Interpretation of Automated Perimetry for Glaucoma by Neural Network

Michael H. Goldbaum,* Pamela A. Sample,* Halbert White,§ Brad Côté,* Paul Raphaelian,* Robert D. Fechtner,‡ and Robert N. Weinreb*  

Purpose. Neural networks were trained to interpret the visual fields from an automated perimeter. The authors evaluated the reliability of the trained neural networks to discriminate between normal eyes and eyes with glaucoma.

Methods. Inclusion criteria for glaucomatous and normal eyes were the intraocular pressure and the appearance of the optic nerve; previous visual fields were not used. The authors compared the backpropagation learning method used by automated neural networks to those used by two specialists in glaucoma to classify the central 24° automated perimetric visual fields from 60 normal and 60 glaucomatous eyes.

Results. The glaucoma experts and a trained two-layered network were each correct at approximately 67%. The average sensitivity of this test was 59% for the two glaucoma specialists and 65% for the two-layered network. The corresponding specificities were 74% and 71% for the specialists and the two-layered network, respectively. The experts and the network were in agreement about 74% of the time, which indicated no significant disagreement between the methods of testing. Feature analysis with a one-layered network determined the most important visual field positions.

Conclusions. The authors conclude that a neural network can be taught to be as proficient as a trained reader in interpreting visual fields for glaucoma. Invest Ophthalmol Vis Sci. 1994;35:3362–3373.

We are trying to automate the process that occurs when a trained physician, after seeing many examples, diagnoses glaucoma based on characteristic abnormalities in a visual field. Whenever a physician diagnoses a disease or selects a treatment based on manifestations and laboratory test results, he is making a classification decision. Classification is an aspect of pattern recognition that permeates the activities of the physician. A medical expert classifies a new pattern as belonging to the class of diagnosis A if it is most similar to the patterns that experience has shown are representative of diagnosis A. When the data (e.g., visual field) come from prespecified groups (e.g., normal eyes or eyes with glaucoma), the learning is supervised; therefore, the classification is a form of supervised pattern recognition.¹

The success of a classification depends on the input, the method of classifying, and the interpretation of the output. The full set of features used as input to discriminate among the classes comprises a pattern. The pattern of features on which a classification decision is based may include measurements, such as the luminance of the stimulus in a particular position in the visual field; descriptions, such as the eye tested (right or left); or the presence or absence of a manifestation, such as a glaucomatous-appearing optic nerve. Careful choice of the features will create the greatest separation of the classes and the most accurate discrimination among the classes.

Machine methods of classification include expert systems,² multivariate statistical classifiers (such as discriminant analysis), and neural networks.³,4 There is a difference in the way that the knowledge is devel-
Neural Network Interpretation of Fields for Glaucoma

opened in these methods. An expert system is taught by incorporating into its knowledge base the beliefs of an expert about the prevalence of each class and the conditional probabilities of the presence of each manifestation for a given class. The knowledge base is refined by adjusting the beliefs as teaching samples are presented to the system. The learning in statistical classifiers and neural networks occurs when the classifier estimates the frequency of the features in the teaching samples for each class.

During the process of learning from training samples, the membership in one of \( G \) given groups is known to the classifier. Once the classifier has completed its training, it can interpret sample patterns presented to it. The test samples also belong to the same \( G \) classes, but the class identity, which is not known beforehand, must be interpreted by the trained classifier. We investigated how well the interpretation of visual fields (samples) can be automated by a neural network (the classifier).

As a human classifier, the skilled clinician learns from experience or is taught patterns in the visual field to classify a field as normal or glaucomatous. Generally, attention is paid to higher level patterns such as an arcuate scotoma, nasal step, Bjerrum defect, baring of the blind spot, constriction of the visual field, and splitting of fixation. With the advent of automated static perimetry, the human observer has learned that defects in the visual field at a more primitive level of features also may be correlated with glaucoma. As an example of primitive features, the fields are then evaluated by looking for clusters of three adjacent test sites with >5 dB depression, or one site with \( \geq 10 \) dB depression.

