Development and Comparison of Automated Classifiers for Glaucoma Diagnosis Using Stratus Optical Coherence Tomography

Mei-Ling Huang¹ and Hsin-Yi Chen²

PURPOSE. To develop and compare the ability of several automated classifiers to differentiate between normal and glaucomatous eyes based on the quantitative assessment of summary data reports from Stratus optical coherence tomography (OCT; Carl Zeiss Meditec Inc., Dublin, CA) in a Chinese population in Taiwan.

METHODS. One randomly selected eye from each of 89 patients with glaucoma and each of 100 age- and sex-matched normal individuals were included in the study. Measurements of glaucoma variables (retinal nerve fiber layer thickness and optic nerve head analysis results) were obtained by Stratus OCT. With the Stratus OCT parameters used as input, receiver operator characteristic (ROC) curves were generated by three methods, to classify eyes as either glaucomatous or normal: linear discriminant analysis (LDA), Mahalanobis distance (MD), and artificial neural network (ANN). The area under the ROC curve was optimized by principal component analysis (PCA). Classification accuracy was determined by cross validation.

RESULTS. The average visual field mean deviation was $-0.7 \pm 0.6$ dB in the normal group and $-2.7 \pm 1.9$ dB in the glaucoma group. The areas under the ROC curves were 0.824 (LDA), 0.849 (MD), 0.821 (ANN), 0.915 (LDA with PCA), 0.991 (MD with PCA), and 0.874 (ANN with PCA).

CONCLUSIONS. With Stratus OCT parameters used as input, automated classifiers show promise for discriminating between glaucomatous and normal eyes. MD measured from multivariate data can predict the severity of glaucoma through the construction of a measurement space. After PCA, implementation results show that the Mahalanobis space created by MD surpasses LDA and ANN in diagnosing glaucoma. (Invest Ophthalmol Vis Sci. 2005;46:4121-4129) DOI:10.1167/iovs.05-0069

Alteration in the structural appearance of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL) usually precedes the development of reproducible glaucomatous achrnomic¹-⁵ and blue-on-yellow⁶ visual field defects. Identification of these changes is important in the diagnosis and treatment of glaucoma at early to moderate stages. ¹ Assessment of RNFL thickness and ONH topographies by clinical evaluation is subject to disagreement between even experienced observers.⁸-¹⁰ Structural measurement of the ONH and RNFL offers the prospect of improving the early detection of glaucoma.⁷ Several imaging instruments, such as the confocal scanning laser ophthalmoscope (CSLO), scanning laser polarimetry (SLP), and optical coherence tomography (OCT), have been designed to measure peripapillary RNFL thickness and optic disc topography objectively and quantitatively.¹¹-¹⁶ The ability to detect glaucoma using these instruments has been widely described and discussed.¹¹-¹⁶ Optical coherence tomography (OCT) is the optical equivalent of ultrasonography with high in vivo resolution.²²-²⁴ The latest generation of Stratus OCT also provides objective, quantitative, and reproducible measurements of the retina, RNFL thickness, and ONH.²⁵-²⁶

Much of the management of glaucoma depends on the diagnosis and risk of progression.²⁷ Diagnosis based on medical image data is common in medical decision-making. For group classification, linear discriminant analysis (LDA) is a well-established method which combines input parameters into a discriminant function to classify patients into different groups and is commonly used in glaucoma.²⁸,²⁹ In an attempt to classify eyes effectively as glaucomatous or healthy, Heidelberg Retina Tomograph (HRT) parameters have been used as LDA input.²⁸ Artificial neural networks (ANNs) have been applied in ophthalmology to interpret and classify visual fields (Goldbaum MH, et al. IOVS 1990;31:ARVO Abstract 2471),³⁰,³¹ detect visual field progression,³² assess structural data from the ONH,³³ and identify noise from visual field information.³⁴ Neural networks are a subset of machine learning classifiers. The terminology has been changed in the artificial intelligence community to the latter term to include classifiers that “learn” but do not necessarily mimic a simple neural pathway of the brain, as neural networks do. Moreover, Mahalanobis distance (MD) is suitably scaled and used for measuring the degree of abnormality on a measurement scale. Each abnormal condition is considered unique, and abnormal conditions are not considered a separate population.³⁵

The purpose of the present study was to develop and compare the best of the three automated classifiers for glaucoma diagnosis based on the summary data reports from the Stratus OCT. The study compared the performances of LDA, MD, and ANN in differentiating glaucomatous from normal eyes in a Taiwanese Chinese population by comparing the area under the ROC for each classification method. PCA was performed on 25 Stratus OCT parameters. The resultant data sets were reprocessed by LDA, MD, and ANN with PCA factor numbers from 1 to 25 to generate the best performance for each classifier.

