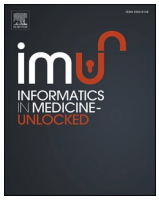
Informatics in Medicine Unlocked 38 (2023) 101208

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



An innovative approach based on machine learning to evaluate the risk factors importance in diagnosing keratoconus

Abin Daniel Zorto [a,*], Mhd Saeed Sharif [a], Julie Wall [a], Arun Brahma [b], Ahmed Ibrahim Alzahrani [c], Nasser Alalwan [c]

[a] Intelligent Technologies Research Group, UEL, University Way, London, E16 2RD, UK
[b] Manchester University NHS Foundation Trust, Manchester Royal Eye Hospital, Oxford Rd, Manchester, M13 9WL, UK
[c] Computer Science Department, Community College, King Saud University, Riyadh, 11437, Saudi Arabia

ARTICLE INFO

Keywords:
Keratoconus, machine learning, risk factors
Deep learning

ABSTRACT

Background and objective: Keratoconus is a non-inflammatory corneal condition affecting both eyes and is present in one out of every 2,000 people worldwide. The cornea deforms into a conical shape and thins, resulting in high-order aberrations and gradual vision loss. Risk factor analysis in the degradation of keratoconus is under-researched.

Methods: This research work investigates and uses effective machine learning models to gain insight.
into how much the risk factors of a patient contribute towards the progressive stages of keratoconus, as well as how significant these factors are in the creation of an accurate prediction model. This research demonstrates the value of machine learning approaches on a clinical dataset. This research paper employs several machine learning algorithms to classify the patients' stage of keratoconus using clinical information, such as measurements of the cornea's topography, elevation, and pachymetry taken using pentacam equipment at Sydney's Vision Eye Institute Chatswood.

Results: Eight different machine learning techniques were investigated over three variations of a dataset and achieved an average accuracy of 68, 80, and 90% for the risk factor, pentacam, and cumulative datasets, respectively. The results show a significant increase in accuracy and a 97% increase in AUC upon the addition of risk factor data compared to the models trained on pentacam data alone. The machine learning methods shown in this paper outperform those in current research.

Conclusions: This research highlights the importance of machine learning methods and risk factor data in the diagnosis of keratoconus and highlights the patient's primary optical aid as the strongest risk factor. The goal of this research is to support the work of ophthalmologists in diagnosing keratoconus and providing better care for the patient.

1. Introduction

In recent years, we have observed an increase in the use of machine learning (ML) and artificial intelligence for the diagnosis and monitoring of diseases. It has established itself as an indispensable resource for detecting and assessing trends in medicine and research. Keratoconus is one such condition within the ophthalmic industry that ought to be examined using ML approaches. Keratoconus is a non-inflammatory corneal disease that can affect both eyes and is present in 1 out of every 2,000 patients globally. It is the deformation and thinning of the cornea into a conical shape, leading to optical aberrations and progressive vision loss [1]. The disease frequently manifests itself in adolescence and progresses to a state of stabilisation by the time the patient reaches their forties, resulting in a significant reduction in quality of life. The most common diagnostic procedures for keratoconus are corneal topography and corneal tomography. This, in conjunction with clinical evaluation parameters and a systematic classification approach, such as the Amsler- Krumeich Grade 1–4 keratoconus

Abbreviations: AUC, Area Under Curve; KC, Keratoconus.

* Corresponding author.
E-mail addresses: u2091940@uel.ac.uk (A.D. Zorto), S.Sharif@uel.ac.uk (M.S. Sharif), J.Wall@uel.ac.uk (J. Wall), arun.brahma@cmft.nhs.uk (A. Brahma), ahmed@ksu.edu.sa (A.I. Alzahrani), nalalwan@ksu.edu.sa (N. Alalwan).

https://doi.org/10.1016/j.imu.2023.101208
Received 11 January 2023; Received in revised form 1 March 2023; Accepted 2 March 2023
Available online 8 March 2023

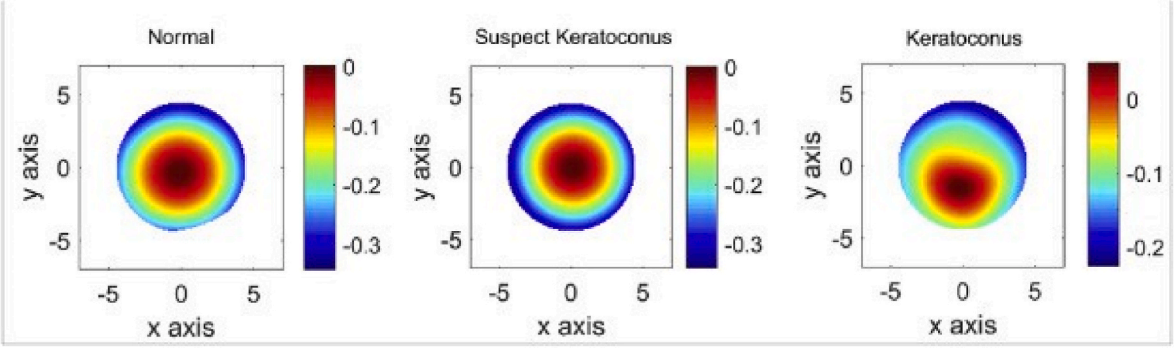


Fig. 1. Keratoconus topography maps showing normal, keratoconus suspect, and moderate keratoconus [5].

classification system, has been standard practice in the diagnosis of keratoconus and other corneal illnesses over the past several decades [2]. The pentacam illustrates several parameters, such as corneal curvature, thickness, and elevation data, through the use of colour-coded maps via corneal imagery, as shown in Fig. 1. The maps are displayed along with the corresponding numerical assessment of the patient's cornea [3].

The Placido ring and Scheimpflug imaging are utilized to acquire the topographic results. Aside from that, optical coherence tomography (OCT) can be used to give additional quantitative and qualitative information that can be used in the diagnosis process. When combined with other tomographic information, such as corneal thickness, this information can be used to classify the patient's stage of degeneration [1].

Although there is currently no consensus on the exact cause of keratoconus, there are several risk factors, both environmental and genetic, that have been identified as being of high importance in the development of keratoconus in individuals. Gender, ethnicity, general health, atopy hyper-tension, hay fever, known eye history, eye rubbing, family history of keratoconus, and primary optical aids are just a few of the risk factors to consider [1]. These factors may provide more insight into what is most important in recognising and diagnosing not only keratoconus but also other corneal disorders in the future. When compared to other ocular disorders, there are far fewer ML experiments conducted on the cornea [4]. The majority of the research available employs deep learning (DL) models, specifically convolutional neural networks, and applies them to corneal topography and tomography maps with elevation, curvature, and thickness metrics, to detect the presence and severity of keratoconus in patients. The current research typically comprises of image datasets with fewer than 400 images, , even though, DL models require a robust dataset to create a generalizable model [4]. This shows the benefit of using predictive algorithms that do not require a large dataset to find a relationship between their input data and their output. Another factor considered in this work is the computational cost of training and optimising accurate deep neural networks. This research work uses a range of less computationally intensive machine learning techniques to classify the data to produce a robust set of results and increase the likelihood of incorporation into the clinical sector. The objective of this research is to show the importance of machine learning methods and risk factor data in the diagnosis of keratoconus. We believe this study can help guide clinicians on the type of data that is of the most importance when assessing a patient and giving an accurate diagnosis. This study also has the potential to highlight to individuals without a diagnosis if there exists a need to go for an assessment.

