Current Keratoconus Detection Methods Compared With a Neural Network Approach

Michael K. Smolek and Stephen D. Klyce

Purpose. Four videokeratographic methods for keratoconus detection were compared with a neural network approach.

Methods. A classification neural network for keratoconus screening was designed to detect the presence of keratoconus (KC) or keratoconus suspects (KCS); a separate cone severity network graded the severity of conelike topography patterns consistent with KC or KCS. Three hundred TMS-1 examinations (Tomey) were randomly divided into training and test sets. Ten topographic indexes were network inputs. Nine categories were used: normal, astigmatism, KC, KCS, contact lens-induced warpage, pellucid marginal degeneration, photorefractive keratectomy, radial keratotomy, and penetrating keratoplasty. KC was subdivided into KC1 (mild), KC2 (moderate), and KC3 (advanced). There were three outputs for the classification network (KC, KCS, and OTHER); target output values of 0 = OTHER, 0.25 = KCS, 0.75 = KC1, 0.75 = KC2, and 1.0 = KC3 were used for the severity network.

Results. The best-trained classification network had 100% accuracy, specificity, and sensitivity for the test set. The severity network had mean outputs (±standard deviation) of OTHER = 0.02 ± 0.02, KC1 = 0.71 ± 0.05, KC2 = 0.52 ± 0.17, KC2 = 0.74 ± 0.12, and KC3 = 0.91 ± 0.15. The severity network output for all categories was well correlated to the keratoconus prediction index (R = 0.892, P < 0.0001). The classification network had an overall accuracy and specificity significantly better (P < 0.005) than the Klyce/Maeda keratoconus index (KCI) test, the Rabinowitz test (K & I-S), and simulated keratometry (average Sim K). However, there were no significant differences in keratoconus sensitivity between the classification network, KCI, and K & I-S. The sensitivity and specificity of average Sim K were significantly worse than those of the other tests. The classification network had significantly better sensitivity (P < 0.001) and specificity (P = 0.025) for KCS detection than the K & I-S.

Conclusions. The neural networks completely distinguished KC from KCS and from topographies that resembled KC. The network approach equaled the sensitivity of currently used tests for keratoconus detection and outperformed them in terms of accuracy and specificity. Investig Ophthalmol Vis Sci. 1997;38:2290-2299.

Keratoconus is a disease characterized by a noninflammatory cone-shaped protrusion of the corneal surface (ectasia), stromal thinning in the region of the cone apex, irregular corneal astigmatism, and a number of other clinical signs and symptoms arising from structural and biochemical changes in the cornea. The etiology of the disease remains unknown, although there is evidence of genetic inheritance and possible linkage with systemic disease and circumstantial evidence that environmental factors may be involved. Additionally, there are inherent, anisotropic structural features in the corneal stroma that may influence the position, form, and magnitude of the ectasia in eyes in which keratoconus develops. Biochemically, keratoconus has been associated with proteolytic breakdown and incomplete repair of the corneal extracellular matrix. Proteases secreted by stromal keratocytes (gelatinase A) have been shown to be elevated in unscarred keratoconus tissue samples. There also is evidence of elevated lysosomal enzyme levels in the conjunctival epithelium of keratoconus cases, suggesting that an epithelial-based biochemical mechanism may initiate the disease.
steepleing of the cornea, such that some keratoconus diagnoses might be more accurately defined as corneal warpage, particularly where external factors have been implicated. For example, at least one traumatically altered postoperative cornea has been reported to have a conelike ectasia, complete with many of the traditional clinical signs of keratoconus. Although associations between contact lens wear or eye rubbing and the diagnosis of keratoconus are intriguing, there is no evidence of a direct cause and effect. However, it is known that a poor contact lens fit may induce reversible localized peripheral steepening and corneal molding artifacts, and intense, repetitive eye rubbing could theoretically warp the cornea. Furthermore, Wilson et al. hypothesized that an increased epithelial release of interleukin-1 alpha into the stroma, which might be induced by contact lens wear or eye rubbing, could facilitate keratocyte apoptosis and thus mediate keratoconus-related thinning and degeneration of the stromal matrix. A sensitive screening technique that detects the earliest and mildest manifestations of ectasia, along with a quantitative index of the severity of the progression, should provide important clues to the true etiology of keratoconus.