METHODS

Subjects

One randomly selected eye from each of 89 patients with glaucoma and each of 100 age- and sex-matched normal individuals were
included in the study. Informed consent was obtained from all participants, and the study was approved by the Institutional Review Board of the China Medical University Hospital. The research adhered to the tenets of the Declaration of Helsinki. Subjects with a best-corrected visual acuity of less than 20/40, a spherical equivalent of more than ±5.0 D, and a cylinder correction of more than 3.0 D were excluded. Patients with peripapillary atrophy were also excluded. All subjects underwent a complete ophthalmic examination, including slit lamp biomicroscopy, measurement of intraocular pressure, stereoscopic fundus examination, and standard full-threshold automated perimetry (30-2 mode; Humphrey Field Analyzer, model 750 [HFA]; Carl Zeiss Meditec, Inc., Dublin, CA).

Inclusion criteria for the patients with glaucoma included an intraocular pressure higher than 22 mm Hg, an open angle, and an early reproducible glaucomatous visual field defect in the absence of any other abnormalities to explain the defect, and a mean deviation of more than −6 dB. The patients with glaucoma were recruited from a group of patients with high-tension type open-angle glaucoma who had received at least 6 months’ regular follow-up at the glaucoma service at China Medical University Hospital between June 2003 and October 2004.

Inclusion criteria for normal subjects included no history of eye disease, no family history of glaucoma, intraocular pressure lower than 21 mm Hg when measured by Goldmann applanation tonometry, and normal optic disc appearance based on clinical stereoscopic examination. A normal result on the glaucoma hemifield test and corrected SD (HFA, program 30-2) within normal limits were required. Subjects with normal eyes were volunteers from the staff at the China Medical University Hospital.

**Visual Field Testing**

Achromatic automated perimetry was performed with an HFA, using the central full-threshold visual field testing program 30-2. Visual field reliability criteria included fixation losses and false-positive and -negative rates of less than 20%. The evaluation of glaucomatous visual field defects was made based on the following liberal criteria: two or more contiguous points with a pattern deviation sensitivity loss of $P < 0.01$, three or more contiguous points with sensitivity loss of $P < 0.05$ in the superior or inferior arcuate areas, or a 10-dB difference across the nasal horizontal midline at two or more adjacent locations and an abnormal result on the glaucoma hemifield test.66

**Stratus OCT Imaging**

The Stratus OCT (ver. A 2.0; Carl Zeiss Meditec Inc.) consisted of an infrared-sensitive video camera to provide a view of the scanning probe beam on the fundus, a low-coherence interferometer as light source and detection unit, a video monitor, a computer, and an image-analysis system. The basic principles and technical characteristics of Stratus OCT have been described extensively.37,38 The Stratus OCT delineates intraretinal and cross-sectional anatomy with axial resolution of ≤10 μm and transverse resolution of 20 μm. The Stratus OCT software package includes 18 scan-acquisition protocols and 18 analysis protocols. Together, they enable the optic disc, RNFL, and macula to be analyzed with a single instrument. The OCT protocol in our study included a regular 3.4-mm circular scan to determine RNFL thickness and a fast ONH radial scan to measure optic disc topography. All scans were completed in a single session by a trained operator after pupil dilatation with tropicamide 1% to achieve a minimum pupillary diameter of 6 mm.

The OCT protocol in our study included a regular 3.4-mm circular scan to determine RNFL thickness. The RNFL thickness 3.4-mm scan protocol consisted of three separate circular scans with a diameter of 3.4 mm centered on the optic disc, each of which consisted of 512 A-scans obtained in 1.28 seconds. The total scanning time for the three scans is between 1 and 2 minutes, as each B-scan is briefly reviewed and saved before a subsequent scan can be recorded. The results were obtained from the mean of three scans. The OCT images were taken in each eye with the same parameters by the same examiner. The RNFL thickness report included the OCT image, the fundus image, and the thickness chart. Circular diagrams showed quadrant (temporal, superior, nasal, and inferior) thickness and clock-hour RNFL thickness (11 o’clock, superior temporal: 45–75°; 1 o’clock, superior nasal: 105–135°; 7 o’clock, inferior temporal: 285–315°; 9 o’clock, temporal: 345–15°). The average RNFL thickness was the average thickness along the entire circumference of the optic disc.

The fast ONH radial scan protocol consisted of six linear scans crossing the optic scan. This protocol acquires six 4-mm radial scans in 1.92 seconds. The machine automatically determined the edge of the ONH as the end of the retinal pigment epithelium-choriocapillaris. This could be manually corrected in cases in which the machine did not identify the edge correctly. A straight line connected the edges of the retinal pigment epithelium-choriocapillaris, and a parallel line was constructed 150 μm anteriorly. The structure below this line was defined as the disc cup, and the structure above the line was defined as the neuroretinal rim. OCT ONH analysis measured the following: vertically integrated rim area volume, horizontally integrated rim width, disc area, rim area, cup/disc (C:D) area ratio, C-D horizontal ratio, and C-D vertical ratio.

