UNSUPERVISED LEARNING WITH INDEPENDENT COMPONENT ANALYSIS CAN IDENTIFY PATTERNS OF GLAUCOMATOUS VISUAL FIELD DEFECTS

BY Michael Henry Goldbaum MD MS Medical Informatics

ABSTRACT

Purpose: We previously reported the use of clustering by unsupervised learning with machine learning classifiers to segment clusters of patterns in standard automated perimetry (SAP) for glaucoma. In this study, the process of unsupervised learning by independent component analysis decomposed SAP field patterns into axes, and the information represented by these axes was evaluated.

Methods: SAP fields were obtained with the Humphrey Visual Field Analyzer on 189 normal eyes and 156 eyes with glaucomatous optic neuropathy (GON) determined by masked review with stereoscopic optic disc photos. The variational Bayesian independent component analysis mixture model (vB-ICA-mm) partitioned the SAP fields into the most informative number of clusters. Simultaneously, it learned an optimal number of maximally independent axes for each cluster.

Results: The most informative number of clusters was two. vB-ICA-mm placed 68.6% of the SAP fields from eyes with GON in a cluster labeled G and 98.4% of the fields from eyes with normal optic discs in a cluster labeled N. Cluster G optimally contained six axes. Post hoc analysis of patterns generated at –1 SD and +2 SD from the cluster G mean on the six axes revealed defects similar to those identified by experts as indicative of glaucoma. SAP fields associated with an axis showed increasing severity as they were located farther in the positive direction from the cluster G mean.

Conclusions: vB-ICA-mm represented the SAP fields with patterns that were meaningful for glaucoma experts. This process also captured severity in the patterns uncovered. These findings should validate vB-ICA-mm as a data mining technique for new and unfamiliar complex tests.

Trans Am Ophthalmol Soc 2005:103:270-280

COMPUTATIONAL OPHTHALMOLOGY

The quote "We are drowning in information and starving for knowledge" by Yale librarian Rutherford D. Rogers concerning the enormous amount of publications also describes the problem faced by physicians interpreting tests for the diagnosis and management of diseases. Enormous amounts of data are being generated in medical research and medical care, and we wish to make sense of it. There is potential for the medical community to increase its knowledge by discovering important patterns that may be hidden in these data. This pattern discovery can be accomplished by learning from the data. This thesis demonstrates how we can learn from data in a discipline of ophthalmology and how that knowledge can help us in the management of disease.

Advances in computational power today put on our desktops the equivalent of a supercomputer from just a few years ago. The central processing unit of desktop computers has increased in speed 31,250 times from 128 kHz in 1980 to 4 gigahertz today. The internal memory in the computer that allows extremely rapid insertion and retrieval of data has enlarged 100,000 times from 4 kilobytes (kB) in 1980 to 4 gigabytes today. Disk storage space for large complex programs and for immense amounts of data has expanded one million times from the 400 kB floppy disk in 1980 to half a terabyte today. Over the next 10 years, it is expected that these capacities will continue to increase at the same rate. This tremendous augmentation in computational capacity enables us to benefit from the routine application of computationally intense analyses that were previously impractical.

VISUAL FIELD PATTERNS

This study focuses on the problem faced by physicians interpreting tests for the diagnosis and management of diseases. The discipline of glaucoma is particularly rich with data-intense tests that are currently being applied or developed to determine the status of glaucoma in an eye. Diagnosing and staging glaucoma both require the use of several tests, including structural and functional assessment of the optic nerve. One of these, visual field testing, has been employed by experts for several decades as the prime indicator for both diagnosis of glaucoma and detection of progression. This wide expertise can be used to evaluate the ability of machine learning classifiers to find information that can be valuable in clinical practice. The hypothesis proposed is that machine learning classifiers can learn useful information about visual field patterns. This information can be validated against the prevailing experience of experts. Such validation would give credence to the application of the same process to new and unfamiliar tests.

Visual fields have been used to diagnose glaucoma since 1856.[1] In 1889, Bjerrum began to uncover patterns of visual field defects, such as a comet-shaped arcuate scotoma or a nasal step scotoma, with quantitative perimetry.[2] Consequently, for more than 110 years, generations of experts in glaucoma have accumulated knowledge to recognize patterns of visual field defects that indicate glaucoma.

From the Department of Ophthalmology, University of California, San Diego, La Jolla, California, and the Veterans Administration San Diego Health Services, San Diego, California. This work was supported by grant EY13235 from the National Institutes of Health.

With the advent of standard automated perimetry (SAP) and the development of statistical field analysis packages like Statpac (Carl Zeiss Meditec Inc, Dublin, California), the depth of defect within the patterns of loss and the relationship of adjacent test locations to each other could be quantified. We applied machine-learning data-mining techniques to uncover visual field patterns associated with glaucoma and to compare these patterns with the subjective qualitative and semiquantitative patterns evolved from experience by human experts.

LEARNING CLASSIFIERS

There are two ways that classifiers can learn from data concerning medical conditions. One way is for classifiers to learn to diagnose disease, predict outcomes, and look for change.[3-8] Another way is by finding useful patterns in large groups of patients. In the current study, we applied variational Bayesian independent component analysis mixture model (vB-ICA-mm) to subdivide SAP fields performed in a cohort of normal subjects and patients with glaucomatous optic neuropathy (GON), to find patterns of interest in the SAP fields, and to propose how this knowledge can improve medical care.

A classifier may learn from training examples the knowledge it needs to make decisions. Specifically, a classifier may learn to diagnose glaucoma by distinguishing abnormal SAP fields from normal fields after being trained with a set of SAP fields labeled with the correct diagnosis. This process is called *learning with a teacher* or *supervised learning*. In addition, supervised learning is commonly used to train classifiers to predict outcome or detect change

When the diagnosis is not supplied with each of the training samples, the objective of the learning algorithm no longer is classification or diagnosis. The goal can be to organize the input data into meaningful structures or groups of patterns. This process is called *learning without a teacher* or *unsupervised learning*. For our purposes, the goal can be to organize data, such as the set of visual fields, into differing clusters of patterns whose members are similar. Hence, another name given to the process is *clustering,* and the investigation of the clusters segmented from the data is *cluster analysis.*

Machine learning classifiers impose fewer constraints about the data than statistical classifiers, permitting better adaptation to the data and thus better organization. Unsupervised learning with these newer classifiers has the potential of grouping the patterns in the data in a manner that is more useful than that done by statistical methods.

