Evaluating the Performance of Various Machine Learning Algorithms to Detect Subclinical Keratoconus

Ke Cao, Karin Verspoor, Srujana Sahebjada, and Paul N. Baird

1 Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia
2 Department of Surgery, Ophthalmology, The University of Melbourne, Melbourne, Victoria, Australia
3 Department of Computing and Information Systems, The University of Melbourne, Melbourne, Victoria, Australia

Correspondence: Paul N. Baird, NHMRC Senior Research Fellow, Department of Surgery, Ophthalmology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Royal Victorian Eye and Ear Hospital, 32 Gisborne Street, East Melbourne, Victoria 3000, Australia. e-mail: pbaird@unimelb.edu.au

Introduction

Keratoconus (KC) is a common corneal condition characterized by progressive corneal thinning that results in corneal protrusion, reduced vision, and potential blindness. Prevalence of KC ranges from 0.17 in 1000 in the United States to 47.89 in 1000 in Saudi Arabia. The reported prevalence appears to have increased rapidly wherein only 1:2000 cases (United States) were reported in 1986 but as many as 1:375 (Netherlands) in 2016; although this may reflect improvements in imaging. Three articles reported KC prevalence in Iran, showing an increase from 1:126 in 2013 to 1:42 in 2014, to 1:32 in 2018. Similarly, the prevalence of KC in Israel increased from 1:43 in 2011 to 1:31 in 2014. A recent meta-analysis that analyzed more than 7 million participants from 15 countries reported the prevalence of KC as 1 in 725.

The onset of the disease is usually in the teens to early adulthood and our recent findings indicate that quality of life of KC patients is substantially lower than that of patients with later onset eye diseases such as age-related macular degeneration or diabetic retinopathy. This highlights the significant long-term morbidity associated with the condition. Management for KC follows an orderly transition from glasses/contact lenses to corneal transplantation as the condition becomes more severe.

Purpose: Keratoconus (KC) represents one of the leading causes of corneal transplantation worldwide. Detecting subclinical KC would lead to better management to avoid the need for corneal grafts, but the condition is clinically challenging to diagnose. We wished to compare eight commonly used machine learning algorithms using a range of parameter combinations by applying them to our KC dataset and build models to better differentiate subclinical KC from non-KC eyes.

Methods: Oculus Pentacam was used to obtain corneal parameters on 49 subclinical KC and 39 control eyes, along with clinical and demographic parameters. Eight machine learning methods were applied to build models to differentiate subclinical KC from control eyes. Dominant algorithms were trained with all combinations of the considered parameters to select important parameter combinations. The performance of each model was evaluated and compared.

Results: Using a total of eleven parameters, random forest, support vector machine and k-nearest neighbors had better performance in detecting subclinical KC. The highest area under the curve of 0.97 for detecting subclinical KC was achieved using five parameters by the random forest method. The highest sensitivity (0.94) and specificity (0.90) were obtained by the support vector machine and the k-nearest neighbor model, respectively.

Conclusions: This study showed machine learning algorithms can be applied to identify subclinical KC using a minimal parameter set that are routinely collected during clinical eye examination.

Translational Relevance: Machine learning algorithms can be built using routinely collected clinical parameters that will assist in the objective detection of subclinical KC.
progresses from mild/moderate to severe stages respectively. KC is the most common indication for corneal transplantation globally, and accounts for ~30% of corneal grafts (Australian Corneal Graft registry). Collagen crosslinking treatment, which stiffens the cornea, has been available as a treatment to slow KC progression for several years. However, collagen crosslinking treatment using the standard Dresden protocol requires a minimum of 400-μm corneal thickness and is only suitable for patients in early (subclinical) stages of KC; early detection is therefore a prerequisite for this treatment. Once KC progresses, patients may require corneal transplantation. Detecting the subclinical stage of KC is clinically challenging because (1) subjects are asymptomatic; (2) do not produce detectable signs at routine clinical examination using slit lamp, retinoscopy, or keratometry; and (3) the advanced corneal topographic systems that can detect subclinical KC are not always available in all optometric/primary eye care practices. Thus, a number of challenges currently exist with regard to reliable detection of subclinical KC.