In Section 2, the related work on ML and DL methods for keratoconus diagnosis and the literature on keratoconus risk factors will be reviewed. In Section 3, some light will be shed on the classification methods used to assess the stages of degradation in a patient. The public patient data used by Hallett, N. et al. [1], and how it was manipulated will be discussed, along with its limitations and delimitations. The ML algorithms that are used for the classification problem will be described, and the architecture of the models used will be shown in this section. The results of the report will be outlined in Section 4 and analyzed in Section 5. Lastly, in Section 6, we will present a set of concluding statements and provide recommendations based on the findings of the study.

2. Related work

Computer-aided diagnosis (CAD) makes use of ML algorithms. These algorithms are trained on several diagnostic samples derived from medical test results along with the opinions of experts' diagnoses to assist medical professionals in anticipating and detecting diseases going forward [6]. Machine learning has become increasingly popular in keratoconus. It has the effect of providing reliable and unbiased diagnosis, which is important when diagnosing patients early on since early intervention with treatments like corneal crosslinking can slow down the degradation of the cornea, avoiding the need for a corneal transplant altogether [7].

Zéboulon et al. [8], used a large dataset with 3,000 maps and achieved a highly accurate predictive model to classify patients, but the model's capability to predict different stages of the disease's progression has not yet been examined. Machine learning algorithms have rarely been used in clinical practice in the real world [9]. This could be attributed to a lack of large patient populations to confirm results, the use of diverse imaging technologies, participant groups made up of people of various ethnicities, medical professionals' general acceptability of predictive modelling for the detection of illnesses, and their relative consistency among clinicians. In their 2000 study, Chastang et al. [10], developed a binary decision tree technique based on corneal topography indices to distinguish clinically obvious keratoconus from the normal cornea. In their 2005 study, Twa et al. [11] modelled the corneal surface using a seventh-order Zernike polynomial as a method for differentiating between keratoconus and normal corneas. Smolek and Klyce [12] in their 1997 study, proposed a neural network approach that utilises corneal topography indices for keratoconus di-agnostics. All the above studies aim to accurately identify individuals with the condition of keratoconus.

All of these approaches relied solely on the cornea's anterior topography. There is a shortage of data on the impact of merging data from multiple devices for ML models used to identify keratoconus. Data on posterior corneal curvature and pachymetry were obtained and utilized to evaluate corneal features as technology advanced [13]. In clinical and subclinical keratoconus, Pinero et al. measured corneal vol-ume, pachymetry, and the relationship between the anterior and posterior shape of the cornea [14]. Fernández Pérez, Valero Marcos, and Martínez Peña [15] showed that using corneal equipment and instruments, such as the Pentacam, in combination with risk factors can lead to the detectionof subclinical keratoconus, but at the cost of a higher false-positive detection rate. Other types of information, such as the corneal epithelial thickness map obtained by OCT, are rapidly

becoming acknowledged as important for the diagnosis of keratoconus [16], particularly early keratoconus [17]. This illustrates that combining data from various devices and considering a wider range of parameters may help to enhance the early diagnosis of keratoconus. According to studies [18], ML models that used data from the Pentacam had greater collective sensitivity and specificity in detecting clinical keratoconus and subclinical keratoconus from control eyes than models that used data from other imaging devices. This is possibly due to the Pentacam's ability to produce a larger set of data than other devices.

2.1. Risk factors

Demographic information, including age and gender, as well as other possible risk factors for keratoconus, including eye rubbing and family history, may help in the diagnosis of keratoconus [7]. As there are so few established risk factors for keratoconus, much of the current research, including that of Sharif et al. [19], is uncertain as to how precisely these variables could impact the diagnosis accuracy of the ML models that are now in use. The inclusion of relevant risk variables into future ML models for diagnostic evaluation seems like an acceptable first step. This research examines the associations between keratoconus and gender, race/ethnicity, health status, history of eye disease, history of eye rubbing, genetics, and primary optical aid.

2.1.1. Gender

Differences between genders in keratoconus have been studied in several ways. The results differ across studies, with some demonstrating prevalence in women, others demonstrating higher prevalence in men, and still others demon-strating no significant gender prevalence [20].

2.1.2. Ethnicity

Keratoconus is a worldwide condition with varying levels of frequency among ethnic communities in the same geographical region. Numerous studies have shown that both ethnicity and environmental factors play important roles in the progression of the illness. Until recently, it was thought that keratoconus afflicted people of all ethnicities equally [21]. Various studies, however, have shown that there is a disparity in prevalence based on ethnicity. For example, Lebanon has a high frequency of 3.3%, India's rural parts have a prevalence of 2.3%, while the north of Denmark, Finland, and Russia have a lower rate [20]. Keratoconus was found to be substantially more common among Indians in Singapore than among Malays or Chinese [22]. Another study revealed that there was a 3.18% prevalence of keratoconus among the study's sample [23].

2.1.3. General health

Kumming [24] asserts that because Ehlers-Danlos Syn-drome (EDS) is a collagen abnormality of the body, it has the potential to disrupt corneal collagen and, as a result, have an impact on the progression of keratoconus. Keratoconus is quite rare in EDS patients. A total of 72 patients with EDS were studied by McDermott [25], and only one patient manifested keratoconus.

Marfan syndrome is caused by mutated X-chromosome genes, it is a condition in which patients have a greater prevalence of keratoconus than in the general population. However, In a study by Maumenee [26] keratoconus was not a feature in any of the 160 examined patients with Marfan syndrome. Osteogenic imperfecta is a hereditary bone disorder that affects bone collagen. Keratoconus has been reported more in individuals that possess this disease [27].

Sharif [28] looked at how common mitral valve prolapse was in 50 people with keratoconus and found that it was much more common (53% vs. 7%) in the keratoconus group. These results are different from those of Street [29] who discovered no statistically significant difference in the frequency of mitral valve prolapse between keratoconus patients and controls. The results of a research study in which 50 people with keratoconus underwent cardiac echography revealed the following: Mitral valve abnormalities were discovered in 22% of patients, which is four times more than the general population (prolapse was discovered in 4% of patients mitral insufficiency was discovered in 10% of patients, and mitral regurgitation was discovered in 8% of patients). The patients in this group of individuals who had prolapses were those who had the illness at an advanced stage [30].

Cullen [31] found that 5.5% of the 143 Down syndrome patients he studied had keratoconus, and he has found that patients who have experienced ocular traumatism are more likely to develop acute keratoconus. Transplant patients have a greater incidence of complications after surgery (such as suture failure, trauma, rejection, and others) [30]. Keratoconus prevalence in patients with Down syndrome appears to differ according to the ethnic group analyzed, according to the results of this study [32].

The relationship between keratoconus and diabetes has been investigated, however, the prevalence of diabetes in people with keratoconus is lower than in the overall population. On the other hand, those suffering from diabetes experience a less severe version of keratoconus [33]. It has been proposed that biochemical property alterations of the cornea in diabetics are caused by abnormal glycosylation of corneal stromal collagen fibres [30]. Diabetes may benefit individuals with keratoconus because it activates autocross linking in the corneal stroma. This mechanism prevents the cornea from becoming biomechanically weaker, which may explain why diabetes is beneficial for individuals with keratoconus.

