Determining Progression in Glaucoma Using Visual Fields

Andrew Turpin[1], Eibe Frank[2], Mark Hall[2], Ian H. Witten[2], and Chris A. Johnson[1]

[1] Discoveries in Sight, Devers Eye Institute, Portland OR USA
{aturpin, cjohnson}@discoveriesinsight.org

[2] Department of Computer Science
University of Waikato
Hamilton, New Zealand
{eibe, mhall, ihw}@cs.waikato.ac.nz

Abstract. The standardized visual field assessment, which measures visual function in 76 locations of the central visual area, is an important diagnostic tool in the treatment of the eye disease glaucoma. It helps determine whether the disease is stable or progressing towards blindness, with important implications for treatment. Automatic techniques to classify patients based on this assessment have had limited success, primarily due to the high variability of individual visual field measurements.
The purpose of this paper is to describe the problem of visual field classification to the data mining community, and assess the success of data mining techniques on it. Preliminary results show that machine learning methods rival existing techniques for predicting whether glaucoma is progressing—though we have not yet been able to demonstrate improvements that are statistically significant. It is likely that further improvement is possible, and we encourage others to work on this important practical data mining problem.

1 Introduction

Glaucoma, a disease that affects the optic nerve, is one of the leading causes of blindness in the developed world. Its prevalence in populations over the age of 40 of European extraction is about 1 to 2%; it occurs less frequently in Chinese populations, and much more often in people descending from West African nations [9]. There are several different types of glaucoma, but all share the characteristic that structural damage to the optic nerve tissue eventually leads to loss of visual function. If left untreated, or treated ineffectively, blindness ultimately occurs.

The cause of glaucoma involves many factors, most of which are poorly understood. If diagnosed early, however, suitable treatment can usually delay the loss of vision. If a patient continues to lose visual function after the initial diagnosis, their glaucoma is said to be *progressing*.

Determining progression is of vital importance to patients for two reasons. First, if a patient's vision continues to deteriorate under treatment, the treatment must be altered to save their remaining visual function. Second, progression, or rate of progression, is used as the main outcome measure of clinical research trials involving new glaucoma drugs. If effective medication is to be developed and made available for widespread use, accurate detection of progression is essential. Patient care is as dependent on monitoring the progression of glaucoma as it is on the original diagnosis of the disease.

Unfortunately, there is no universally accepted definition of glaucomatous progression. Various clinical drug trials use different definitions, and criteria used by practicing ophthomologists in vision clinics also differ widely. All agree, however, that *visual field* measurements are an essential tool for detecting progression.

The next section explains the standard technique for measuring visual fields. Following that, we describe standard methods that have been used to determine progression. In Section 4 we introduce a dataset that has been assembled for the purpose of testing different progression metrics, and describe the data mining methods that we have applied. Then we summarize the results obtained, using the standard technique of pointwise linear regression as a benchmark. Surprising results were obtained using a simple 1R classifier, while support vector machines equaled and often outperformed the benchmark. We also investigated many other techniques, including both preprocessing the data by smoothing, taking gradients, using t-statistics, and physiologically clustering the data; and different learning schemes such as decision trees, nearest-neighbor classifiers, naive Bayes, boosting, bagging, model trees, locally weighted regression, and higher-order support vector machines. Although we were able to obtain little further improvement from these techniques, the importance of the problem merits further study.

2 Visual Field Measurement

The standard visual field assessment currently employed in glaucoma management requires the patient to sit facing a half sphere with a white background that encompasses their entire field of vision. Typically only the central 30° of the visual field is tested. Subjects are instructed to fixate in the center of the hemisphere and press a button whenever they see a small white light flash in the "bowl." Lights of varying intensities are flashed in 76 locations of the visual field, and the minimum intensity at which the patient can see the target is recorded as their *threshold*. It is not feasible to spend more than about 10 seconds determining patient thresholds at each location because fatigue influences the results, but consistent threshold measurements are possible—particularly in reliable patients. Thresholds are scored on a logarithmic decibel scale. At any particular location a score of 0 dB indicates that the brightest light could not be seen (blindness), while 35 dB indicates exceptional vision at that point.