Traditionally defined keratoconus occurs in approximately 4 to 600 persons per 100,000 (0.004% to 0.6%) in the general population, but a considerably higher percentage (5.7% to 10%) of patients with previously undiagnosed clinical keratoconus or keratoconus-like topographic patterns seek relief from their irregular astigmatism at contact lens and refractive surgery clinics. This high percentage may not be caused solely by self-selection for treatment; it also can be credited at least in part to the sensitivity of videokeratography as a keratoconus screening tool. Given the relatively high percentage of surgical candidates exhibiting videokeratographic signs of a conelike feature, some have argued that minor conelike features seen in videokeratography are a natural condition consistent with an otherwise normal cornea. However, in at least one refractive surgery study, all the cases exhibiting a conelike feature not associated with contact lens–induced warpage were later determined to be true cases of clinical keratoconus.

Although it is reassuring that videokeratography is a sensitive measure of the keratoconus pattern, "keratoconus suspects" (KCS) have become problematic in preoperative screening. We define keratoconus suspects as corneas with a videokeratographic pattern of localized steepening but none of the traditional clinical signs of keratoconus, nor any other circumstances that might explain the topography pattern (e.g., trauma or contact lens wear). Some refractive procedures, such as radial keratotomy, are contraindicated when keratoconus is present. Despite the lack of knowledge about keratoconus suspects, surgical procedures are being performed on these corneas; however, long-term follow-up is needed to confirm the acceptability of the results.

Keratoconus suspects may be considered to be preclinical patterns of keratoconus when the disease is clinically diagnosed in the fellow eye. Nevertheless, fellow eye observations are not reliable for ruling out keratoconus in the suspect eye, because the suspect eye may show the first indication of the disease in the patient. Thus, a screening test should not rely on fellow eye data. A second indicator that a keratoconus suspect may progress to clinical keratoconus is a familial history of the disease, however, details concerning the mode of inheritance and the degree of penetrance are still being evaluated. Finally, follow-up videokeratographic examinations that show a continuous progression in the direction of increasing cone development may suggest keratoconus, but this effect is not considered diagnostic without the emergence of other clinical signs.

In recent years, significant advances have been made in characterizing corneal topography with quantitative corneal indexes. Topographic analysis with these indexes has been refined and automated through the use of multivariate and discriminant analysis, expert systems, and neural networks to generate keratoconus screening algorithms. A particular advantage of the neural network approach is its ability to respond using the same terminology a clinician prefers, rather than requiring the clinician to learn the nuances of yet another computer-generated index.

The current study reports the development of a pair of neural networks. One network detects and classifies clinical keratoconus and keratoconus suspects from among a variety of potentially confounding topographic patterns. A second network quantifies the severity of any conelike feature that matches the topographic pattern of clinical keratoconus or keratoconus suspects. The results are compared with currently available topographic keratoconus detection methods.

METHODS

Three hundred TMS-1 examinations (Tomey USA, Cambridge, MA) were collected from medical records at the LSU Eye Center. Research followed the tenets of the Declaration of Helsinki, informed consent was obtained from patients, and videokeratography has institutional review board approval. The examinations were critically reviewed for quality of the topographic image (focus and alignment) and whether the clinical interpretation was specific for a single corneal condition. Diagnostic classification for all maps was obtained from medical records (except for keratoconus suspects, which were classified by the authors based on topographic appearance after review of the medical records). The examinations were evenly and randomly...
divided by number and category into a training set of 150 examinations and a test set of 150 examinations. As indicated in Table 1, the examinations included nine different corneal categories: normal (NRM; \(<1.5\) D cylinder), with-the-rule astigmatism (WTR; \(\geq1.5\) D cylinder), keratoconus, keratoconus suspects, contact lens–induced corneal warpage (CLW), pellucid marginal degeneration (PMD), photorefractive keratectomy (PRK), radial keratotomy (RK), and keratoplasty (PKP).

The keratoconus examinations were subdivided by videokeratography using a semiquantitative method: clinical KC maps were graded as mild (KC1), moderate (KC2), or advanced (KC3). Specifically, a map contour power threshold at or below 55 D was used to differentiate mild keratoconus from moderate and advanced cases. Advanced and moderate keratoconus cases then were differentiated subjectively by the presence of disrupted contour steps in the location of cone apex in advanced cases and undisrupted contour steps in the maps of moderate cases. The authors independently agreed on the final subclassification of keratoconus. This grading system was somewhat arbitrary, but it was necessary to provide severity information to train the neural network.