STRUCTURE OF DATA REPRESENTATION

The decomposition of the data by different types of unsupervised learning yields different structures. Instead of clusters, the current study evaluates a different structure that relies on axes. *Component analysis* projects the data within each cluster located in multidimensional space onto axes that meaningfully represent the data. Although the axes do not produce clusters, representation of the data with axes still yields useful information about the patterns in the data. A further refinement, *principal component analysis*, projects *d*-dimensional data on a lower dimension subspace of *s* orthogonal axes. Though the expectation is that the orthogonal axes are independent, in reality the axes may not be independent. *Independent component analysis* (ICA) was chosen because it incorporates a measure of independence to produce axes that are maximally independent.[9,10] There may be data distributions where components are nonlinearly related or clustered such that they are difficult to describe by one ICA model. The *ICA mixture model*[9] is a nonlinear ICA technique that extends the linear ICA method by learning multiple ICA models and weighting them in a probabilistic manner. The ICA mixture model settles on the optimal number of axis sets simultaneously with the generation of the axes (Figure 1, A and B).

Representing all the data (eg, normal and glaucoma eyes together) with a single set of axes may produce suboptimal representation of the data; multiple sets of axes are more likely to match the local conditions (which, in this study, turned out to be a set mostly of normal eyes and a set mostly of eyes with glaucoma). Instead of relying on a chosen number of clusters with fixed dimensionality, the mixture of ICA model learns the dimensionality and number of classes. The exact computation of the marginal likelihood is computationally intractable; thus, variational Bayesian approximation techniques were employed.[11] The variational Bayesian framework helps to capture the number of axes in the local axis set and reduces the computational complexity (by bounding intractable integrals). The amalgamation of all these processes is the *variational Bayesian ICA mixture model* (vB-ICA-mm) [6] that was applied in this study to fields from healthy eyes and eyes with glaucoma in order to identify patterns of field loss associated with this diagnosis.

METHODS

PARTICIPANT SELECTION AND TESTING

Participants

This study employed the visual field set from the first 12 years of the ongoing National Eye Institute–sponsored longitudinal Diagnostic Innovations in Glaucoma Study of visual function in glaucoma. Normal participants for this study were recruited from the community, staff, and spouses or friends of patients. Primary open-angle glaucoma patients were recruited from the Hamilton Glaucoma Center. Informed consent was obtained from all participants. The study was approved by the Institutional Review Board of the University of California, San Diego, and it followed the tenets of the Declaration of Helsinki.

Exclusion criteria for both groups included unreliable visual fields (defined as a value more than 33% for either fixation loss, false-negative errors, or false-positive errors),[12] angle abnormalities on gonioscopy, diseases other than glaucoma that could affect the visual fields, and medications known to affect visual field sensitivity. Subjects with a best-corrected visual acuity worse than 20/40, spherical

equivalent outside ±5.0 diopters, and cylinder correction greater than 3.0 diopters were excluded. Poor-quality stereoscopic photographs of the optic nerve head also served as an exclusion for the glaucoma category. A family history of glaucoma was not an exclusion criterion.

Inclusion criteria for the normal category required that subjects have normal dilated eye examinations, open angles, and no evidence of visible GON. Normal optic discs had a cup-to-disc ratio asymmetry of 0.2 or less; intact rims without hemorrhages, notches, excavation; and an absence of nerve fiber layer defects in the adjacent retina. Normal subjects had intraocular pressures of 22 mm Hg or less and no history of elevated intraocular pressure. If both of the eyes met the inclusion criteria, only one of the eyes was selected at random to ensure independence between eyes.