Fig. 1 shows visual field measurements taken on the right eye of a progressing glaucoma patient at an interval of five years. Each number is a dB threshold value. There are 76 locations in each field, separated by 6° of visual angle. Note the blind spot (readings $\leq$1 dB) at $(15, -3)$ where the optic nerve exits the eye. In a left eye this blind spot occurs at $(-15, -3)$—assuming the patient remains fixated on the center spot for the duration of the test. For convenience in data processing, the asymmetry between eyes is removed by negating left eye x-coordinates, so that all data is in a right eye format.

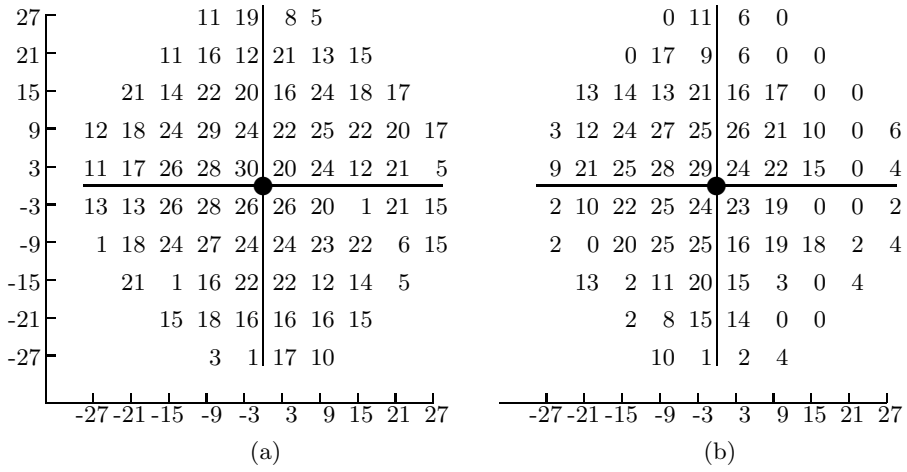

Fig. 1. Visual fields for a progressing right eye. (a) Baseline measurement, (b) the same eye five years later. The axes measure degrees of visual angle from the center fixation mark.

To assist the diagnosis of glaucoma, it is useful to compare the thresholds recorded in a visual field to a database of normal thresholds. To monitor progression, visual field measurements must be compared from one visit to the next, seeking locations whose thresholds have decreased. In Fig. 1 it is obvious that visual sensitivity around $(15, 15)$ and $(9, -15)$ has decreased from Fig. 1(a) to Fig. 1(b). Visual fields are quite noisy, however, so thresholds from a large series of visits are usually required to distinguish true progression from measurement noise. For example, the threshold at location $(-9, -27)$ in Fig. 1 increases from 3 dB to 10 dB, which in this case is probably because the original estimate of 3 dB is low. The entire field can fluctuate from visit to visit ("long term variability" in the ophthomology literature) depending on factors such as the patient's mood and alertness, as well as physiological factors like blood pressure and heart rate. Moreover, each location can vary during the test procedure ("short term

variability") depending on a patient's criterion for pressing the response button, fatigue, learning effects, and mistakes.

The challenge is to detect progression as early as possible, using the smallest number of sequential visual fields.

3 Determining Progression

Glaucoma patients are usually tested at yearly or half-yearly intervals. When presented with visual fields from successive visits, the ophthomologist's task is to decide whether change has occurred, and if so whether it indicates glaucoma or is merely measurement noise. This paper casts the decision as a classification problem: patients must be classified as *stable* or *progressing* based solely on their visual field measurements. Several automatic techniques exist to aid the clinician in this task; they can be divided into three broad classes.

The first group bases the classification on "global indices," which average information across all locations of the visual field. The most commonly used such measure is *mean deviation*, which averages the difference between measured thresholds and a database of thresholds in non-glaucomatous eyes over the visual field. Each location is weighted, based on the variability of normal thresholds. The spatial variance of this measure (referred to as *pattern standard deviation*) is also used clinically. A third global technique assigns scores to locations based on their threshold values and those of their immediate neighbors, and sums the scores into a single measure of the visual field [13]. Studies have shown that regression on global indices alone is a poor indicator of progression [2, 6, 20, 22].