The keratoconus suspect group was selected based on the absence of any clinical signs, symptoms, or medical history that might point to the presence of keratoconus or any other condition (e.g., contact lens wear) in that eye, even though the videokeratographic appearance bore a striking resemblance to that of an eye with a minor conelike ectasia. Keratoplasty, pellucid marginal degeneration, contact lens–induced corneal warpage, and keratoconus suspect corneas may have topographies that resemble true keratoconus; some normal, astigmatic, contact lens–induced corneal warpage, mild keratoconus, and decentered radial keratotomy and photorefractive keratectomy corneas may have topographies that resemble keratoconus suspects.

As shown in Table 1, 10 corneal indexes were extracted from each TMS-1 examination for use as neural network inputs: differential sector index (DSI), opposite sector index (OSI), center-surround index (CSI), analyzed area (AA), cylinder (CYL), irregular astigmatism index (IAI), the steep axis of simulated keratometry (SK1), surface regularity index (SRI), surface asymmetry index (SAI), and the standard deviation of corneal power (SDP). These indexes have been described in the literature. The complete set of indexes for each TMS-1 examination comprised a single data record for use with the training or test sets.

Two completely independent networks were designed for this study using Brainmaker Professional (version 3.1; California Scientific Software, Nevada City, CA). The first network detected and classified the presence of either clinical keratoconus or keratoconus suspects apart from all other confounding corneal topographies; the three output categories were symbolic (KC, KCS, and OTHER). Brainmaker internally converted these symbolic representations to numeric values and subsequently normalized the outputs to a range between 0 (negative response) and 1 (positive response) for each category. A second neural network was trained to grade the severity of a cone pattern specific to the keratoconus and keratoconus suspect examinations used in the training set. The output was numeric, with the desired target values specified as 0

### Table 1. Mean (±SD) of Test Set Indexes Used as Neural Network Inputs

<table>
<thead>
<tr>
<th>Category (N)</th>
<th>DSI</th>
<th>OSI</th>
<th>CSI</th>
<th>SK1</th>
<th>CYL</th>
<th>AA</th>
<th>IAI</th>
<th>SRI</th>
<th>SAI</th>
<th>SDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRM (14)</td>
<td>1.729± 0.707± 0.250± 43.900± 0.600± 82.300± 0.381± 0.254± 0.262± 0.814±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTR (13)</td>
<td>0.585± 0.377± 0.183± 1.119± 0.242± 3.041± 0.057± 0.154± 0.094± 0.218±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KC (6)</td>
<td>3.800± 2.850± 0.100± 45.400± 1.280± 83.000± 0.400± 0.478± 0.597± 1.383±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCS (11)</td>
<td>0.949± 1.302± 0.486± 2.349± 0.496± 7.725± 0.018± 0.149± 0.214± 0.256±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCI (11)</td>
<td>6.897± 5.627± 0.900± 47.800± 1.591± 80.600± 0.401± 0.950± 1.611± 2.791±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCS (13)</td>
<td>2.153± 2.654± 0.887± 1.926± 0.969± 6.156± 0.074± 0.414± 0.583± 0.602±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCS (9)</td>
<td>12.911± 8.033± 1.967± 59.300± 9.333± 50.400± 0.816± 2.454± 4.950± 6.278±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKP (13)</td>
<td>5.556± 5.316± 2.966± 4.795± 4.319± 15.710± 0.094± 0.324± 2.274± 1.826±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKP (13)</td>
<td>8.318± 6.900± 0.878± 47.100± 6.130± 80.300± 0.464± 1.231± 1.221± 3.171±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRK (22)</td>
<td>4.296± 3.789± 0.907± 4.626± 3.278± 10.970± 0.078± 0.441± 0.616± 1.383±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RK (17)</td>
<td>9.592± 3.051± 0.450± 49.600± 9.138± 43.800± 0.680± 1.620± 1.785± 3.808±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RK (17)</td>
<td>8.152± 2.117± 1.431± 3.608± 3.682± 14.377± 0.131± 0.450± 1.030± 1.220±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RK (17)</td>
<td>1.453± 1.346± 0.961± 2.088± 0.947± 8.251± 0.115± 0.300± 0.463± 0.988±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLW (20)</td>
<td>1.724± 1.090± 1.207± 39.800± 0.932± 86.600± 0.418± 0.566± 0.409± 1.551±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLW (20)</td>
<td>0.632± 0.847± 1.027± 2.091± 0.462± 9.420± 0.064± 0.197± 0.222± 0.889±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLW (20)</td>
<td>2.860± 1.080± 0.030± 45.400± 1.240± 78.000± 0.400± 0.720± 0.650± 1.050±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLW (20)</td>
<td>0.804± 0.578± 0.411± 1.990± 1.030± 6.460± 0.081± 0.339± 0.445± 0.266±</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Both networks were trained separately to an error tolerance of 0.1, such that all records in the training set were correctly categorized with respect to the intended output target pattern with that level of error or less. A single hidden neuron layer was used. To find the ideal number of hidden neurons that minimized error, 71 separate experiments were run with each network in which the hidden neuron number varied from 5 to 75. Training was allowed to progress for an indefinite number of runs until the error tolerance was achieved. After training was completed for each hidden layer modification, the average error using the training set was determined. Average error varied with hidden neuron number because for some architectures, more records produced less error than the training tolerance value. Using lowest average error as the optimal architecture criterion, we determined that 55 hidden neurons were needed for the classification network and 34 hidden neurons for the cone severity network.

