Special Issue

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

Ke Cao[1,2], Karin Verspoor[3], Srujana Sahebjada[1,2,*], and Paul N. Baird[2,*]

[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

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.

Introduction

Keratoconus (KC) is a common corneal condition characterized by progressive corneal thinning that results in corneal protrusion,[1] reduced vision and potential blindness. Prevalence of KC ranges from 0.17 in 1000 in the United States[2] to 47.89 in 1000 in Saudi Arabia.[3] The reported prevalence appears to have increased rapidly wherein only 1:2000 cases (United States) were reported in 1986[4] but as many as 1:375 (Netherlands) in 2016[5]; although this may reflect improvements in imaging.[6] Three articles reported KC prevalence in Iran, showing an increase from 1:126 in 2013,[7] to 1:42 in 2014[8], to 1:32 in 2018[9]. Similarly, the prevalence of KC in Israel increased from 1:43 in 2011[10] to 1:31 in 2014[11]. A recent meta-analysis that analyzed more than 7 million participants from 15 countries reported the prevalence of KC as 1 in 725[12].

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[13]. 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

progresses from mild/moderate to severe respectively. KC is the most common indication for corneal transplantation globally[14], and accounts for ~30% of corneal grafts (Australian Corneal Graft registry)[15]. Collagen crosslinking treatment, which stiffens the cornea, has been available as a treatment to slow KC progression for several years[16,17]. 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[18–27]. The majority of these studies have used a single machine learning method such as regression analysis[28-30], a tree-based method[25,31–33], ensemble method[34,35], discriminant function analysis[24,36,37], support vector machine[19,23,38], or neural network[20,21,39–42]. 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[18,19,22,40,43]. 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[44–46].

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[47–50]. 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[51] 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 χ^2 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 method[53] makes a decision by searching through the whole dataset for the k most similar instances;
(4) Discriminant function analysis (e.g., linear discriminant analysis[37]) finds a combination of variables that will discriminate between the categories;
(5) Support vector machine[38] translates data into another space where a plane ("hyperplane") maximally separates disparate data groups from itself;
(6) Regularization methods (e.g., lasso regression[54]) 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 inputs[39].

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 diagnosis[55–59];
(b) These are the most commonly used algorithms in the previously published KC studies[18,21,23-26,28,29,32,33,40,41];
(c) They represent eight distinct learning functions[31,35,37,54,60–62] in machine learning and thus it is worthwhile to empirically compare the performance of these algorithms for our dataset.

Machine Learning Analysis

The caret package[63] (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.

Comparison of Machine Learning Algorithms

The following steps were performed in a loop for each of the methods:

(1) Data for each eye, with the corresponding label indicating whether the eye was subclinical KC or not, was imported into RStudio package;
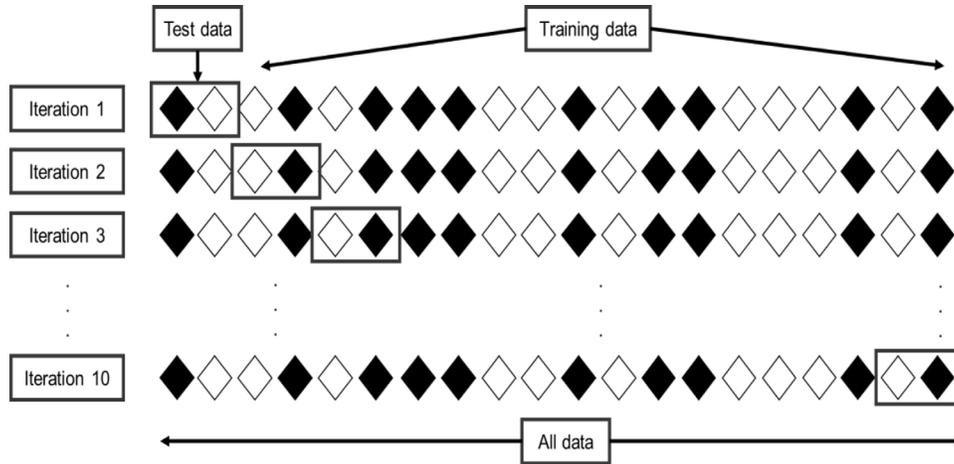
Figure 1. The 10-fold cross validation for analysis of test data. Twenty rhombuses are randomly partitioned into 10 subsets, with two rhombuses in each subset. Of the 10 subsets, one subset is retained as the validation data, and the remaining nine subsets are used to train the model. This cross-validation process is then repeated 10 times. In summary, cross-validation combines measures of 10 fitness and provide an average.