Machine learning models have been applied to detect KC at different clinical stages with a number of these presented as specific to a particular tomographic or topographic imaging system. The majority of these studies have used a single machine learning method such as regression analysis, a tree-based method, discriminant function analysis, support vector machine, or neural network. Parameters derived from a particular topographic or tomographic imaging system were collected in these studies, and established the machine learning models without selecting important parameter combinations. We therefore lack knowledge as to the performance characteristics of different machine learning methods to the same dataset and the evaluation of the same machine learning method to various parameter combinations. This would most readily be addressed by applying a number of machine learning algorithms to the same dataset and comparing their results.

Moreover, the indicated studies that have used machine learning in KC use a number of parameters related to corneal measures, they do not include other clinical measures such as axial length (AL), spherical equivalent or demographic parameters, which are reported to have an association with KC and may also have a role when establishing a clinically accepted detection model for KC.

We therefore wished to evaluate the performance of a range of different machine learning methods on a subset of subclinical dataset recruited in Australia as a part of the Australian Study of Keratoconus (ASK).

These machine learning methods were further systematically trained and tested to explore various combinations of commonly used corneal topographic parameters together with clinical and demographic parameters to identify a best performing machine learning model to detect subclinical KC from control eyes.

### Methods

#### Subjects

This is a substudy of ASK that was established to better understand the clinical, genetic environmental risk factors for KC. The study protocol was approved by the Royal Victorian Eye and Ear Hospital Human Research and Ethics Committee (Project #10/954H). The protocol followed the tenets of the Declaration of Helsinki and all privacy requirements were met.

Subclinical KC patients were recruited from public and private clinics at the Royal Victorian Eye and Ear Hospital and private consulting rooms and optometry clinics in Melbourne, Australia. All patients were provided with a patient information sheet, consent form, privacy statement, and patient rights. A comprehensive eye examination was undertaken for each patient and KC was diagnosed clinically.

Subclinical KC was defined as those eyes with abnormal corneal topography, including inferior-superior localized steepening or asymmetric bowtie pattern. These eyes had no detectable clinical signs on slit-lamp biomicroscopy and retinoscopy examination. Subjects with other ocular diseases such as corneal degenerations and dystrophies, macular disease, and optic nerve disease (e.g., optic neuritis, optic atrophy) were excluded from the study. KC subjects were recruited from ASK, whereas controls were recruited from the “GEnes in Myopia” study where a similar recruitment protocol was used. The control group consisted of refractive error subjects with no ocular disease that may affect refraction in the eyes including amblyopia (greater than a two-line Snellen difference between the eyes), strabismus, visually significant lens opacification, glaucoma, or any other corneal abnormality. Individuals with connective tissue disease such as Marfan’s or Stickler syndrome were also excluded from the study. The latter conditions were identified by the individual’s medical history obtained via a general questionnaire.

#### Eye Examination

The anterior segment was assessed using slit-lamp biomicroscopy examination and refraction was performed on each eye using a Nidek auto refractor.
AL was recorded for each participant using a noncontact partial coherence interferometry with an IOL Master optical biometer (Carl Zeiss, Oberkochen, Germany). The corneal topographic measurement parameters were obtained on all subjects using a Pentacam corneal tomographer (Oculus, Wetzlar, Germany). The subjects were required to remove their contact lenses, if worn, at least 24 hours before examination. The results used in the study were from the four-map selectable display of Pentacam results incorporating front and back elevation maps, along with front sagittal curve and pachymetry. These maps were chosen to highlight the inferior decentration of the corneal apex on both the front and back surface, which assisted in the detection of KC. Mean corneal curvature (Km) was calculated automatically by the device as the mean value of horizontal and vertical central radial curvatures in the 3-mm zone. The detailed methodology of eye examination for KC can be found elsewhere.47 Nine parameters that classically represent KC, including AL, SE, mean front corneal curvature (front Km), mean back corneal curvature (back Km), central corneal thickness (CCT), corneal thickness at the apex (CTA), corneal thickness at the thinnest point (CTT), anterior chamber depth (ACD), and corneal volume (CV), were included in this study to build the algorithms to detect subclinical KC.