The Rabinowitz test (K & I-S values), the average power of the steep and flat axes of simulated keratometry, the keratoconus prediction index (KPI), and the Klyce/Maeda keratoconus index (KCI), were obtained for the independent test set using the TMS-2 for Windows videokeratography system (software version W1.2) for comparison with the results of the neural network detection scheme. The Rabinowitz test uses a central K power threshold of 48.7 D and an I-S threshold of 1.9 to detect clinical keratoconus-like patterns, and K values of 47.2 to 48.7 or I-S values of 1.4 to 1.9 as the range for detecting keratoconus suspect-like patterns. 36 These TMS-2 criteria differ from those used in a previous study comparing keratoconus detection methods with the TMS-1. 37 For simulated keratometry, the powers of the steep and flat axes were averaged for all examinations in the test set. The mean of the averaged simulated keratometry of the normal test set group was determined to be 43.53 ± 1.02 standard deviations; that value plus two standard deviations (45.57 D) was used as the criterion for keratoconus detection. There has been no criterion established in the literature by which simulated keratometry alone can be used to screen for keratoconus suspects and therefore should report 0% for these examinations.

RESULTS

The test set was passed through the best-trained networks and the results were analyzed. The classification network was 100% accurate, sensitive, and specific using winner-take-all scoring at a 0.5 output response for keratoconus suspects and therefore should report 0% for these examinations.
level. The output response was strong, with 144 of 150 maps having a value of 0.95 or above. The lowest correct output was 0.69 in the KC category. Figure 1 shows the achieved output response for all cases in the test set.

The severity grading network, which operated independently of any result from the classification network, had output responses that were significantly different from each other for OTHER, KCS, and each of the three KC categories ($P < 0.001$; Fig. 2). The achieved mean severity outputs (± standard deviation) were OTHER = 0.02 ± 0.02, KCS = 0.21 ± 0.05, KC1 = 0.52 ± 0.17, KC2 = 0.74 ± 0.12, and KC3 = 0.91 ± 0.15. When the severity network response range was examined in terms of the classification network's three output categories, the severity range was found to be OTHER = 0 to 0.11, KCS = 0.15 to 0.29, and KC = 0.51 to 1.0.

When the test set output of the severity network was compared to KPI, there was a statistically significant correlation between the two measures ($R = 0.892; P < 0.0001$; Fig. 3A). Whereas the severity network made full use of the scale from 0 to 1, the KPI measure only used half this range for this test data; KPI was never normalized to a range between 0 and 1 when it was initially derived by discriminant analysis. KPI could not distinguish the KCS group from the OTHER category, nor could KPI distinguish a significant difference between moderate (KC2) and advanced (KC3) keratoconus. In contrast, the severity network differentiated among all the subcategories of KC, KCS, and OTHER ($P < 0.001$). Plotting of the raw data shows, however, that KPI and the severity network have overlapping distributions across severity grades (Fig. 3B); this effect may be partly the result of limitations in our subjective grading of KC maps into three subcategories.

