Artificial intelligence for detecting keratoconus

Cochrane Library

Cochrane Database of Systematic Reviews

Artificial intelligence for detecting keratoconus (Review)

Vandevenne MMS, Favuzza E, Veta M, Lucenteforte E, Berendschot TTJM, Mencucci R, Nuijts RMMA, Virgili G, Dickman MM

Vandevenne MMS, Favuzza E, Veta M, Lucenteforte E, Berendschot TTJM, Mencucci R, Nuijts RMMA, Virgili G, Dickman MM.
Artificial intelligence for detecting keratoconus.
Cochrane Database of Systematic Reviews 2023, Issue 11. Art. No.: CD014911.
DOI: 10.1002/14651858.CD014911.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Artificial intelligence for detecting keratoconus

Magali MS Vandevenne[1], Eleonora Favuzza[2], Mitko Veta[3], Ersilia Lucenteforte[4], Tos TJM Berendschot[1], Rita Mencucci[2], Rudy MMA Nuijts[1], Gianni Virgili[2,5], Mor M Dickman[1]

[1]University Eye Clinic Maastricht, Maastricht University Medical Center (MUMC+), Maastricht, Netherlands. [2]Department of Neurosciences, Psychology, Pharmacology and Child Health, University of Florence, Florence, Italy. [3]Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands. [4]Department of Statistics, Computer Science and Applications «G. Parenti», University of Florence, Florence, Italy. [5]Queen's University Belfast, Belfast, UK

Contact: Mor M Dickman, mor.dickman@mumc.nl.

Editorial group: Cochrane Eyes and Vision Group.
Publication status and date: New, published in Issue 11, 2023.

Citation: Vandevenne MMS, Favuzza E, Veta M, Lucenteforte E, Berendschot TTJM, Mencucci R, Nuijts RMMA, Virgili G, Dickman MM. Artificial intelligence for detecting keratoconus. *Cochrane Database of Systematic Reviews* 2023, Issue 11. Art. No.: CD014911. DOI: 10.1002/14651858.CD014911.pub2.

ABSTRACT

Background

Keratoconus remains difficult to diagnose, especially in the early stages. It is a progressive disorder of the cornea that starts at a young age. Diagnosis is based on clinical examination and corneal imaging; though in the early stages, when there are no clinical signs, diagnosis depends on the interpretation of corneal imaging (e.g. topography and tomography) by trained cornea specialists. Using artificial intelligence (AI) to analyse the corneal images and detect cases of keratoconus could help prevent visual acuity loss and even corneal transplantation. However, a missed diagnosis in people seeking refractive surgery could lead to weakening of the cornea and keratoconus-like ectasia. There is a need for a reliable overview of the accuracy of AI for detecting keratoconus and the applicability of this automated method to the clinical setting.

Objectives

To assess the diagnostic accuracy of artificial intelligence (AI) algorithms for detecting keratoconus in people presenting with refractive errors, especially those whose vision can no longer be fully corrected with glasses, those seeking corneal refractive surgery, and those suspected of having keratoconus. AI could help ophthalmologists, optometrists, and other eye care professionals to make decisions on referral to cornea specialists.

Secondary objectives

To assess the following potential causes of heterogeneity in diagnostic performance across studies.

- Different AI algorithms (e.g. neural networks, decision trees, support vector machines)
- Index test methodology (preprocessing techniques, core AI method, and postprocessing techniques)
- Sources of input to train algorithms (topography and tomography images from Placido disc system, Scheimpflug system, slit-scanning system, or optical coherence tomography (OCT); number of training and testing cases/images; label/endpoint variable used for training)
- Study setting
- Study design
- Ethnicity, or geographic area as its proxy
- Different index test positivity criteria provided by the topography or tomography device
- Reference standard, topography or tomography, one or two cornea specialists
- Definition of keratoconus
- Mean age of participants

• Recruitment of participants
• Severity of keratoconus (clinically manifest or subclinical)

Search methods

We searched CENTRAL (which contains the Cochrane Eyes and Vision Trials Register), Ovid MEDLINE, Ovid Embase, OpenGrey, the ISRCTN registry, ClinicalTrials.gov, and the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). There were no date or language restrictions in the electronic searches for trials. We last searched the electronic databases on 29 November 2022.

Selection criteria

We included cross-sectional and diagnostic case-control studies that investigated AI for the diagnosis of keratoconus using topography, tomography, or both. We included studies that diagnosed manifest keratoconus, subclinical keratoconus, or both. The reference standard was the interpretation of topography or tomography images by at least two cornea specialists.

Data collection and analysis

Two review authors independently extracted the study data and assessed the quality of studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. When an article contained multiple AI algorithms, we selected the algorithm with the highest Youden's index. We assessed the certainty of evidence using the GRADE approach.

Main results

We included 63 studies, published between 1994 and 2022, that developed and investigated the accuracy of AI for the diagnosis of keratoconus. There were three different units of analysis in the studies: eyes, participants, and images. Forty-four studies analysed 23,771 eyes, four studies analysed 3843 participants, and 15 studies analysed 38,832 images.

Fifty-four articles evaluated the detection of manifest keratoconus, defined as a cornea that showed any clinical sign of keratoconus. The accuracy of AI seems almost perfect, with a summary sensitivity of 98.6% (95% confidence interval (CI) 97.6% to 99.1%) and a summary specificity of 98.3% (95% CI 97.4% to 98.9%). However, accuracy varied across studies and the certainty of the evidence was low.

Twenty-eight articles evaluated the detection of subclinical keratoconus, although the definition of subclinical varied. We grouped subclinical keratoconus, forme fruste, and very asymmetrical eyes together. The tests showed good accuracy, with a summary sensitivity of 90.0% (95% CI 84.5% to 93.8%) and a summary specificity of 95.5% (95% CI 91.9% to 97.5%). However, the certainty of the evidence was very low for sensitivity and low for specificity.

In both groups, we graded most studies at high risk of bias, with high applicability concerns, in the domain of patient selection, since most were case-control studies. Moreover, we graded the certainty of evidence as low to very low due to selection bias, inconsistency, and imprecision.

We could not explain the heterogeneity between the studies. The sensitivity analyses based on study design, AI algorithm, imaging technique (topography versus tomography), and data source (parameters versus images) showed no differences in the results.

Authors' conclusions

AI appears to be a promising triage tool in ophthalmologic practice for diagnosing keratoconus. Test accuracy was very high for manifest keratoconus and slightly lower for subclinical keratoconus, indicating a higher chance of missing a diagnosis in people without clinical signs. This could lead to progression of keratoconus or an erroneous indication for refractive surgery, which would worsen the disease.

We are unable to draw clear and reliable conclusions due to the high risk of bias, the unexplained heterogeneity of the results, and high applicability concerns, all of which reduced our confidence in the evidence.

Greater standardization in future research would increase the quality of studies and improve comparability between studies.

PLAIN LANGUAGE SUMMARY

How accurate is artificial intelligence for diagnosing keratoconus?

Key messages

• The included studies suggest that artificial intelligence (AI) can identify keratoconus. This may lead to early detection and prevention of vision loss.
• Estimates were similar for different types of AI algorithms.
• We have little confidence in the evidence; there is a need for more research on this topic.

What is keratoconus and why is (early) diagnosis so important?

Keratoconus is a disease of the cornea (the clear window at the front of the eye) that affects people between the ages of 10 and 40 years. In those affected, the cornea weakens and thins over the years, gradually bulging into the typical cone-like shape, which leads to reduced vision. Glasses can resolve this problem in the early stages of keratoconus, but no longer offer a satisfying solution as the disease becomes more severe. Early diagnosis is imperative to ensure follow-up and treatment and thus prevent loss of vision.