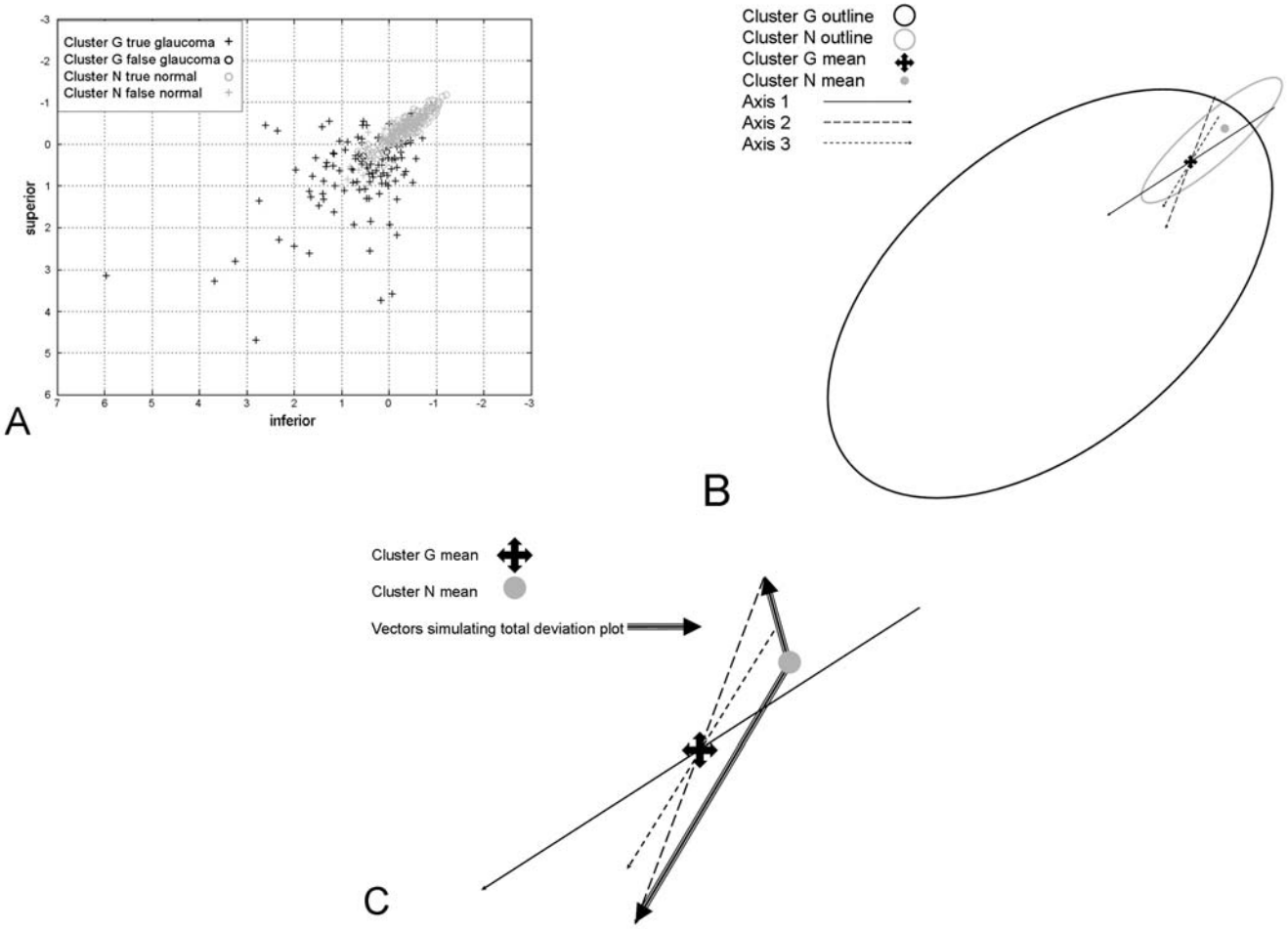

FIGURE 1

A, Two-dimensional scatter plot of participants with (+) and without (o) glaucomatous optic neuropathy (GON). Units are in standard deviations away from the mean of cluster G. B, Two-dimensional projection of 53-dimension space. vB-ICA-mm converged to two clusters, named from post hoc analysis the glaucoma cluster (cluster G) and normal cluster (cluster N, see text). The glaucoma cluster decomposed into six axes through the glaucoma mean. Axes 1 through 3 are shown; axes 4 through 6 are not depicted to avoid clutter in the figure. The positive direction of each axis moves farther from the normal mean. C, Relationship of points on axes in cluster G to the mean of the normal cluster. Large arrows represent vectors from cluster N mean to a point on the negative side and a point on the positive side of axis 2. Gray scales at the tips of these vectors simulate the gray scale of total deviation plots. Increasing severity away from the normal mean most closely matches increasing deviation from the glaucoma mean on the positive side of the glaucoma mean.

Glaucomatous Optic Neuropathy As an Indicator of Glaucoma

Since the goal was to analyze SAP visual fields, visual fields were not used to determine if an eye was glaucomatous or not. The classification of an eye as glaucomatous or normal and the labeling of its visual field for the post hoc analysis of the results were based

on the appearance of the optic disc. The designation of the optic disc as glaucomatous or normal was accomplished with masked evaluations by two independent graders of a stereoscopic disc photograph taken within 6 months of the visual field. Inconsistencies between the graders' evaluations were resolved by consensus or through adjudication by a third masked evaluator. Color simultaneous stereoscopic photographs were obtained using a Topcon camera (TRC-SS, Topcon Instrument Corp of America, Paramus, New Hampshire) after maximal pupillary dilation. These photos were taken within 6 months of the field in the data set. Stereoscopic disc photographs were evaluated for all eyes with the exception of a subset of normal subjects (95 eyes) for whom photography had not been performed early in the collection of the normal database. All normal subjects had no evidence of optic disc damage during dilated slit-lamp indirect ophthalmoscopy with a hand-held 78-diopter lens. The final selection of eyes totaled 345, including 189 normal eyes (mean age in years ± SD, 50.0 ± 6.7) and 156 eyes with GON (62.3 ± 12.4).

Visual Field Testing

All subjects had automated full-threshold standard visual field testing with the Humphrey Visual Field Analyzer (HFA, Carl Zeiss Meditec Inc, Dublin, California) with program 24-2 or 30-2. Though most clinicians use the 24° 24-2 program to test for glaucoma, some of the subjects were in other studies that required field testing to be done with the 30° 30-2 program. The visual field locations in the 30-2 fields that are not in 24-2 fields are at the edge of the field and do not contribute much to the accurate diagnosis of glaucoma. These locations were deleted from the 30-2 field data and displays to make the SAP data consistent.

REPRESENTATION OF DATA

Input for Classifier

The absolute sensitivity (in decibels) of the 52 visual field locations (L) plus age formed a vector in 53-dimensional input space for each of the 345 SAP fields of normal and glaucomatous eyes. The 52 threshold values were extracted from the HFA using the Peridata 6.2 program (Peridata Software GmbH, Huerth, Germany). Each feature vector was $\mathbf{x} = (L1,\ldots, L54, \text{age})$, excluding locations L18 and L31, because they fell in the normal blind spot. Age was included because both normal and glaucomatous SAP fields are affected by age, and age was used in some previous studies that incorporated supervised learning.[3-5,7,8]

Partitioning Data Into Clusters and Adjusting Axes

The unsupervised learning was performed with the variational Bayesian ICA mixture model. This method is the core of the new direction in data exploration we propose. A detailed, mathematically rigorous description of this method is available in the Appendix. A general description follows. Starting with 345 subjects evenly distributed in c clusters and with random initialization of axes, the axes and the probability of each cluster were learned. The number of clusters, c, was increased from $c = 1$ until $c = 5$, seeking the value beyond which no further gain in information would be obtained. Subjects belonging to the same cluster defined the mean and templates for that cluster. For each SAP field, its cluster assignment was then recomputed according to its likelihood value given by ICA. ICA alone would have sought one set of axes. The mixture model of ICA created clusters and sought an optimal set of axes for each cluster. Bayesian learning was applied to overcome overfitting by maximum likelihood estimation. The cluster assignments and axis adjustments were iterated until no further change occurred in the cluster assignment.[6,9] The possibility of ending up in a high local minimum of the error surface was minimized by selecting the model with the best marginal likelihood value from 100 different random initializations, thus ending up as close to the global minimum as possible.

POST HOC ANALYSIS

Validation of Structure

The process was validated by observing whether the structure obtained by vB-ICA-mm was appropriate for the data. As described in the "Results" section, the data were best represented by two clusters. Because the data contained visual fields from normal and glaucomatous eyes, we first evaluated the two clusters, created without knowledge of the diagnosis, for their proportion of normal and glaucomatous eyes. The second validation came from evaluation of the SAP patterns in each cluster represented by the axes uncovered by ICA. The representation by these axes of patterns similar to those discovered by generations of glaucoma experts through over a century of accumulated experience would validate the structure uncovered by unsupervised learning with machine learning classifiers.

Placing Fields About an Axis

Experience with cluster analysis methods that do only clustering demonstrates that analysis of members inside the clusters gives us some understanding of the data organization achieved by unsupervised learning.[13] Instead of creating subclusters, as was reported with variational Bayesian mixture of factor analysis mixture model,[13] vB-ICA-mm created axes within the clusters. The axes were analyzed by generating patterns in SAP field input space at specific points along the axes created by the vB-ICA-mm. We wished not to ignore potential information in the SAP fields. Information was also sought by examining the SAP fields in addition to examining patterns generated at points on the axes. The SAP fields in a cluster generated by the vB-ICA-mm were thus organized into clouds around the axes created by the vB-ICA-mm, and these clouds were analyzed.

To analyze the individual SAP fields in cluster c, each field was assigned to an axis. In 53-dimensional space, the angle was calculated at the cluster c centroid between the vector for any individual SAP field and the vector of each of the axes. The individual field was assigned to the axis with which the SAP field vector had the smallest angle. This created a cloud of points in space about each axis, with the axis running up the center of the cloud (Figure 2). One can imagine the shape of each cloud to approximate a "hypercone" (multidimensional manifestation of a cone in three-dimensional space) expanding away from the glaucoma mean. Each

axis had a positive cloud and a negative cloud. Each field was assigned to one axis in one direction, and there was no overlap in the clouds. This allowed us to look at the fields falling within each cloud and hopefully to determine what was similar about those fields in one cloud and how they differed from fields falling within the other clouds.