(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[63,64]. 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.

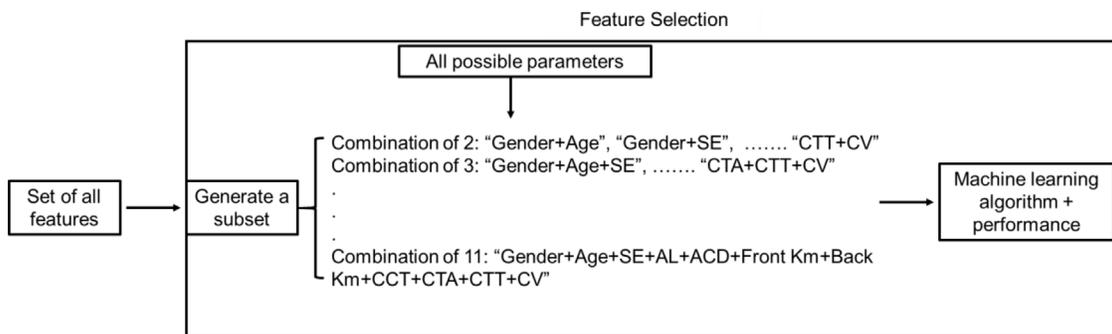(3) A 10-fold cross-validation method was used to validate the performance of each model.

Figure 2. Training machine learning models with different parameter 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.

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

	N	Mean Age (SD)	% Female	*P* Value
Subclinical KC	49	30.37 (12.53)	24.5	<0.01
Control	39	36.08 (11.91)	64.1	<0.01

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[65]. 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[66]. Precision defines how correctly the proportion of case identifications was achieved by the model[67].

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 fit, 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

	Subclinical KC Eyes	Control Eyes	*P* Value
SE, D (SD)	−2.20 (3.32)	−6.59 (4.87)	<0.01
AL, mm (SD)	24.44 (1.48)	26.62 (2.21)	<0.01
ACD, mm (SD)	3.59 (0.60)	3.67 (0.43)	0.09
Front Km, D (SD)	42.45 (1.38)	43.22 (2.09)	0.01
Back Km, D (SD)	−6.03 (1.02)	−6.22 (0.34)	0.60
CCT, μm (SD)	511.20 (45.82)	531.74 (31.49)	0.02
CTA, μm (SD)	511.90 (46.60)	531.87 (31.07)	0.01
CTT, μm (SD)	487.67 (82.22)	528.97 (31.56)	<0.01
CV, mm^3 (SD)	61.22 (21.13)	59.02 (4.81)	0.27

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

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

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[68]. 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[69].

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%

translational vision science & technology

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 0.94 (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[70]. The result from this study is a good start for providing a machine learning based model to assist clinicians to identify KC in early/subclinical form and reducing the economic burden of the condition.

The Pentacam imaging system that we used is a sensitive device for detecting subtle corneal curvature and pachymetry changes that have high reproducibility and repeatability[71]. For the purpose of better clinical interpretability, we analyzed only commonly available Pentacam corneal parameters but also included other routinely measured parameters that are of primary relevance in keratoconus detection to assess how they would alter the models. One of the main limitations of previous studies that have used machines learning techniques is that the models that were built were specific to the instrument that was used. However, the parameters available for each machine may vary. For example, in the study by Lopes et al.[22] used 18 parameters derived from the Pentacam in their random forest model (sensitivity 0.85, specificity 0.97), but several Pentacam-derived indices (e.g. index of surface variance, index of vertical asymmetry), which were only available from the Pentacam machine, were included in their model. Thus, their model could only be applied in clinics with a Pentacam and not exported

to other machines. To address this issue, we assessed all possible combination sets of parameters to test in three dominant machine learning algorithms with the aim of achieving a high degree of identifying subclinical KC from controls with the minimum number of parameters. Based on the results, we demonstrated that this approach could identify smaller subsets of parameters and increase their performance of machine learning models compared to using all parameters.

Another common feature of most of the studies published on machine learning techniques and subclinical KC is the definition used for classifying these eyes. Subclinical KC was defined as the normal fellow eye of unilateral KC[21,23-26,30]. 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 (Table 4).