The diagnosis of keratoconus is based on an eye exam (measuring the eye and evaluating the cornea with a vertical beam of light and a microscope) and imaging (computer-assisted techniques that create three-dimensional pictures or maps of the cornea). Interpreting the images can be challenging, especially in primary eye care settings and in the early stages of the disease. Not recognizing keratoconus could lead to worsening of the disease and worsening of vision. For example, people at risk of developing keratoconus who undergo refractive surgery (surgery to correct their vision) could end up with worse vision.

What is artificial intelligence and how can it help detect keratoconus?

Detecting keratoconus based on images is challenging, especially for untrained clinicians. AI gives machines the ability to adapt, reason, and find solutions. Algorithms can be developed and trained to analyse images of the cornea and recognize keratoconus. These tests could help ophthalmologists, optometrists, and other eye care professionals to make a diagnosis and refer people with keratoconus to cornea specialists in time to preserve their vision. There are many different types of algorithms, but they all distinguish between healthy eyes and keratoconus based on images of the cornea.

What did we want to find out?

The aim of the review was to find out whether AI can correctly diagnose keratoconus in people seeking refractive surgery and people whose vision can no longer be corrected fully with glasses.

What did we do?

We searched for studies that investigated the accuracy of AI for diagnosing keratoconus, preferably in people seeking refractive surgery or people whose vision can no longer be corrected fully with glasses. We compared and summarized the results of the studies to calculate two measures of accuracy: sensitivity (the ability of AI to correctly identify keratoconus) and specificity (the ability of AI to correctly rule out keratoconus). The closer sensitivity and specificity were to 100%, the better the algorithm.

What did we find?

We found 63 studies that used three different units (eyes, participants, and images) to analyse the accuracy of AI for detecting keratoconus: 44 studies analysed 23,771 eyes, four studies analysed 3843 participants, and 15 studies analysed 38,832 images.

The accuracy of AI for detecting manifest keratoconus (keratoconus that can be detected through a clinical examination) was high. If 1000 people were tested, 30 people with keratoconus would be correctly referred to a cornea specialist, and none would be missed. Of the remaining 970 people (without keratoconus), only 17 would be wrongly referred. These people would receive additional non-invasive tests to verify whether they had keratoconus.

The accuracy of AI for detecting early keratoconus was lower. If 1000 people were tested, nine people with keratoconus would be correctly referred to a cornea specialist and one would be missed. If this person received refractive surgery, it would aggravate the disease and worsen their vision. Of the remaining 990 people (without keratoconus), 941 would be reassured that they did not have the disease and would receive refractive surgery or glasses; 49 people would be wrongly referred.

The evidence suggests that AI may be good at detecting manifest keratoconus but may not be ideal for screening early keratoconus.

What are the limitations of the evidence?

We have little confidence in the evidence on the accuracy of AI for detecting manifest keratoconus, and we have little to no confidence in the evidence related to early keratoconus. There were problems with how the studies were conducted, which may result in AI appearing more accurate than it really is.

How up-to-date is this evidence?

The evidence is up-to-date to 29 November 2022.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings: artificial intelligence for the detection of keratoconus in refractive surgery candidates and people with refractive errors

Review Question	What is the diagnostic accuracy of AI algorithms in the detection of keratoconus in people presenting with refractive errors, people seeking corneal refractive surgery, or people suspected of having keratoconus?
Population	People presenting with refractive errors, especially those whose vision can no longer be corrected fully with glasses, people seeking corneal refractive surgery, or people suspected of having keratoconus
Index test	AI algorithms e.g. neural network, logistic regression, support vector machine, etc. analysing topography and tomography images
Target condition	Keratoconus
Reference standard	Topography and tomography images interpreted by at least two cornea specialists
Action	(Early) referral of people suspected of having keratoconus to a cornea specialist by ophthalmologists, optometrists, and other eye care professionals.
Quantity of evidence	63 studies

Outcome	Effect (95% CI)	Number of participants (studies)	Test result	Number of results per 1000 participants tested		Certainty of evidence
				Real clinical setting*	Included studies**	
Manifest keratoconus	**Summary sensitivity** **98.6%** (97.6% to 99.1%)	21,330 (54)	True positive	30 (29 to 30)	493 (488 to 496)	⊕⊕## **Low**[a]
			False negative	0 (0 to 1)	7 (5 to 12)	
	Summary specificity **98.3%** (97.4% to 98.9%)	29,189 (54)	True negative	954 (945 to 959)	492 (487 to 495)	⊕⊕## **Low**[a]
			False positive	16 (11 to 27)	9 (6 to 13)	

Subclinical keratoconus	**Summary sensitivity** **90.0%** (84.5% to 93.8%)	2758 (28)	True positive	9 (8 to 9)	225 (211 to 235)	⊕### **Very low**[b]
			False negative	1 (1 to 2)	25 (16 to 39)	
	Summary specificity **95.5%** (91.9% to 97.5%)	6750 (28)	True negative	945 (911 to 970)	716 (687 to 731)	⊕⊕## **Low**[a]
			False positive	45 (20 to 79)	34 (19 to 63)	

*Estimated prevalence in the real clinical setting was 3% for manifest keratoconus and 1% for subclinical keratoconus.
**Prevalence calculated from the included studies was 50% for manifest keratoconus and 25% for subclinical keratoconus.

AI: artificial intelligence; **CI:** confidence interval.

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

[a] Downgraded one level for risk of bias (high risk of selection bias and other biases due to case-control design with indirectness) and one level for inconsistency (sensitivity varies widely across studies).
[b] Downgraded one level for risk of bias (high risk of selection bias and other biases due to case-control design with indirectness), one level for inconsistency (sensitivity varies widely across studies), and one level for imprecision (wide CIs for sensitivity).

BACKGROUND

Target condition being diagnosed

Keratoconus is an ectatic degenerative disorder of the cornea, usually affecting both eyes. Ultra-structural examination of the human cornea ex vivo has revealed disruption and loss of the native collagen network, leading to biomechanical instability and severe corneal thinning (Hayes 2012; Meek 2005). The disease is generally progressive in nature, resulting in the cornea taking a typical cone shape. This causes myopia and irregular astigmatism, impairing visual acuity.

Normally, keratoconus begins during puberty and gradually progresses until the person affected is in their 30s. It usually progresses more rapidly in people younger than 17 years and stabilizes with age (Ferdi 2019). The reported prevalence of keratoconus varies among studies (Hashemi 2020). This may be due to several reasons, such as different diagnostic criteria, different diagnostic methods, change in testing rates over time, genetic variation, or environmental differences.

The pathophysiology of keratoconus is not well understood; however, both environmental and genetic factors seem to play a role (Rabinowitz 2021). One risk factor that has been investigated extensively is eye rubbing; others include the wearing of contact lenses and allergic disease. Research on the genetic contribution to keratoconus suggests a possible association (Rabinowitz 2021). However, diagnostic genetic testing for keratoconus is not currently available.

Some people who undergo refractive surgery may be at risk of developing iatrogenic keratoectasia (i.e. weakening of the biomechanical stability of the cornea due to the surgery), which leads to a keratoconus-like ectasia. Although this is not a frequent occurrence (Giri 2017), the consequences can be sight-threatening, so it is crucial to detect corneas at risk of developing the condition. Possible risk factors are irregular topography and thin corneal pachymetry (Giri 2017).