Quality assessment of Stratus OCT scans was determined by an experienced examiner. Good-quality scans had to have focused ocular fundus images, the signal strength had to be greater than 6, and a centered circular ring around the optic disc had to be present. The RNFL thickness was determined by the difference in distance between the vitreoretinal interface and a posterior boundary, based on a pre-defined reflectivity signal level. A good-quality scan was defined as one with a signal-to-noise ratio of >35, 100% accepted A-scans, and well-delineated NFL in the scan images. Accepted A-scan percentage and signal-to-noise ratio are the scanned image quality indexes that Stratus OCT provides in the analysis printout. These measurements enable the user to make judgments about the technical quality of the particular scan. Only images meeting the criteria of a good-quality scan were selected for further analysis, and the OCT scans of the participants in our study met the criteria of a good-quality scan. ONH images were excluded when the machine incorrectly determined the edge of the ONH as the end of the retinal pigment epithelium-choriocapillaris in automatic mode; if the ONH image was unacceptable; and when images could not be analyzed in the version A 2.0 software (Carl Zeiss Meditec, Inc.), such as those with very small C:D ratios. Therefore, the disc margin in our study was automatically defined in all eyes.

We selected the average RNFL thickness, quadrant thickness (temporal, superior, nasal, and inferior), 12-clock-hour (30° sector) RNFL thicknesses, and ONH analysis results (vertical integrated rim area, horizontal integrated rim area, disc area, cup area, rim area, cup-to-disc area ratio, cup-to-disc horizontal ratio, cup-to-disc vertical ratio) as our 25 input parameters.

The perimetry and OCT examinations were all performed within a maximum period of 2 weeks. If the tests were conducted on the same day, the perimetry examination was performed first.

**Data Normalization**

Because the variables in medical diagnosis data have different units, and variances of variables can be compared only if the variables are measured using the same scale, data normalization is necessary. A correlation matrix based on standardized values of all observations was used in our study. Each feature of the original data set was normalized to have zero mean and unit variance by dividing the mean corrected data by the respective SD before further processing. The benefit of preprocessing normalized data is that it cancels out interdata variations.

**Linear Discriminant Analysis**

The idea underlying this classification method is to develop a linear combination, $F$, of $n$ variables as $F = \beta_1x_1 + \beta_2x_2 + \ldots + \beta_nx_n$ with...
values for $\beta_1$, $\beta_2$, \ldots, $\beta_p$ chosen so as to maximize the discrimination between two groups. This widely used nonparametric statistical method is based on a projection of the multidimensional feature vector $x$ onto an optimal chosen vector $w=(w^1x, \ldots, w^p x)$ to maximize the ratio of between-class scatter to within-class scatter. Detailed descriptions of the statistical procedures involved in the assessment and use of a LDA model can be found in the literature.\textsuperscript{39}

**Mahalanobis Distance**

MD\textsuperscript{40} is a generalized distance, that can be considered a single measure of the degree of divergence in the mean values of different characteristics of a population by considering the correlations between the variables. MD is effective in determining the similarity between a set of values from an unknown sample and a set of values measured from a collection of known samples. MD is superior to other multidimensional distances, such as Euclidean distance, because it considers distribution of the points.\textsuperscript{35} Mahalanobis space is a database containing the means, standard deviations, and correlation structure of the variables in the reference group. The selection of variables in constructing the Mahalanobis space is important. The more homogeneous variables there are in the Mahalanobis space, the smaller the MD. MD can be used to separate the original pattern of data into homogeneous and heterogeneous groups. Thus, MD can be viewed as an indicator that classifies observations into healthy and unhealthy groups when applied to medical examination. Moreover, it can be used to predict a new observation (unknown sample) in a set of variables measured from the collected sample base.

The definition of the MD is a squared distance ($D^2$) calculated for the $j$th observation in a sample of size $n$ with $p$ variables using the following formula:

$$D^2 = \sum_{i=1}^{p} \frac{(z_{ij} - m_i)^2}{s_i^2}$$

where $z_{ij}$ is $(X_{ij} - m_j)/s_j, i = 1$ to $p$ and $j = 1$ to $n$; $z_{ij}$ is $(z_1, z_2, \ldots, z_p)$; $X_{ij}$ is the $i$th characteristic in the $j$th observation; $m_i$ is the mean of the $i$th characteristic; $s_i$ is the SD of the $i$th characteristic; and $T^{-1}$ is the inverse of the correlation matrix.

The steps necessary for constructing a Mahalanobis space are described as follows: (1) Define the variables that determine the healthiness of a condition; (2) collect the data on all the variables from the healthy group; (3) compute the standardized values of the variables of the healthy group; and (4) compute the MD of all observations using the inverse of the correlation matrix. Based on the Mahalanobis space constructed from the healthy group, the MD for the unhealthy group are calculated by using the same inverse of the correlation matrix.\textsuperscript{35}