Humans perform exceptionally well at pattern recognition compared to any currently available statistical technique or expert system. Recently, attempts have been made to emulate the type of pattern analysis that occurs in the human brain. These attempts are based on neural network models because they use processing elements that may exhibit many characteristics of the brain. They can learn from experience. They can generalize what they have learned from previous examples to identify new ones. They can find essential correlations in complex data that may also contain irrelevant information. They perform well when some of the information is incomplete or incorrect.

Perhaps the best studied learning method for neural networks is backpropagation. This method uses a large number of training examples (e.g., visual fields) to extract higher-order statistical information that allows new examples to be interpreted.

MATERIALS AND METHODS

Selection Criteria for Normal and Glaucomatous Eyes

We developed neural networks to automate the interpretation of visual fields, and we evaluated the reliability of the networks to discriminate between normal and glaucomatous eyes. The tenets of the Declaration of Helsinki were adhered to, informed consent was obtained, and the study was approved by the University of California, San Diego human subjects committee. There were 60 normal eyes and 60 glaucomatous eyes. All 60 glaucomatous eyes (100%) and 33 normal eyes (55%) were from the Glaucoma Center at UCSD. The subjects with glaucoma were selected sequentially between August 1988 and September 1990. The remainder of the data on normal eyes in determining age-matched normal visual field sensitivity surfaces was from the normative database used by Humphrey Instruments (San Leandro, CA) for the Visual Field Analyzer. All individuals were at least 40 years old with visual acuity was 20/40 or better. Lenses were sufficiently clear to provide a sharp view of the optic nerve head and fovea and to allow visual acuity of 20/40 or better. The right or left eye was randomly selected. All individuals were questioned and examined for diseases other than glaucoma that could cause defects in the visual field. The retnas of each individual from UCSD were examined with indirect ophthalmoscopy, and the optic nerves were examined stereoscopically with a 78 D Volk (Mentor, OH) lens and a Haag Streit (Bern, Switzerland) slit lamp by an expert in glaucoma (RNW). Each of the patients in the normal group developed for the Humphrey normative database had funduscopic and optic nerve head examinations (personal communication, M. Patella, 1990). Glaucoma was diagnosed based on the characteristic appearance of the optic nerve head (cup–disk ratio of at least 0.7, notching of the disk rim, or progressive cupping on sequential examinations) and intraocular pressure exceeding 22 mm Hg on two separate visits. The appearance of the visual field by other methods, such as tangent field or Goldmann perimetry, was not used to select subjects for the normal group and the group with glaucoma. Intraocular pressure of normal eyes was 21 mm Hg or less. We tested the central 24° of the visual field with 6° of spatial resolution using...
FIGURE 1. Two-layered neural network. During the learning phase, each teaching case is presented at the input layer and is processed until a value appears at the output layer. The value at each unit of the output layer is compared to the desired value in the matched teaching unit to determine the error. The total error is processed backward to adjust the weights to reduce the error. All the teaching cases are repeatedly presented to the input layer until the error is sufficiently small. A test case is only processed forward to provide a classification decision.

Backpropagation
In the backpropagation learning method, the network is organized into a layer of input units, a layer of output units, and one or more layers of hidden units between (Fig. 1). In a network with a single hidden layer, each unit in the input layer connects to each unit of the hidden layer, and each unit in the hidden layer connects to each unit of the output layer. Every connection has a weight value. The activity of each unit in the input layer is one of the low-level features in the pattern. The total input to each of the units in the hidden layer is the sum of the weighted output from all the units in the input layer below. The output of each of these hidden units is a specified function, commonly a logistic (S-shaped) function. Similarly, the total input to a unit in the output layer is a sum of the weighted outputs from the hidden layer below.