The treatment for keratoconus depends on the severity of the disease. In the initial stage, the aim of treatment is usually to correct visual acuity with glasses or specialized contact lenses. However, these treatments do not cure keratoconus. As the disease progresses, visual acuity can worsen to the point that glasses no longer offer a satisfactory solution. Corneal cross-linking has been used since 2003 to stop the progression of keratoconus, but this treatment cannot reverse visual impairment (Sykakis 2015). Before the introduction of corneal cross-linking, the only treatment to cure keratoconus was corneal transplantation. Despite the development of cross-linking, keratoconus is still one of the most common reasons for corneal transplantation (Kelly 2011; Röck 2018). Thus, the diagnosis of keratoconus may help to avoid poor visual outcomes and possible corneal transplantation, especially if the diagnosis is made early.

Keratoconus diagnosis is based primarily on corneal topographic and tomographic analysis in people presenting with refractive errors, especially those whose vision can no longer be fully corrected with glasses, and those seeking corneal refractive surgery. A global consensus committee of ophthalmology experts concluded that "abnormal posterior ectasia, abnormal corneal thickness distribution, and clinical non-inflammatory corneal thinning are mandatory findings to diagnose keratoconus" (Gomes 2015). However, applying this definition in practice is not straightforward: because the consensus mentions no cut-offs or parameters, the definition is open to the interpretation of the specialist. Ocular findings that may point to early keratoconus include abnormal keratometry readings and a distorted red reflex when using an ophthalmoscope, both of which indicate an irregular cornea. Detecting keratoconus at an early stage may be challenging, as people affected are often asymptomatic, and there are few or no clinical signs. In later stages of the disease, clinical signs are visible during slit-lamp examination and include stromal thinning, conical protrusion of the cornea at the apex, Fleischer's ring, and Vogt's striae (Zadnik 1996). Another challenge in the diagnosis of keratoconus is detecting an at-risk cornea or subclinical keratoconus in people seeking corneal refractive surgery. Iatrogenic keratoectasia due to biomechanical decompensation may occur in these people if the disease is not detected (Giri 2017).

Currently, there is no accurate and objective method to detect keratoconus. An artificial intelligence (AI)-based tool for keratoconus detection could help ophthalmologists, optometrists, and other eye care professionals to make decisions on referral to cornea specialists.

AI is a growing field within ophthalmology, and is expected to play an important role in the diagnosis and characterization of eye diseases in the future. There has been an increasing interest in the application of AI methods for diseases of the anterior segment (Ting 2020). This review will seek to determine if AI is a valid tool for diagnosing keratoconus as an aid for ophthalmologists.

Index test(s)

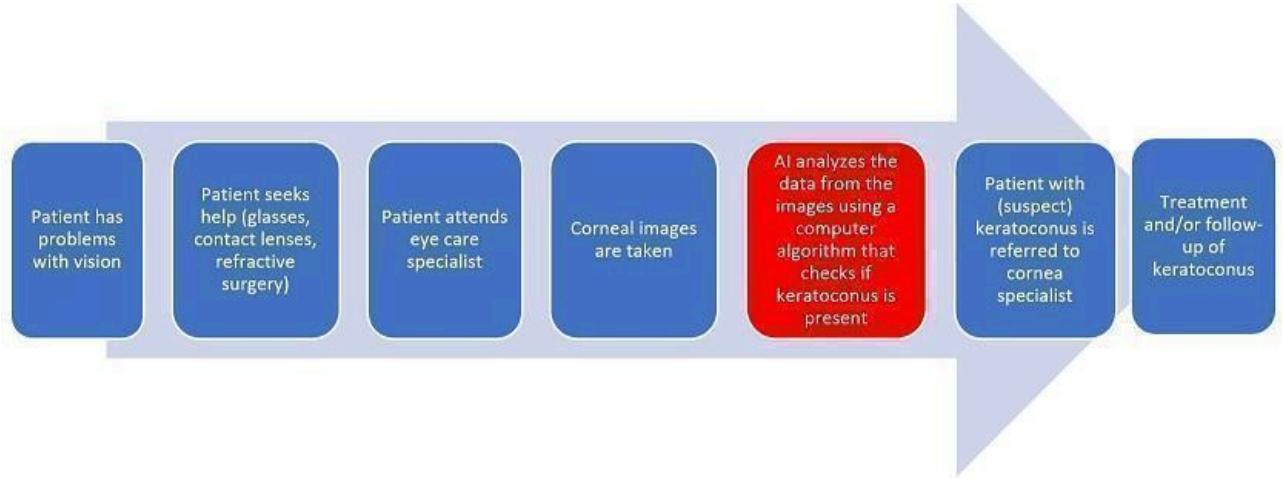
This review will evaluate the application of AI in the diagnosis of keratoconus. AI methods are already contributing to many aspects of human life and society, ranging from home automation, smart assistants (e.g. 'Siri', 'Google Assistant'), and self-driving cars to facial recognition and automatic detection of 'fake news' on social media. There has been notable progress with the use of AI in the field of medical image analysis, including applications in ophthalmology (Ting 2019).

AI provides machines with the capability to adapt, reason, and find solutions. Machine learning is a subdiscipline of AI that enables machines to learn from data and experience through algorithms. Examples of machine learning algorithms are support vector machine, random forest, and decision tree. Deep learning is a subdiscipline of machine learning that uses neural networks, much like the human brain. It teaches machines to learn through pattern recognition and even to improve themselves (LeCun 2015).

Initially, most AI research in ophthalmology focused on the posterior segment. Studies have investigated multiple deep learning applications for several common ophthalmic diseases, including diabetic retinopathy (Abràmoff 2016; Gargeya 2017; Gulshan 2016; Ting 2017), age-related macular degeneration (Grassmann 2018; Ting 2017), glaucoma (Shibata 2018), and retinopathy of prematurity (Brown 2018). More recent research has concentrated on the development of deep learning applications for the anterior segment, in particular keratoconus (Ting 2020).

In keratoconus, the AI algorithm analyses images of the cornea using a computer to determine whether the disease is present (see Figure 1). There are different devices for producing these non-invasive corneal images, which are called topography or tomography images. Most devices (e.g. Scheimpflug-based devices, optical coherence tomography (OCT)) take both tomography and topography images. However, some devices (e.g. Placido disc devices) only take topography images. The image is uploaded to a computer, where the algorithm performs a series of analyses to come to a decision on whether keratoconus is present.

Figure 1. Clinical pathway.



The first step in developing an algorithm is collecting a representative data set for keratoconus, which includes topographic or tomographic images of both keratoconus and healthy eyes. The data set is then divided into training, validation, and test sets. The training set is used to determine the parameters or features of keratoconus via an optimization procedure. The validation set is used for model selection (e.g. determining the best neural network architecture) and monitoring for overfitting (i.e. the algorithm is only applicable to the data on which it was trained). The independent test set is used for evaluation of the model (i.e. determining the performance of the model on unseen data). In principle, the test set should only be used once, after the model is developed and trained. When these three phases are completed, the algorithm will theoretically be able to differentiate keratoconus eyes from healthy eyes.

Each AI algorithm has its own grading system to classify keratoconus and healthy eyes. Depending on the goal of the AI tool (screening or diagnosis), the thresholds of sensitivity and specificity will differ.