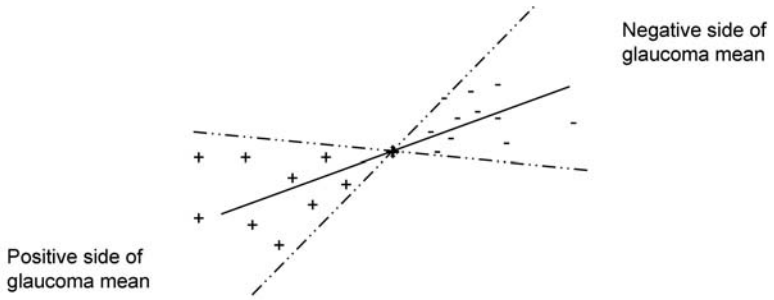

FIGURE 2

Distribution of SAP fields assigned to an axis. Each axis has a cloud around the positive axis and a cloud around the negative axis. The broken lines indicate the boundary surface for the cloud. Individual absolute sensitivity plots are distributed around the axis they are closest to, inside the cloud on either side of the glaucoma cluster mean. The vector that connects the normal cluster mean with an individual field simulates the total deviation plot.

Visual Field Interpretation

To determine if the structure disclosed by the unsupervised learning was meaningful to the study of visual fields in glaucoma, a post hoc classification of fields was done. Two visual field experts were masked to the mean patterns generated by the vB-ICA-mm along the positive and negative directions of each of the resultant axes. They were then asked to review the actual gray-scale printouts from the SAP visual fields arranged along each axis according to their distance in standard deviation from the mean of cluster *c*, to discern if a consistent pattern of field loss was seen in each cloud and, if so, to describe the pattern within each of the resulting groups of fields. The best descriptions were arrived at by consensus.

RESULTS

POST HOC ANALYSIS

After convergence with unsupervised learning, post hoc analysis found cluster 1 to contain 107 eyes with GON and three eyes that had normal optic discs. Cluster 2 contained 186 eyes with normal optic discs and 49 eyes with GON. Without knowing the diagnosis during the learning phase, vB-ICA-mm placed 68.6% of the eyes with GON in cluster 1 and 98.4% of the eyes with normal optic discs in cluster 2. For ease of reading, from this point on, cluster 1 will be called cluster G, and cluster 2 will be called cluster N.

CLUSTERS AND AXES CREATED BY VB-ICA-MM

With a data set of 345 eyes containing SAP fields of 189 normal eyes or 156 eyes with GON, the model which performed best was that with two clusters of one and six axes, respectively. To determine the optimal number of axes within each global cluster, the contribution of each axis for reconstructing the input was plotted against the axis number choices provided by the vB-ICA-mm. In cluster G, the contribution value dropped to near zero after six axes and was zero after 12 axes (Figure 3). Examination of axes 7 through 12 indicated little additional information, and these axes may have represented noise. Hence, the most informative number of axes in cluster G was six. This means most of the information in the visual field data could be described using six maximally independent axes. In cluster N, the contribution of the axis dropped to zero after one axis; hence, only one axis was necessary to represent cluster N.

Figure 1, B shows in a two-dimensional projection the distribution of clusters G and N collapsed from the 53 dimensions, with the vB-ICA-mm–derived axes superimposed. The two clusters overlapped, and each cluster appeared to have an ellipsoidal distribution. Each axis in cluster G passed through the centroid of that cluster in 53-dimensional space. The findings of the vB-ICA-mm were related to clinical practice by examining specific points along each axis, at –1 and +2 SD away from the centroid (Figure 4). The mean pattern generated at the centroid of cluster G is also displayed in Figure 4. A similar examination of the axis in cluster N was not done, because it was the only axis.

Pattern Display

Simulation of Total Deviation Plot in Generated Patterns. Patterns generated at a particular point in the 53-dimensional space are equivalent to absolute sensitivity patterns. Visual field experts do not evaluate the absolute sensitivity values plotted in the visual field printout. Instead, they typically rely on total deviation (TD) or pattern deviation (PD) plots supplied by the Statpac analysis, which

take into account the deviation from age-matched healthy eyes (TD) and the effect on the fields of global factors such as cataract (PD). Thus the numerical TD plot was simulated by subtracting these generated absolute sensitivity patterns from the mean absolute sensitivity pattern at the centroid of cluster N, the cluster holding nearly all of the normal eyes. Then the simulated numerical TD pattern was converted into a gray-scale pattern. Another way to consider these derived plots is as 53-dimensional vectors originating at the centroid of the mainly normal cluster N and ending at specified points along any of the six axes in cluster G. Figure 1, C shows graphically how the generated patterns on the axes are made to simulate the TD plot of Statpac.

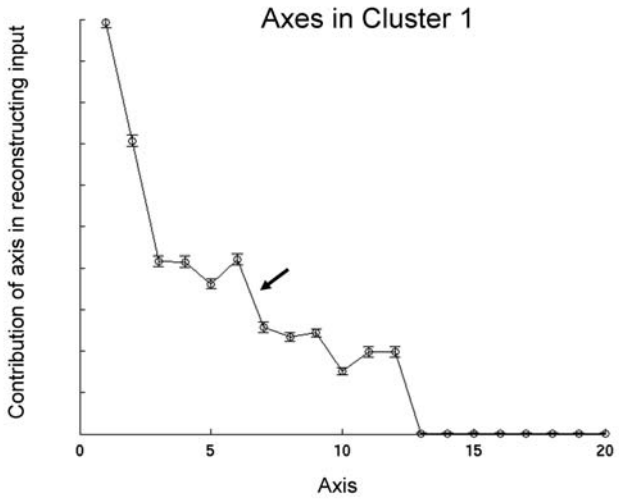

FIGURE 3

Contribution of axis versus axis number. The norm of $\mathbf{A}_m$, on the y-axis represents the contribution of axis *m* in reconstructing the input. Note the small contribution of axes 7 through 12. There was no contribution after 12 axes.

Pattern Shift Along Each Axis. The centroid of cluster G from the vB-ICA-mm result could be considered the mean glaucoma visual field (Figure 4), and the centroid of cluster N could be considered the mean normal visual field. Although cluster N contained 49 eyes with GON along with 98.4% of the normal eyes, neither the two visual field experts nor supervised and unsupervised learning algorithms noted field abnormalities in the majority of these 49 eyes.[4,13] The mean glaucoma field was not quite uniformly depressed, with greater depression than average at the nasal step zone, a slightly greater depression than average in the superior hemifield, and the least depression just inferior to fixation (Figure 4).