**Artificial Neural Network**

Machine learning classifiers usually use a form of supervised learning. Supervised learning refers to systems that are trained, instead of programmed, by a set of examples that are input–output pairs.\textsuperscript{41} The input is the data and the output is the classification made by the machine learning method. During training, the classifier is told whether it is correct or incorrect, based on a gold standard, and, after each run-through, it adjusts its internal parameters to arrive at more correct responses. This process is repeated until the classification performance does not improve. After training, the goal is that the machine classifier has learned and can correctly classify new input data that were not part of the original training sets. An attractive aspect of these classifiers is their ability to learn complex patterns and trends in data and to create decision rules adaptively, without the constraints imposed by statistical classifiers.\textsuperscript{12,43} such as LDA. Multilayer perceptron (MLP), a feed-forward, back-propagation network, is the most frequently used ANN technique in glaucoma research.\textsuperscript{45} An ANN can be considered to be composed essentially of a multitude of elemental computing elements (called neurons) organized as a network, similar to the way in which neurons are believed to be interconnected in the brain.\textsuperscript{46} An ANN consists of one input layer, one output layer, and one or more hidden layers that extract essential information during learning. The parameters are adjusted iteratively with a supervised learning process during which the network learns to associate input vectors with the appropriate pattern class. The gradient steepest descent algorithm with back-propagation of error and a momentum term were used during training in this study. Using this method, the neural network classifier is trained to detect a relationship between input (Stratus OCT parameters) and a predefined gold-standard diagnosis by comparing its prediction with the labeled diagnosis and by learning from its mistakes.

**Principal Component Analysis**

PCA was used as a data-reduction technique and was used for explaining the variance–covariance structure through a few linear combinations of original variables. These linear combinations represented the new coordinate system obtained by rotating the original system with maximum variability, to allow a simpler description of the covariance structure. PCA characterizes most of the variance, while greatly reducing the input data set to a few orthogonal variables. PCA projects the set of observed data onto a subspace expanded by the principal components. The projection onto the principal component subspace maximizes the separation in data clusters. The components are artificial variates designed to maximize variance accounted for, not for interpretability. To aid in interpreting, varimax rotation is applied.

**Analysis of Specificity and Sensitivity**

A simple way to assess the performance of classifiers is to compare their average misclassification rate. In biomedical data, the true-positive data, also known as sensitivity, is the fraction of positively labeled test data classified as positive. This is the area under the positive class density curve, to the right of the decision threshold. The true-negative rate, known as specificity, is the fraction of negatively labeled test data classified as negative. The ability of a test to discriminate diseased cases from normal cases is evaluated using ROC curve analysis.\textsuperscript{45} ROC curve is a plot of sensitivity versus specificity. The area under the ROC curve summarizes the quality of classification over a wide range of misclassification costs.\textsuperscript{46} In this study, the area under the ROC curves to classify eyes as glaucomatous or healthy were determined for all techniques.

**Cross-Validation**

A learning classifier should be trained well enough on its training set to be able to generalize new examples. Cross-validation was used to determine the classification accuracy in an unbiased manner. A full cross-validation was performed as follows: First, normal and glaucomatous groups were randomly divided into 10 sets. Then, one normal set was combined with one patient set to form each of the 10 partitions of the dataset. One set became the testing set, and the remaining nine sets were used as training set. During the training of the neural network, another set was treated as a stopping set, to avoid overtraining, and the remaining eight sets were used as the training set. The training-test process was repeated until each set was appointed as a testing set.

**Limitations of Chosen Classifiers**

Traditional statistical multivariate techniques usually require some assumptions; however, it is hard to satisfy all of them simultaneously. The MD and ANN do not require any assumptions regarding the distribution of input variables. The advantage of MD is it considers the correlation among all variables. Because the Mahalanobis space is calculated from healthy data, the selection of healthy samples is important. Therefore, a further study with a larger sample size is needed. ANN techniques incur one general criticism. They are is too complex to allow the interaction of important variables to be identified and measured. ANN is used for pattern recognition. It is therefore nece-
sary to randomize the patterns so that an ANN distinguishes all types of patterns in the system. Furthermore, there is no definite means to set the number of hidden layers in an ANN, which can affect the ultimate results.

**RESULTS**

**Demographic Data**

The demographic details are presented in Table 1. The mean age was 45.9 ± 13.2 years in the normal group and 42.1 ± 12.3 years in the glaucoma group. There was no significant difference in age or refraction status between the two groups (P > 0.05). There was a significant difference in mean deviation in the normal group (−0.7 ± 0.6 dB) and the glaucoma group (−2.7 ± 1.9 dB; P < 0.0001).

**Table 1.** Subject Demographics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Glaucoma</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender Male, n (%)</td>
<td>54 (54)</td>
<td>48 (53.9)</td>
<td></td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>46 (46)</td>
<td>41 (46.1)</td>
<td></td>
</tr>
<tr>
<td>Age (y)</td>
<td>45.9 ± 13.2</td>
<td>42.1 ± 14.3</td>
<td>0.070</td>
</tr>
<tr>
<td>Mean deviation (dB)</td>
<td>−0.7 ± 0.6</td>
<td>−2.7 ± 1.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Refraction (Ds)</td>
<td>−1.0 ± 1.6</td>
<td>−1.3 ± 1.9</td>
<td>0.236</td>
</tr>
</tbody>
</table>

Age, mean deviation, and refraction are expressed as the mean ± SD.

* Compared by t-test.