2.1.4. Hay fever and hypertension

Statistically, there was a higher prevalence of asthma ($P = 0.0002$) and hay fever ($P = 0.007$) in the keratoconus group compared to the control group [34]. The number of times people wipe their eyes increased among those who had asthma, hay fever, or eczema keratoconus disease. Tomey TMS-2, on the other hand, showed no significant connection with the hayfever demographics when measured against corneal power [34]. The findings of Naderan et al. [35], imply that keratoconus may be associated with other disorders, often including vernal keratoconjunctivitis and hypertension.

2.1.5. Known eye history

A possible link between keratoconus and floppy eyelid syndrome has been suggested by some studies [36], although there is currently no conclusive evidence to support this and further research or data may be required. Leber Congenital Amaurosis is a congenital illness that causes poor vision from infancy onward. In comparison to other blinding diseases, this condition appears to be more related to keratoconus [37]. Although a genetic component could be the source of this condition [30].

Keratoconus is only seldomly related to corneal dys-trophies, according to the literature [30]. Several publications have reported an association between keratoconus and macular dystrophy [38]. Granular dystrophy is linked with keratoconus, and lamellar transplantation is the treatment of choice for keratoconus when dystrophy worsens [30]. Only a small number of individuals have been diagnosed with posterior polymorphous corneal disease (PPCD), a very uncommon autosomal dominant bilateral degenerative membrane dystrophy, but PPCD has also been associated with keratoconus in a small number of documented instances [30].

2.1.6. Eye rubbing

In a recent study by Ref. [39], researchers found strong evidence that keratoconus is linked to the habit of rubbing one's eyes. People who did not have the condition and did not wear contact lenses were asked to massage their eyes in a controlled way for 60 s. In a second study, tears were collected before and after rubbing the eyes, and it was found that the levels of MMP-13, IL-6, and TNF were significantly higher after rubbing the eyes. Multiple studies have shown that the constant rubbing of the eyes, which is common in people with keratoconus, may help the

disease get worse by keeping the levels of proteases, protease activity, and inflammatory mediators high.

2.1.7. Family history (genetics)

Genetic determinants in keratoconus are well supported by research findings [20]; According to the bulk of published research [21], the disease is inherited in an autosomal dominant way with variable expression and includes mild forms such as keratoconus fruste and moderate irregular astigmatism, among others. Several instances in the scientific literature show recessive inheritance, but none give clear evidence that three generations of the condition were reviewed or that minor variations of the illness were sought to be included in the pedigree analysis. Formal genetic analyses are needed to find out exactly how the different types of keratoconus are passed down and what role genetic factors may play in the cause of the disease, even though most research points to a dominant mode of inheritance [21].

2.1.8. Primary optical aid

A total of 199 patients were evaluated; 53 (27%) had a history of wearing contact lenses before the diagnosis of keratoconus, 68 (34%) had no history of wearing contact lenses before or after the diagnosis of keratoconus, and 78 (39%) had an existing history of keratoconus before wearing contact lenses [40]. When it came to wearing contact lenses before diagnosis with keratoconus, the 53 patients had an average of 12.2 years (with a range of 5.5–22 years) of experience, averaging 15.2 h each day. Although the study acknowledges that it is conceivable that keratoconus would still have progressed regardless of contact lens use, it also acknowledges that exceedingly mild symptoms of keratoconus may have been overlooked on the first test. It is considered that the wearing of contact lenses causes trauma to the cornea that, in most circumstances, is not severe enough to cause keratoconus to develop [40].

3. Methodology

3.1. Staging method

In addition to identifying keratoconus eyes as a discrete category, several research studies classified keratoconus eyes into clinical phases and employed machine learning (ML) methods to identify each stage independently, as seen in Fig. 2. The study reveals a variety of classification criteria for keratoconus eyes. Kamiya et al. [42], used the Amsler-Krumeich (AK) classification technique for Grades 1–4, which strongly emphasises keratometry but also often combines refraction and pachymetry. measures, as shown in Table 3 [43]. Using a distinct classification method, dubbed RETICS, Blazquez et al. [44], and Bolarin et al. [45]. The physical representations of keratoconus staging are illustrated in Fig. 3.

[45] grouped eyes into Grades I–IV plus. The AK technique relies heavily on distant visual acuity that has been adjusted [43]. The AK classification may be used to identify and monitor the evolution of keratoconus. The Sandali classification scheme and this classification approach are the two most often used classification models in the corneal industry, according to Macsai [40]. For consistency with the gathered data samples, this study employs Grades 1–4 according to the AK grading system. Other grades and methodologies are employed in current research for severity classification, but this work uses Grades

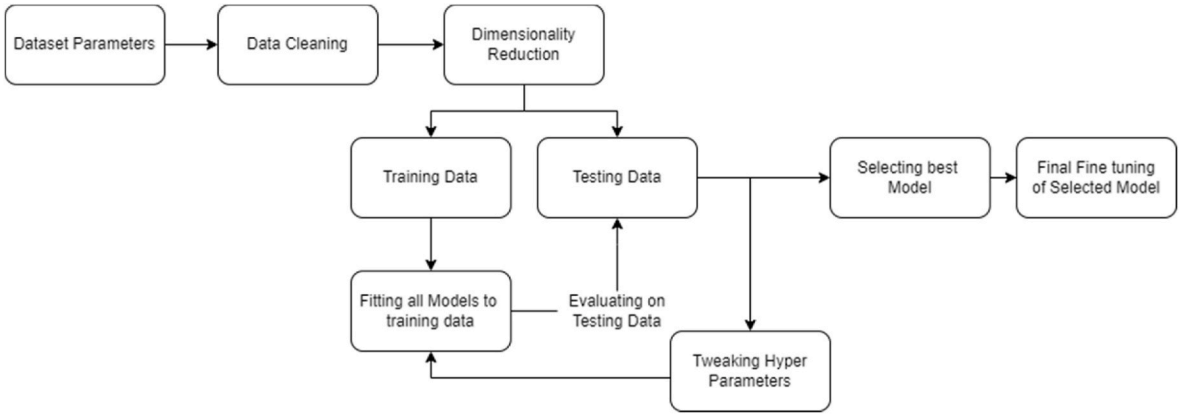


Fig. 2. Flow chart of steps taken for the analysis of KC stages for each dataset.