Topography was previously considered the best method for diagnosing keratoconus, but according to current guidelines, corneal tomography is now the gold standard (Gomes 2015). Topography only analyses the anterior corneal surface, whereas tomography analyses both anterior and posterior corneal surfaces and can create three-dimensional images, resulting in improved accuracy. In clinical practice, diagnosis of keratoconus involves both tomography and topography parameters, including maximum keratometry, minimal pachymetry, astigmatism, and asphericity. These show only a moderate correlation with keratoconus (Kanellopoulos 2013a; Kanellopoulos 2013b; Lopes 2012; Sedghipour 2012). Most devices also provide objective indices to aid diagnosis, including the keratoconus index, the index of surface variance, and the inferior-superior index. However, these parameters and indices individually do not provide sufficient information, but must be combined and interpreted together (Martínez-Abad 2017). Unfortunately, not all ophthalmologists, optometrists, or eye care professionals have these diagnostic skills. A second issue is the intra- and interobserver variability in the diagnosis of keratoconus (Brunner 2018; Flynn 2016). AI could be a solution to both problems, as it can easily combine tomography and topography parameters and indices based on an enormous amount of data, and it is not affected by diagnostic variability. It could help young ophthalmologists, ophthalmologists in non-academic centres, optometrists, and other eye care specialists to diagnose the disease early and refer affected people to a cornea specialist. In this way, follow-up can start earlier and specialists can detect any progression before visual loss.

Clinical pathway

The clinical pathway to diagnosing keratoconus is based on clinical examination, which includes visual acuity testing, slit-lamp examination of the anterior segment, and corneal imaging (all non-invasive tests). Corneal imaging is performed in people presenting with refractive errors, especially those whose vision can no longer be corrected fully with glasses, those seeking corneal refractive surgery, and those referred by ophthalmologists, optometrists, or other eye care professionals because of suspected keratoconus.

The different methods of corneal imaging include Placido topography, Scheimpflug tomography, and slit-scanning tomography. Interpretation of the images can be challenging, and the signs of keratoconus can be subtle for general ophthalmologists, optometrists, and other eye care professionals. In current practice, the ophthalmologist will analyse the corneal images, looking for patterns and evaluating device-dependent parameters such as keratometry, elevation, and pachymetry parameters (see Appendix 1). As the global consensus mentions no cut-offs in the definition of keratoconus, specialists need to rely on their knowledge and experience, which means the diagnosis is subjective.

After being diagnosed with keratoconus, the person affected will need regular follow-up visits to check if the disease progresses. The global consensus document states that treatment is essential when there is documented clinical progression, defined as a consistent change in at least two of the following parameters where the magnitude of the change is above the normal noise of the testing system (Gomes 2015).

- Steepening of the anterior corneal surface
- Steepening of the posterior corneal surface
- Thinning or an increase in the rate of corneal thickness change from the periphery to the thinnest point

As with the definition of keratoconus, these criteria are open to interpretation. The consensus document mentions no cut-offs, time intervals, or specific parameters.

A missed diagnosis of keratoconus could lead to delayed treatment, poor visual outcome, and a greater risk for corneal transplantation, all of which impact on patients' quality of life, especially because the disease normally affects young people who are active and in their primary income-earning years.

The same corneal images that are analysed by clinicians can be uploaded to a computer and analysed by an AI algorithm. AI based on a large ophthalmic data set can achieve high accuracy in distinguishing a normal cornea from a keratoconus cornea by analysing the topography or tomography images (Lin 2019; Lopes 2019). Since the global consensus does not give an accurate definition of keratoconus or keratoconus progression, AI could be helpful in making this decision. It could help with early diagnosis of keratoconus so that affected people can be monitored and any progression can be detected sooner. Once progression is detected and confirmed by a cornea specialist, the patient would receive corneal cross-linking to halt the deterioration of the disease, which in turn would lead to a better visual prognosis and lower risk of corneal transplantation. Since the cornea specialist is still responsible for the diagnosis, the first role of AI would be as triage to make decisions on referral.

To implement an AI algorithm in clinical practice, it needs to be efficient and able to analyse images in a few seconds. It should give one clear indication of whether keratoconus is present.

Devices that measure biomechanical properties, such as the Corvis ST or the Ocular Response Analyzer (ORA), were not included in this review.

Rationale

AI is a rapidly growing field in ophthalmology, with numerous new developments in the detection of keratoconus (Ting 2020). It is important to have reliable evidence regarding the accuracy of these developments. This review will give a clear overview of the different AI detection tools and their accuracy.

Corneal imaging devices are becoming increasingly sophisticated, and with the help of AI algorithms, they can detect keratoconus earlier. AI uses a vast amount of data to learn characteristic features of keratoconus. It can process thousands of images in a short time to learn how to detect the disease, whereas an ophthalmologist needs years of practice. AI will help ophthalmologists, optometrists, and eye care professionals in the diagnosis of keratoconus, potentially leading to earlier diagnoses. This is beneficial for patients because they may have a better visual outcome, which would improve their quality of life. There are also important financial consequences, in terms of reduced healthcare costs and personal costs.

Nevertheless, AI has its limitations. The accuracy of the algorithms relies on the generalizability of the training sets. If training sets do not contain sufficient data or sufficiently varied data, the algorithms could miss diagnoses due to inadequate learning (LeCun 2015).

One narrative review published in 2019 suggested that AI may be a reliable tool (Lin 2019). Another review discussed AI in the anterior segment and mentioned the detection of keratoconus (Ting 2020). However, neither of these previous reviews determined the reliability of the included studies.

There is a need for a reliable overview of current knowledge on the different existing AI algorithms and an analysis of their accuracy.

OBJECTIVES

To assess the diagnostic accuracy of artificial intelligence (AI) algorithms for detecting keratoconus in people presenting with refractive errors, especially those whose vision can no longer be fully corrected with glasses, those seeking corneal refractive surgery, and those suspected of having keratoconus. AI could help ophthalmologists, optometrists, and other eye care professionals to make decisions on referral to cornea specialists.

Secondary objectives

To assess the following potential causes of heterogeneity in diagnostic performance across studies.

- Different AI algorithms (e.g. neural networks, decision trees, support vector machines)
- Index test methodology (preprocessing techniques, core AI method, and postprocessing techniques)
- Sources of input to train algorithms (topography and tomography images from Placido disc system, Scheimpflug system, slit-scanning system, or optical coherence tomography (OCT); number of training and testing cases/images; label/endpoint variable used for training)
- Study setting
- Study design
- Ethnicity, or geographic area as its proxy
- Different index test positivity criteria provided by the topography or tomography device
- Reference standard, topography or tomography, one or two cornea specialists
- Definition of keratoconus
- Mean age of participants
- Recruitment of participants
- Severity of keratoconus (clinically manifest or subclinical)

METHODS

Criteria for considering studies for this review

Types of studies

We included cross-sectional studies and diagnostic case-control studies (either prospective or retrospective).

We organized the included studies based on the main characteristics of the AI methodology (preprocessing techniques, core AI method, and postprocessing techniques), data that were used to train the model (patient inclusion criteria, number of training and testing cases/images, label/endpoint variable used for training), and evaluation (evaluation metric, reported performance on the independent test set).

Participants

We aimed to include people with refractive errors:

- whose vision could not be fully corrected with glasses; or
- who were seeking refractive surgery; or
- who had suspected keratoconus (for whom a decision was to be made on referral to cornea specialists).

However, research in this field is still in its early stages, and we accepted studies that did not satisfy this optimal definition of participants, such as case-control studies that included people with keratoconus and healthy controls based on different sets of criteria.

As keratoconus can progress until the fourth decade of life, we included participants up to the age of 50 years.