Although KPI is available to the clinician as a corneal index, there are three additional methods by which a clinician can attempt to detect keratoconus using the TMS-2: the Rabinowitz test (modified K & I-S values), the Klyce/Maeda test (KCI), and the average of the steep and flat simulated keratometry values. The test set was analyzed using each of these methods and compared with the results from the classification network (Table 2). Sensitivity was not significantly different for the classification network, the Rabinowitz test, and the Klyce/Maeda test; however, all of these tests were significantly more sensitive than the average simulated keratometry method (Table 3). The classification network was significantly more specific than any of the other tests ($P = 0.005$ or less; Table 4), and both the Klyce/Maeda and Rabinowitz tests were significantly more specific than average simulated keratometry. In terms of overall accuracy, the classification network outperformed all other tests ($P = 0.002$ or less; Table 5). The Klyce/Maeda test performed significantly better than the Rabinowitz test in terms of overall accuracy because of the poor performance of the Rabinowitz test on KCS, which the Klyce/Maeda test does not even attempt to classify. Surprisingly, there was no significant difference between the Rabinowitz test and the average simulated keratometry in terms of overall accuracy, owing to the relatively poor
performance of the Rabinowitz keratoconus suspect criterion.

**DISCUSSION**

The classification network had no incorrect responses to any record in the test set. A few records had output responses lower than 0.9 (target = 1.0), but the responses were still well above the test threshold of 0.5 (see Fig. 1). It is important to understand why 0.5 rather than the training threshold of 0.1 was used in the classification network scoring. First, computing the testing performance at a 0.1 scoring threshold would give a distorted summary of what the network had actually been able to learn from the training set; the testing standard cannot be held to the same standard used for training. Therefore, we compensated by giving the network credit for being partially certain of the correct response. Second, use of the 0.5 value ensured that there were no null answers, but only correct or incorrect responses. Finally, we made the neural network respond as if the classification decision were similar to a forced-choice test in which the desired answer was that the topographic pattern was either present or not present for each of the categories; this can best be achieved using the 0.5 criterion.

The severity network also performed well, producing mean values for each category that were not significantly different from the intended target values \( P < 0.001 \); see Fig. 2). There was a slight roll-off in the response near a value of 1.0 (the KC3 target value) because the network outputs cannot distribute themselves normally about a target value that is also the output limit. In retrospect, it would have been more appropriate to choose a target value less than 1.0 for KC3. If desired, the severity network can also be used to classify keratoconus; in this case, cutoff values determined from the study are OTHER < 0.15 \( \leq \) KCS < 0.3 \( \leq \) KC.

The two cases of KC missed by the KCI test (i.e., false-negative cases) were atypical in that they had KPI values high enough to be classified initially as keratoconus, but the expert system rules changed the outcome to nonkeratoconus. One case had a low Sim K2 value of 37.9 D; the other case had a low CSI value of -2.61. These two exceptional cases should not be construed as sufficient reason to redefine the KPI discriminant threshold.

Despite the superficial topographic similarity between keratoconus and pellucid marginal degeneration cases, the severity and classification neural networks correctly detected differences in their respective patterns of the corneal indexes. In comparison, KPI tended to be too high for a few of the pellucid marginal degeneration cases, indicating the likelihood of keratoconus (see Fig. 3B). Furthermore, four of the eight false-positive cases for keratoconus that occurred with the KCI test were pellucid marginal degeneration.

**TABLE 2. Validation of Videokeratographic Detection of Keratoconus (KC) or Keratoconus Suspects (KCS) Using the Test Set**

<table>
<thead>
<tr>
<th>Method</th>
<th>True Positives (sensitivity)*</th>
<th>False Positives (specificity)†</th>
<th>Total Correct (accuracy)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KC</td>
<td>KCS</td>
<td>KC</td>
</tr>
<tr>
<td>Total correct</td>
<td>33/33 (100)</td>
<td>6/6 (100)</td>
<td>0/117 (100)</td>
</tr>
<tr>
<td>Classification net</td>
<td>TMS-2 modified IS</td>
<td>TMS-2 KCI§</td>
<td>Simulated keratometry§</td>
</tr>
<tr>
<td></td>
<td>33/33 (100)</td>
<td>1/6 (16.67)</td>
<td>31/33 (93.94)</td>
</tr>
<tr>
<td></td>
<td>13/117 (88.89)</td>
<td>12/144 (91.67)</td>
<td>23/117 (80.34)</td>
</tr>
<tr>
<td></td>
<td>0/144 (100)</td>
<td>121/150 (80.67)</td>
<td>140/150 (93.33)</td>
</tr>
<tr>
<td></td>
<td>150/150 (100)</td>
<td>121/150 (80.67)</td>
<td>140/150 (93.33)</td>
</tr>
</tbody>
</table>