Each axis passed through the mean glaucoma visual field at the centroid of cluster G. The six axes were, by definition, maximally independent. Hence, the patterns represented by each axis were maximally different. The farther along the axis away from the glaucoma mean, the greater was the deviation from the *glaucoma* mean. The axis direction was considered to be positive if the distance from the *normal* mean field always increased as the shift from the glaucoma mean increased along the axis. As the distance from the glaucoma mean increased in the negative direction, the distance of a point on the axis to the normal mean initially decreased until the minimum distance of the axis to the normal mean was reached (Figure 1, B). Thereafter, the distance from a point on the axis to the normal mean increased as the distance from the glaucoma mean increased.

Analysis of Generated Patterns and SAP Fields in Axis-Derived Clouds

Extraction of Information From Vectors. The information collected from each axis in cluster G was the generated field pattern on the axis at 0 SD, –1 SD, and +2 SD displayed as TD-like plots along with the actual printed Statpac gray-scale plots for each of the SAP fields assigned to the axis (Figure 4). The generated plot at 0 SD, being the mean pattern at the centroid of cluster G, was common to all axes. There were no individual SAP fields placed at or near the centroid, with the closest fields being 0.4 to 0.5 SD from the centroid on the negative side of axes 1 through 5 (there were no fields on the negative side of axis 6).

On the negative side of the axis, the mean distance from the centroid for all axes was –0.75 SD. Hence, we rounded down and analyzed the generated pattern in this direction at –1 SD. On the positive side, the mean distance from the centroid was 2.50 SD; thus, we rounded down to analyze the generated pattern on the axis at +2 SD.

Table gives the average, minimum, and maximum distance in standard deviation in the positive and negative directions along each axis from the cluster G mean of the 110 SAP fields assigned to the axes.

TABLE. AVERAGE, MINIMUM, AND MAXIMUM DISTANCE (SD) IN THE POSITIVE AND NEGATIVE DIRECTIONS ALONG EACH AXIS FROM GLAUCOMA MEAN OF THE 110 SAP FIELDS ASSIGNED TO THE AXES IN THE GLAUCOMA CLUSTER

AXIS		1	2	3	4	5	6	TOTAL
Negative side of glaucoma mean	**n**	**18**	**7**	**8**	**15**	**4**	**0**	**52**
	average	*–0.9*	*–0.8*	*–0.7*	*–0.8*	*–0.6*		
	min:max	–1.5:–0.5	–1.0:–0.5	–1.0:–0.4	–1.3:–0.4	–0.6:–0.4		
Positive side of glaucoma mean	**n**	**7**	**9**	**10**	**16**	**5**	**11**	**58**
	average	*+1.8*	*+2.6*	*+2.5*	*+1.8*	*+3.9*	*+2.5*	
	min:max	+0.8:+6.4	+1.0:+4.5	+0.6:+4.9	+1.2:+2.9	+1.4:+6.1	+1.4:+4.1	
Total		**25**	**16**	**18**	**31**	**9**	**11**	**110**

min:max = minimum and maximum values in range; n = number.

Visual Fields Associated With Each Axis. The printed Statpac TD plot for each study eye in cluster G was labeled with the axis to which that visual field was assigned (see previous section "Placing Fields About an Axis") and with the distance of that field from the glaucoma mean in units of standard deviation. For each axis, the visual field experts evaluated the printed fields ranked from most minus to most plus and looked for the common pattern elements in the fields assigned to that axis.

Axis 1: At –1 SD, the vB-ICA-mm–generated field had a mild generalized depression that was slightly deeper superiorly and nasally (Figure 4). The negative side of this axis had the least correlation between the generated fields and the SAP fields and the least in common among the SAP fields. The 16 SAP fields ranged from –1.5 to –0.5 SD. The common feature for the 16 SAP fields was the tendency to have a single small depression, but that depression varied in location among the fields (superior nasal step, superior paracentral focus, superior arcuate, enlarged blind spot, and temporal wedge). The generalized depression in the generated field represented an average of all the spot depressions in the individual SAP fields on the negative side. At +2 SD, the generated field had a greater generalized depression deepest in both superior and inferior nasal steps and a slightly greater-than-average depression in the inferior hemifield. The seven SAP fields on the positive side ranged from +0.8 to +6.4 SD. They had multiple deep defects involving three or four quadrants with no consistent patterns. The field defects became larger or deeper as the distance of the field increased from the glaucoma mean. The farthest SAP field at +6.4 SD was too advanced to show a pattern other than severe depression everywhere.

Axis 2: The generated field at –1 SD had mild inferior hemifield depression with the greatest depression at the nasal steps (Figure 4). The six SAP fields ranged from –0.9 to –0.5 SD. They also had mild arcuate defects inferiorly or a focus of depression at a nasal step. The generated field at +2 SD revealed a strong superior hemifield defect augmented nasally. The nine SAP fields ranged from +1.0 to +4.5 SD. They also displayed mostly superior nasal arcuate and hemifield defects. The defects became more severe as the distance from glaucoma mean increased (Figure 5).

Axis 3: The generated field at –1 SD showed a nasal hemifield depression greater along the nasal edge and a little exaggerated at the superior nasal step (Figure 4). The eight SAP fields ranged from –1.0 to –0.4 SD. The common pattern elements in TD plots of the fields in the negative direction were nasal step depressions and arcuate depressions at the nasal edge of the 24° SAP field. The generated field at +2 SD revealed superior hemifield with superior temporal arcuate (wedge) depression combined with a superior nasal step. The 10 SAP fields ranged from +0.6 to +4.9 SD. The fields in the positive direction also tended to combine a superior temporal wedge or a superior outer arcuate depression with a depression at the superior temporal wedge. The defects became more severe as the distance from the glaucoma mean increased.

Axis 4: The generated field at –1 SD showed peripheral depression somewhat greater than central depression (Figure 4). The SAP fields in general had a peripheral ring of defect or focal defects scattered about the periphery. The generated field at +2 SD had a line of depression extending from the blind spot to the nasal step, passing just superior to fixation. To contrast this pattern with the positive pattern on axis 5, this pattern is called the arrow shaft. The central depression was greater than the peripheral depression. The SAP fields on the positive side ranged from +1.2 to +2.8 SD. The depression pattern was horizontal linear from the blind spot to the nasal step, passing just superior to fixation. The defects became deeper or larger as the SAP field was located farther from the glaucoma mean.