Classifiers in the second group treat each location independently of the others. By far the most common approach is to use pointwise univariate linear regression (PWLR) to detect trends in individual locations [2, 8, 20, 22, 23, 27]. Typically this is applied to each individual location, and if the slope of the fitted line is significantly less than zero the patient is classified as *progressing*. Several variations on this theme have been investigated, the most notable being to correct for natural age-related decline in thresholds [27]. Performing PWLR on each of the 76 locations introduces a multiple-comparison problem. This is solved previously using multiple t-tests with a Bonferroni correction, ignoring the fact that locations in the visual field are not strictly independent.

Other pointwise techniques have been investigated, including multivariate regression [20] and using high-order polynomials to fit threshold trends [26]. The *glaucoma change probability* [19] calculates the likelihood that the difference between a threshold and a baseline measure falls outside the 95% confidence limits established by a database of stable glaucomatous visual fields.

All these pointwise techniques can be refined by requiring that a cluster of points show progression, rather than a single location. Alternatively, progression can be confirmed on subsequent tests by requiring that points not only show progression after examining n visual fields, but also when $n + 1$ fields are examined. These techniques are currently employed in combination with the glaucoma change probability as outcome measures in several drug trials [17].

The final group of classifiers falls between the two extremes of global indices and pointwise modeling. These classifiers attempt to take account of the spatial relationship between neighboring locations. Some approaches cluster the locations based on known physiological retinal cell maps [18, 24], and apply regression to the mean of the clusters [20, 22]. Others use neural networks as classifiers, relying on the network to learn any spatial relations [3, 11, 16, and references therein].

It is difficult to compare different approaches, because most studies do not report standard classification metrics. Moreover, patient groups used in experimental studies differ in size, stage of glaucoma, and type of glaucoma. To provide a baseline for comparison, we include results from one of the best-performing statistical methods, pointwise univariate regression, which is specifically tailored to this application [2]. This technique is detailed in Section 4.2 below.

4 Experiments

Diagnosing glaucoma progression is a prime example of a medical classification task that is ideally suited for data mining approaches because pre-classified training data, although arduous to collect, is available. However, apart from neural networks [16] no standard data mining paradigms have been applied to this problem, and there is no evidence that neural networks outperform well-known special-purpose statistical algorithms designed for this application.

We present an empirical evaluation of two standard data mining algorithms, support vector machines [7] and 1R decision stumps [12], and show that they detect progressing glaucoma more accurately than pointwise univariate regression.

4.1 The Data

Data was collected retrospectively from patient charts of 113 glaucoma patients of the Devers Eye Institute, each having at least 8 visual field measurements over at least a 4 year period as part of their regular ophthomologic examination. Unlike many previous studies, no special efforts were made to ensure patient reliability, nor was the quality of the visual fields evaluated. These are typical patient records from a typical clinical situation.

The patients were classed as *progressing*, *stable*, or *unknown* by an expert (author CAJ), based on optic disk appearance and their visual field measurements. The final data set consisted of 64 progressing eyes and 66 stable eyes, each with eight visual field measurements at different points in time. The visual field threshold values were adjusted for age at measurement by 1 dB per decade, so in effect all eyes were from a 45 year old patient. Left eyes were transformed into right eye coordinates.

4.2 The methods

Pointwise univariate linear regression analysis (PWLR) is a well-established diagnostic tool in the ophthomology literature. In an empirical comparison of several statistical methods for detecting glaucoma progression—pointwise univariate regression, univariate regression on global indices, pointwise and cluster-wise multivariate regression, and glaucoma change analysis—it emerged as the most accurate predictor of clinician diagnosis (Table 4 in [20].) It consists of three steps. First, a linear regression function is fitted to each of the 76 locations in a visual field, using all available measurements for the patient (8 in the data we used.) Second, a significance test is performed on those locations with a negative slope to ascertain whether the decrease in sensitivity is statistically significant. Third, a series of visual field measurements is declared to be *progressing* if at least two adjacent test locations are deemed significant, and *stable* otherwise.

In our experiments, we used a one-sided t-test to detect progression [25]. We experimented with different significance levels α and report results for both $\alpha = 0.01$ and $\alpha = 0.05$. We did not adjust for multiple significance tests using the Bonferroni correction (which would result in significantly smaller significance levels) because, like Nouri-Madhavi [20] we have noticed that it fails to detect many cases of progressing glaucoma.

In contrast to this statistical approach, data mining algorithms exploit training data to build a classification model that can diagnose progressing glaucoma. In our experiments two very different data mining approaches turned out to perform well: linear support vector machines and 1R decision stumps. We describe these next. In Section 4.4 we mention other approaches that failed to improve upon them.