With backpropagation, the network learns connection weights by experience. The knowledge in a trained neural network is embodied in the weight values of all the connections. Initially, the values of the weights are randomly assigned numbers. All the patterns in the teaching set make up an epoch. The epoch of patterns is repeatedly presented in random order to the input layer. The weights are adjusted after each pattern is processed until the total error between the actual output and the desired output of each teaching pattern is small. The learning rate, which establishes how rapidly
Neural Network Interpretation of Fields for Glaucoma

we chose absolute threshold sensitivity as our visual field input to the network because we wanted to allow the network to make its own associations with unprocessed data. The value of the first unit was the absolute threshold sensitivity of the most nasal visual field position in the top row of the numeric grid. The subsequent 53 units were the rest of the positions of the visual field read in sequence, nasal to temporal, top to bottom. Because the input value for each unit is constrained between zero and one, the decibel values of the visual field were normalized to range between zero and one. The maximum absolute sensitivity measured in the eyes tested was 48 dB; therefore, the visual field units were normalized by dividing by 48. Age was also considered because it affects the visual field in persons with or without glaucoma. The value for the 55th unit was the age normalized between the minimum age (40 years) and the maximum age (84 years). Unit number 56 was given the value of 0 for the left eye and 1 for the right eye. The hidden layer contained four units. Experimentation disclosed that fewer hidden units lengthened the learning time, and more hidden units decreased the accuracy in classifying new patterns. The output layer had one unit. The expected output was 0 for normal and 1 for glaucoma. The classification of a new pattern in the test set was assigned to normal if the output for that pattern was less than 0.5 and to glaucoma if the output was more than 0.5. The threshold of 0.5 was chosen because it was halfway between the extremes of 0 and 1.

The network was considered taught when the measured error of the output was less than 1% of the error of an untaught network or when the error no longer decreased with further iterations. The number of iterations for learning was decreased by using a fast learning rate of 0.1 for the first 500 iterations, 0.03 for the second 500, and a slow 0.01 for the last 500. We found that 1500 iterations of an epoch was sufficient for all the teaching sets.

A multilayered network, with logistic output and hidden layers of the sort just described, uses curved surfaces in input space tailored to separate the data into the classes. A network with logistic output but without a hidden layer (single layer of connections) separates the data with hyperplanes. By comparing the results of the single layer network, which is analogous to linear discriminant analysis, to our results with the two-layered network, we can assess the improvement afforded by our nonlinear separation techniques. To provide this comparison, we trained and tested a single-layered network in the same fashion as the two-layered network.

In general, the larger the teaching set, the better trained the network becomes and the more accurately it can classify previously unseen data. We tested if the size of our teaching sets affected the performance of the training by removing half the data from each teaching set. The test sets remained the same. The identical starting weights were used for each of the full-size and half-size training sets.

Evaluation of Network Classification Performance

The apparent error rate of a taught network, obtained when the network is tested against its own teaching set, gives an overly optimistic estimate of accuracy because the correct classification may depend on specific properties of a pattern rather than general properties of the class. The "honest error rate" of a taught network applied to a test set of patterns it has not seen before demonstrates the ability of the network to generalize its knowledge for classifying and is the equivalent of a masked study. Cross validation efficiently uses all the patterns as both teaching patterns and test patterns at different times. The entire set of patterns is divided into k partitions. The first partition is set aside as the test set, and the other k-1 partitions are combined to make the teaching set. The process is repeated, with each of the other partitions having an opportunity to be the test set.

We divided our group of 120 subjects into 10 mutually exclusive partitions, each having 6 eyes with glaucoma and 6 eyes without glaucoma. For each teaching set, the errors were counted in the matching test set. The error rate was the total number of errors in all 10 test sets divided by 120.

Trained Human Reader

When comparing classifiers, in this case a neural network and two trained human readers, it is desirable to present each classifier with the same or nearly the same data. The network trained and classified on the absolute sensitivity. Because human readers do not usually rely on absolute sensitivity alone, two specialists in glaucoma (RNW and RDF) were presented with a printout of the visual field that contained the numeric grid (absolute sensitivity), defect depth (loss of sensitivity surface), gray-scale representation, patient information, test parameters, test reliability indices, and total quadrant threshold values for each subject. The individual's name and the date of the test were masked from the experts. Each expert classified the visual field
as glaucomatous or nonglaucomatous. The visual fields, incorrectly classified by either the two-layered network or the experts, were analyzed to determine the reason for the misclassification.