**Table 2.** Stratus OCT Glaucoma Variables Included in the 25-Input Set

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Glaucoma</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average RNFL thickness (µm)</td>
<td>113.0 ± 13.8</td>
<td>94.1 ± 14.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Temporal quadrant</td>
<td>84.9 ± 16.8</td>
<td>82.4 ± 21.3</td>
<td>0.375</td>
</tr>
<tr>
<td>Superior quadrant</td>
<td>135.5 ± 17.2</td>
<td>114.4 ± 24.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Nasal quadrant</td>
<td>87.5 ± 20.6</td>
<td>66.5 ± 18.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inferior quadrant</td>
<td>142.9 ± 19.1</td>
<td>113.3 ± 22.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Clock-hour segment thickness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>131.6 ± 24.4</td>
<td>113.4 ± 30.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>11 (superior-temporal)</td>
<td>146.3 ± 24.0</td>
<td>126.6 ± 29.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>10</td>
<td>99.9 ± 21.3</td>
<td>94.8 ± 25.9</td>
<td>0.142</td>
</tr>
<tr>
<td>9 (temporal)</td>
<td>69.5 ± 14.5</td>
<td>67.7 ± 21.1</td>
<td>0.493</td>
</tr>
<tr>
<td>8</td>
<td>86.5 ± 19.9</td>
<td>84.9 ± 27.2</td>
<td>0.682</td>
</tr>
<tr>
<td>7 (inferior-temporal)</td>
<td>160.1 ± 24.5</td>
<td>126.4 ± 35.0</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6</td>
<td>151.2 ± 27.3</td>
<td>119.7 ± 34.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>5</td>
<td>118.5 ± 25.3</td>
<td>94.0 ± 25.0</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>4</td>
<td>86.2 ± 21.7</td>
<td>64.8 ± 19.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3 (nasal)</td>
<td>76.9 ± 22.0</td>
<td>58.0 ± 18.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2</td>
<td>100.5 ± 25.0</td>
<td>77.6 ± 24.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>1</td>
<td>128.9 ± 23.9</td>
<td>103.6 ± 29.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ONH analysis result</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical integrated rim area (mm²)</td>
<td>0.50 ± 0.51</td>
<td>0.31 ± 0.22</td>
<td>0.001</td>
</tr>
<tr>
<td>Horizontal integrated rim width (mm)</td>
<td>1.77 ± 0.24</td>
<td>1.57 ± 0.27</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Disc area (mm²)</td>
<td>2.48 ± 0.39</td>
<td>2.57 ± 0.57</td>
<td>0.203</td>
</tr>
<tr>
<td>Cup area (mm²)</td>
<td>0.70 ± 0.34</td>
<td>1.14 ± 0.68</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Rim area (mm²)</td>
<td>1.77 ± 0.38</td>
<td>1.48 ± 0.41</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cup/disc area ratio</td>
<td>0.28 ± 0.12</td>
<td>0.42 ± 0.18</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cup/disc horizontal ratio</td>
<td>0.57 ± 0.14</td>
<td>0.70 ± 0.16</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cup/disc vertical ratio</td>
<td>0.47 ± 0.11</td>
<td>0.66 ± 0.67</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Data are expressed as the mean ± SD.

* Compared by t-test.

**Classification Results**

Table 3 summarizes the sensitivities, specificities, and areas under the ROC curves for 25 individual parameters. The inferior quadrant thickness was the best individual parameter for differentiating between normal and glaucomatous eyes (ROC area, 0.832). The average RNFL thickness was the second best parameter (ROC area, 0.825). In addition, the 7-clock-hour segment thickness ranked third (ROC area, 0.773).

The areas under the ROC curves were LDA (0.824), MD (0.849), and ANN (0.821). The areas were enhanced after PCA. The number of factors from 1 to 25 were used to perform LDA, MD, and ANN techniques. Figures 4 and 5 show the ROC curves for LDA, MD, and ANN classification techniques, with and without PCA. The areas under the ROC curves were LDA with PCA (0.915), MD with PCA (0.991), and ANN with PCA (0.874). The differences between the areas under the ROC curves were compared between six classifiers. The t-test revealed significant difference in ROC curves between groups of MD with PCA and LDA (P < 0.0001), groups of MD with PCA and MD (P < 0.0001), groups of MD with PCA and ANN (P < 0.0001), and groups of MD with MD and ANN (P < 0.0001).
0.0001), groups of MD with PCA and LDA with PCA (P < 0.0001), and groups MD with PCA and LDA with PCA (P < 0.0001). Areas under the ROC curves (±SD; with sensitivities at 80% and 90% specificities) for all classification techniques applied with and without PCA are displayed in Table 4.

In our study, LDA, MD, and ANN techniques were applied to 25 parameters in the full input set with and without PCA. This investigation showed that the classification technique can be used to discriminate glaucomatous from normal eyes. All six classification techniques were able to discriminate glaucoma-
tous from normal eyes (ROC area, >0.810). MD with PCA was the best classifier for differentiating between normal and glaucomatous eyes (ROC area, 0.991).

