use of these classifiers in a new data set would be consistent with our previous findings, we assessed the presence of GON in the 43 (of 114) eyes that showed confirmed abnormal fields based on one or more classifiers. Two of these eyes had photographs that could not be assessed because of poor quality or missing information, leaving 41 eyes for the analysis. Based on the SVMg results, 66% (27/41) of these eyes had a glaucomatous optic disc at baseline or within 1 year thereafter. An additional 16% (5/41) showed development of GON sometime during follow-up, although 22% (9/41) showed no discernible evidence of GON during the course of the study. Table 4 shows the GON results, by classifier, in eyes identified as having converted visual fields. The table shows that SVMg identified as many eyes as SAGE and that it identified mostly, although not always, the same eyes. However, SVMg showed the best agreement with the presence of GON at 94% (32/34). The percentage of eyes with confirmed abnormal fields identified by the other classifiers as having GON was 81% (26/32) with SVMI, 81% (21/26) with MoG, 75% (24/32) with QDF, 75% (18/24) with MGG, and 74% (25/34) with SAGE.

DISCUSSION

Current methods for assessing change in visual fields fall into three main categories.^{23,24} Subjective assessment of a series of visual field printouts based on clinical judgment is perhaps most common. Trend analysis on a series of fields is useful if several (seven or more) field test results are available.²⁵ Finally, some form of event analysis to identify change in a single visual field relative to a reference is often used. Event analysis can be based on a system that scores both baseline and subsequent fields. A change in score then signals possible progression. Another form of event analysis is predicated on statistically based change. An example of this is the glaucoma change probability analysis available on the Humphrey Visual Field Analyzer (Humphrey Instruments).²⁶ This analysis signals

TABLE 3. Percent Agreement in Classification of the 114 Eyes

	SVMI	SVMg	MoG	QDF	MGG
SAGE	88 (0.711)	88 (0.716)	90 (0.761)	95 (0.874)	89 (0.738)
SVMI	—	96 (0.912)	87 (0.685)	84 (0.616)	88 (0.666)
SVMg	—	—	85 (0.630)	84 (0.623)	86 (0.649)
MoG	—	—	—	90 (0.749)	96 (0.877)
QDF	—	—	—	—	88 (0.677)
MGG	—	—	—	—	—

Kappa statistic is in parentheses.

TABLE 4. Results of Simultaneous Stereophotograph Analysis for Presence of GON in Eyes with Confirmed Abnormal Fields

	SVMI	SVMg	MoG	QDF	MGG	SAGE
Confirmed abnormal fields	32	34	26	32	24	34
GON >1 year before conversion by field	21	27	15	18	14	24
GON within ± 1 year of field conversion	2	2	6	5	4	0
GON >1 year after field conversion	3	3	0	1	0	1
No evidence of GON during study	6	2	5	8	6	9

$n = 41$.

change in individual test locations after taking into account the fluctuation found in a group of patients with stable glaucoma.

To differentiate true change from random fluctuation, evidence for change in visual fields should be confirmed on subsequent visual fields. For example, the Ocular Hypertension Treatment Study (OHTS),²⁷ the Advanced Glaucoma Intervention Study (AGIS),²⁸ the Collaborative Initial Glaucoma Treatment Study (CIGTS),²⁹ and the Early Manifest Glaucoma Trial (EMGT)³⁰ require that change from baseline be observed in three consecutive follow-up visual field analyses before change is verified. In each of these studies, a different algorithm was used for identifying change in their differing patient samples. There is no gold standard for change in visual fields.

This lack of a gold standard influenced the present study as well. In our initial work to determine the classifiers, we used the presence or absence of GON as our surrogate gold standard for glaucoma. The advantage of this was that it eliminated the bias in comparing classifiers with each other, Statpac, or the glaucoma experts that would be present if the fields themselves were used in the training. If a chosen set of criteria for field abnormality were used as the gold standard, it would probably include some elements of Statpac or expert judgment. This would, by definition, make those criteria perform the best. Using GON, a definite marker for glaucoma, eliminated this confounder. This said, some of the classifiers performed comparably to the best glaucoma expert.