RESULTS

On a personal computer with a 25 MHz 80486 central processing unit and a 80487 mathematics coprocessor, the backpropagation learning time for 108 cases in a teaching set was 12 minutes. The time for testing one visual field was estimated to be 1 msec. The classification of an unknown visual field appeared to be almost instantaneous.

The normal subjects ranged in age from 50 to 80 years (63 ± 8.6 years), and the patients with glaucoma ranged in age from 41 to 84 years (69 ± 9.2 years). The inclusion criteria for the group with glaucoma and the normal group were based solely on the intraocular pressure and the appearance of the optic nerve head. This resulted in visual fields in the normal group that ranged from normal in appearance to abnormal with multiple regions of defect. In the group with glaucoma, visual fields ranged from normal in appearance to profound visual field loss with a central island of vision. We compared the classification accomplished by human experts with that performed by backpropagation in a two-layered network. To ascertain the effect of the size of the teaching set and the presence of a hidden layer in the network, we compared the two-layered network trained on a full set of data to a two-layered network trained on half the data and to a single-layered network trained on a full set of data, respectively.

Table 1 summarizes the results of the visual field classifiers. With the classification threshold set at 0.5, the network was slightly more specific than sensitive, falling at a percentile rank of 56.7 from the normal end of the sample (Figs. 2 and 3). Because the human experts also tended to be more specific, this balance lent credence to the comparison of the two classifiers. The receiver operating characteristic curve (Fig. 3) shows the continuous representation of the tradeoff between the true-positive rate and the false-positive rate as the cutoff for diagnosing glaucoma is moved from 0 to 1.

The likelihood ratio is the factor by which the odds of a diagnosis are changed with knowledge of the test results.29 From these values, the likelihood ratio of the visual field as an independent test for glaucoma (LRc = the probability of a visual field being classified as glaucoma given that the person has glaucoma divided by the probability of the visual field being classified as glaucoma given that the person does not have glaucoma) was 2.3 for the experts, 2.3 for the two-layered network, 1.2 for the single-layered network, and 1.6 for the half-trained network. (Table 1). The likelihood ratio of the visual field by itself for normal (LR N = P[VF|-.G]/P[.VF|G] = specificity/[1 - sensitivity]) was 1.8 for the experts combined, 2.0 for the two-layered network, 1.2 for the single-layered network, and 1.5 for the half-trained network. With the groups with glaucoma and the groups without glaucoma combined, the percentage of agreement between the two experts was 75%; between expert 1 and the two-layered network, it was 72%; and between expert 2 and the two-layered network, it was 74%. The McNemar statistic of disagreement between two methods of testing30 found no significant disagreement between expert 1 and expert 2 or between either expert and the network.

DISCUSSION

Suitability of Data and Classifiers

Because the pattern by automated perimetry is not independent of the patterns by other methods of visual fields for the same eye, we think that using a human expert's interpretation of a visual field for the selection of class members for the purpose of evaluating automated perimetry will affect the distribution of patterns studied. Furthermore, using visual fields evaluated by human experts to select candidates prevents the machine classifier from performing better
Neural Network Interpretation of Fields for Glaucoma

10-Partition Histogram of 2-layer Neural Network Output

FIGURE 2. Histogram of the distribution of output values from the two-layered neural network. The output suggests how much the neural network considered that the tested visual field resembles a glaucomatous field defect. Whether correct or incorrect, most of the output values were close to 0 or 1 as a result of the learning algorithm. The vertical line in the center indicates the 0.5 threshold used to develop the sensitivity and specificity values.

than human experts. By relying on measures of glaucoma other than the visual field, we can allow each classifier to use whatever methods or visual field information it chooses to classify a pattern and then compare the performance of the classifiers.