Likelihood ratios (LRs) with their 95% confidence intervals (CIs) were calculated for six automated classifiers, with arbitrary cutoffs (Tables 5, 6, 7). The scales for different classifiers were different. For example, in MD, we had only positive values, whereas in LDA, the values were negative and positive. Therefore, the cutoffs were not the same (Tables 5, 6, 7). LDAs of less than −1.0 (LR > 10) were as associated with large effects on posttest probability of disease, whereas the other test ranges were associated with small to moderate effects. LDA with PCA of less than −1.0 (LR > 10) but greater than 1.0 (LR < 0.1) were associated with large effects on posttest probability of disease, whereas the other test ranges were associated with small to moderate effects. ANN values greater than 1.0 (LR > 10) were associated with large effects on posttest probability of disease, whereas the other test ranges were associated with small to moderate effects. For ANN with PCA, values of less than −0.5 (LR ≥ 24) were associated with large effects on posttest probability of disease, whereas the other test ranges were associated with small to moderate effects. For MD, results greater than 4.0 (LR > 10) were associated with large effects on posttest probability of disease, whereas the other test ranges were associated with small to moderate effects. MD with PCA between 1.5 and 2.0 (LR = 4.5) were associated with small effects, whereas the other test ranges were associated with large effects.

**DISCUSSION**

Recently, studies have been undertaken to evaluate the performance of glaucoma detection using Stratus OCT (Carl Zeiss Meditec, Inc.). In our previous report, we differentiated between normal and glaucomatous eyes in a Taiwan Chinese population based on the summary data reports from Stratus OCT, by using a linear discriminant analysis with forward and backward selection to determine the best combination of parameters for discriminating between glaucomatous and healthy eyes. In the present study, we increased the sample size and applied other automated classifiers to reconfirm OCT in diagnosing glaucoma. Our results showed that it is possible to differentiate between glaucomatous and normal eyes by analyzing the input parameters with Stratus OCT.

MD uses simple measures of descriptive statistics and is not probabilistic in nature. As far as we know, our study is the first to apply MD to differentiate glaucomatous from normal eyes. Taguchi and Rajesh used MD to discriminate between a

![Figure 3. Use of PCA to determine the optimum ROC area for ANN. The area under the ROC curve (y-axis) is shown as a function of the number of PCA factors. The largest area was 0.874, with a principal component factor of 3.](https://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933437/ on 08/28/2018)

![Figure 4. ROC curves and areas for optimized LDA, MD, and ANN. The areas under the ROC curves for three classifiers are not statistically significant.](https://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933437/ on 08/28/2018)
healthy group and those with liver disease. From their results, the MDs ranged from 0.3784 to 2.3581 in the healthy group and from 7.7274 to 135.6978 in the unhealthy groups. The classification rate in Taguchi et al was 100%. Su and Li\textsuperscript{51} compared the classification performance of MD and ANN on liver disease. The correct classification rates were 95.52% and 89.55% by MD and ANN, respectively, in their study. In addition, Samek et al.\textsuperscript{52} applied MD pattern recognition to evaluate laser-induced breakdown spectroscopy (LIBS) spectra recorded from teeth. They achieved close to 100% identification, and only one sample was misinterpreted. The accuracy of MD to differentiate glaucoma from normal eyes in our study was 97.66%.

A few studies have been conducted to evaluate the discriminant powers of OCT in diagnosing glaucoma. Sanchez-Galeana et al.\textsuperscript{11} determined from OCT summary data reports that the sensitivity and specificity ranged from 76% to 79% and 68% to 81%, respectively, for discriminating between early to moderate glaucomatous and normal eyes. In a study by Greaney et al.,\textsuperscript{7} four sectors in order of the most discriminating were identified by stepwise discriminating analysis: temporal to superior (45–75°), inferior (265–295°), temporal (345–15°) and superior to temporal (15–45°); they resulted in an area under the ROC curve 0.88. The area under the ROC curves for the earlier versions of the OCT ranged from 0.79 to 0.94, depending on the parameters and characteristics of the population evaluated.\textsuperscript{53–56} In a study by Medeiros et al.,\textsuperscript{47} the inferior quadrant thickness had the highest ROC area (0.92) among the total parameters. Our result also demonstrated that the maximum area under the ROC curve was 0.832 in inferior quadrant thickness. The most plausible reason that individual parameters from Stratus OCT were not good enough is that most patients in our glaucoma group had early glaucoma (mean deviation, $-2.7 \pm 1.9$ dB). However, after the application of automated classifiers, the area under the ROC curve increased to 0.991.

There were some limitations in this cross-sectional study. First, the particular mathematical models that we chose are unlikely to be the only ones that could be applied. Comparisons among other classification methods should be made to yield the best models for improving the discriminant power of OCT. Second, comparisons across studies are very difficult.