Axis 5: The generated field at –1 SD on the negative axis had generalized depression greater in the superior hemifield, centrally, and at the superior nasal step (Figure 4). The four SAP fields ranged from –0.6 to –0.4 SD. The patterns were not uniform, with a tendency for more depressions superiorly and at the superior nasal step. The generated field at +2 SD on the positive axis was

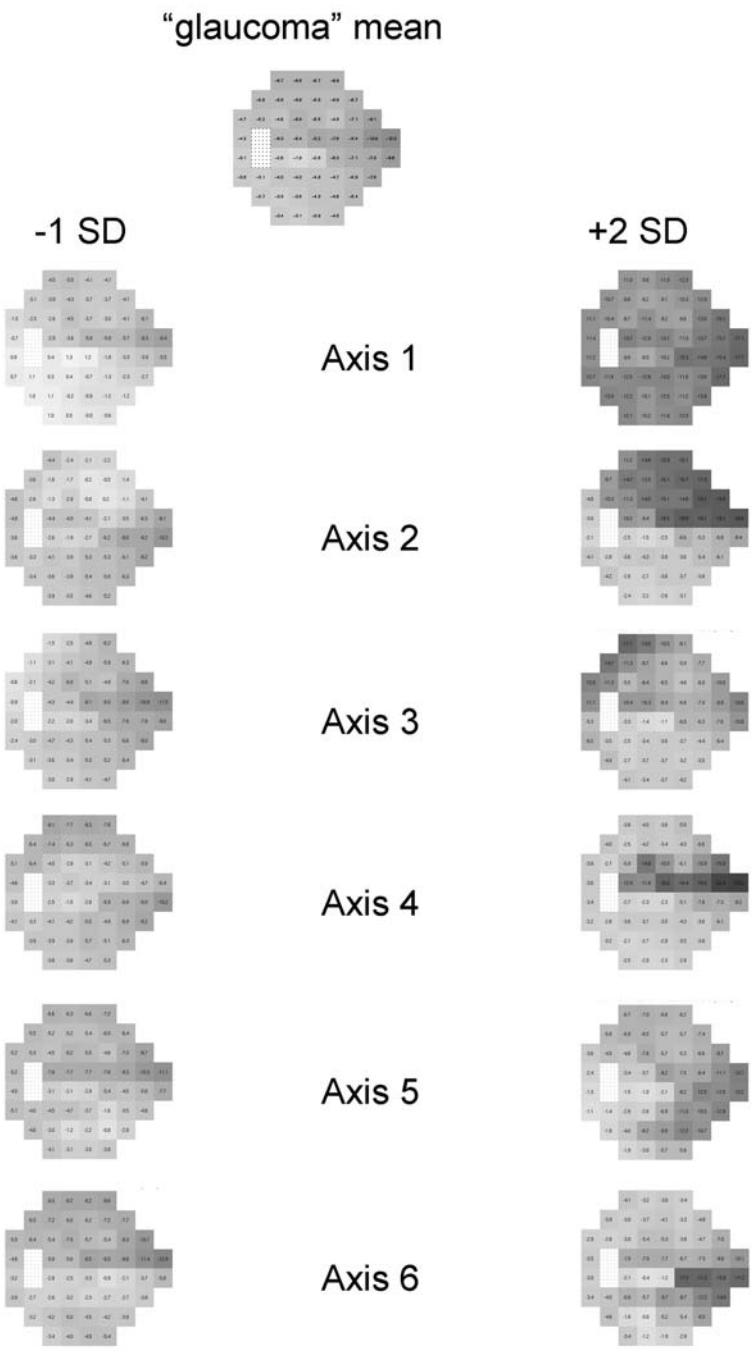

FIGURE 4

Generated patterns on axes. Patterns generated at cluster G (mostly GON) mean, –1 SD on each axis in cluster G, and +2 SD on each axis in cluster G, with the plots displayed as if they were gray-scale images of total deviation plots of the left eye.

weighted along the inferior (more) and superior (less) nasal arcuate edge. To contrast this pattern with that on the positive side of axis 4, this is called the arrowhead. The five SAP fields ranged from +1.3 to +6.1 SD. The defects had the arrowhead at the nasal periphery or were inferior nasal. Except for one field, there was a sense of increasing defect as the distance increased from the glaucoma mean.

Axis 6: The generated field at –1 SD combined mildly greater defects at the superior edge, superior nasal, and superior temporal central (Figure 4). There were no SAP fields on the negative side of axis 6. It is not known whether a larger or different data set would have fields in the space defined by the negative side of axis 6 or whether no real fields would ever exist in this region. As with the other axes, the pattern on the negative sides of the axis tended to be the inverse of the pattern on the positive side. The generated pattern at +2 SD on the axis had a linear depression from inferior to fixation to the inferior nasal step. The 11 SAP fields ranged from

+1.4 to 4.1 SD. The defects were linear from inferior to fixation to the inferior nasal step or were mostly inferior nasal quadrant. The defects tended to broaden or deepen as the distance increased from the glaucoma mean.

In summary, the mean patterns generated by the vB-ICA-mm in general were in good agreement with those identified by experts in this set of fields.

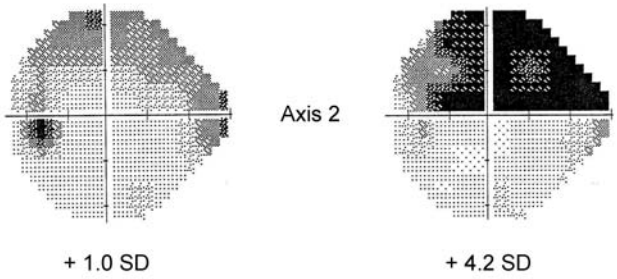

FIGURE 5

Demonstration of severity along an axis. Representative gray scale of two SAP fields from axis 2 showing increasing defect as the distance from cluster G mean increases in the positive direction. Notice that the SAP field gray scale matches the pattern of the generated field at +2 SD in Figure 4 and that the +4.2 SD pattern is deeper but not different in shape from the +1.0 SD pattern.

DISCUSSION

REPRESENTATION OF DATA STRUCTURE

vB-ICA-mm settled on two clusters. Without knowing the diagnosis during the learning phase, vB-ICA-mm placed 68.6% of the eyes with GON in cluster G (true positive rate) and 98.4% of the eyes with normal optic discs in cluster N (true negative rate). In a previous report,[4] mixture of Gaussian, a comparable classifier that had knowledge of the correct diagnosis during the learning phase (supervised learning), had a sensitivity of 67% when specificity was restricted to 100%. Consequently, the cluster structure obtained with unsupervised learning by vB-ICA-mm correlated well with that obtained by supervised learning.

In cluster G, most of the information in the visual field data from glaucoma patients could be described using six maximally independent axes. The need for six axes means that there were several patterns. In cluster N, one axis was sufficient to describe all the data, indicating that the normal-appearing fields were essentially uniform.

CHANGE ALONG AN AXIS

Why should we be interested in generating axes to represent the data when we can generate clusters of patterns that are internally similar but differ from each other?[11,14] The generated field patterns and the analyzed SAP fields did not include age and were in 52-dimensional space. I will discuss the shapes in three-dimensional space, because it is easier to grasp cognitively. Cluster G can itself be segmented into clouds. The elongated clouds that are created around axes within cluster G are likely to have a different type of similarity of its members than clusters close to spherical in shape. Clusters of glaucoma fields created by unsupervised learning with variational Bayesian mixture of factor analysis were somewhat spherical.[13] With that type of clustering, it is possible for one spherical cluster to have mostly mild defects and another cluster to have mostly advanced defects.