Index tests

We included studies reporting accuracy data for automated diagnostic tests. All AI algorithms developed to analyse corneal topography or tomography for detecting keratoconus were eligible.

Target conditions

The target condition was keratoconus of any stage. When studies reported accuracy for multiple severity levels, we prioritized data from participants with at least mild severity. In fact, fruste keratoconus is generally non-progressive, or very slowly progressive.

Reference standards

The reference standard for keratoconus is topography or tomography. These non-invasive examinations are routinely performed on people who are seeking refractive surgery or people referred to an ophthalmologist for suspected keratoconus. Two or more cornea specialists should independently analyse and interpret the corneal images. We accepted studies that used only one cornea specialist for diagnosis as a low-quality reference standard.

Topography examines the anterior corneal surface. The Placido disc system is a device that uses topography. Concentric rings of light are projected on the cornea. Thousands of points along these concentric rings are analysed, and these data are translated to the curvature of the anterior corneal surface (Fan 2018). The main parameters measured by Placido systems are maximum keratometry, steep keratometry, flat keratometry, and astigmatism (see Appendix 1).

Tomography examines both the anterior and posterior corneal surfaces. The Scheimpflug system uses a single rotating Scheimpflug camera (e.g. Pentacam (Oculus GmbH, Wetzlar, Germany)), a single rotating Scheimpflug camera combined with Placido disc topography (Sirius, CSO, Italy), or a dual-Scheimpflug camera with Placido disc technology incorporated to improve curvature information on the central cornea (e.g. the Galilei (Ziemer, Biel, Switzerland)). Another device that examines both anterior and posterior corneal surfaces is the slit-scanning system; this is an elevation-based method for the assessment of topography and tomography (e.g. the Orbscan IIz (Bausch & Lomb, Rochester, NY)). Multiple complimentary slits are used to perform an assessment of the corneal surface. In addition to keratometry (which is also measured by the Placido systems), the Scheimpflug system and slit-scanning system measure corneal elevation and pachymetry (see Appendix 1).

OCT also examines both the anterior and posterior corneal surfaces. Anterior segment OCT (AS-OCT) uses low-coherence interferometry to assess the cornea and the anterior segment. The low-coherent light is emitted and split by an interferometer into a reference beam and a probe beam (Wojtkowski 2010). The probe beam is backscattered from the different corneal layers. The echo time delay is measured and transformed into two- or three-dimensional images by the OCT (Subhash 2013). A new instrument called the MS-39 (CSO, Italy) combines Placido disc corneal topography with high-resolution OCT-based anterior segment tomography. It measures keratometry, elevation, pachymetry, and other parameters.

Search methods for identification of studies

Electronic searches

The Cochrane Eyes and Vision Information Specialist searched the following electronic databases and trials registries on 29 November 2022. There were no restrictions on language or date of publication.

- Cochrane Central Register of Controlled Trials (CENTRAL; 2022, Issue 11), which contains the Cochrane Eyes and Vision Trials Register, in the Cochrane Library (searched 29 November 2022; Appendix 2)
- MEDLINE Ovid (1946 to 29 November 2022; Appendix 3)
- Embase Ovid (1980 to 29 November 2022; Appendix 4)
- System for Information on Grey Literature in Europe (OpenGrey; 1995 to 29 November 2022; Appendix 5)
- ISRCTN registry (www.isrctn.com/editAdvancedSearch; searched 29 November 2022; Appendix 6)
- US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (www.clinicaltrials.gov; searched 29 November 2022; Appendix 7)
- World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp; searched 29 November 2022; Appendix 8)

Searching other resources

We searched the reference lists of the review's included studies.

Data collection and analysis

Selection of studies

Two review authors (MV and EF) independently evaluated the records retrieved by the searches using the online review management software Covidence. They screened the titles of the records and eliminated those that were clearly ineligible. The same two review authors then assessed the full-text articles of the remaining records against our inclusion criteria. They resolved any disagreements by discussion or, if necessary, by involving the other review authors.

Data extraction and management

The two review authors independently extracted the following data from the selected articles using a standardized data collection form.

- Study design
- Study population
- Definition of keratoconus
- Reference standard
- Index tests
- Description of architecture and training mechanisms
- The ground truth (one observer versus multiple observers)
- Size of data sets used
- Data required to fill in a 2 × 2 diagnostic contingency table for each index test

We compared the data collected independently by the two review authors, and resolved any discrepancies through discussion and consensus. If we needed to obtain further data from a paper, or if there were missing data, we tried to contact the study author for further information.

When an article reported multiple AI algorithms, we selected the algorithm with the highest Youden's index. We are aware that this selection could inflate accuracy, especially in smaller studies, and we highlighted this as a limitation. However, we considered this decision acceptable in this early stage of research as it could also reduce redundancy. Examples of algorithms are random forest, support vector machine, decision tree, and neural network.

We used GRADEpro GDT to create a summary of findings table (GRADEpro GDT).

Assessment of methodological quality

Two authors (MV and EF) independently assessed the included studies for bias using the revised Quality Assessment of Diagnostic Accuracy Studies tool (QUADAS-2), as described in Chapter 9 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Reitsma 2009). The QUADAS-2 tool has four assessment domains: patient selection, index test, reference test, and flow and timing. Each domain has signalling questions to assess the risk of bias. The tool also assesses applicability for the first three domains.

During the quality assessment process, we decided to add an item that is specific to AI studies (Appendix 9). In the domain 'Index test', we added the question 'Was the model designed in an appropriate manner?'. We considered a study at low risk of bias if data from a single participant were reserved to only one data partition, parameters were tuned, and the optimal model was selected. We considered a study at high risk of bias if data from a single participant were not reserved to only one data partition, parameters were not tuned, and the optimal model was not selected. When the design of the model was unclear, and we could not determine the above-mentioned properties, we considered the study at unclear risk. In the protocol for this review, we stated that the 'Concerns regarding applicability' question in the domain 'Reference standard' ('Are there concerns that the target condition as defined by the reference standard does not match the review question?') was not applicable to this review (Vandevenne 2021); however, we corrected this during quality assessment. Additionally, in the domain 'Flow and timing', we removed the question 'Was there an appropriate interval between index test(s) and reference standard?', as it was not applicable to this review. The reference test and index test were performed on the same corneal images or parameters, so the interval between the index and reference test is irrelevant.

Statistical analysis and data synthesis

We conducted all statistical analysis and data synthesis in accordance with Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Macaskill 2010).

Initially, we presented data in a 2 × 2 table, showing cross-classification of the index test result versus the reference standard outcome. For each index test, in all studies, we calculated the sensitivity and specificity with a 95% confidence interval (CI). To visually evaluate the variation in calculations of sensitivity and specificity, we used Review Manager 5 (RevMan 5) to generate coupled forest plots and present studies in receiver operating characteristic (ROC) plots (Review Manager 2020).

Since AI studies were unlikely to give a definite threshold that would be comparable across studies, we had planned to use a hierarchical summary ROC (HSROC) model and estimate the average sensitivity at fixed specificity values according to cut-offs for terciles of specificity (Macaskill 2010). However, we found accuracy was nearly perfect in the vast majority of studies, which clustered close to the upper-left corner of the ROC space. Therefore, we pooled data using a bivariate model, which is equivalent to an HSROC model in absence of covariates (Harbord 2007). We conducted analyses using the 'metadas' user-written command in SAS software, as recommended in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Takwoingi 2022).

We had planned to conduct direct comparisons between the index tests (different types or data sources for AI) if sufficient data were available. However, few studies presented paired data for this comparison, so we decided to explore heterogeneity between studies in subgroup analyses. We conducted these analyses with a test covariate in the bivariate model as suggested in Chapter 10 of the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Takwoingi 2022).