We attempted to reduce selection bias by using the spectrum of visual fields of patients with glaucoma in a consultative glaucoma practice and a blend of age-matched normal subjects from the population. The scatterplot of the corrected pattern standard deviation and the mean defect show the spectrum of patients with glaucoma and normal subjects (Fig. 4). We did not restrict the visual fields to include only normal-appearing fields for nonglaucomatous eyes and only fields with defects for the group with glaucoma. Consequently, in our set of patients with glaucoma, approximately one-third of the eyes had visual fields that appeared normal to the trained human observers and to the neural network. Similarly, about one-third of the subjects without glaucoma had abnormal fields that could be interpreted as glaucomatous. Other reports described similar or better sensitivity and specificity of the visual field in glaucoma.\textsuperscript{10,31-34} It may be that there was greater dissimilarity between the glaucomatous and nonglaucomatous visual fields used for these reports. For example, any study that uses visual fields by any means as a selection criterion for the group with glaucoma and the normal group will show better sensitivity and specificity of its classifiers because all the patients with glaucoma will have visual field patterns that fit some definition of a glaucomatous pattern, and all the normal patients will have normal visual fields or abnormalities defined as nonglaucomatous in appearance. Some reports may have overestimated the accuracy of classification with visual fields by using the apparent error rate rather than the honest error rate to determine sensitivity and specificity.\textsuperscript{34}

Performance of Network and Human Classifiers of Visual Fields

Initially, the human experts were given visual fields with absolute sensitivities to evaluate. However, the human experts were not used to evaluating automated perimetry in this manner, which hampered performance. The human experts classified patterns of defect depth in this study because they did not want to compare neural networks to humans handicapped with reading methods not commonly used. On the other hand, the neural network learned from patterns of absolute sensitivities plus age to reduce the effects of filtering and to allow the network to make its own correlations.
FIGURE 3. Receiver operating characteristic curve of a two-layered neural network output. The vertical and horizontal lines inside the graph indicate the false-positive rate (1 - specificity) and the true-positive rate (sensitivity) at the threshold output value of 0.5.

The two experts were similar in overall correctness but differed in the strictness with which the criteria for glaucoma were applied. Of the total population of subjects, approximately two-thirds were correctly classified by both the expert and the two-layered network. The single-layered network performed no better than chance alone, indicating that the data were not linearly separable. Figure 2 represents the distribution of the neural network values for the normal and glaucomatous visual fields. It illustrates an overlap whereby the patterns of about one-third of the fields from the group with glaucoma resembled the patterns of about one-third of fields from the normal group (Fig. 5). In our data set, the age-matched visual field as an independent test was an imperfect predictor of glaucoma. In other words, the pattern of features (age-matched visual field) from automated perimetry left considerable overlap in the clusters of glaucomatous and nonglaucomatous eyes; there was incomplete separation.

Most of the misclassifications by either the experts or the two-layered network were with the subset of visual fields in which the patterns from the glaucomatous eyes resembled the patterns from the nonglaucomatous eyes. The percentage of agreement between the two experts and between each expert and the two-layered network ranged between 72% and 75%. This similarity suggests that the data, rather than the classification method, limited the accuracy of the classification.

Methods That May Improve Classifier Performance

Theoretically, it is possible to improve classification performance by presenting the classifiers with a better input pattern or by using a better method of classification. Suggested methods of selecting better input patterns include teaching with a different set of visual fields from glaucomatous and nonglaucomatous eyes, filtering to enhance some aspect of the data, or increasing the experience of the network. The teaching patterns we used are representative of the types of

FIGURE 4. Scatterplot of pattern standard deviation (x-axis) against the mean defect (y-axis) to illustrate the distribution of normal and glaucomatous fields. Note the overlap that occurred because subjects were chosen by intraocular pressure and optic nerve appearance and not by visual field. The mean value (x, y) of these indices were (3.14, -1.18) for the normal eyes and (4.93, -5.18) for the eyes with glaucoma.
Neural Network Interpretation of Fields for Glaucoma

cases an ophthalmologist encounters. The patients with glaucoma came from a consultative glaucoma practice, and those without glaucoma were from our clinic patients without glaucoma and subjects without glaucoma that Humphrey collected to generate their age-matched normal surfaces.