Linear support vector machines (LSVM) construct a hyperplane in the input space to classify new data. All data on one side of the hyperplane is assigned to one class; all that on the other side to the other class. Unlike hyperplanes constructed by other learning algorithms—for example, the perceptron—support vector machine hyperplanes have the property that they are maximally distant from the convex hulls that surround each class (if the classes are linearly separable.) Such a hyperplane is called a "maximum-margin" hyperplane, and is defined by those feature vectors from the training data that are closest to it. These are called "support vectors."

Support vector machines are very resilient to overfitting, even if the feature space is large—as it is in this application—because the maximum-margin hyperplane is very stable. Only support vectors influence its orientation and position, and there are usually only a few of them. If the data is not linearly separable, for example, because noise is present in the domain, learning algorithms for support vector machines apply a "soft" instead of a "hard" margin, allowing training instances to be misclassified by ending up on the "wrong" side of the hyperplane. This is achieved by introducing an upper bound on the weight with which each support vector can contribute to the position and orientation of the hyperplane.

The glaucoma data is very noisy: it is often hard even for experts to agree on the classification of a particular patient. Thus our experiments impose a low

upper bound on the support vectors' weights, namely 0.05. To learn the support vector machines we employed the sequential minimal optimization algorithm [21] with the modifications described in [14]. An implementation of this algorithm is included in the WEKA machine learning workbench [28].

The most straightforward way to apply support vector machines to this problem is to use each individual visual field measurement as an input feature. With 8 visual field measurements and 76 locations for each one, this produces 608 features. However, proceeding this way discards valuable information—namely the fact that there is a one-to-one correspondence between the 76 locations for different visual field measurements. A more promising approach is to use the per-location degradation in sensitivity as an input feature, yielding 76 features in total. This produced consistently more accurate results in our experiments. We experimented with three different ways of measuring degradation: (a) using the slope of a univariate regression function constructed from the 8 measurements corresponding to a particular location, (b) using the value of the t-statistic for that slope, and (c) simply taking the difference between the first and the last measurement. Surprisingly, methods (a) and (b) did not result in more accurate classifications than (c). Thus the experimental results presented in Section 4.3 are based on method (c).

A slightly more sophisticated approach is to sort the per-location differences into ascending order before applying the learning scheme. This is motivated by the fact that different patients exhibit progressing glaucoma in different locations of the visual field. A disadvantage is that it prevents the learning scheme from exploiting correlations between neighboring pixels. With this approach, the learning scheme's first input feature is the location exhibiting the smallest decrease in sensitivity, and its 76th feature is the location exhibiting the largest decrease. The median is represented by the 38th feature. In our experiments, the sorted differences produced more accurate predictions. All the results presented in Section 4.3 are based on this approach.

Compared to a support vector machine, a 1R decision stump is a very elementary classification scheme that simply determines the single most predictive feature and uses it to classify unknown feature vectors. Numeric features are discretized into intervals before applying the scheme. If the value of the chosen feature of a test instance falls into a particular interval, that instance is assigned the majority class of the training examples in this interval. The discretization intervals for a particular feature are constructed by sorting the training data according to the value for that feature and merging adjacent feature vectors of the same class into one interval. To prevent overfitting, one additional constraint is employed: the majority class in a particular interval must be represented by a certain minimum number of feature vectors in the training data. Holte [12] recommends a minimum of 6 as the threshold: this is what we used in our experiments.

As with support vector machines, we used the sorted differences in sensitivity between the first and last measurement for each location as a set of 76 input features for 1R. Applying this single-attribute scheme on the unsorted differences

is not appropriate: classifications would be based on the single test location in the series of visual field measurements that is the most predictive one across all patients in the training data. It is unlikely that the same location is affected by progressing glaucoma in every patient. As we show in the next section, 1R tends to choose a feature for prediction that is close to the median per-location difference.

4.3 The Results

Tables 1 and 2 summarize the experimental results that we obtained by applying the three different techniques to our dataset consisting of 130 visual field sequences. For PWLR we show results for two different significance levels α, namely 0.05 and 0.01. All performance statistics are estimated using stratified 10-fold cross-validation [28]. The same folds are used for each method. Standard deviations for the 10 folds are also shown. Note that PWLR does not involve any training. Thus cross-validation is not strictly necessary to estimate its performance. However, it enables us to compare PWLR to the other methods and lets us test potential performance differences for statistical significance.