Investigations of heterogeneity

To investigate heterogeneity, where data were available, we added covariates in a meta-regression, using the sources presented in the Objectives. We used all covariates as categorical variables.

Sensitivity analyses

We conducted a sensitivity analysis by excluding studies that used images as the unit of analysis, since sample sizes could exceed the number of participants by several times.

Certainty of the evidence assessment

We graded the certainty of evidence for each outcome using the GRADE approach and following Chapter 12 of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (Leeflang 2022). GRADE considers five domains: risk of bias, indirectness, inconsistency, imprecision, and publication bias. We explained all decisions to downgrade the certainty of evidence using footnotes to the summary of findings tables. We decided post hoc to adopt a threshold of 0.95 as desirable sensitivity to assess imprecision in GRADE, given a triage test role. We set the threshold for specificity at 0.90, considering the workload generated at low prevalence.

Assessment of reporting bias

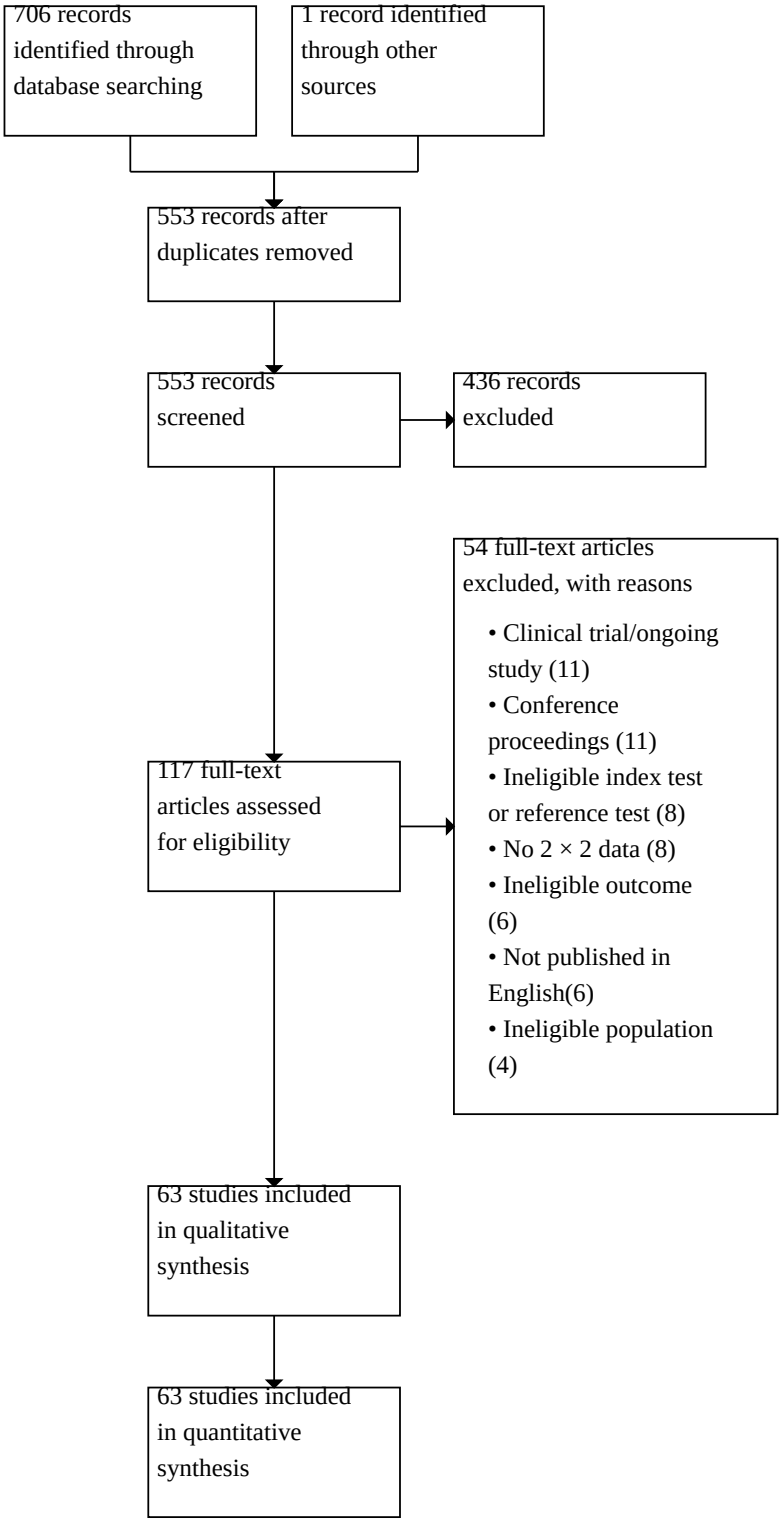
We assessed reporting biases when a study protocol was available. We attempted to maximize data collection by using comprehensive search methods and contacting study authors when we needed further information to reach a decision on study eligibility.

RESULTS

Results of the search

The searches yielded a total of 707 records (Figure 2). After deduplication, we assessed the titles and abstracts of 553 records, of which we considered 436 to be clearly irrelevant. We retrieved and read the full-text articles of the remaining 117 records, excluding 54 articles for different reasons (see Figure 2). We included 63 studies (63 reports) in the final qualitative synthesis.

Figure 2. Flow diagram illustrating study selection process.



Included studies

Twelve studies had a prospective design. Fifty-eight were case-control studies; the remaining five were cross-sectional studies. Only one study had randomized allocation; the rest were non-randomized or had an unclear sampling. In the protocol, we described the study population as "patients with refractive errors, whose vision cannot be fully corrected with glasses, patients seeking refractive surgery or patients suspected of keratoconus, for whom a decision is to be made on referral to cornea specialists" (Vandevenne 2021). Of the 63 studies, only 17 included refractive surgery candidates, and only one included referred patients. The remaining studies included people with diagnosed keratoconus and healthy controls. More extensive details on these articles are available in the Characteristics of included studies table.

Excluded studies

Of the 54 articles excluded during full-text analysis, 11 were study protocols and 11 were conference proceedings. We contacted the study authors, but no additional data were available. Eight articles had an ineligible index test or reference test (e.g. devices that measure the biomechanical properties of the cornea). Eight studies had no 2 × 2 data available to compute sensitivity and specificity. We only included studies published in English, so we excluded six articles published in a different language. Four studies reported ineligible outcomes, and four studies included ineligible populations (e.g. allergic eye disease).

Methodological quality of included studies

Figure 3 and Figure 4 give an overview of the methodological quality of the included studies.

Figure 3. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies.

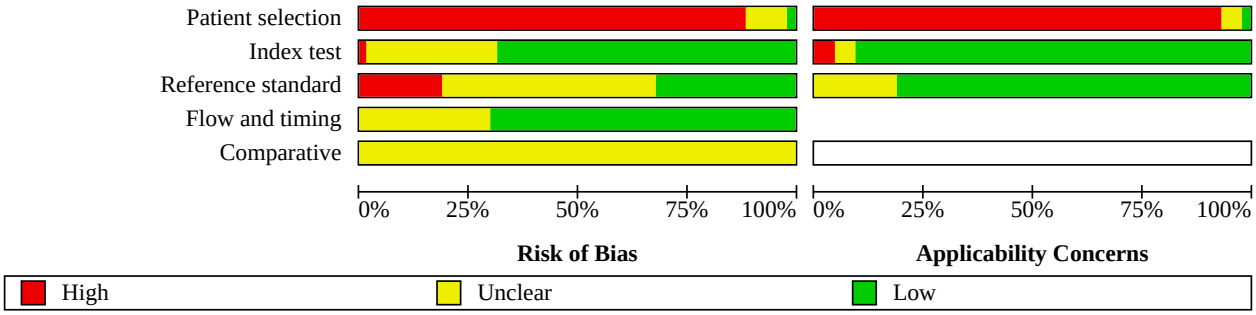


Figure 4. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study.