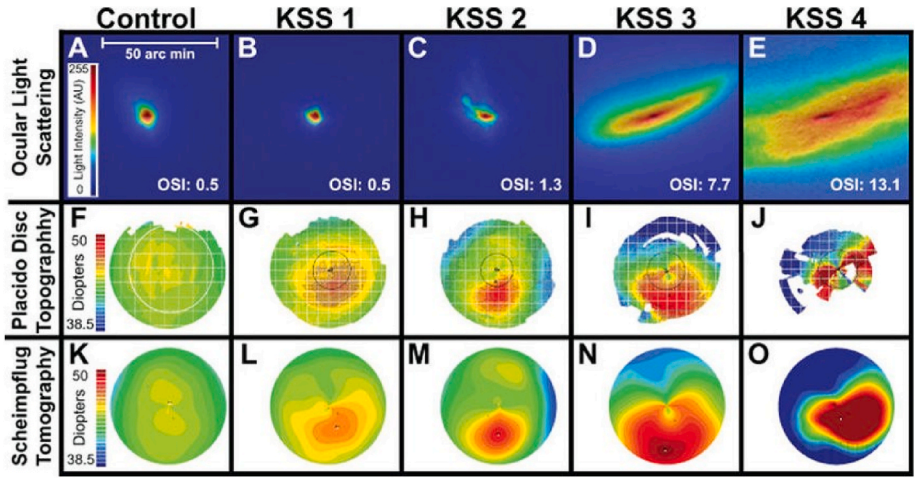


Fig. 3. Physical representation of keratoconus Staging [41].

1–4 according to the AK grading scheme. In individuals with keratoconus, the AK and Sandali classifications are the most often used systems; however, they cannot be used interchangeably since they indicate various corneal abnormalities. Specifically, the AK classification is superior at identifying individuals in the early stages of disease and following their progression through time. In contrast, the Sandali classification, which is based on the examination of anterior segment optical coherence tomography (AS-OCT) images, is useful for the diagnosis and followup of patients with advanced stages of keratoconus, particularly when selecting a surgical approach. This demonstrates the usefulness of the AK model as a more accurate classification model for machine learning (ML) algorithms, which are also reliable in earlystage diagnosis [40].

3.2. Dataset

Most of the research in this field is concerned with image datasets for the prediction of keratoconus, such as this publicly available dataset [46] provided by Al-Timemy et al. [2], But the data used for this research is numerical. Between 2014 and 2017, data [47] on 124 keratoconus patients was gathered at the Vision Eye Institute Chatswood (VEIC) in Sydney, Australia. Table 1 and Table 2 contain information that has been extracted to be used as variables of consideration for the training process. Based on clinical expertise, all patients were classified using the AK classification, which takes into account the presence and severity of scarring, mean central keratometry (MCK), minimum corneal thickness (MCT), myopia, and astigmatism (MA). Table 3 illustrates how patients are classified using the AK categorization system [1]. Table 4 shows the distribution of cases for each grade.

Table 1
Risk Factor Data [1].

Variable
Atopy Ethnicity
Primary optical aid Gender
Eye history Diabetes Family history Allergy Hypertension General Health Eye rubbing

Table 2
Pentacam and clinician data [1].

Variable
Steep keratometry
Flat keratometry Thinnest pachymetry Sphere
Cylinder Refractive axis X-axis
y-axis Pachymetry
Amsler-Krumeich (AK) classification

Table 3
Amsler–Krumeich (AK) classification scheme [1].

Grade	Characteristic
1	Eccentric steeping MCK 48.00D MA 5.00D
2	Absence of scarring MCT 400 μm MCK 53.00D MA 5.00–8.00D
3	Absence of scarring MCT 300–400 μm MCK 53.00D MA 8.00–10.00D
4	Central corneal scarring MCT 200 μm MCK 55.00D Refraction not measurable

Table 4
Number of cases.

Grade	Cases
1	140
2	51
3	24
4	22
Total	**237**

3.2.1. Dataset manipulation and delimitation

The collected dataset contains metrics that are collected according to the assessment of the clinician. For this paper, these data points are not advisable and were removed from the analysis as they do not fall in line with the goal of using data readily available from the patient and the Pentacam (or any other appropriate topography device). The dataset was also split into two sections for the training and testing of the model, with the training section comprising 90% of the entire dataset and the test section 10%.

3.2.2. Data analysis

Quantitative analysis was carried out on the dataset to assess the correlation between the various attributes within it. These correlations will be separated according to the source of the data to provide a useful analysis of the significance of each attribute. The dataset will be analyzed in three different sections: the risk factor data readily available from the patient, the pentacam data readily available from the device, and the dataset comprising both parts to see the relative effect of the dataset sections on their own. Within this field of research, the sample size of data for applications of this nature is usually much larger to make the predictive model more generalizable. However, this project works on a relatively small dataset with the assumption that the ML techniques will offset this drawback. In addition, the demographic characteristics of the dataset are evaluated to give richer interpretations of the data and opportunities for development.

3.3. Machine learning vs deep learning

Machine learning and deep learning are two popular subsets of artificial intelligence (AI) that rely on algorithms to make predictions or decisions based on input data. While both machine learning and deep learning can be effective on large datasets, their performance may vary when dealing with smaller datasets [48]. Machine learning algorithms use statistical techniques to find patterns in the data and make predictions. They typically work well with smaller datasets because they are designed to be more interpretable and computationally efficient than deep learning algorithms.

ML models can provide insight into why a particular prediction was made, which can be useful for decision-making in many applications [49]. In contrast, DL algorithms use complex neural networks to learn patterns in the data. They require a large amount of data to train and can be computationally intensive [48]. As a result, DL models are typically more effective on larger datasets. However, they can also be less interpretable than machine learning models, particularly when dealing with smaller datasets [49].

3.4. Machine learning algorithms and model parameters

This study creates its models using the most popular machine learning methods from the scikitlearn Python package and this section will go into detail on what model parameters were used.

3.4.1. Multinomial logistic regression

- lbfgs solver
- 10,000 maximum iterations

3.4.2. Nearest neighbour classification

- Neighbour count set to 4

3.4.3. Support vector machine for classification

- Kernel set to linear
- 0.025 regularisation term

3.4.4. Decision tree

- The maximum depth of the tree is set to 5 items.

3.4.5. Random Forest

- Max tree depth is set to 5.
- The maximum number of trees in the forest has been set to ten.
- The maximum number of features to consider is set to 1.

3.4.6. Multilayer perceptron

- Alpha value set to 5
- Max tree number in the forest set to 10
- Maximum number of features to consider set to 1

3.4.7. AdaBoost decision tree

- All parameters left as default

3.4.8. Quadratic discriminant analysis

- All parameters left as default

4. Results and analysis

4.1. Risk factor correlation data

The correlation heatmap of the risk factor dataset, as shown in Fig. 4 was created with quantitative analysis using Python, and it highlights how strongly all the risk factors correlate to the resulting class of the diagnosis as well as to each other factor (see Fig. 3). There are some notable data points within the heatmap with a correlation of 0.2 and above, which are as follows.

- The relatively high correlation of 0.5 between atopy and hay fever
- The correlation between atopy, hypertension, hay fever, and family history (genetics)
- Hypertension has a notable link to Race with a value of 0.2
- There is also a notable correlation between gender and eye rubbing (0.24).

The most significant contributor to the class assessment process according to the risk factor correlation data is the primary optical aid of the patient, this is followed by the gender of the patient and the runner-up to this is eye rubbing. The only truly notable contributor according to the data is the primary optical aid of the patient. The average correlation of the attributes to the assessment is 0.092, which can be seen as insignificant.