* Values are number (percent sensitivity) of correct classifications/total number of keratoconus or keratoconus suspect cases.
† Values are number of false-positive classifications/total number of nonkeratoconus or nonkeratoconus suspect classifications; where \( 1 - \text{false-positive ratio} \times 100 = \text{specificity.} \)
‡ Includes KCS detection where applicable.
§ This method does not specify a criterion for keratoconus suspect detection.

**TABLE 3. Significance of McNemar’s Test Comparing Keratoconus Screening Sensitivity: \( P \) Values**

<table>
<thead>
<tr>
<th></th>
<th>Rabinowitz K and I-S</th>
<th>Klyce/Maeda KCI</th>
<th>Average Sim K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification network</td>
<td>NS</td>
<td>NS</td>
<td>0.005</td>
</tr>
<tr>
<td>Rabinowitz K and I-S</td>
<td>—</td>
<td>NS</td>
<td>0.019</td>
</tr>
<tr>
<td>Klyce/Maeda KCI</td>
<td>—</td>
<td>—</td>
<td>0.023</td>
</tr>
</tbody>
</table>

NS = not significant.

**TABLE 4. Significance of McNemar’s Test Comparing Keratoconus Screening Specificity: \( P \) Values**

<table>
<thead>
<tr>
<th></th>
<th>Rabinowitz K and I-S</th>
<th>Klyce/Maeda KCI</th>
<th>Average Sim K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification network</td>
<td>&lt;0.001</td>
<td>0.005</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Rabinowitz K and I-S</td>
<td>—</td>
<td>NS</td>
<td>0.037</td>
</tr>
<tr>
<td>Klyce/Maeda KCI</td>
<td>—</td>
<td>—</td>
<td>0.001</td>
</tr>
</tbody>
</table>

NS = not significant.
cases (see Table 2). KPI and KCI were not designed specifically to discriminate pellucid marginal degeneration from keratoconus; therefore, these types of false-positive results are to be expected with KCI and KPI.

Only the neural networks and the Rabinowitz method attempt to detect keratoconus suspects using interpretation of videokeratography. The Rabinowitz test incorrectly detected 12 of 144 corneas from the OTHER category as keratoconus suspects (91.67% specificity), but curiously failed to detect five out of six keratoconus suspect corneas correctly (16.6% sensitivity). The cases that were false positives for KCS included contact lens–induced corneal warpage, pellucid marginal degeneration, photorefractive keratectomy, radial keratotomy, and normal corneas. In comparison, the classification network had 100% keratoconus suspect sensitivity and specificity. Whereas the Rabinowitz test appeared to assume that keratoconus suspects were merely a lesser form of keratoconus (in terms of K and I-S value magnitudes), the neural network method established a unique category for keratoconus suspects and then trained the network to detect keratoconus suspects using a large set of corneal indexes.

Although KCI does not provide a keratoconus suspect output (keratoconus suspects should give a KCI value of 0%), four of the eight false-positive cases of KC identified by KCI were keratoconus suspects. This proved to be detrimental in that it lowered the specificity of the KCI test to true keratoconus. KCI outputs for KCS cases were 20% to 36% likelihood of keratoconus, a low but nonetheless affirmative response for a pattern consistent with true keratoconus. This result was in contrast to the mean KPI value for the KCS category, which was found not to be significantly different from the mean value of the OTHER category (see Fig. 3A). Apparently the expert rules of the KCI test tend to override the KPI result that correctly suggests that a keratoconus suspect is something other than true keratoconus.