Statistical Analysis

Data were analyzed with RStudio (version 1.1.456) for Windows. All statistical tests were considered significant when the P value was less than 0.05. A χ² test was used to compare gender between groups, and a Wilcoxon signed-rank test was applied to test the difference in age and other clinical characteristics, including SE, AL, front Km, back Km, CCT, CTA, CTT, ACD, and CV.

Machine Learning Methods

Eight machine learning algorithms, including random forest, decision tree, logistic regression, support vector machine, linear discriminant analysis, multilayer perceptron neural network, lasso regression, and k-nearest neighbor were applied to build classification models to differentiate subclinical KC from control subjects. Briefly,

(1) Simple regression methods (e.g., logistic regression) learn a mapping from input variables (X) to an output variable (Y) with $Y = f(X)$;

(2) Tree-based methods (e.g., decision tree)52 involve building a decision making tree with “if this then that” logic; ensemble methods (e.g., random forest)35 and combine several machine learning techniques into one model (i.e., random forest constructs multiple trees);

(3) k-nearest neighbor method53 makes a decision by searching through the whole dataset for the k most similar instances;

(4) Discriminant function analysis (e.g., linear discriminant analysis57) finds a combination of variables that will discriminate between the categories;

(5) Support vector machine38 translates data into another space where a plane (“hyperplane”) maximally separates disparate data groups from itself;

(6) Regularization methods (e.g., lasso regression54) add a penalty to optimize outcomes; and

(7) Neural networks (e.g., multilayer perceptron neural network) process across multiple layers of interconnected nodes, each computing a non-linear function of the sum of their inputs39.

We chose these eight methods for the following reasons:

(a) These algorithms are eight of the most commonly used machine learning algorithms that aid in health care diagnosis55–59;

(b) These are the most commonly used algorithms in the previously published KC studies18,21,23-26,28,29,32,33,40,41;

(c) They represent eight distinct learning functions31,35,37,54,60–62 in machine learning and thus it is worthwhile to empirically compare the performance of these algorithms for our dataset.

Comparison of Machine Learning Algorithms

The caret package63 (Classification And REgression Training) in R was used to perform all of the machine learning processes, including training and assessing the performance of models. The “train” function in the caret package was also applied for tuning the hyperparameters in each model. During each of the training, the train function generated a candidate set of parameter values, and the function picked the tuning parameters associated with the best accuracy.
(2) Each machine learning method was respectively trained to differentiate subclinical KC from control eyes using all 11 parameters;

(3) To validate the results of each model, a 10-fold cross-validation method was used on the full dataset, wherein the data was split into 10 subsets (folds), each representing 10% of the data. On each iteration, a model was trained using nine of these folds (90% of the data) and tested on the remaining fold, repeatedly, 10 times across the folds to assess the performance of the methods as the output. In this way, each fold serves as held-out test data for a model trained on the other nine folds, and the average performance across the 10 folds was measured. This represents a standard evaluation paradigm for small datasets. Figure 1 shows an example of 10-fold cross-validation.

Selection of Parameter Combinations

Models that achieved the highest performance in at least one of the evaluation metrics were used for subsequent analysis. The following steps were performed in a loop for each of the methods:

(1) Each combination of the considered parameters, ranging from two variables (e.g., age, gender) up to all 11 variables, a total of 2036 different combinations were considered.

(2) Each method was trained to differentiate subclinical KC eyes from control eyes using each combination of parameters described previously; Figure 2 represents a flowchart for training machine learning models with different parameter combinations.

(3) A 10-fold cross-validation method was used to validate the performance of each model.

Figure 1. The 10-fold cross validation for analysis of test data. Twenty rhombuses are randomly partitioned into 10 subsets, with two rhombuses in each subset. Of the 10 subsets, one subset is retained as the validation data, and the remaining nine subsets are used to train the model. This cross-validation process is then repeated 10 times. In summary, cross-validation combines measures of 10 fitness and provide an average.

Figure 2. Training machine learning models with different parameters sets. All possible combination of 11 parameters, from combination of two (e.g., gender and age, gender and SE) parameters to combination of 11 parameters, were used as input respectively to train machine learning models to differentiate subclinical KC eyes from control eyes.
Machine Learning to Detect Subclinical Keratoconus

Table 1. Demographic Data for all the Subjects Included in the Study

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Age (SD)</th>
<th>% Female</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subclinical KC</td>
<td>49</td>
<td>30.37 (12.53)</td>
<td>24.5</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Control</td>
<td>39</td>
<td>36.08 (11.91)</td>
<td>64.1</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

KC, keratoconus; SD, standard deviation.