4.2. Pentacam correlation data

The correlation heatmap shown in Fig. 4 highlights how strongly all the pentacam features correlate to the resulting class of the diagnosis as well as to each other factor. We can see from the data that there is a higher overall correlation between the attributes in this section of the dataset compared to the risk factors section of the dataset. Within this section of the dataset, there are more notable correlations between the attributes and the assessed class. The keratometry readings are the highest correlators, but it is important to note the corneal thickness measurements (pachymetry readings) and refractive measurements also show high correlation. The average correlation of the attributes to the assessment is 0.43, which can be seen as very significant.

4.3. Demographic diversity of data

The demographic diversity of the dataset is illustrated in Fig. 5 showing a skew towards Male participants (blue portion of gender diversity pie chart), In the dataset, ethnic diversity is mostly biased towards three ethnicities.

4.4. Model performance evaluation

As we can see in Fig. 6 the prediction capability of the risk factor dataset has a lower average accuracy than that of the pentacam dataset, which itself has a lower accuracy than the cumulative dataset. This is to be expected, as the addition of useful data to a model will generally result in better performance. The data shows that the addition of risk factor data improved model performance by 12.5%, which is a notable performance boost. The MLP model is the only one to drop in accuracy when combining both data sections; this may be due to the depth of the network and its incompatibility with a high number of attributes but a

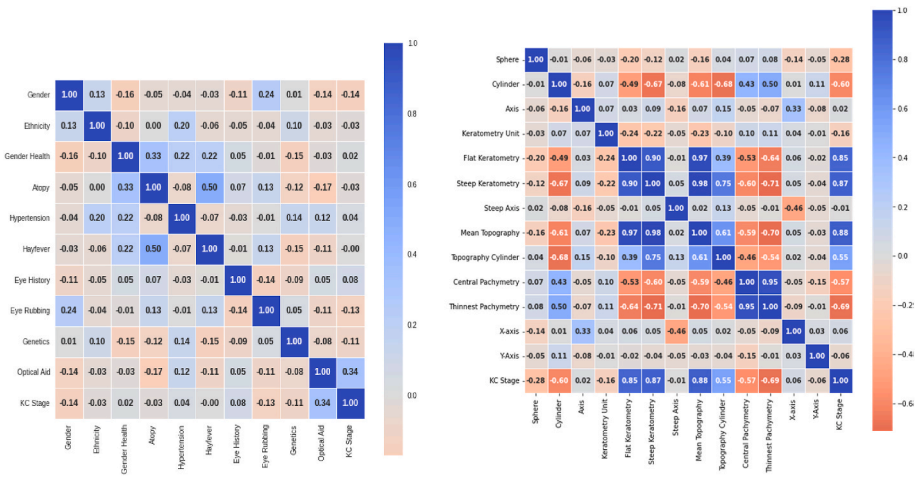


Fig. 4. Correlation data (risk factor heatmap left, pentacam heatmap right).

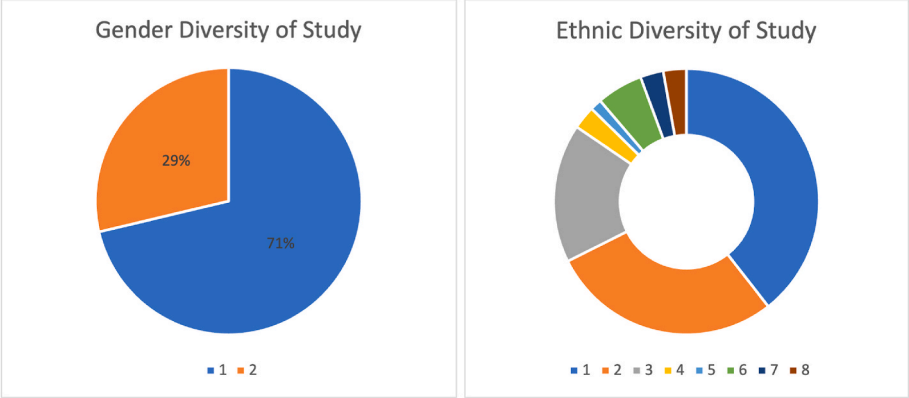


Fig. 5. Demographic diversity.

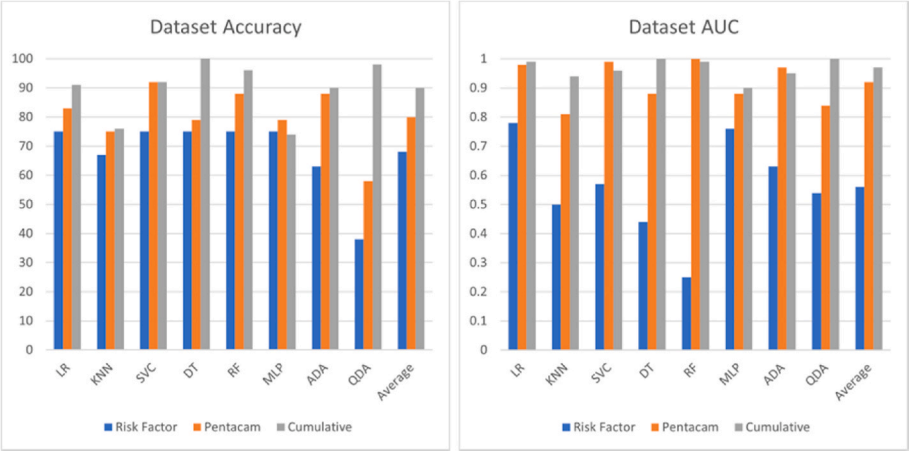


Fig. 6. Model performance (Accuracy left, Area under curve (AUC) right).

relatively small amount of data. We can say this illustrates the suitability of ML models for fields in which there is a small amount of data available for training. Several models performed fairly well at diagnosis with only risk factor data, but when we include average AUC in our analysis as well as Pentacam attributes, we gain more insight into what the most suitable model to use would be in any given scenario. In Fig. 6 the results show a lower average AUC for the risk factor versions of the models compared to the Pentacam versions of the models. This phenomenon has similarly been observed in a recent study, where Perez et al. [15], illustrated that the use of corneal equipment and techniques such as the Pentacam in combination with risk factors can result in the detection of subclinical keratoconus, but at the expense of a higher rate of false-positive detections. This is a limitation of the risk factor data, which tends to underdiagnose severity in patients. More work can be done in tuning the algorithms towards a lower false positive detection rate.

Table 5 illustrates the three key factors in determining the significance of each section of the clinical assessment. Looking at the correlation data, we see that the Pentacam data has a very significant correlation to the assessment, compared to the risk factor data, which shows little relevance. However, as previously mentioned, the 12.5% overall increase in performance upon adding risk factor data indicates that there is a noticeable relevance to the assessment. The results in Table 5 also show a slight increase of 5% in average AUC with the addition of risk factor data. Table 6 and 7 illustrate how little the risk factor helps in improving the precision and recall of a predictive model.

Table 5
Overall comparison of results between different sections of the dataset.

Section	Risk Factor	Pentacam	Cumulative
Correlation	0.09	0.43	–
Macro Average AUC	0.56	0.92	0.97
Accuracy	68	80	90

Table 6
Precision data for the created machine learning models.

Model	Risk Factor	Pentacam	Cumulative
LR	0.56	0.86	0.91
K Neighbors	0.60	0.76	0.75
SVC	0.56	0.96	0.93
DT	0.56	0.89	1.00
RF	0.56	0.96	0.87
MLP	0.56	0.77	0.77
ADA	0.54	0.83	0.88
QDA	0.52	0.80	0.98

Table 7
Recall data for the created machine learning models.