Representing the data with axes can give us different insights. The generated patterns represented by the axes differed from each other. The patterns resembled classic glaucoma patterns, such as altitudinal hemifield depression, defects in the arcuate region, reduced sensitivity adjacent to fixation, and temporal wedge. Frequently, there was a combination of classic defects in an axis. The pattern on the negative side of an axis tended to be the inverse of the pattern on the positive side, for example, central versus peripheral, superior versus inferior, temporal versus nasal. Another observation was for each axis in general to capture not only a unique pattern but also mild and severe forms of that pattern.

The axes in the glaucoma cluster pass through the glaucoma mean. But an increase in the severity of a field pattern with respect to the normal mean should more closely match the progression of glaucoma away from normal. It is difficult to model the data in a way that captures severity along axes coming out of the normal mean; the severity along ICA axes with respect to the glaucoma mean serves as a surrogate for the sense of severity of a field from normal. Axes that are close to perpendicular to the vector from the normal mean to the glaucoma mean will least capture the sense of severity, and axes that are close to parallel to the vector from the normal mean to the glaucoma mean will most capture the sense of severity. A question arises whether moving in the positive direction along an axis represents increasing severity of the same pattern, or whether moving in the positive direction shows increasing severity by adding new patterns. This issue is addressed in a separate report that detects worsening of disease by tracking the position of serial fields different from those used in this study. Those fields were tracked within the vB-ICA-mm space developed in this study.[15] That report describes increasing severity as worsening of the same pattern rather than the addition of new patterns. Another question is the

effect of the age difference between the normal and glaucoma participants. In cluster analysis done on the same participants, the age difference was found not to influence the patterns.[13] Since severity, which is incorporated in each axis, is not a factor that distinguishes the axes, the age difference between the normal and glaucoma participants also does not affect the representation of data created with vB-ICA-mm. The representation of patterns is based on glaucoma and not age.

VALIDATION OF METHOD

vB-ICA-mm produces both clusters and representative axes within the clusters. Without knowing the goals of the analysis of the glaucoma and normal fields, and without the fields being labeled with an indication of glaucoma, the clustering accomplished with vB-ICA-mm separated the SAP fields into two clusters. Cluster N had 98.4% of the SAP fields from normal eyes, and cluster G had 68.6% of the fields from eyes with GON. If the goal had been classification instead of data organization, these values would have been specificity and sensitivity, respectively. The cluster structure obtained with unsupervised learning by vB-ICA-mm correlated well with that obtained by supervised learning.

The patterns generated along the ICA axes matched the descriptions of the SAP fields provided by the two visual field experts in this study. More important, the patterns generated along the ICA axes were similar to those discovered by generations of human experts with decades of experience analyzing visual fields in eyes with glaucoma. This finding validates the vB-ICA-mm as a method of separating out patterns of defects consistent with glaucoma.

In summary, the vB-ICA-mm process can be relied on to analyze a complex test for glaucoma in a meaningful way. The data were separated with unsupervised learning by vB-ICA-mm into normal and glaucoma clusters as well as could be accomplished with machine learning classifiers trained with supervised learning on the diagnosis and by glaucoma experts. The patterns disclosed by the axes representing the glaucoma cluster correlated to patterns discovered by years of human experience with glaucomatous visual fields. The validation of vB-ICA-mm on tests we understand indicates that this approach to data mining should teach us the patterns of significance in tests we do not yet understand. Lastly, the type of data representation achieved by ICA axes captures the sense of severity, giving this type of analysis the potential of detecting increasing severity of disease in a quantitative manner.[15]

ACKNOWLEDGMENTS

I would like to acknowledge Pamela A. Sample, PhD, and Robert N. Weinreb, MD, at the Hamilton Glaucoma Center of the University of California, San Diego, for the data sets of normal subjects and of patients with glaucomatous optic neuropathy; Pamela A. Sample, PhD, and Catherine Boden, PhD, at the Hamilton Glaucoma Center for the post hoc analysis of the visual fields assigned to each axis; Te-Won Lee, PhD, at the Institute for Neural Computation of the University of California, San Diego, for the development of the Independent Component Analysis module for Matlab; and Terrance Sejnowski, PhD, Te-Won Lee, PhD, Kwokleung Chan, PhD, and Zuohua Zhang, PhD, at the Institute for Neural Computation and the Computational Neurobiology Laboratory of the Salk Institute for the processing of our data with the variational Bayesian independent component analysis mixture model.

APPENDIX

The application of the variational Bayesian independent component analysis (ICA) mixture model in this project is adapted from the formal presentation in Chan and coworkers.[6] vB-ICA-mm automatically determines the number of clusters and dimensions (axes) of each cluster. The input data (52 field locations plus age for each eye) are denoted by $\mathbf{x}$, in 53-dimensional space, and vB-ICA-mm models the data density by clusters, $p(\mathbf{x}) = \sum_c P(c)\,p(\mathbf{x}\,|\,c)$, where $P(c)$ is the probability mass of cluster c and $p(\mathbf{x}|c)$ is the probability density of $\mathbf{x}$ within cluster c. Within each cluster c, the data are modeled by the linear combination of independent sources, $\mathbf{x}^c = \mathbf{A}^c\mathbf{s}^c + \boldsymbol{\upsilon}^c + \boldsymbol{\sigma}^c$, where $\mathbf{A}^c = (\mathbf{A}_1^c, \mathbf{A}_2^c, \cdots, \mathbf{A}_n^c)$ is the mixing matrix of the independent axes $\{\mathbf{A}_1^c, \cdots, \mathbf{A}_n^c\}$, $\mathbf{s}^c = (s_1^c, \cdots, s_n^c)^T$ are the activation coefficients along the axes, and $\boldsymbol{\upsilon}^c$ is the centroid, $\boldsymbol{\sigma}^c$ the noise, and n the dimensionality (number of axes) of cluster c. The noise is modeled by Gaussian distribution with zero mean and covariance $\boldsymbol{\psi}^c$. So the distribution of x in any cluster can be written as $p(\mathbf{x}\,|\,c) = p(\mathbf{x}\,|\,\mathbf{A}^c, \boldsymbol{\upsilon}^c, \boldsymbol{\psi}^c) = \int N(\mathbf{x}\,|\,\mathbf{A}^c\mathbf{s}^c + \boldsymbol{\upsilon}^c, \boldsymbol{\psi}^c)\,p(\mathbf{s}^c)\,d\mathbf{s}^c$, where N denotes Gaussian distribution, the activation coefficients $\mathbf{s}^c$ are assumed to be independent, and the density of each source, s_m^c, is modeled by k mixtures of Gaussian, $p(s_m^c) = \sum_k \pi_{mk}^c N(s_m^c\,|\,\phi^c{}_{mk}, \beta_{mk}^c)$, where each π_{mk}^c is a mixture weight and N denotes Gaussian distribution whose mean is ϕ_{mk}^c and covariance is β_{mk}^c. The prior for mixing matrix $\mathbf{A}^c$ is also Gaussian with zero mean and covariance $\boldsymbol{\alpha}$: $p(\mathbf{A}_{nm|}^c\,|\,\alpha_m^c) = N(\mathbf{A}_{nm}^c\,|\,0, \alpha_m^c)$.