Evaluation Criteria

Accuracy, sensitivity, specificity, area under the curve (AUC) and precision were the measures that were used to evaluate the performance of each model as these are typically used in health care settings. Accuracy determines the ability of the model to correctly classify the cases and controls, sensitivity represents the ability of the model to identify the cases from the case group (true positive), and specificity is the ability of the model to identify controls from the given control group (true negative) under investigation. AUC represents how much a model is capable of distinguishing between two groups. The higher the AUC, the better the model is at classifying cases as cases and controls as controls.

The criteria for evaluating the performance of each model was optimized from 0 to 1 with values greater than 0.90 defined as the highest performing model, followed by those ranging between 0.8 and 0.9 classified as good performers, between 0.5 and 0.79 as moderate performers, and finally those less than 0.5 considered as poor performing methods in the present study.

Results

Demographics

A total of 88 subjects consisting of 49 subclinical KC eyes, and 39 control eyes were available for analysis. There were significantly (P < 0.01) more males in the subclinical group (37 males, 75.5%) compared with the control group (14 males, 35.9%). The mean age of the subclinical KC group was 30.37 ± 12.53 years and the control group was 36.08 ± 11.91 years. The subclinical KC patients were significantly younger compared to the control group (P < 0.05). Demographic data for all subjects are presented in Table 1.

Considering individual parameters, there were significant differences between subclinical KC and control eyes for spherical equivalent (P < 0.01), AL (P < 0.01), front Km (P = 0.01), as well as corneal thickness-related parameters (CCT, P = 0.02; CTA, P = 0.01; CTT, P < 0.01). As expected, corneal thickness in the subclinical KC group was significantly thinner than those in the control group. However, subclinical KC eyes tended to have significantly flatter cornea when compared with control eyes (P = 0.01). Moreover, control eyes showed more myopic and longer AL than subclinical KC (P < 0.01). This may due to the control group being recruited from the GEnes in Myopia study. There was no significant difference in back Km (P = 0.60), ACD (P = 0.09), and CV (P = 0.27) between the groups (Table 2).

Comparison of Machine Learning Algorithms

Table 3 shows the performance of each of the eight machine learning methods that were used in

Table 2. Clinical Characteristics of all Eyes Used in the Analysis by Individual Parameter

<table>
<thead>
<tr>
<th></th>
<th>Subclinical KC Eyes</th>
<th>Control Eyes</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE, D (SD)</td>
<td>−2.20 (3.32)</td>
<td>−6.59 (4.87)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>AL, mm (SD)</td>
<td>24.44 (1.48)</td>
<td>26.62 (2.21)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ACD, mm (SD)</td>
<td>3.59 (0.60)</td>
<td>3.67 (0.43)</td>
<td>0.09</td>
</tr>
<tr>
<td>Front Km, D (SD)</td>
<td>42.45 (1.38)</td>
<td>43.22 (2.09)</td>
<td>0.01</td>
</tr>
<tr>
<td>Back Km, D (SD)</td>
<td>−6.03 (1.02)</td>
<td>−6.22 (0.34)</td>
<td>0.60</td>
</tr>
<tr>
<td>CCT, μm (SD)</td>
<td>511.20 (45.82)</td>
<td>531.74 (31.49)</td>
<td>0.02</td>
</tr>
<tr>
<td>CTA, μm (SD)</td>
<td>511.90 (46.60)</td>
<td>531.87 (31.07)</td>
<td>0.01</td>
</tr>
<tr>
<td>CTT, μm (SD)</td>
<td>487.67 (82.22)</td>
<td>528.97 (31.56)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>CV, mm³ (SD)</td>
<td>61.22 (21.13)</td>
<td>59.02 (4.81)</td>
<td>0.27</td>
</tr>
</tbody>
</table>