Ruiz et al.[23] 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.[21] 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.[26] 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.[30] 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

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

*Subclinical KC was defined as normal fellow eye of uniliteral KC.[21,23–26,30]
#KC of mild and moderate severity was considered.[18]
†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

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

References

1. Sharif R, Bak-Nielsen S, Hjortdal J, Karamichos D, Progress in retinal and eye research. 2018;67:150–167.

2. Reeves SW, Ellwein LB, Kim T, Constantine R, Lee PP. Keratoconus in the Medicare population. *Cornea*. 2009;28:40–42.

3. Torres Netto EA, Al-Otaibi WM, Hafezi NL, et al. Prevalence of keratoconus in paediatric patients in Riyadh, Saudi Arabia. *Br J Ophthalmol*. 2018;102:1436–1441.

4. Kennedy RH, Bourne WM, Dyer JA. A 48-year clinical and epidemiologic study of keratoconus. *J Ophthal*. 1986;101:267–273.

5. Godefrooij DA, De Wit GA, Uiterwaal CS, Imhof SM, Wisse RP. Age-specific incidence and prevalence of keratoconus: a nationwide registration study. *Am J ophthal*. 2017;175:169–172.

6. Serdarogullari H, Tetikoglu M, Karahan H, Altin F, Elcioglu M. Research v. Prevalence of keratoconus and subclinical keratoconus in subjects with astigmatism using pentacam derived parameters. *J Ophthalmic Vis Res*. 2013;8:213.

7. Hashemi H, Beiranvand A, Khabazkhoob M, et al. Prevalence of keratoconus in a population-based study in Shahroud. *Cornea*. 2013;32:1441–1445.

8. Hashemi H, Khabazkhoob M, Yazdani N, Ostadimoghaddam H, Norouzirad R, Amanzadeh K, et al. The prevalence of keratoconus in a young population in Mashhad, Iran. Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists). 2014;34:519–527.

9. Hashemi H, Heydarian S, Yekta A, et al. High prevalence and familial aggregation of keratoconus in an Iranian rural population: a population-based study. *Ophthal Physiol Optics*. 2018;38:447–455.

10. Millodot M, Shneor E, Albou S, Atlani E, Gordon-Shaag A. Prevalence and associated factors of keratoconus in Jerusalem: a cross-sectional study. *Ophthal Epidemiol*. 2011;18:91–97.

11. Shneor E, Millodot M, Gordon-Shaag A, et al. Prevalence of keratoconus among young Arab students in Israel. 2014;3:9.

12. Hashemi H, Heydarian S, Hooshmand E, et al. The prevalence and risk factors for keratoconus: a systematic review and meta-analysis. *Cornea*. 2020;39:263–70.

13. Olivares Jimenez JL, Guerrero Jurado JC, Bermudez Rodriguez FJ, Serrano Laborda D. Keratoconus: age of onset and natural history. *Optom Vision Sci*. 1997;74:147–151.

14. Röck T, Bartz-Schmidt KU, Röck DJPo. Trends in corneal transplantation at the University Eye Hospital in Tübingen, Germany over the last 12 years: 2004–2015. *PloS One*. 2018;13:e0198793.

15. Williams KA, Keane MC, Coffey NE, Jones VJ, Mills RA, Coster DJ. The Australian Corneal Graft Registry 2018 Report. 2018.

16. Tan P, Mehta JS. Collagen crosslinking for keratoconus. *J Ophthalmic Vis Res*. 2011;6:153–154.

17. Godefrooij DA, Mangen MJ, Chan E, O'Brart DPS, Imhof SM, de Wit GA, et al. Cost-effectiveness analysis of corneal collagen crosslinking for progressive keratoconus. *Ophthalmology*. 2017;124:1485–1495.

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

corneal topography. J Biomed Informatics. 2002;35:151–159.

19. Arbelaez MC, Versaci F, Vestri G, Barboni P, Savini G. Use of a support vector machine for keratoconus and subclinical keratoconus detection by topographic and tomographic data. *Ophthalmology*. 2012;119:2231–2238.

20. Issarti I, Consejo A, Jimenez-Garcia M, Hershko S, Koppen C, Rozema JJ. Computer aided diagnosis for suspect keratoconus detection. *Computers Biol Med*. 2019;109:33–42.