In the present study, we avoided a gold standard for determining the sensitivity of each of the different classifiers and instead simply compared them. Therefore, the true sensitivity of the determinations is unknown. Whether these improvements in early detection are valid and whether they will be clinically useful remains to be seen. However, we can make a strong argument that the classifiers are indeed seeing something consistent with glaucomatous visual field loss for two reasons. First, they identified the same eyes that were later identified by traditional methods for assessing field abnormality. Second, they identified the same eyes that had GON or later development of GON. However, we must stress that our conclusions are based on the assumption that characteristic visual field abnormality determined by traditional methods in conjunction with GON indicates glaucoma.

With regard to specificity, there is an advantage to using a large longitudinal data set from normal eyes to determine which criteria identify glaucomatous change in visual fields from normal to early abnormality. This approach yielded five commonly used traditional criteria for SAGE, each highly specific individually, with only one confirmation required. Even in combination, they maintained a specificity of 96%. When cut-off values for each of the machine learning classifiers were also set this way, fair comparisons among methods were possible. Use of the longitudinal normal data set to select criteria for change is also supported by the high level of agreement among the various classifiers, and by the presence of GON in a high percentage of the eyes identified as changed, especially with SVMg.

The identification of change from a normal to abnormal visual field should be considered within the context of glau-

coma progression and available treatment options. In general, the time course of glaucoma is slow and the need for early intervention requires assessment of many factors in addition to the vision loss, including the patient's age, family history, other risk factors, and quality of life. Some elderly patients with a newly found loss of vision may expect to live out their lives without any noticeable change in performance or quality of life. However we cannot, as yet, accurately predict the likely rate of change for each individual. Younger patients at higher risk may show rapid change, and early detection, and intervention may significantly prolong good vision.

Although, current treatment for glaucoma involves lowering of intraocular pressure to target levels and ongoing follow-up for evidence of the success or failure of treatment, the advent of better medical and surgical therapies, genetic marking, and neuroprotective agents will most likely influence this treatment paradigm. The earlier detection of vision loss by machine learning classifiers, and their use in clinical trials to provide quantifiable and comparable evaluation of the data across sites could be very important in the accurate assessment of these new therapies. In our study, the machine learning classifiers detected visual field abnormality, on average, 4 years before traditional SAGE classification. In theory, their use could significantly shorten the time of clinical trials assessing small changes in SAP visual fields over time.

The use of appropriate machine learning classifiers may be even more important in studies in which other methods are used to measure visual function or optic nerve structure. Some of these newer methods, such as SWAP, frequency-doubling technology perimetry, and confocal scanning laser ophthalmoscopy, have already been used in clinical trials.^{31,32} Clinicians are not as familiar with these tests as they are with SAP. Interpretation of their results is therefore more difficult. In addition, the analysis packages available with the newer visual function tests are modifications of those developed for SAP and therefore may not be optimal. The use of machine learning classifiers with these newer tests should improve their utility and shorten clinical trial durations, although this remains to be studied.

In summary, we found that machine learning classifiers were able to identify confirmed change in visual fields of eyes with OHT substantially earlier than more traditional methods of analysis of SAP results.

References

1. Poggio T, Shelton CR. Machine learning, machine vision, and the brain. *AI Magazine*. 1999;20:37-58.
2. Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall; 1999:297.
3. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol 1. Cambridge, MA: MIT Press; 1986:Chap 8.
4. Goldbaum MH, Sample PA, Chan K, et al. Comparing machine classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci*. 2002;43:162-168.