P value- values of Wilcoxon signed-rank test

ACD, anterior chamber depth; AL, axial length; back Km, mean back corneal curvature; CCT, central corneal thickness; CTA, corneal thickness at the apex; CTT, corneal thickness at the thinnest point; CV, corneal volume; front Km, mean front corneal curvature; KC, keratoconus; SD, standard deviation; SE, spherical equivalent.
Table 3. Comparison of the Eight Machine Learning Algorithms Using Different Performance Indicators

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random forest</td>
<td>0.87</td>
<td>0.88</td>
<td>0.85</td>
<td>0.96</td>
<td>0.89</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>0.86</td>
<td><strong>0.92</strong></td>
<td>0.78</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>K-nearest neighbors</td>
<td>0.73</td>
<td>0.61</td>
<td><strong>0.88</strong></td>
<td>0.73</td>
<td>0.88</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.81</td>
<td>0.84</td>
<td>0.77</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>Linear discriminant analysis</td>
<td>0.81</td>
<td>0.84</td>
<td>0.78</td>
<td>0.89</td>
<td>0.83</td>
</tr>
<tr>
<td>Lasso regression</td>
<td>0.84</td>
<td>0.86</td>
<td>0.83</td>
<td>0.91</td>
<td>0.88</td>
</tr>
<tr>
<td>Decision tree</td>
<td>0.80</td>
<td>0.82</td>
<td>0.78</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>Multilayer perceptron neural network</td>
<td>0.52</td>
<td>0.80</td>
<td>0.20</td>
<td>0.51</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The number in bold indicates the highest value obtained for each performance indicator.

this analysis. In our dataset, amongst all the methods, random forest method presented the highest performance for AUC (0.96) and had a good accuracy (0.87) and precision (0.89) while support vector machine had the highest sensitivity (0.92) and the k-nearest neighbor model was good for specificity (0.88). On the other hand, the multilayer perceptron neural network showed poor performance in our dataset with a specificity of 0.2 and a precision of 0.44. Other models had moderate to good performance ranging from 0.51 to 0.89.

Selection of Parameter Combinations

From the previous step, we confirmed that random forest, support vector machine and k-nearest neighbor methods fit better than the other methods by using the 11 input parameters for all machine learning classifiers.

We then tested all possible parameter combinations with these three methods. In our dataset, the following models had the best performances using a minimal parameter set:

(A) The greatest AUC was obtained using a minimal parameter set of gender, SE, front Km, CTT, and CV (AUC of 0.97, using the random forest method).
(B) The highest sensitivity was obtained with the parameter set of SE, ACD, back Km, CCT, and CTT (sensitivity of 0.94, using support vector machine method).
(C) The highest specificity was obtained with the parameter set of age, SE, AL, CTA, and CTT (specificity of 0.90, using k-nearest neighbor method).

Discussion

We compared the eight commonly used machine learning techniques and their performance in distinguishing subclinical KC eyes from control eyes using an Australian dataset. This is the first study to evaluate and compare the performance of such a wide range of machine learning techniques and present their efficacy in detecting subclinical KC. It is also the first time to develop algorithms with a great amount of parameter combinations to achieve the most parsimonious performing machine learning model to detect subclinical KC.

Machine learning algorithms are computational methods that allow us to efficiently navigate complex data to arrive at a best-fit model. The performance of different machine learning algorithms strongly depends on the nature of the data and the task being explored, and thus the correct choice of algorithm is best determined through experimentation.

In our dataset, using 11 parameters (age, gender, SE, AL, ACD, front Km, back Km, CCT, CTA, CTT, CV), the random forest model had a highest performance for AUC (0.96), which means clinically it has a good measure of differentiating subclinical KC from the control eyes. Conversely, multilayer perception neural network had an AUC near to 0.5, reflecting that this model has no discrimination capacity to distinguish subclinical KC and control eyes. The random forest model also achieved good performance for accuracy (0.87) in our dataset (i.e., clinically it can correctly classify 87% of subclinical KC and control eyes). Moreover, the precision of the random forest model is 0.89, translating to 89% of subclinical KC eyes classified by the random forest model are real subclinical KC eyes.

In addition, the support vector machine model reached 0.92 of sensitivity, showing a 92%
probability of correctly identifying subclinical KC eyes, and k-nearest neighbor had an 88% chance of correctly identifying control eyes (specificity of 0.88).