21. Kovacs I, Mihaltz K, Kranitz K, et al. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with preclinical signs of keratoconus. *J Cataract Refractive Surg*. 2016;42:275–283.

22. Lopes BT, Ramos IC, Salomao MQ, Guerra FP, Schallhorn SC, Schallhorn JM, et al. Enhanced Tomographic Assessment to Detect Corneal Ectasia Based on Artificial Intelligence. *Am J Ophthal*. 2018;195:223–32.

23. Ruiz Hidalgo I, Rodriguez P, Rozema JJ, Ni Dhubhghaill S, Zakaria N, Tassignon MJ, et al. Evaluation of a machine-learning classifier for keratoconus detection based on Scheimpflug tomography. Cornea. 2016;35:827–832.

24. Saad A, Gatinel D. Topographic and tomographic properties of forme fruste keratoconus corneas. *Invest Ophthal Visual Sci*. 2010;51:5546–5555.

25. Smadja D, Touboul D, Cohen A, et al. Detection of subclinical keratoconus using an automated decision tree classification. *Am J Ophthal*. 2013;156:237–246.

26. Ucakhan OO, Cetinkor V, Ozkan M, Kanpolat A. Evaluation of Scheimpflug imaging parameters in subclinical keratoconus, keratoconus, and normal eyes. *J Cataract Refractive Surg*. 2011;37:1116–1124.

27. Ventura BV, Machado AP, Ambrosio R, Jr., et al. Analysis of waveform-derived ORA parameters in early forms of keratoconus and normal corneas. *J Refractive Surg (Thorofare, NJ: 1995)*. 2013;29:637–643.

28. Mahmoud AM, Roberts CJ, Lembach RG, Twa MD, Herderick EE, McMahon TT. CLMI: the cone location and magnitude index. *Cornea*. 2008;27:480–487.

29. Qin B, Chen S, Brass R, et al. Keratoconus diagnosis with optical coherence tomography-based pachymetric scoring system. *J Cataract Refractive Surg*. 2013;39:1864–1871.

30. Hwang ES, Perez-Straziota CE, Kim SW, Santhiago MR, Randleman JBJO. Distinguishing highly asymmetric keratoconus eyes using combined Scheimpflug and spectral-domain OCT analysis. *Ophthalmology*. 2018;125:1862–71.

31. Podgorelec V, Kokol P, Stiglic B, Rozman IJJoms. Decision trees: an overview and their use in medicine. 2002;26:445–463.

32. Chastang PJ, Borderie VM, Carvajal-Gonzalez S, Rostene W, Laroche L. Automated keratoconus detection using the EyeSys videokeratoscope. *J Cataract Refractive Surg*. 2000;26:675–683.

33. Twa MD, Parthasarathy S, Roberts C, Mahmoud AM, Raasch TW, Bullimore MA. Automated decision tree classification of corneal shape. *Optom and Vision Sci*. 2005;82:1038–1046.

34. Ambrosio R, Jr., Lopes BT, Faria-Correia F, et al. Integration of Scheimpflug-Based Corneal Tomography and Biomechanical Assessments for Enhancing Ectasia Detection. *Journal of refractive surgery (Thorofare, NJ: 1995)*. 2017;33:434–443.

35. Breiman L. Random forests. *Machine Learn*. 2001;45:5–32.

36. Saad A, Guilbert E, Gatinel D. Corneal enantiomorphism in normal and keratoconic eyes. *J Refractive Surg (Thorofare, NJ: 1995)*. 2014;30:542–547.

37. Silva APD, Stam A. Discriminant analysis. *Reading and understanding multivariate statistics*; Washington, DC, US: APA. 1995:277–318.

38. Hearst MA. Support vector machines. J IEEE Intelligent Sys. 1998;13:18–28.

39. Ramchoun H, Idrissi MAJ, Ghanou Y, Ettaouil M. Multilayer perceptron: architecture optimization and training with mixed activation functions. *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*; Tetouan, Morocco. 3090427: ACM; 2017. p. 1–6.

40. Carvalho LA. Preliminary results of neural networks and zernike polynomials for classification of videokeratography maps. *Optom and Vision Sci*. 2005;82:151–158.

41. Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. *Invest Ophthal Visual Sci*. 1995;36:1327–1335.

42. Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Invest Ophthal Visual Sci*. 1997;38:2290–2299.

43. Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations. *Clinics (Sao Paulo)*. 2010;65:1223–1228.

44. Rafati S, Hashemi H, Nabovati P, et al. Demographic profile, clinical, and topographic characteristics of keratoconus patients attending at a tertiary eye center. *J Curr Ophthalmol*. 2019;31:268–74.

45. Rodriguez LA, Villegas AE, Porras D, Benavides MA, Molina J. Treatment of six cases of advanced ectasia after LASIK with 6-mm Intacs SK. *J Refractive Surg (Thorofare, NJ: 1995)*. 2009;25:1116–1119.

46. Ernst BJ, Hsu HYJE, Lens c. Keratoconus association with axial myopia: a prospective biometric study. *Eye & Contact Lens*. 2011;37:2–5.

47. Sahebjada S, Xie J, Chan E, Snibson G, Daniel M, Baird PN. Assessment of anterior segment parameters of keratoconus eyes in an Australian population. *Optom and Vision Sci*. 2014;91:803–809.

48. Abolbashari F, Mohidin N, Ahmadi Hosseini SM, Mohd Ali B, Retnasabapathy S. Anterior segment characteristics of keratoconus eyes in a sample of Asian population. *Cont Lens Anterior Eye*. 2013;36:191–195.

49. Emre S, Doganay S, Yologlu S. Evaluation of anterior segment parameters in keratoconic eyes measured with the Pentacam system. *J Cataract Refract Surg*. 2007;33:1708–1712.

50. Piñero DP, Nieto JC, Lopez-Miguel A. Characterization of corneal structure in keratoconus. *J Cataract Refractive Surg*. 2012;38:2167–2183.

51. Baird PN, Schäche M, Dirani M. The GEnes in Myopia (GEM) study in understanding the aetiology of refractive errors. *Progress in Retinal and eye Res*. 2010;29:520–42.

52. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *Med Syst*. 2002;26:445–63.

53. Cover T, Hart P. Nearest neighbor pattern classification. *J IEEE Trans Inf Theor*. 2006;13:21–27.

54. Tibshirani R. Regression shrinkage and selection via the lasso. . 1996;58:267–288.

55. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascu Neurology*. 2017;2:230–43.

56. Kim SM, Kim Y, Jeong K, Jeong H, Kim JJU. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography*. 2018;37:36.

57. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One*. 2017;12:e0177726.

58. Tsao HY, Chan PY, Su EC. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinformatics*. 2018;19:283.

59. Yang JJ, Li J, Shen R, et al. Exploiting ensemble learning for automatic cataract detection and grading. *Comput Methods Programs Biomed*. 2016;124:45–57.

60. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 1967;13:21–7.

61. Ramchoun H, Idrissi MA, Ghanou Y, Ettaouil M. Multilayer perceptron: architecture optimization and training. *International Journal of Interactive Multimedia and Artificial Intelligence*. 2016;4:26–30.

62. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Sys Their Applications*. 1998;13:18–28.

63. Kuhn M. Building predictive models in R using the caret package. *J Stat Software*. 2008;28:1–26.

64. Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808. 2018.

65. Zhu W, Zeng N, Wang N. life sciences B, Maryland. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: Health Care and Life Sciences, Baltimore, Maryland*. 2010;19:67.

66. Sonego P, Kocsor A, Pongor S. ROC analysis: applications to the classification of biological sequences and 3D structures. *Briefings Bioinform*. 2008;9:198–209.

67. Hossin M, Sulaiman M. A review on evaluation metrics for data classification evaluations. *Int J Data Mining Knowledge Manage Proc*. 2015;5:1.

68. Deo RCJC. Machine learning in medicine. *Deo RC. Circulation*. 2015;132:1920–30.

69. Hall M. PhD Thesis, Correlation-based feature selection for machine learning. 1999. New Zealand Department of Computer Science, Waikato University.

70. Chan E, Baird PN, Vogrin S, Sundararajan V, Daniell MD, Sahebjada S. Economic impact of keratoconus using a health expenditure questionnaire: a patient perspective. *Clin Exp Ophthalmol*. 2019;48:287–300.

71. Dienes L, Kránitz K, Juhász É, Gyenes A, Takács Á, Miháltz K, et al. Evaluation of intereye corneal asymmetry in patients with keratoconus. *A Scheimpflug Imaging Study*. 2014;9:e108882.