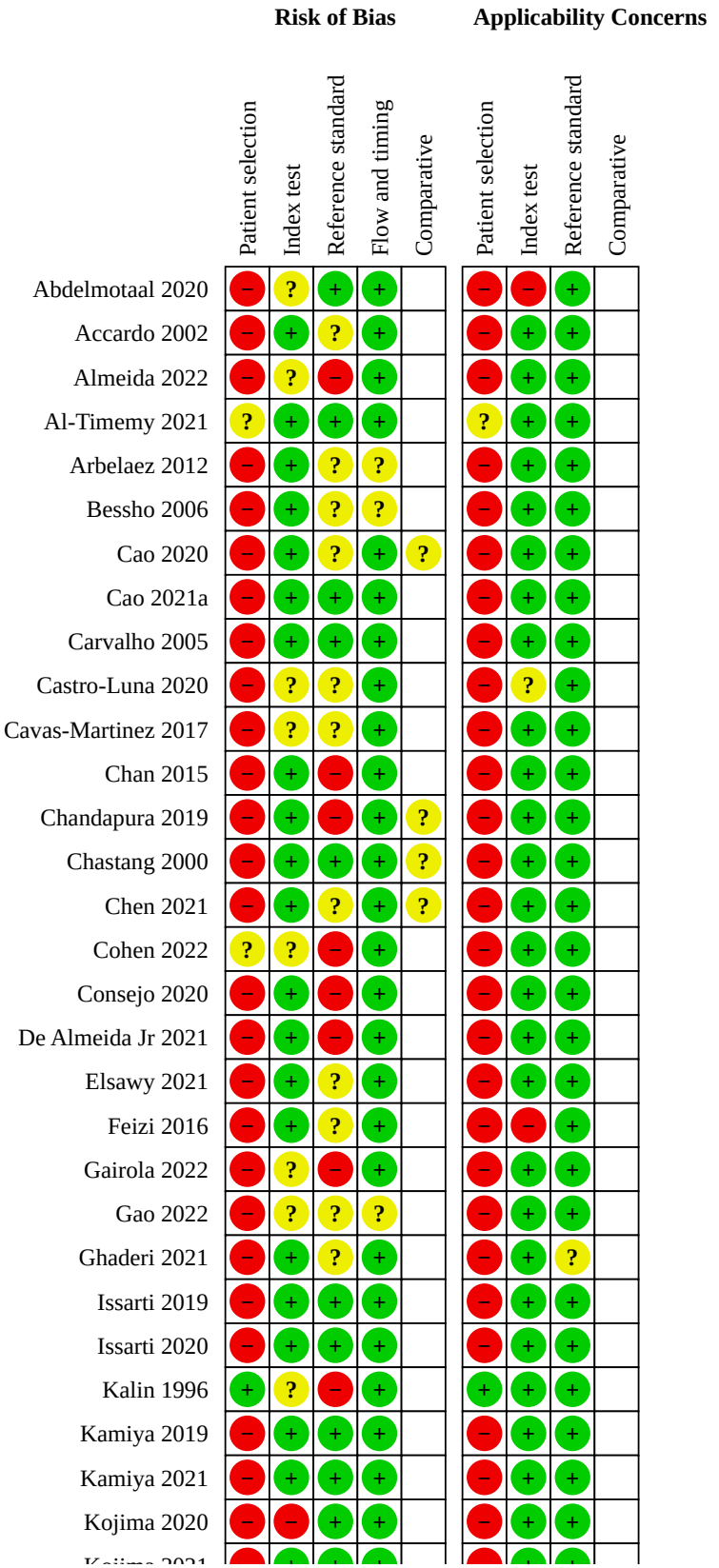
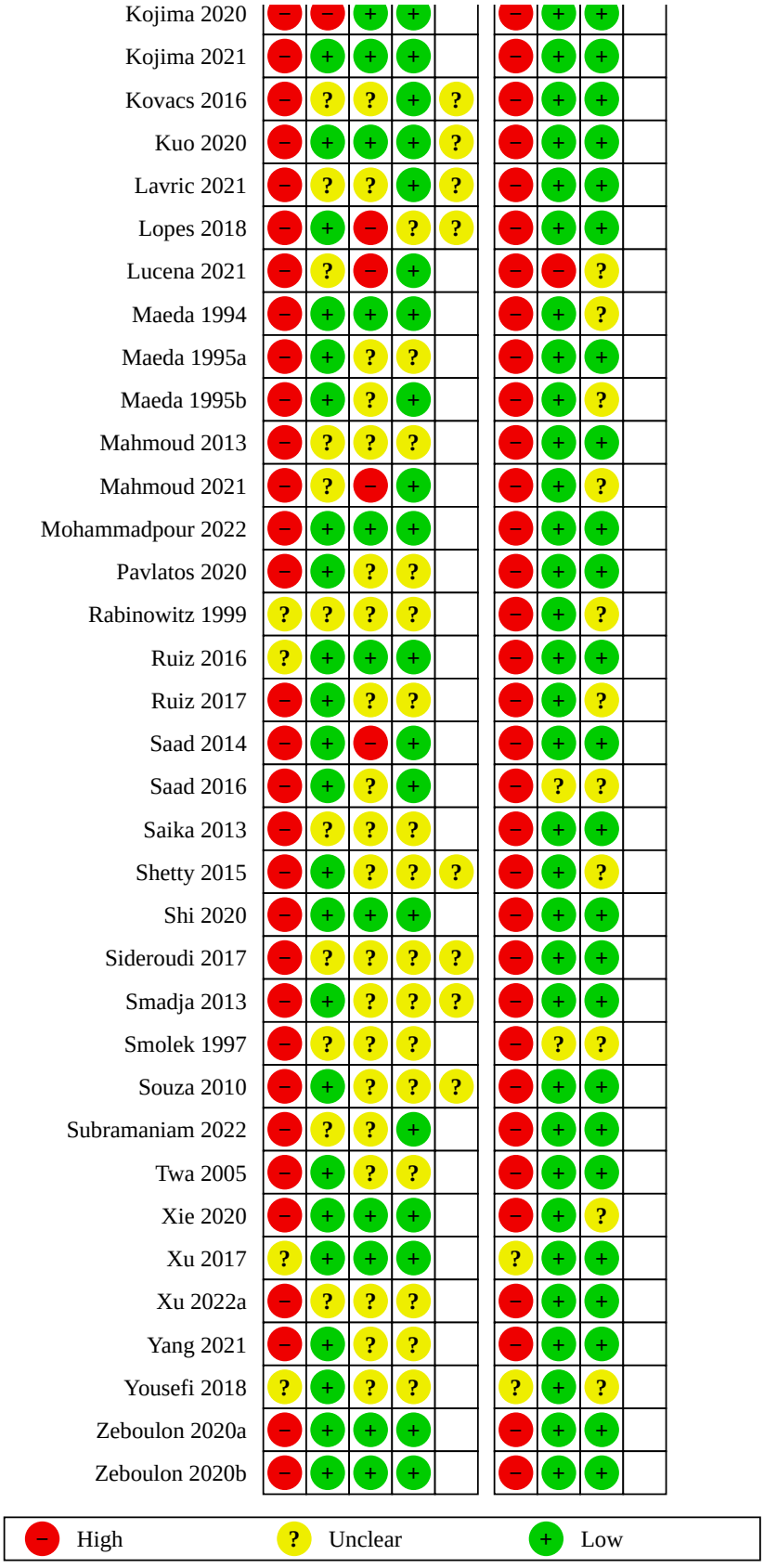