Often maximum likelihood estimation overfits data, and Bayesian learning overcomes overfitting by introducing priors. The priors introduced for the parameters $\boldsymbol{\pi}, \boldsymbol{\varphi}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\upsilon}, \boldsymbol{\psi}, P(c)$ are D, N, Γ, Γ, N, Γ, D, respectively, where N, Γ, and D are Gaussian, Gamma, and Dirichlet distributions, respectively.

To illustrate variational Bayesian learning, it is convenient to collect all the parameters and call them $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\varphi}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\upsilon}, \boldsymbol{\psi}, P(c)\}$. Given all priors for these parameters, denoted by $p(\boldsymbol{\theta})$, the marginal likelihood is given by $p(\mathbf{x}) = \int p(\mathbf{x}\,|\,\boldsymbol{\theta})\,p(\boldsymbol{\theta})\,d\boldsymbol{\theta}$. This Bayesian approach automatically performs model selection. But exact Bayesian learning is rarely computationally tractable, because it is hard to

obtain the posterior distribution of parameters $p(\boldsymbol{\theta} \mid \mathbf{x})$. We need a simpler function, $q(\boldsymbol{\theta})$, to approximate the true posterior of parameters, and the log marginal likelihood is lower bounded by

$$\log p(\mathbf{x}) \geq \int q(\boldsymbol{\theta}) \log p(\mathbf{x} \mid \boldsymbol{\theta}) d\boldsymbol{\theta} + \int q(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} .$$

We introduced separable distribution over the parameters, and the closed form of $q(\boldsymbol{\theta})$ was obtained, but with different parameters. The learning of vB-ICA-mm was accomplished as follows:

For each number of clusters, *c* (*c*=1,2,…5), each number of axes, *m* (*m*=up to 20), vB-ICA-mm did the following:

Step 1: Initialize the parameters of $q(\boldsymbol{\theta})$.

Step 2: Derive the new axes by re-estimating the parameters of $q(\boldsymbol{\theta})$ with $q(\boldsymbol{\theta})$ fixed.

Step 3: Based on the new axes, recalculate the cluster probability of each datum and update the parameters of $q(\boldsymbol{\theta})$.

The vB-ICA-mm then iterated between steps 2 and 3 until the axes and the cluster probability stop changing (convergence).

Readers are referred to Chan and coworkers[6] for further details of the functional form for priors, close functional form of $q(\boldsymbol{\theta})$, and the learning rules.

As a local method, vB-ICA-mm is likely to get stuck at local minima. Random initialization is a way to overcome the local minima problem by permitting selection of the best local minimum. Consequently, the vB-ICA-mm was set to repeat the initial randomization as in step 1 100 times, so that for each *c* and *m*, there were 100 models. All the above was repeated while simultaneously varying the number of clusters and the number of axes. The final model was chosen by comparing all the models based on their marginal likelihood values (the larger the value, the better) and the classification accuracy.

As described in the "Results" section, the best model we got for this analysis was with two clusters: one cluster with only one axis and the other cluster with six axes. vB-ICA-mm assigned a posterior probability, $p(c \mid \mathbf{x}) \propto p(\mathbf{x} \mid c)P(c)$, to each SAP field, which enabled the program to determine the class of SAP with maximum posterior probability. Each cluster was then examined and labeled according to the majority of GON and normal data points (eyes) within: the cluster with one axis that contained mostly normal eyes was called cluster N, and the cluster with six axes that contained mostly eyes with GON was called cluster G. Within cluster G, each data point was projected onto the six axes and was assigned into the cloud of the axis with which the vector for that data point had the smallest angle at the centroid of cluster G.

On a desktop computer with 2GH Pentium 4 processor running Matlab with the independent component analysis code, training time was about 15 minutes for generating clusters and axes. The testing time for each field was small part of a second.

REFERENCES

1. von Graefe A. Über die Untersuchung des Gesichtsfeldes bei amblyopischen Affectionen [Examination of the visual functions in amblyopic affections]. *Graefes Arch Ophthalmol* 1856;2:258-298.
2. Bjerrum J. Om en tilføjelse til den saedvanlige synsfeltun-undersügelse samt om synsfeltet ved glaukom. *Nord Ophthalmol Tidsskr* 1889;2:141-185.
3. Goldbaum MH, Sample PA, White H, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci* 1994;35:3362-3373.
4. Goldbaum MH, Sample PA, Chan K, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci* 2002;43:162-169.
5. Sample PA, Goldbaum MH, Chan K, et al. Using machine learning classifiers to identify glaucomatous change earlier in standard visual fields. *Invest Ophthalmol Vis Sci* 2002;43:2660-2665 [erratum *Invest Ophthalmol Vis Sci* 2003;44:1813].
6. Chan K, Lee T-W, Sejnowski TJ. Variational learning of clusters of undercomplete nonsymmetric independent components. *J Machine Learn Res* 2002;3:99-114.
7. Brigatti L, Hoffman BA, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol* 1996;121:511-521.
8. Bowd C, Chan K, Zangwill LM, et al. Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci* 2002;43:3444-3454.
9. Lee T-W, Lewicki MS, Sejnowski TJ. ICA mixture models for unsupervised classification of non-Gaussian sources and automatic context switching in blind signal separation. *IEEE Trans PAMI* 2000;22:1078-1089.
10. Hyvarinen A, Karhunen J, Erkki O. *Independent Component Analysis*. New York: Wiley-Interscience; 2001.
11. MacKay DJC. Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems* 1995;6:469-505.
12. Bickler-Bluth M, Trick GL, Kolker AE, et al. Assessing the utility of reliability indices for automated visual fields. Testing ocular hypertensives. *Ophthalmology* 1989;96:616-619.
13. Sample PA, Chan K, Boden C, et al. Using unsupervised learning with variational Bayesian mixture of factor analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci* 2004;45:2596-2605.
14. Henson DB, Spenceley SE, Bull DR. Spatial classification of glaucomatous visual field loss. *Br J Ophthalmol* 1996;80:526-531.
15. Sample PA, Boden C, Zhang Z, et al. Unsupervised machine learning with Independent Component Analysis to identify areas of progression in glaucomatous visual fields. *Invest Opthalmol Vis Sci* 2005;46:3684-3692.