5. Anderson DR, Patella VM. *Automated Static Perimetry*. 2nd ed. New York: Mosby; 1999:103-120.
6. Asman P, Heijl A. Glaucoma hemifield test: automated visual field analysis. *Arch Ophthalmol*. 1992;110:812-819.
7. Johnson CA, Sample PA, Cioffi GA, Liebmann JR, Weinreb RN. Structure and function evaluation (SAFE): I. Criteria for glaucomatous visual field loss. *Am J Ophthalmol*. In press.
8. Vapnik VN. *Statistical Learning Theory*. New York: John Wiley & Sons; 1998.
9. Vapnik VN. *The Nature of Statistical Learning Theory*. 2nd ed. New York: Springer-Verlag; 2000.
10. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*. 1998;2: 121-167.
11. Keerthi SS, Shevade SK, Bhattacharyya C, Murth KRK. A fast iterative nearest point algorithm for support vector machine classifier design. *IEEE Trans Neural Networks*. 2000;11:1124-1136.
12. Platt JC. Sequential minimal optimization: a fast algorithm for training support vector machines. In: Scholkopf B, Burges CJC, Smola AJ, eds. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press; 1998:185-208.
13. Osuna E, Freund R, Girosi F. Training support vector machines: an application to face detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Los Alamitos, CA: IEEE Computer Society Press. 1997; 130-136.
14. Dumais ST, Platt JC, Heckerman D, Sahami M. Inductive learning algorithms and representations for text categorization. In: Gardarin G, French JC, Pissinou N, Makki K, Bouganim L, eds. *Proceedings of the CIKM '98 7th International Conference on Information and Knowledge Management*. New York, NY: The Association for Computing Machinery. 1998:148-155.
15. LeCun Y, Jackel LD, Bottou L, et al. Learning algorithms for classification: a comparison on handwritten digit recognition. *Neural Networks: Proceedings of the CTP-PBSRI Joint Workshop on Theoretical Physics*. Singapore: World Scientific; 1995:261-276.
16. Mangasarian OL, Street WN, Wolberg WH. Breast cancer diagnosis and prognosis via linear programming. *Operations Res*. 1995;43: 570-577.
17. Lee T-W, Lewicki MS. The generalized Gaussian mixture model using ICA. In: Pajunen P, Karhunen J, eds. *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'00)*. Espoo, Finland: Otamedia. 2000:239-244.
18. Lee T-W, Lewicki MS, Sejnowski TJ. Unsupervised classification with non-Gaussian sources and automatic context switching in blind signal separation. *IEEE Trans Pattern Anal Mach Intell*. 2000;22:1078-1089.
19. Burges CJC. Geometry and invariance in kernel based methods. In: Scholkopf B, Burges CJC, Smola AJ, eds. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press; 1998:89-116.
20. Jordan M, Ghahramani Z, Jaakkola T, Saul OL. An introduction to variational methods in graphical models 1. In: Jordan MI, ed. *Learning in Graphical Models*. Cambridge, MA: MIT Press; 1999: 105-161.
21. Rabiner L, Juang OB-H. *Fundamentals of Speech Recognition, Vol 1. Signal Processing Series*. Englewood Cliffs, NJ: Prentice-Hall; 1993.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
23. Sample PA, Bosworth CF, Weinreb RN. The loss of visual function in glaucoma. *Semin Ophthalmol*. 2000;15:182-193.
24. Spry PGD, Johnson CJ. Identification of progressive glaucomatous visual field loss. *Surv Ophthalmol*. In press.
25. Fitzke FW, Hitchings RA, Poinosawmy D, McNaught AL, Crabb DP. Analysis of visual field progression in glaucoma. *Br J Ophthalmol*. 1996;80:40-48.
26. Morgan RK, Feuer WK, Anderson DR, Statpac 2 glaucoma change probability. *Arch Ophthalmol*. 1991;109:1690-1692.
27. Keltner JL, Johnson CJ, Spurr JO, Kass MA, Gordon MO, and OHS Group. Confirmation of visual field abnormalities in the Ocular Hypertension Treatment Study (OHTS). *Arch Ophthalmol*. 2000; 118:1187-1194.
28. Advanced Glaucoma Intervention Study. 2. Visual field test scoring and reliability [see comments]. *Ophthalmology*. 1994;101:1445-1455.
29. Leske MC, Heijl A, Hyman L, Bengtsson B. Early Manifest Glaucoma Trial: design and baseline data. *Ophthalmology*. 1999;106:2144-2153.
30. Lichter PR, Mills RP, CIGTS Study Group: Quality of life study: determination of progression. In: Anderson DR, Drance SM, eds. *Encounters in Glaucoma Research 3: How to Ascertain Progression and Outcome*. Amsterdam: Kugler Publications; 1996:149-163.
31. Sample PA. Short-wavelength automated perimetry: its role in the clinic and for understanding ganglion cell function. *Prog Retinal Eye Res*. 2000;19:369-383.
32. Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol*. 1999;117:573-583.