Retrospective analysis of the significance and mean sensitivity of the neural network inputs was performed using Brainmaker software tools. Analysis revealed that for the classification network, SAI, CSI, and OSI were ranked 1, 2, and 3 in relative importance for classifying clinical keratoconus. These same indexes had the opposite ranking (3, 2, 1) for classifying keratoconus suspects (Fig. 4A). The network input for the steep axis of simulated keratometry (SKI) also had a positive influence in classifying keratoconus and keratoconus suspects, but to a lesser degree. The classification for OTHER showed a strong negative influence for SAI, CSI, OSI, and SKI as expected. It was surprising to find that DSI had only a minor influence on keratoconus suspect classification and no influence on keratoconus classification, because this index was designed to be sensitive to meridionally oriented localized steepening, which is typical for keratoconus. We believe the lack of influence can be explained by the lack of a significant difference in DSI among keratoconus, keratoconus suspects, pellucid marginal degeneration, and keratoplasty cases (see Table 1). Cylinder (CYL) had a negative influence in classifying keratoconus and keratoconus suspects because many corneas in the OTHER category also exhibited regular astigmatism and because there was a fair amount of variance in the CYL value within each category (see Table 1).

For the severity network, the indexes that had a positive effect were, in decreasing order of importance, OSI, CSI, SKI, SAI, IAL, and SDP (Fig. 4B). Surface regularity index, DSI, and CYL had a negative impact on the network’s ability to grade corneas, whereas AA had no influence on the grading decision. As with the classification network, the lack of influence by DSI was unexpected.

The sensitivity of the Rabinowitz method to keratoconus increased from 96% in a previous study to 100% in this study. This change was attributed to a lowering in the K & I-S threshold settings in the TMS-2 software. Surprisingly, the increased sensitivity did not have a detrimental effect on specificity, which also increased from 85% to approximately 89%. This value was still lower than desirable and resulted in many false-positives. Six of the false-positives were obvious postsurgical corneas, but the others were five pellucid marginal degeneration cases, one case of contact lens–induced warpage, and one case of keratoconus suspect. Even for experienced clinicians, these categories can be difficult to discriminate from true keratoconus using videokeratography alone. One might argue that the five pellucid marginal degeneration cases mistaken for keratoconus indicates that the Rabinowitz test is a generalized ectasia detector, but seven additional pellucid marginal degeneration cases were not detected as keratoconus.

The sensitivity of KCI was reduced from 98% in the earlier study to approximately 94% in this study. This was because only one additional keratoconus map was not correctly detected and the keratoconus suspect group had been slightly reduced in overall number. KCI specificity also was reduced from 99% to approximately 98%, primarily because KCI detected four keratoconus suspect maps as being true keratoconus. Simulated keratometry was the least ef-

### Table 5. Significance of McNemar’s Test Comparing Keratoconus Screening Accuracy*: P Values

<table>
<thead>
<tr>
<th>Test Type</th>
<th>K and I-S</th>
<th>KCI</th>
<th>Average Sim K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rabinowitz K and I-S</td>
<td>0.002</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Klyce/Maeda KCI</td>
<td>—</td>
<td>0.006</td>
<td>NS</td>
</tr>
<tr>
<td>Classification network</td>
<td>—</td>
<td>—</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* Includes KCS detection where applicable.

NS = not significant.

---

**Accuracy** of keratoconus (in terms of K and I-S value magnitude of 0%), four of the eight false-positive cases of KC ranked 1, 2, and 3 in relative importance for classifying keratoconus suspects and dien trained die network to detect keratoconus suspects using a large set of corneal indexes. These same indexes had the opposite ranking (3, 2, 1) for classifying keratoconus suspects (Fig. 4A). The network input for the steep axis of simulated keratometry (SKI) also had a positive influence in classifying keratoconus and keratoconus suspects, but to a lesser degree. The classification for OTHER showed a strong negative influence for SAI, CSI, OSI, and SKI as expected. It was surprising to find that DSI had only a minor influence on keratoconus suspect classification and no influence on keratoconus classification, because this index was designed to be sensitive to meridionally oriented localized steepening, which is typical for keratoconus. We believe the lack of influence can be explained by the lack of a significant difference in DSI among keratoconus, keratoconus suspects, pellucid marginal degeneration, and keratoplasty cases (see Table 1). Cylinder (CYL) had a negative influence in classifying keratoconus and keratoconus suspects because many corneas in the OTHER category also exhibited regular astigmatism and because there was a fair amount of variance in the CYL value within each category (see Table 1).