Model	Risk Factor	Pentacam	Cumulative
LR	0.75	0.83	0.91
K Neighbors	0.67	0.75	0.76
SVC	0.75	0.92	0.92
DT	0.75	0.79	1.00
RF	0.75	0.88	0.96
MLP	0.75	0.79	0.74
ADA	0.63	0.88	0.90
QDA	0.38	0.58	0.98

5. Discussion

Discrepancies arose with regard to the most significant risk factors, i.e., the significance of primary optical aids and gender, in assessing the severity of keratoconus and current research in the literature. The results from the experiments on risk factor data contrast with the current research on these factors, which may lead us to believe that the models created may have suffered from data bias due to an over-representation of certain elements in the dataset, causing the models to favour the overrepresented elements. More clinical research on these factors may be necessary to debunk or support any correlation between these factors and the assessment of the disease.

5.1. Keratoconus grade performance

Based on the data shown in Fig. 7 and Table 8 the model's performances in staging the degradation of keratoconus differ according to each stage. The accuracy of the models trained on the risk factor dataset, all show a general ability to classify the severity of the diagnosis as grade 1, but this is essentially due to the models underdiagnosing every case as grade 1. The only algorithm that shows more robust behaviour, in this case, is the quadratic discriminant analysis (QDA) algorithm, as it correctly classified some cases as grade 3 keratoconus but had more incorrect predictions for grade 1 cases. Overall, the models show poor performance on the risk factor section of this dataset. On the Pentacam dataset, the models showed a decrease in accuracy for detecting grade 1 cases but also showed a much more robust level of prediction, with each model showing better performance on stages other than the first stage, leading to an overall increase in accuracy across the stages. Random Forest was the best-performing algorithm for the Pentacam dataset across the stages, as well as the most robust model. Meanwhile, QDA, which has shown the most robustness on the risk factor dataset, had trouble classifying the fourth stage of the disease accurately. When both datasets are combined, the overall performance of the models increases again, but less so than the difference between the risk factor dataset and the Pentacam dataset. All models show a satisfactory level of robustness on the cumulative data, but the decision tree and QDA models show great robustness and performance, with the decision tree model performing better overall.

5.2. F score for model selection

The F score is a weighted harmonic mean of precision and recall. It is a class-balanced accuracy metric. When false negatives and false positives are important in the prediction process, the F-score is used. The findings demonstrate that we are more adept at predicting grade 1 cases than any other grade. This is likely due to unbalanced classes and the fact that the models in the risk factor section and some in the Pentacam section overfit and have a high tendency to underdiagnose the severity of keratoconus to grade 1 keratoconus. Since practically every instance in the data is classified as grade 1, the accuracy of predicted grade 1 cases is almost equal to the accuracy of all cases; hence, the recall of the data dominates the overall accuracy measure.

Table 8
Accuracy % of the created machine learning models.

Model	Risk Factor	Pentacam	Cumulative
LR	75	83	91
K Neighbors	67	75	76
SVC	75	92	92
DT	75	79	100
RF	75	88	96
MLP	75	79	74
ADA	63	88	90
QDA	38	58	98

The adjusted F-score is a variation of the F-score that allows us to place a higher value on precision or recall if they are more relevant to our application. For this research, an F 0.5 score is used to give more priority to the precision value.

$$F_\beta = \left(1+\beta^2\right) \times \frac{\text{precision} \times \text{recall}}{\left(\beta^2 \times precision\right) + \text{recall}} \quad (1)$$

where β is = 0.5.

Table 9
F 0.5 Score for the created machine learning models.

Model	Risk Factor	Pentacam	Cumulative
LR	0.59	0.85	0.91
K Neighbors	0.61	0.75	0.74
SVC	0.59	0.94	0.93
DT	0.59	0.78	1.00
RF	0.59	0.87	0.84
MLP	0.59	0.75	0.78
ADA	0.55	0.83	0.87
QDA	0.48	0.72	0.98

Table 10
SVC Regularisation Tuning for the selected model.

Regularisation Parameter	F Score	Accuracy
0.002	0.88	88
0.02	0.92	91
0.2	0.92	92
2	0.98	98
20	**0.99**	**99**
200	0.96	96
2000	0.96	96

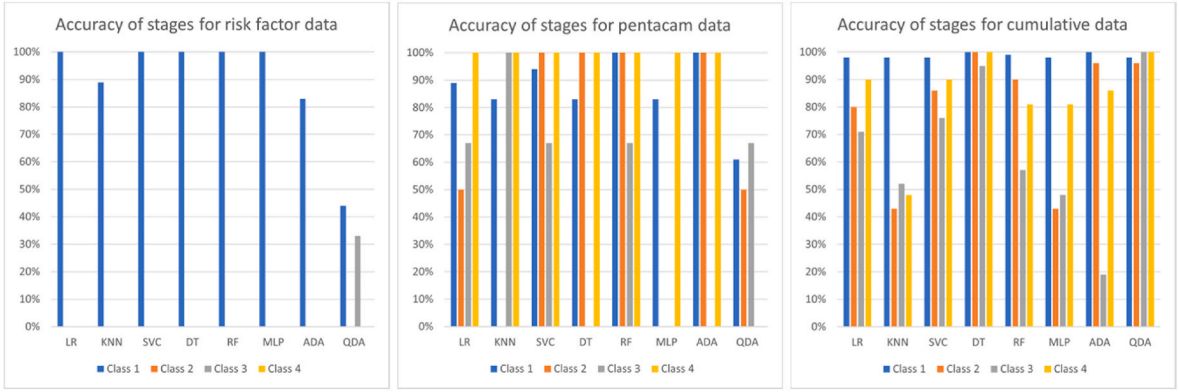


Fig. 7. KC Stage performance (risk factor left, pentacam middle, cumulative right).

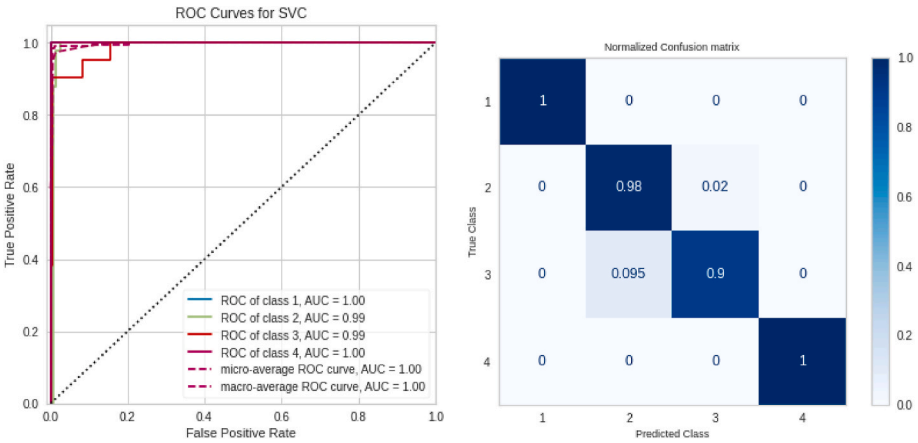


Fig. 8. SVC model performance (ROC curve (left), confusion matrix (right)).