Table 1 shows the estimated percentage of correct classifications. Note that 50.8% is the accuracy achieved by a classifier that always predicts the majority class (i.e. "stable glaucoma"). The five different columns correspond to different numbers of visual fields: for the column labeled "8 VFs" we used all 8 visual field measurements corresponding to a particular patient to derive a classification (and for training if applicable), for the column labeled "7 VFs" we used the first 7, etc. The classification problem gets harder as fewer visual field measurements become available, and corresponds to an earlier diagnosis.

Table 2 shows estimated performance according to the kappa statistic [1]. This statistic measures how much a classifier improves on a chance classifier that assigns class labels randomly in the same proportions. It is defined by:

$$\kappa = \frac{p_c - p_r}{1 - p_r}, \tag{1}$$

where p_c is the percentage of correct classifications made by the classifier under investigation, and p_r is the corresponding expected value for a chance classifier. Following the convention established by Landis and Koch [15], values of kappa above 0.8 represent excellent agreement, values between 0.4 and 0.8 indicate moderate-to-good agreement, and values less than 0.4 represent poor agreement.

The first observation is that the accuracy of the different methods depends on the number of visual field measurements that are available. With the exception of PWLR given a setting of $\alpha = 0.05$ (for reasons explained below), all methods achieve kappa values greater than 0.4 for 7 and 8 VFs, and exhibit a decline in performance as fewer VFs become available. For 6 and fewer VFs all estimates of kappa are below 0.4.

The second observation is that the performance of PWLR depends on an appropriately chosen significance level. According to a paired two-sided t-test,[1]

[1] Significant at the 0.05%-level.

Table 1. Percent correct, and standard deviation, for different numbers of visual field measurements (estimated using stratified 10-fold cross-validation)

Method	8 VFs	7 VFs	6 VFs	5 VFs	4 VFs
PWLR (α=0.01)	75.4±8.7	72.3±10.4	62.3±9.2	60.8±9.2	53.9±6.3
PWLR (α=0.05)	61.5±8.1	60.8±8.5	61.5±8.9	60.8±8.5	63.1±8.7
1R	80.0±9.0	78.5±12.5	64.6±13.2	61.5±11.5	56.2±10.3
LSVM	75.4±9.5	72.3±8.3	68.5±6.7	61.5±8.9	63.9±12.1

PWLR with $\alpha = 0.01$ performs significantly better than PWLR with $\alpha = 0.05$ for 8 and 7 VFs (according to both percent correct and kappa); for 6 and 5 VFs there is no significant difference between the two parameter settings; and for 4 VFs $\alpha = 0.05$ significantly outperforms $\alpha = 0.01$ according to the percent correct measure.

The reason for this result is that $\alpha = 0.05$ is too liberal a significance level if 7 or 8 VFs are available: it detects too many decreasing slopes in the series of visual field measurements, consequently classifying too many glaucoma patients as progressing. On the other hand, $\alpha = 0.01$ is too conservative if only a few VF measurements are present: with 4 VFs it almost never succeeds in diagnosing progressing glaucoma.

The third observation is that LSVM does not share this disadvantage of PWLR. It does not require parameter adjustment to cope with different numbers of visual field measurements. For 7 and 8 VFs it performs as well as PWLR with $\alpha = 0.01$ and significantly better than PWLR with $\alpha = 0.05$; for 6 VFs it performs better than both; and for 4 VFs it performs as well as PWLR with $\alpha = 0.05$ and significantly better than PWLR with $\alpha = 0.01$.

The fourth observation is that the 1R-based method is the best-performing one for 8 and 7 VFs. For 6 and less VFs it appears to be less accurate than LSVM. However, the only differences that are statistically significant occur for 7 and 8 VFs between 1R and PWLR using $\alpha = 0.05$. Interestingly, if 7 or 8 VFs are available, 1R consistently bases its predictions on the 33rd-largest per-location difference. Forcing the scheme to use the median per-location difference—the 38th-largest difference—slightly decreases performance. Unfortunately, due to lack of additional independently sampled data, we could not test whether this is a genuine feature of the domain or just an artifact of the particular dataset we used.