### Table 4. Comparison of Sensitivities at Particular Specificities and Comparison of Area under the ROC Curve

<table>
<thead>
<tr>
<th>Technique</th>
<th>ROC Area Total (±SE)</th>
<th>Sensitivity of 80%</th>
<th>Sensitivity of 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>0.824 (±0.032)</td>
<td>72.8</td>
<td>58.0</td>
</tr>
<tr>
<td>MD</td>
<td>0.859 (±0.050)</td>
<td>79.0</td>
<td>56.6</td>
</tr>
<tr>
<td>ANN</td>
<td>0.821 (±0.033)</td>
<td>74.1</td>
<td>71.6</td>
</tr>
<tr>
<td>LDA with PCA</td>
<td>0.915 (±0.022)</td>
<td>82.7</td>
<td>76.5</td>
</tr>
<tr>
<td>MD with PCA</td>
<td>0.991 (±0.008)</td>
<td>100.0</td>
<td>98.8</td>
</tr>
<tr>
<td>ANN with PCA</td>
<td>0.874 (±0.027)</td>
<td>81.5</td>
<td>66.7</td>
</tr>
</tbody>
</table>

### Table 5. Interval Likelihood Ratios and 95% CI for LDA Classifiers

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Interval LRs</th>
<th>Subjects in Each Category n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\leq -2.0$</td>
<td>15.00 (10.8–20.8)</td>
<td>16 (9)</td>
</tr>
<tr>
<td>$&gt;-2.0$</td>
<td>11.50 (8.40–15.75)</td>
<td>25 (15)</td>
</tr>
<tr>
<td>$&gt;-1.0$</td>
<td>1.19 (0.90–1.57)</td>
<td>46 (27)</td>
</tr>
<tr>
<td>$&gt;0.0$</td>
<td>0.40 (0.30–0.54)</td>
<td>35 (20)</td>
</tr>
<tr>
<td>$&gt;1.0$</td>
<td>0.20 (0.15–0.27)</td>
<td>42 (25)</td>
</tr>
<tr>
<td>$\geq 2.0$</td>
<td>0.17 (0.12–0.24)</td>
<td>7 (4)</td>
</tr>
<tr>
<td>LDA with PCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\leq -2.0$</td>
<td>Infinity (NA)</td>
<td>16 (9)</td>
</tr>
<tr>
<td>$&gt;-2.0$</td>
<td>Infinity (NA)</td>
<td>25 (15)</td>
</tr>
<tr>
<td>$&gt;-1.0$</td>
<td>1.47 (1.31–1.96)</td>
<td>42 (25)</td>
</tr>
<tr>
<td>$&gt;0.0$</td>
<td>0.43 (0.33–0.58)</td>
<td>43 (25)</td>
</tr>
<tr>
<td>$&gt;1.0$</td>
<td>0.06 (0.04–0.08)</td>
<td>37 (21)</td>
</tr>
<tr>
<td>$\geq 2.0$</td>
<td>0</td>
<td>8 (5)</td>
</tr>
</tbody>
</table>

**Figure 5.** ROC curves and areas for optimized LDA with PCA, MD with PCA, and ANN, with PCA. The area under the ROC curve for MD with PCA was statistically more significant than the areas under the ROC curves for ANN with PCA and LDA with PCA.
because of differences in population demographics and the definition and severity of glaucoma and because very few studies have been conducted to investigate RNFL thickness and disc topography together, using automated classifiers method. Third, the sample used in studies is usually a clinic-based population of patients with glaucoma. These patients have been identified on the basis of particular patterns of structural and functional abnormality that meet preconceived notions that bias the outcome of comparison.37 Besides, the selection bias from OCT images, such as those with peripapillary atrophy or some optic disc shape that could not be analyzed by Stratus software version A 2.0, making those individuals poor candidates for OCT examination. Therefore, the selection bias did exist in this study. Furthermore, the Stratus OCT software in our laboratory was different from that used in other laboratories in Western countries. The A 2.0 version, unlike the A 3.1 version, does not have an internal normative database. Therefore, we were unable to calculate the interval LRs for each parameter. Instead, interval LRs were calculated for the six automated classifiers. The usefulness of a diagnostic test is influenced by the proportion of patients suspected of having the target disorder whose test results have high (>10) or very low (<0.1) LRs, thus greatly affecting the probability of disease.28 As indicated in Tables 5, 6, and 7, this proportion was 24% for LDA, 50% for LDA with PCA, 10% for MD, 92% for MD with PCA, 4% for ANN, and 42% for ANN with PCA. The proportion of large effects of MD with PCA were much higher than the effects of other classifiers. A multilevel LR for glaucoma of <0.1 with an MD of >2.0 indicates that MD < 2.0 almost exclusively occurred in the healthy group. A multilevel LR for glaucoma of 71 with an MD > 4.0 indicates that an MD > 4.0 occurs over 70 times more often in patients with glaucoma than in healthy persons. These data suggest that glaucoma subjects rarely have an MD < 2.0 and the healthy subjects almost never have an MD > 4.0 (Table 6). Selection of other cutoffs may result in different proportions. Larger sample sizes would provide more precise and robust estimations of LRs using smaller intervals of the range of possible test values. However, our results can be used as the basis for further improving the diagnostic accuracy of glaucoma in the Taiwan Chinese population in the near future.

In summary, automated classifiers showed promise for differentiating glaucomatous from normal eyes in the Taiwan Chinese population, by using summary data from Stratus OCT. Although the result was good, clinicians should be cautious when accepting this classification as a reliable indicator of diagnosis of glaucoma and should integrate the Stratus OCT result into the entire clinical picture when diagnosing glaucoma.