5.3. Selecting a model for tuning

Based on the results shown in Table 9, the algorithm selected as the most suitable model for this research is the support vector classifier due to its higher overall F score across the three datasets. This model was chosen to undergo further fine-tuning, and after varying the regularisation parameter of the algorithm, a model was created with a 0.99 F-score and 99% accuracy with a regularisation term of 20 as shown in Table 10. The ROC curve and confusion matrix of the model can also be seen in Fig. 8 with an AUC of 1.00.

5.4. Research comparison

Based on the results in Table 11, we can see that when we compare the performance of the DL models used by Hallett et al. [1], to the ML models used in this work, the ML models have a better overall performance than the DL models when using the same dataset with a small sample size.

When we look at the performance of the best model in this research work compared to some of the state-of-the art models in current research in Table 12, we can see it possesses similar or greater performance than the models shown in the table. The SVC model presented in this research exhibits favourable performance despite its very small sample size when compared to the datasets with which it is to be evaluated, such as the data obtained by Kamiya et al. [50] from Kisato University (KU), and Malyugin et al. [51] from S. Fyodorov Eye Microsurgery Complex Head Office (SFEMCHO).

Table 11
Research comparison of average model performance between machine learning and deep learning models trained on the same dataset.

Model	Hallett [1].	Cumulative
Macro Average AUC	0.90	0.97
Accuracy	77	90

Table 12
Research comparison of results from best-performing keratoconus models from different authors.

Models	AUC	Accuracy	Dataset	Technique
Our Work	1.00	99	VEIC	ML
Hallett et al.	0.89	82	VEIC	DL
Kamiya et al.	0.94	97	KU	DL
Malyugin et al.	0.97	97	SFEMCHO	ML

6. Conclusion

After conducting this study, we can comfortably say that the addition of risk factor data to the overall dataset increases the performance of the predictive models by a significant amount. The risk factors average correlation to the assessment is 0.09 which brought about a 12.5% increase in accuracy when employed. Comparing the results presented in this paper to those in the literature, there is a notable average overall increase in performance, illustrating the effectiveness of ML algorithms on small datasets over DL algorithms. Most research is focused on the detection of keratoconus and not the diagnosis of the stage of severity, so the literature used to support this study may not be wholly applicable but may be sufficient to serve the purpose of this paper. In the future work, we are aiming to expand the data and include participant diversity. Finally, as the data for the diagnosis of this disease remains scarce, ML techniques should receive more favour as they demonstrate a notable increase in performance compared to DL techniques. With tuning, the efficacy of these models can significantly improve, further increasing the likelihood of their implementation in realworld diagnosis. We aim to assist ophthalmologists in their work and to deliver improved patient care.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is funded by the researchers supporting project number (RSP2023R157), King Saud University, Riyadh, Saudi Arabia.

References

[1] Hallett Nicole, Yi Kai, Dick Josef, Hodge Christopher, Sutton Gerard, Wang Yu Guang, You Jingjing. Deep learning based unsupervised and semi-supervised classification for keratoconus (CornealAI). January 2020.
[2] Al-Timemy Ali H, Mosa Zahraa M, Alyasseri Zaid, Lavric Alexandru, Marcelo M Lui, Hazarbassanov Rossen M, Yousefi Siamak. A hybrid deep learning construct for detecting keratoconus from corneal maps. Transl. Vis. Sci. Technol. December 2021;10(14). 16.
[3] Fırat Murat, Çınar Ahmet, Çankaya Cem, Tuncer Fırat İlknur, Tuncer Taner. Prediction of pentacam image after corneal cross- linking by linear interpolation technique and U-NET based 2D re- gression model. Comput Biol Med July 2022; 146:105541.
[4] Cao Ke, Verspoor Karin, Sahebjada Srujana, Baird Paul N. Accuracy of machine learning assisted detection of keratoconus: a systematic review and Meta-Analysis. J Clin Med Res January 2022;11(3).

[5] Issarti Ikram, Consejo Alejandra, Jiménez-García Marta, Her- shko Sarah, Koppen Carina, Rozema Jos J. Computer aided diagnosis for suspect keratoconus detection. Comput Biol Med June 2019;109:33–42.

[6] Shehab Mohammad, Abualigah Laith, Shambour Qusai, Abu-Hashem Muhannad A, Yousef Shambour Mohd Khaled, Alsalibi Ahmed Izzat, Gandomi Amir H. Machine learning in medical applications: a review of state-of-the-art methods. Comput Biol Med June 2022;145:105458.

[7] Yousefi Siamak, Yousefi Ebrahim, Takahashi Hidenori, Hayashi Takahiko, Tampo Hironobu, Inoda Satoru, Arai Yusuke, As- bell Penny. Keratoconus severity identification using unsupervised machine learning. PLoS One November 2018;13 (11):e0205998.

[8] Zéboulon Pierre, Debellemanière Guillaume, Bouvet Magalie, Gatinel Damien. Corneal topography raw data classification using a convolutional neural network. Am J Ophthalmol November 2020;219(33–39).

[9] Kelly Christopher J, Karthikesalingam Alan, Suleyman Mustafa, Corrado Greg, King Dominic. Key challenges for delivering clinical impact with artificial intelligence. BMC Med October 2019;17(1):195.

[10] Chastang PJ, Borderie VM, Carvajal-Gonzalez S, Rostène W, Laroche L. Automated keratoconus detection using the EyeSys videokeratoscope. J Cataract Refract Surg May 2000;26(5):675–83.

[11] Twa Michael D, Parthasarathy Srinivasan, Roberts Cynthia, Mahmoud Ashraf M, Raasch Thomas W, Bullimore Mark A. Automated decision tree classification of corneal shape. Optom Vis Sci December 2005;82(12):1038–46.

[12] Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. Invest Ophthalmol Vis Sci October 1997;38(11):2290–9.

[13] Ambrósio Jr Renato, Ruiz Simonato Alonso, Allan Luz, Guillermo Coca Velarde Luis. Corneal-thickness spatial profile and corneal-volume distribution: tomographic indices to detect kerato- conus. J Cataract Refract Surg November 2006;32(11):1851–9.

[14] Piñero David P, Alió Jorge L, Alesón Alicia, Escaf Vergara Munir, Miranda Mauricio. Corneal volume, pachymetry, and correlation of anterior and posterior corneal shape in subclinical and different stages of clinical keratoconus. J Cataract Refract Surg May 2010;36(5):814–25.

[15] Fernández Pérez Joaquín, Valero Marcos Almudena, Martínez Peña Fran- cisco Javier. Early diagnosis of keratoconus: what difference is it making? Br J Ophthalmol November 2014;98(11):1465–6.

[16] Silverman Ronald H, Urs Raksha, RoyChoudhury Arindam, Archer Timo- thy J, Gobbe Marine, Reinstein Dan Z. Combined tomog- raphy and epithelial thickness mapping for diagnosis of keratoconus. Eur J Ophthalmol March 2017;27(2): 129–34.

[17] Venkateswaran Nandini, Galor Anat, Wang Jianhua, Karp Carol L. Optical coherence tomography for ocular surface and corneal diseases: a review. Eye Vis (Lond) June 2018;5(13).