4.4 Other Approaches

During the course of our experiments we tried many other learning schemes and data pre-processing regimes:

- Replacing the above classifiers with decision trees, nearest-neighbor rules, naive Bayes, model trees, and locally weighted regression [28]. We also tried

Table 2. Value of kappa statistic, and standard deviation, for different numbers of visual field measurements (estimated using stratified 10-fold cross-validation)

Method	8 VFs	7 VFs	6 VFs	5 VFs	4 VFs
PWLR (α=0.01)	0.51±0.18	0.45±0.2	0.24±0.18	0.2±0.17	0.06±0.13
PWLR (α=0.05)	0.24±0.16	0.23±0.15	0.23±0.19	0.23±0.15	0.25±0.19
1R	0.59±0.19	0.56±0.25	0.28±0.27	0.23±0.22	0.12±0.2
LSVM	0.50±0.19	0.44±0.17	0.36±0.14	0.23±0.18	0.28±0.24

higher-order support vector machines, which are able to represent non-linear class boundaries [7].

- Performance enhancing wrappers applied to the above classifiers such as boosting [28], bagging [28], and additive regression [10].
- A custom stacking approach where several classifiers are built using the difference between the ith and the (i–1)th visual field and a meta-classifier to arbitrate between the predictions of these base classifiers.
- Smoothing the measurements for each location using a rectangular filter before applying the learning scheme (with varying filter sizes.)
- Using physiologically clustered data where the 76 test locations are reduced into several cluster based on physiological criteria [24].
- Using simulated data [23] in addition to data based on real patients.
- Focusing on the N locations with the largest decrease in sensitivity and using only those points, along with some of the surrounding ones, as the input to the learning scheme.
- Using the number of per-location-decreases as feature value instead of the difference between the first and the last measurement.
- Adding the difference between the mean of the per-location measurements for the first and the last visual field as an extra attribute.

None were able to improve on the results reported above.

5 Discussion

Determining whether a glaucoma patient has deteriorating vision on the basis of visual field information is a challenging, but important, task. Each visual field assessment provides a large number of attributes, and the task lends itself to data mining approaches. We hope that exposing the problem to the data mining community will stimulate new approaches that reduce the time required to detect progression.

The results of our experiments show that both 1R and support vector machines appear to improve on pointwise univariate regression, a method that is commonly used for glaucoma analysis. Among all methods tested, 1R produced the best results if a series of eight or seven visual field assessments is available for each patient. If six or fewer measurements are available, a linear support vector

machine appears to be the better choice, and we found it to be the most reliable method across different numbers of measurements. Compared to pointwise univariate regression analysis, whose performance depends strongly on an appropriately chosen significance level, this stability is an outstanding advantage.

We have focused on white-on-white visual field data measured in 76 locations. However, many other attributes are available that assist in progression analysis, and these may help learning schemes to yield more accurate results. For example, there are other visual field tests that detect glaucoma earlier and signal progression more rapidly, such as blue-on-yellow perimetry and Frequency Doubling Technology perimetry. The latter is particularly promising because it appears less variable than white-on-white perimetry [5].

Supplementary data is often available from measures of the structure of the optic nerve head. For example, confocal scanning laser tomography obtains 32 optical sections of the optic nerve with a laser and uses them to generate a 3-D topographic map [4]. This allows changes in nerve head shape, which may signal progression of glaucoma, to be quantified. However, these measures are very noisy, and it remains to be seen whether they can increase accuracy.

A serious impediment to research is the effort required to gather clean data sets. Not only is it arduous to collect clinical data, but significant expert time is needed to classify each patient. Computer simulation of visual field progression, which generates data that closely models reality, may offer an alternative source of training data [23]. An added advantage is that simulated visual fields are known to be progressing or stable by design, so the classification operation introduces no noise into the data.