We chose to use the unfiltered absolute sensitivity of the numeric format plus age as input because we wanted the network to find its own correlations. Computer filtering of the data in automated perimetry produces indices such as cluster analysis, glaucoma hemifield test, cross-meridional test, and probability maps. For a review, see Asman. These indices, which are added to add the separability of the classes, can be additional feature inputs for either the human or the machine classifiers. Non-Gaussian, population-based probability density curves individualized for each test position yield a defect map that may allow human or machine classifiers to discriminate the two classes better. It is likely that neural networks and experts in glaucoma will perform better if they are given the probability maps and indices because both types of classifiers performed in a similar manner with the data given in this study. Future studies can test this hypothesis and perform feature analysis to indicate the importance of each new index.

The networks trained on the full-size teaching sets performed better than the networks trained on the half-size teaching sets. Some patterns may have been few, and thus there may have been a quantization effect caused by the partitioning of data. This effect is unpredictable. Improved performance afforded by larger teaching sets suggests that the network's performance can be improved with experience, especially when it is taught new patterns that it had misclassified.

Though the two-layered network with backpropagation has at least equaled the human expert in learning how to classify visual fields, conceivably there is a neural network learning algorithm that is more suited to the classification of visual fields.

Decision-Making by Machine and Human Classifiers

Automated interpretation of visual fields may be realized with various machine classifiers, such as expert systems, statistical classifiers, or neural networks. There are advantages to an expert system, the most important of which is its ability to create an audit trail to explain how it has arrived at a diagnosis or why it is requesting the value of a particular feature (the next best rule). An expert system has been reported that uses high-level features of the visual field, such as nasal steps or Bjerrum defects, to classify visual fields. The unprocessed data provided by the automated perimeter are low-level features that do not lend themselves to the development of rules. In contrast, the various indices developed from automated perimetry constitute high-level features that could add to the knowledge of an expert system to assist its interpretation of visual fields.

Neural networks have been shown to perform well compared to statistical classifiers that use discriminant or regression analysis for similar classification tasks. Neural networks are capable of finding complex relationships in the data to help make the correct classification. No matter how successful a classifier may be, clinicians do not like to put faith in a black box that gives an answer without an explanation. The knowledge base is the weights of the connections. Unfortunately, it is difficult to understand what a multilayered network is doing with these weights to make the classification. The analysis of the reasoning process of taught networks is a current area of investigation in neural network research.

Analysis of Visual Field Features by Neural Network

To derive some understanding of how the visual field data are used by the network, we analyzed the weights of the single-layered network. This information should be useful even though the data were not linearly separable because the weights represent the optimal separation of the classes with a hyperplane. Because each input unit is connected directly to the output layer, the weights can give some indication of how each position in the visual field is used in the decision process. Figure 6 shows the mean weight in the trained network for each position in the visual field for all 10 training sets used for cross-validation. The weights are displayed after the position of the input values in the visual field. If the sum of the product of the inputs (absolute threshold values) and their connection weights are negative, the network will classify the field as normal. A positively weighted sum will yield a classification of glaucoma.

Depressions in the visual field reduce the threshold value. Therefore, depressions in positions of high minus weights (the lowest quartile, \( < -5 \)) will move the weighted sum in a positive direction toward the diagnosis of glaucoma. Depressions in the visual field in which the weights are high plus (the highest quartile, \( > +4 \)) push the weighted sum away from the diagnosis of glaucoma. In nonglaucomatous eyes, visual field depressions either do not occur, happen because of artifact, or follow a nonglaucomatous pattern. We interpret depressions in locations of high plus weights as decreasing the significance of depressions in the sites of high minus weights. The field positions with weights close to zero have little effect in the classification process.