[18] Chen Davie, Lam Andrew KC. Reliability and repeatability of the pentacam on corneal curvatures. Clin Exp Optom March 2009;92(2):110–8.

[19] Sharif Rabab, Bak-Nielsen Sashia, Hjortdal Jesper, Karamichos Dimitrios. Pathogenesis of keratoconus: the intriguing therapeu- tic potential of prolactin-inducible protein. Prog Retin Eye Res November 2018;67(150–167).

[20] Alió Jorge L. Keratoconus: recent advances in diagnosis and treatment. Cham: Springer; 2017.

[21] Rabinowitz Yaron S. Keratoconus. *Surv. Ophthalmol.* January 1998;42(4):297–319.

[22] Shneor Einat, Millodot Michel, Gordon-Shaag Ariela, Essa Maron, Adel Barbara. Prevalence of keratoconus among young arab students in Israel, vol. 3; April 2014. p. 9–14. 1.

[23] Pan Chen-Wei, Cheng Ching-Yu, Sabanayagam Charumathi, Chew Merwyn, Lam Janice, Marcus Ang, Wong Tien Y. Ethnic variation in central corneal refractive power and steep cornea in asians. Ophthalmic Epidemiol April 2014;21 (2):99–105.

[24] Kuming BS, Joffe L. Ehlers-Danlos syndrome associated with keratoconus. a case report. S Afr Med J August 1977;52(10):403–5.

[25] McDermott ML, Holladay J, Liu D, Puklin JE, Shin DH, Cowden JW. Corneal topography in Ehlers-Danlos syndrome. J Cataract Refract Surg September 1998;24 (9):1212–5.

[26] Maumenee IH. The eye in the marfan syndrome. Trans Am Ophthalmol Soc 1981; 79:684–733.

[27] Beckh U, Schönherr U, Naumann GO. [autosomal dominant keratoconus as the chief ocular symptom in lobstein osteogenesis imperfecta tarda]. Klin. Monbl. Augenheilkd. April 1995;206(4):268–72.

[28] Sharif KW, Casey TA, Coltart J. Prevalence of mitral valve prolapse in keratoconus patients. J R Soc Med August 1992;85(8):446–8.

[29] Debra A Street, Vinokur Eric T, Waring George O, Pollak Scott J, Clements Stephen D, Perkins John V. Lack of association between keratoconus, mitral valve prolapse, and joint hypermobility. Ophthalmology February 1991;98(2):170–6.

[30] Barbara Adel. Textbook on keratoconus: new insights. In jaypee. jaypee; 2012.

[31] Cullen JF, Butler HG. Mongolism (DOWN'S syndrome) and keratoconus. Br J Ophthalmol June 1963;47:321–30.

[32] Omer Khor. Epidemiology of keratoconus worldwide. Open Ophthalmol J November 2018;12(1):289–99.

[33] Irene C Kuo, Broman Aimee, Pirouzmanesh Ashkan, Melia Michele. Is there an association between diabetes and keratoconus? Ophthalmology February 2006;113 (2):184–90.

[34] Weed KH, MacEwen CJ, Giles T, Low J, McGhee CNJ. The dundee university scottish keratoconus study: demographics, corneal signs, associated diseases, and eye rubbing. Eye April 2008;22(4):534–41.

[35] Naderan Mohammad, Shoar Saeed, Rezagholizadeh Farzaneh, Zolfaghari Ma- soome, Naderan Morteza. Characteristics and associ- ations of keratoconus patients. Contact Lens Anterior Eye June 2015;38(3):199–205.

[36] Negris R. Floppy eyelid syndrome associated with keratoconus. J Am Optom Assoc May 1992;63(5):316–9.

[37] Galin MA, Berger R. Atopy and keratoconus. Am J Ophthalmol June 1958;45(6): 904–6.

[38] Javadi Mohammad-Ali, Rafee'i Alireza Baradaran, Kamalian Nasser, Karimian Farid, Ja'farinasab Mohammad-Reza, Yazdani Shahin. Concomitant keratoconus and macular corneal dystrophy. Cornea July 2004;23(5):508–12.

[39] Balasubramanian Sivaraman A, Pye David C, Willcox Mark DP. Effects of eye rubbing on the levels of protease, protease activity and cytokines in tears: relevance in keratoconus. Clin Exp Optom March 2013;96(2):214–8.

[40] Macsai MS, Varley GA, Krachmer JH. Development of keratoconus after contact lens wear. patient characteristics. Arch Ophthalmol April 1990;108(4):534–8.

[41] Leonard Anthony P, Gardner Scott D, Rocha Karolinne M, Zeldin Evan R, Tremblay David M, George O, Waring, 4th. Double- Pass retina point imaging for the evaluation of optical light scatter, retinal image quality, and staging of keratoconus. J Refract Surg November 2016;32(11):760–5.

[42] Kamiya Kazutaka, Ayatsuka Yuji, Kato Yudai, Fujimura Fusako, Takahashi Masahide, Shoji Nobuyuki, Mori Yosai, Miyata Kazunori. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diag- nostic accuracy study. BMJ Open September 2019; 9(9):e031313.

[43] Alió Jorge L, Vega-Estrada Alfredo, Sanz-Díez Pablo, Peña- García Pablo, Maldonado Miguel. Keratoconus management guide- lines, vol. 4; April 2015. p. 1–39. 1.

[44] Velázquez-Blázquez Jose S, Bolarín José M, Cavas- Martínez Francisco, Alió Jorge L. EMKLAS: a new automatic scoring system for early and mild keratoconus detection. Transl. Vis. Sci. Technol. May 2020;9(2):30.

[45] Bolarín José M, Cavas F, Velázquez JS, Alió JL. A Machine- Learning model based on morphogeometric parameters for RETICS disease classification and GUI development. NATO Adv. Sci. Inst. Ser. E Appl. Sci. March 2020;10(5):1874.

[46] TVST paper images - google drive. https://drive.google.com/drive/folders/1GR9T-p7GWGY_0nI5sm8GdJ4V6qlV4vZ2. Accessed: 2023-2-28.

[47] Kai Yi. Dataset at master · ykiiiiii/CornealAI.

[48] Ongsulee Pariwat. Artificial intelligence, machine learning and deep learning. In 2017 15th *International Conference on ICT and Knowledge Engineering (ICTKE)*; 2017. p. 1–6.

[49] Aron Prenner Julian, Robbes Romain. Making the most of small software engineering datasets with modern machine learning. IEEE Trans Software Eng 2022;48(12):5050–67.

[50] Kamiya Kazutaka, Ayatsuka Yuji, Kato Yudai, Shoji Nobuyuki, Mori Yosai, Miyata Kazunori. Diagnosability of keratoconus using deep learning with placido Disk-Based corneal topography. Front Med October 2021;8:724902.

[51] Malyugin Boris, Sakhnov Sergej, Izmailova Svetlana, Boiko Ernest, Pozdeyeva Nadezhda, Axenova Lyubov, Axenov Kirill, Titov Aleksej, Anna Terentyeva, Zakaraiia Tamriko, Myasnikova Viktoriya. Keratoconus diagnostic and treatment algorithms based on Machine- Learning methods. Diagnostics October 2021;11(10).