We further developed models using random forest, support vector machine, and k-nearest neighbor methods with different parameter combinations to distinguish subclinical KC and control eyes. Our results indicated that using a combination of gender, spherical equivalent, mean front corneal curvature, corneal thickness at the thinnest point, and corneal volume had a good measure of identifying subclinical KC from control eyes (AUC 0.97). In addition to this, a model developed using spherical equivalent, anterior chamber depth, mean back corneal curvature, central corneal thickness, and corneal thickness at the thinnest point had a sensitivity 0.89 (i.e., this model can 94% of the times correctly identify subclinical eyes). Finally, we could also develop a model using age, spherical equivalent, axial length, corneal thickness at the apex, and corneal thickness at the thinnest point, which had the highest specificity and a 90% chance to identify control eyes.

Therefore, our analysis attempted to optimize performance by testing multiple algorithms, comparing the results between algorithms and selecting the appropriate algorithm for clinical practice. Chan et al. recently reported the costs associated with the diagnosis and management of keratoconus represent a significant economic burden to the patient as well as the society.

Subclinical KC was defined as the normal fellow eye of unilateral KC. The current study avoided this limitation by defining subclinical eyes based on their own characteristics. Hence, data labeling was based on the clinical assessment of the eyes, which were then used to train the machine to mimic and build the algorithms that most closely represented the input dataset. Our models are based on a clinically meaningful dataset.

For the same dataset, different machine learning methods have different performance characteristics, which can be applied accordingly based on the clinical requirements. In the present study, we achieved the highest AUC, sensitivity and specificity using the random forest, support vector machine, and k-nearest neighbor. These results were comparable but had better performance in detecting subclinical KC eyes compared with other results in the literature.

Ruiz et al. used a support vector machine method to analyze 22 parameters derived from Pentacam. They found a sensitivity of 0.79 and specificity of 0.98 in discriminating “forme fruste” KC (N = 67) from normal eyes (N = 194). Kovács et al. used 15 unilateral KC and 30 normal KC subjects to construct a model using multilayer perceptron neural network and reported 0.90 sensitivity and specificity. Similarly, Ucakhan et al. used 44 KC and 63 non-KC subjects using logistic regression and reported a sensitivity of 0.77 and specificity of 0.92 to detect subclinical KC from control eyes.

Hwang et al. reported an accuracy of 100%, after training a logistic regression model based on 13 parameters combining measurements from Pentacam and OCT imaging. However, this study indicated that they trained the model with 90 eyes (30 subclinical KC and 60 normal) but did not clarify if the same dataset that was used both for training and testing of the model so it is possible that the same dataset was used in both training and model testing resulting in an artificially higher performance (known as overfitting). We have tested each of our models with subjects that were not included in the training dataset through the
**Table 4.** Details of Previously Published Studies Using Machine Learning Algorithms for the Detection of Subclinical Keratoconus

<table>
<thead>
<tr>
<th>Author and Year</th>
<th>Topography System</th>
<th>Sample Size</th>
<th>Algorithm Used</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kovács et al. 21 (2016)</td>
<td>Pentacam</td>
<td>15 cases(^*)/30 controls</td>
<td>Multilayer perceptron neural network</td>
<td>Sensitivity 0.90; Specificity 0.90</td>
</tr>
<tr>
<td>Ruiz et al. 23 (2016)</td>
<td>Pentacam HR</td>
<td>67 cases(^*)/339 controls</td>
<td>Support vector machine</td>
<td>Sensitivity 0.79; Specificity 0.98</td>
</tr>
<tr>
<td>Hwang et al. 30 (2018)</td>
<td>Pentacam HR and SD OCT</td>
<td>30 cases(^*)/60 controls</td>
<td>Multivariable logistic regression</td>
<td>Sensitivity 1.00; Specificity 1.00</td>
</tr>
<tr>
<td>Smadja et al. 25 (2013)</td>
<td>GALILEI</td>
<td>47 cases(^*)/177 controls</td>
<td>Decision tree</td>
<td>Sensitivity 0.94; Specificity 0.97</td>
</tr>
<tr>
<td>Accardo et al. 18 (2002)</td>
<td>EyeSys</td>
<td>30 cases(^*)/65 controls</td>
<td>Neural network</td>
<td>Sensitivity 1.00; Specificity 0.99</td>
</tr>
<tr>
<td>Saad et al. 24 (2010)</td>
<td>OrbscanIIz</td>
<td>40 cases(^*)/72 controls</td>
<td>Discriminant analysis</td>
<td>Sensitivity:0.93; Specificity:0.92</td>
</tr>
<tr>
<td>Ucakhan et al. 26 (2011)</td>
<td>Pentacam</td>
<td>44 cases(^*)/63 controls</td>
<td>Logistic regression</td>
<td>Sensitivity:0.77; Specificity:0.92</td>
</tr>
<tr>
<td>Ventura et al. 27 (2013)</td>
<td>Ocular Response Analyzer</td>
<td>68 cases(^\dagger)/136 controls</td>
<td>Neural network</td>
<td>AUC: 0.978</td>
</tr>
</tbody>
</table>