References

1. P. Armitage and G. Berry. *Statistical Methods in Medical Research*. Blackwell Scientific Pulications, Oxford, third edition, 1994.
2. M.K. Birch, P.K. Wishart, and N.P. O'Donnell. Determining progressive visual field loss in serial humphrey visual fields. *Ophthomology*, 102(8):1227–1235, 1995.
3. L. Brigatti, K. Nouri-Mahdavi, M. Weitzman, and J. Caprioli. Automatic detection of glaucomatous visual field progression with neural networks. *Archives of Ophthalmol.*, 115:725–728, 1997.
4. R.O. Burk, A. Tuulonen, and P.J. Airaksinen. Laser scanning tomography of localised nerve fibre layer defects. *British J. of Ophthalmol*, 82(10):1112–1117, 1998.
5. B.C. Chauhan and C.A. Johnson. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Investigative Ophthomology and Vision Science*, 40(3):648–656, 1999.
6. B.C. Chauhan, S.M.Drance, and G.R. Douglas. The use of visual field indices in detecting changes in the visual field in glaucoma. *Investigative Ophthomology and Vision Science*, 31(3):512–520, 1990.
7. C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.
8. D.P. Crabb, F.W. Fitzke, A.I. McNaught, D.F. Edgar, and R.A. Hitchings. Improving the prediction of visual field progression in glaucoma using spatial processing. *Ophthomology*, 104(3):517–524, 1997.

9. M. Fingeret and T.L. Lewis. *Primary care of the glaucomas*. Mc Graw Hill, New York, second edition, 2001.
10. J.H. Friedman. Greedy function approximation: A gradient boosting machine. Technical report, Department of Statistics, Stanford University, CA, 1999.
11. D.B. Henson, S.E. Spenceley, and D.R. Bull. Artificial neural network analysis of noisy visual field data in glaucoma. *Art. Int. in Medicine*, 10:99–113, 1997.
12. R.C. Holte. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11:63–91, 1993.
13. J. Katz. Scoring systems for measuring progression of visual field loss in clinical trials of glaucoma treatment. *Ophthomology*, 106(2):391–395, 1999.
14. S. Keerthi, S. Shevade, C. Bhattacharyya, and K. Murthy. Improvements to platt's SMO algorithm for SVM classifier design. Technical report, Dept. of CSA, Banglore, India, 1999.
15. J.R. Landis and G.G. Koch. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2):363–374, 1977.
16. T. Leitman, J. Eng, J. Katz, and H.A. Quigley. Neural networks for visual field analysis: how do they compare with other algorithms. *J Glaucoma*, 8:77–80, 1999.
17. M.C. Leske, A. Heijl, L. Hyman, and B. Bengtsson. Early manifest glaucoma trial: design and baseline data. *Ophthomology*, 106(11):2144–2153, 1999.
18. S. Mandava, M. Zulauf, T. Zeyen, and J. Caprioli. An evaluation of clusters in the glaucomatous visual field. *American J. of Ophthalmol.*, 116(6):684–691, 1993.
19. R.K. Morgan, W.J. Feuer, and D.R. Anderson. Statpac 2 glaucoma change probability. *Archive of Ophthalmol.*, 109:1690–1692, 1991.
20. K. Nouri-Mahdavi, L. Brigatti, M. Weitzman, and J. Caprioli. Comparison of methods to detect visual field progression in glaucoma. *Ophthomology*, 104(8):1228–1236, 1997.
21. J. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods—Support Vector Learning*. MIT Press, Cambridge, MA, 1998.
22. S.D. Smith, J. Katz, and H.A. Quigly. Analysis of progressive change in automated visual fields in glaucoma. *Investigative Ophthomology and Vision Science*, 37(7):1419–1428, 1996.
23. P.G.D. Spry, A.B. Bates, C.A. Johnson, and B.C. Chauhan. Simulation of longitudinal threshold visual field data. *Investigative Ophthomology and Vision Science*, 41(8):2192–2200, 2000.
24. J. Weber and H. Ulrich. A perimetric nerve fiber bundle map. *International Ophthalmology*, 15:193–200, 1991.
25. C.J. Wild and G.A.F. Weber. *Introduction to probability and statistics*. Department of Statistics, University of Auckland, New Zealand, 1995.
26. J.M. Wild, M.K. Hussey, J.G.Flanagan, and G.E. Trope. Pointwise topographical and longitudinal modeling of the visual field in glaucoma. *Investigative Ophthomology and Vision Science*, 34(6):1907–1916, 1993.
27. J.M. Wild, N. Hutchings, M.K. Hussey, J.G.Flanagan, and G.E. Trope. Pointwise univariate linear regression of perimetric sensitivity against follow-up time in glaucoma. *Ophthomology*, 104(5):808–815, 1997.
28. Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, CA, 2000.