For the severity network, the indexes that had a positive effect were, in decreasing order of importance, OSI, CSI, SKI, SAI, IAL, and SDP (Fig. 4B). Surface regularity index, DSI, and CYL had a negative impact on the network’s ability to grade corneas, whereas AA had no influence on the grading decision. As with the classification network, the lack of influence by DSI was unexpected.

The sensitivity of the Rabinowitz method to keratoconus increased from 96% in a previous study to 100% in this study. This change was attributed to a lowering in the K & I-S threshold settings in the TMS-2 software. Surprisingly, the increased sensitivity did not have a detrimental effect on specificity, which also increased from 85% to approximately 89%. This value was still lower than desirable and resulted in many false-positives. Six of the false-positives were obvious postsurgical corneas, but the others were five pellucid marginal degeneration cases, one case of contact lens–induced warpage, and one case of keratoconus suspect. Even for experienced clinicians, these categories can be difficult to discriminate from true keratoconus using videokeratography alone. One might argue that the five pellucid marginal degeneration cases mistaken for keratoconus indicates that the Rabinowitz test is a generalized ectasia detector, but seven additional pellucid marginal degeneration cases were not detected as keratoconus.

The sensitivity of KCI was reduced from 98% in the earlier study to approximately 94% in this study. This was because only one additional keratoconus map was not correctly detected and the keratoconus suspect group had been slightly reduced in overall number. KCI specificity also was reduced from 99% to approximately 98%, primarily because KCI detected four keratoconus suspect maps as being true keratoconus. Simulated keratometry was the least ef-
Neural Networks for Keratoconus Detection

**FIGURE 4.** Mean sensitivity to input indexes. (A) Classification network. For each of the three basic categories (KC, KCS, and OTHER), positive bars indicate indexes that contribute to the classification task in a positive manner and negative bars indicate indexes that make classification more difficult. Indexes near 0, such as the surface regularity index (SRI) and the analyzed area (AA), make little if any contribution to classification. (B) Cone severity network.

Effective detection method for keratoconus, confirming the results of the earlier study. In summary, the overall accuracy of the classification network was significantly better than that of the other methods examined, when they were used at their intended clinical thresholds for keratoconus or keratoconus suspect detection (see Tables 2, 3, 4, 5, and 6). However, the Klyce/Maeda (KCI) test and the Rabinowitz (K & I-S) test were not significantly different from the classification network in terms of...
keratoconus susceptibility. Simulated keratometry was significantly worse in terms of sensitivity than any of the other tests. The classification network was significantly better than the Klyce/Maeda, Rabinowitz, and simulated keratometry tests in terms of keratoconus specificity. There was no significant difference between the Klyce/Maeda and Rabinowitz tests for keratoconus specificity. For keratoconus suspects, the classification network was significantly more sensitive and specific than the Rabinowitz test.

In conclusion, we developed two neural networks that answer two clinical screening questions: "Is keratoconus present?" and "How severe is the keratoconus?" The classification network generally outperforms currently available methods on the TMS-2 videokeratoscope in terms of overall accuracy, but in terms of sensitivity to keratoconus, the KCI and K & IS indexes are as good as the network approach. The neural network method returns fewer false positives for keratoconus among confounding corneal topographies, which is important for a truly automated detection system.

The network method also accurately classifies keratoconus suspect and was found to be significantly more sensitive and specific for these cases than the Rabinowitz method on the TMS-2. We also sought to establish an objective, quantitative definition for keratoconus suspect, and the severity network established statistically validated numeric thresholds that may be useful to clinically monitor and document cone development over time.

There is as yet no gold standard for the clinical diagnosis of keratoconus, although various clinical signs are combined to arrive at a diagnosis. The traditionally accepted clinical signs of keratoconus may vary in their occurrence and outward appearance among patients. Their interpretation is, for the most part, entirely subjective, although if and when the disease progresses, a diagnosis of keratoconus becomes easier and is more likely to be made. Clearly, videokeratography and neural networks can be important tools to supplement the traditional clinical signs of keratoconus. With further clinical testing and independent validation, quantitative videokeratography analysis may become more acceptable as a possible diagnostic test for keratoconus.

### Key Words

- Artificial intelligence
- Keratoconus
- Keratoconus suspect
- Neural network
- Videokeratography

### References

Neural Networks for Keratoconus Detection