\(^*\)Subclinical KC was defined as normal fellow eye of unilateral KC.\(^{21,23-26,30}\)

\(^\dagger\)KC of mild and moderate severity was considered.\(^{18}\)

\(^\dagger\)Grade I and II KC according to the Krumeich severity classification.\(^{27}\)

10-fold cross-validation methodology, which allowed us to evaluate the performance of our model across different (simulated blind) test sets.

The study by Smadja et al.\(^{25}\) used 47 Forme Fruste KC and 177 normal eyes to show that the decision tree algorithm with six parameters from Galilei could achieve a sensitivity of 0.94 and specificity 0.97. Although this performance is somewhat better than the model presented in this study, they included a machine-specific index (e.g., asphericity asymmetry index, opposite sector index), which cannot be applied to other imaging systems. Similarly, Saad et al.\(^{24}\) used 40 Forme Fruste KC and 72 normal eyes to show that discriminant analysis resulted in a sensitivity of 93% and specificity of 92%. There were more than 50 parameters generated from the Orbscan IIz involved in their model, including calculated parameters that could not be repeated by other imaging systems. In contrast to these studies, we used routinely measured clinical parameters and common corneal topographic parameters such as corneal curvature, pachymetry and corneal volume that are not limited to a specific device, providing a real opportunity for our results to be translated and used in different imaging systems.

Several limitations of the current study should be noted. First, we have considered measurements derived only from a single topographic machine (Pentacam) in a single hospital. Further experimentation is required to test whether the models would be effective with data sourced from different machines. Second, the cross-validation strategy we used for evaluation was the most appropriate to allow for simulation of a held-out test data scenario, considering distinct training/test sets. However, this approach still draws the test data from the same underlying sample. Therefore, it would be reassuring to collect more data from our hospital and data from other clinics to allow for more rigorous testing of the generalization capacity and robustness of the best models in the face of patient variation.

**Conclusion**

The current study shows promising results toward detecting subclinical keratoconus from control eyes using parameters that can be collected in a routine clinical eye examination. Results from this study suggested the value of exploring a range of machine learning methods for the detection of subclinical keratoconus.
techniques for modeling the task, and the impact of including a broad range of clinical and demographic features related to keratoconus when developing such approaches and the usefulness of selecting important parameter combinations from a larger parameter set when building machine learning models. Further experimentation will lead to more objective and effective screening strategies for keratoconus and would be a helpful tool in clinical practice.

Acknowledgments

The authors thank the study participants and Eye Surgery Associates, Lindsay and Associates and Keratoconus Australia for assistance with subject recruitment.

This study was supported by the Australian National Health and Medical Research Council (NHMRC) project Ideas grant APP1187763 and Senior Research Fellowship (1138585 to PNB), the Louisa Jean De Bretteville Bequest Trust Account, University of Melbourne, the Angior Family Foundation, Keratoconus Australia, Perpetual Impact Philanthropy grant (SS) and a Lions Eye Foundation Fellowship (SS). The Centre for Eye Research Australia (CERA) receives Operational Infrastructure Support from the Victorian Government.

Disclosure: K. Cao, None; K. Verspoor, None; S. Sahebjada, None; P.N. Baird, None

*SS and PNB contributed equally as senior authors.

References

18. Accardo PA, Pensiero S. Neural network-based system for early keratoconus detection from...