Figure 7 shows clusters of high minus weights in the inferior and superior arcuate zones, immediately
RIGHT

○ = WITHIN 4 DB OF EXPECTED
NO. = DEFECT DEPTH IN DB
26 DB xx = CENTRAL REF LEVEL

FIXATION LOSSES 2/29
FALSE POS ERRORS 0/10
FALSE NEG ERRORS 1/12
QUESTIONS ASKED: 582
FLUCTUATION: 2.64 DB
FOVEA: 28 DB
TEST TIME: 00:18:16
HFA S/N

LEFT

○ = WITHIN 4 DB OF EXPECTED
NO. = DEFECT DEPTH IN DB
34 DB = CENTRAL REF LEVEL

FIXATION LOSSES 1/21
FALSE POS ERRORS 1/12
FALSE NEG ERRORS 0/12
QUESTIONS ASKED: 352
FLUCTUATION: 2.90 DB
FOVEA: 35 DB
TEST TIME: 00:11:15 50°
HFA S/N

A

B
supranasal to fixation and inferonasal to the horizontal, 9° to 15° nasal to fixation (possibly corresponding to a nasal step). The network relies most on the depression of the visual field in these locations to diagnose glaucoma. Trained readers of visual fields also rely on depressions in the arcuate zones and the region of the nasal step. The locations the network found do not correspond to four positions derived by a linear correlation program for predicting glaucoma from a screening test. The asymmetry in the weights in complementary positions on either side of the horizontal meridian give credence to tests such as the glaucoma hemifield test. Apparently, the network considers depression just inferior to fixation by itself and at the superior, inferior, and temporal edge of the field not indicative of glaucoma. An interesting observation is that the network does not regard enlargement of the blind spot to be meaningful, and baring of the blind spot supratemporally works against the diagnosis of glaucoma. In itself, increasing age favors a diagnosis of glaucoma. An important consideration is that the two-layered network may be making complex associations in visual field depressions and age that are not obvious from the analysis of the single-layered network. This may explain the better performance of the

![Figure 5](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933407/) Output of automated perimeter. In each figure, upper field is the gray-scale representation of the visual field. Lower left is the defect depth, with defects less than 5 dB represented by 0. Lower right is numeric grid of absolute sensitivities in decibels of each of the 54 tested positions in the visual field. (A) Normal subject with major nonglaucoma defect has been incorrectly classified as having glaucoma by both the expert and the neural network. (B) Patient with glaucoma with minor defect has been incorrectly classified as normal by both the expert and the neural network. These examples show the type of fields in which there was difficulty in the classification performed by either the expert or the neural network.  

![Figure 6](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933407/) Average weights at each location of visual field from the 10 training sets in 10-partition cross-validation. Positive values are displayed in italics. High negative weights (<25th percentile [< -5]) are indicated by a dark gray background. High positive weights (>75th percentile [> +4]) are indicated by a light gray background.

![Figure 7](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933407/) Surface graphs show location and contour lines at sites of high negative weights and high positive weights. Location of a blind spot and fixation are represented by a black oval and a small black dot, respectively. The network considers visual field depressions at high minus locations to be indicative of glaucoma. The network considers depressions at high plus locations to be representative of artifact or depressions found in some eyes of otherwise normal individuals.
two-layered network. The lack of independence of individual positions forms the basis of cluster analysis.\textsuperscript{39}

The human observer uses certain rules of thumb to perform the classification. An expert at classifying visual fields may use intuition, which cannot be explained. The backpropagation learning method of neural networks is at least as accurate as a human expert in glaucoma in classifying visual fields from an automated perimeter when both are given the same information. Adding indices such as the hemifield test or cluster analysis should improve the performance of both classifiers.\textsuperscript{40} Whereas automated perimetry alone is not sufficient to predict glaucoma, its value (glaucoma or not glaucoma) may itself be used as a single feature that can be combined with other features (e.g., intraocular pressure, optic nerve head appearance, retinal nerve fiber layer appearance, age, race, family history, and gonioscopy) to form an input pattern for a classifier, such as an ophthalmologist, an expert system, or another neural network.

**Key Words**

neural networks, artificial intelligence, pattern recognition, glaucoma, visual fields

**References**

31. Enger C, Sommer A. Recognizing glaucomatous field