### References


14. Choplin NT, Lundy DC, Dreher AW. Differentiating patients with glaucoma from glaucoma suspects and normal subjects by nerve
fiber layer assessment with scanning laser polarimetry. Ophthal-
tomography and scanning laser polarimetry in normal, ocular hy-
terpertensive, and glaucomatous eyes. Am J Ophthalmol. 2000;129:
129–135.
ophthalmoscopy: reproducibility of optic nerve head topographic
measurements with the confocal laser scanning ophthalmoscope.
17. Weinreb RN, Shakeda S, Sample PA, et al. Association between
quantitative nerve fiber layer measurement and visual field loss in
fiber layer thickness in normal and glaucomatous eyes using optical
layer thickness measurement reproducibility using scanning laser
of retinal nerve fiber layer thickness in glaucoma diagnosis. J Glau-
21. Chen HY, Hung PT, Hung TJ. Optical coherence tomography study
on retinal nerve fiber layer thickness in open angle glaucoma.
22. Hee MR, Izatt JA, Swanson EA, et al. Optical coherence tomogra-
a new tool for glaucoma diagnosis. Curr Opin Ophthal-
24. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomogra-
fiber layer thickness measurement reproducibility using optical
fiber thickness, macular thickness, and optic nerve head measure-
ments using Stratus OCT. Invest Ophthalmol Vis Sci. 2004;45:
1716–1724.
learning classifiers for diagnosing glaucoma from standard auto-
and linear discriminant functions for glaucoma detection using
confocal scanning laser ophthalmoscopy of the optic disc. Invest
for early detection of glaucomatous optic disc changes. Br J Ophthal-
30. Goldbaum MH, Sample PA, White H, et al. Interpretation of auto-
mated perimetry for glaucoma by neural network. Invest Ophthal-
31. Mutilukan, E, Keating, K. Visual field interpretation with a personal
tion of glaucomatous visual field progression with neural net-
33. Briggatt I, Hoffman BA, Caprioli J. Neural networks to identify
glaucoma with structural and functional measurements. Am J Ophthal-
34. Henson DB, Spence CE, Bull DR. Artificial neural network ana-
lysis of noisy visual field data in glaucoma. Artif Intell Med. 1997;
35. Taguchi G, Rajesh J. The Mahalanobis-Taguchi Strategy: A Pat-
39:2321–2328.
37. Schuman JS. Appendix A: physical principles of optical coherence
Tomography of Ocular Disease. 2nd ed. Thorofare, NJ: Slack;
col choice affects optical coherence tomography (OCT-3) mea-
39. Kleinbaum DG, Kupper LL, Muller KE. Applied Regression Anal-
ysis and other Multivariate Methods. 2nd ed. Boston: PSW-Kent;
1998.
40. Mahalanobis PC. On the generalized distance in statistics, proceed-
41. Poggio T, Shelton CR. Machine learning, machine vision, and the
42. Haykin S. Multilayer perceptrons. In: Neural Networks: A Com-
prehensive Foundation. 2nd ed. Upper Saddle River, NJ: Prentice-
43. Rumelhart DE, Hinton G, Williams R. Learning representations of
44. Tomatis S, Bono A, Bartoli C, et al. Automated melanoma detection:
multispectral imaging and neural network approach for classifica-
45. Goodenough DJ, Rossmann K, Lusted LB. Radiographic applica-
tions of receiver operating characteristic ROC curve. Radiology.
1974;110:89–95.
46. Chan K, Lee TW, Sample PA, et al. Comparison of machine learn-
ing and traditional classifiers in glaucoma diagnosis. IEEE Trans
VCC scanning laser polarimeter, HRT II confocal scanning laser
ophthalmoscope, and Stratus OCT optical coherence tomography
cohere tomography scanning areas for detection of glaucoma-
49. Chen HY, Huang ML. Discrimination between normal and glauco-
atous eyes using Stratus optical coherence tomography in Tai-
243:894–902.
51. Su CT, Li TS. A Mahalanobis distance based classifier for diagnosis
52. Samek O, Telle HT, Beddows DCS. Laser-induced breakdown
spectroscopy: a tool for real-time, in vitro and in vivo identification
tomography measurement of macular and nerve fiber layer thick-
ness in normal and glaucomatous human eyes. Ophthalmology.
glaucomatous damage on retinal nerve fiber layer thickness be-
tween normal and glaucomatous eyes using the Heidelberg retino-
tomography. GDX nerve fiber analyzer, and optical coherence
55. Williams ZY, Schuman JS, Gamell L, et al. Optical coherence
tomography measurement of nerve fiber layer thickness and the
likelihood of a visual field defect. Am J Ophthalmol. 2002;134:
538–546.
normal and glaucomatous eyes using the Heidelberg retina tomog-
raphy, GDX nerve fiber analyzer, and optical coherence tomogra-
57. Garway-Heath DF, Hitchings RA. Sources of bias in studies of optic
82:986.
Users’ Guide to the Medical Literature: Essentials of Evidence-