Figure 4. (Continued)



Regarding the first domain, 'Patient selection', we considered most studies at very high risk of bias due to their case-control design, as the participants were already diagnosed with keratoconus and did not fit the definition stated in our protocol (Vandevenne 2021). Kalin 1996 was the only study that we judged at low risk of bias in this domain, as it included consecutive people seeking refractive surgery. Six studies were at unclear risk of bias because they provided insufficient information on sampling. With respect to applicability, we considered concern was low in Kalin 1996 only. There was unclear concern for three studies because they provided an insufficient description of the population. We considered concern was high for the remaining 59 studies because they all included people attending cornea services for known disease, were population-based studies, or were registry-based studies.

When evaluating the second domain, 'Index test', we judged Kojima 2020 at high risk of bias because it did not provide an appropriate description of the AI algorithm. We judged 18 studies at unclear risk of bias, mainly because they provided an unclear description of the study design. We considered the remaining 45 studies at low risk of bias.

In the third domain, 'Reference standard', we judged 17 studies at high risk of bias: although almost all reference standard diagnoses preceded the index test diagnoses, none of the 17 studies had two or more cornea specialists interpreting the images. We considered 18 studies at low risk of bias because at least two cornea specialists interpreted the images to diagnose keratoconus. The remaining 28 studies were at unclear risk of bias because they did not clearly state whether the results were interpreted independently. Regarding applicability, we considered there was unclear concern for 13 studies and low concern for the other 50 studies.

Regarding the fourth domain, 'Flow and timing', we judged 23 studies at unclear risk of bias because it was unclear whether all participants had received the same reference standard. In the studies that used a pre-existing database of people with keratoconus, it was unclear how the diagnosis was established. The remaining 43 studies were at low risk of bias.

In the last domain, 'Comparative', we judged risk of bias and applicability concerns in the 12 studies that developed and compared multiple AI algorithms. We considered all 12 studies at unclear risk of bias because none indicated whether the results of the different algorithms were interpreted independently. Only a few mentioned the size of the data set used to validate the different tests. We could not use directly comparative data as these were sparse and difficult to group. Thus, we conducted indirect comparisons between AI algorithms.

Findings

We included 63 studies that developed and investigated the accuracy of an AI algorithm for the diagnosis of keratoconus. The studies were published between 1994 and 2022. The median prevalence of keratoconus across all studies was 45% (interquartile range 28% to 54%, range 6% to 94%).

Characteristics of included studies

Table 1 shows the main characteristics of all the included studies, such as study design, population, sample size, country, instrument, index test, and reference test. Most studies (58) had a case-control design, and the other six studies had a cross-sectional design. Most studies had a single- or multicentre data set of people with keratoconus and controls. The controls were healthy people or refractive surgery candidates in most cases, though some studies included people with other ocular diseases (for details see Table 1). The sample size was often large, which is necessary for the development and training of an AI algorithm. Only five studies included fewer than 100 people (Cao 2020; Carvalho 2005; Castro-Luna 2020; Consejo 2020; Xu 2022a). The studies included data from 21 different countries and represented all continents.

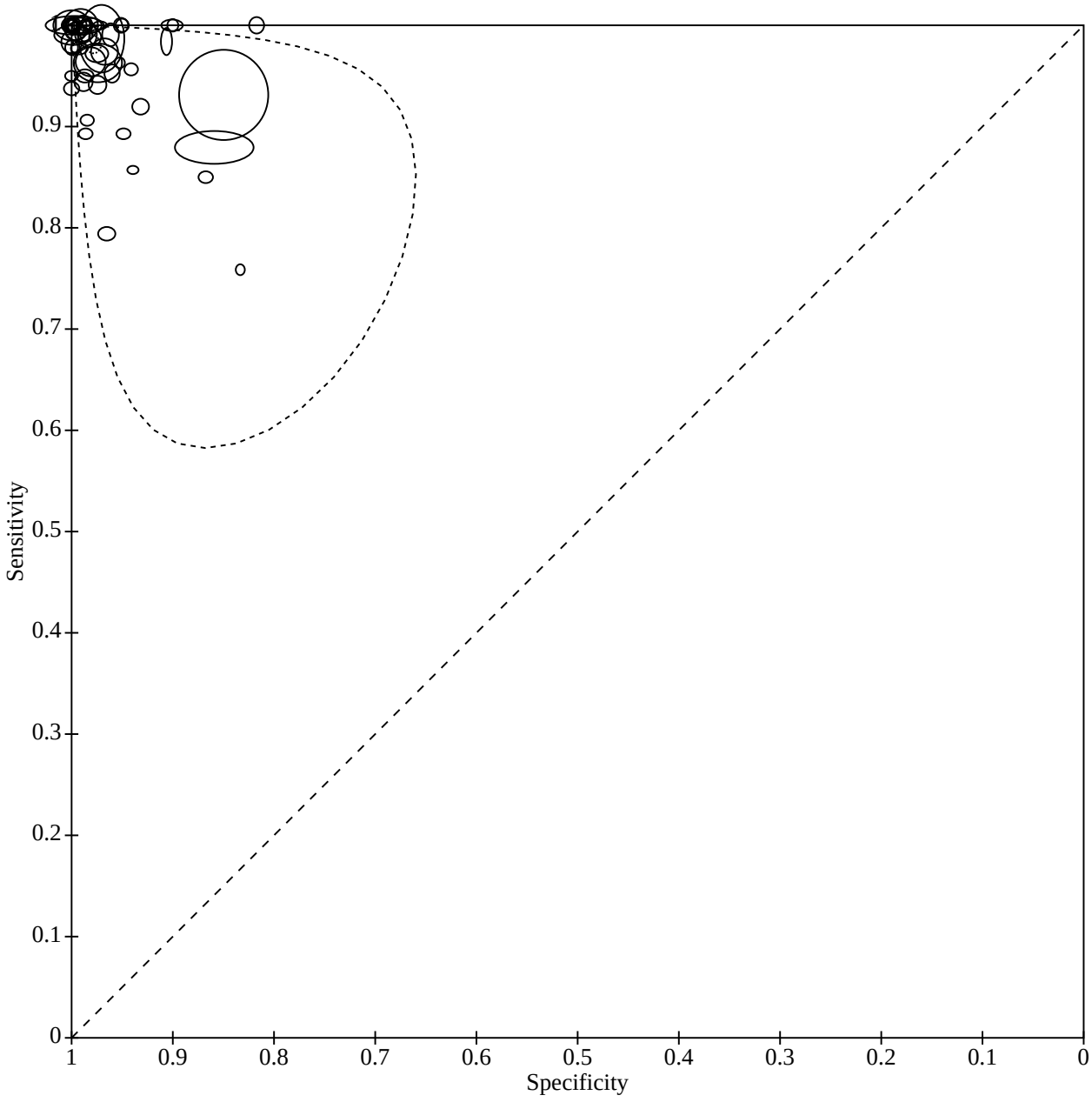
The instruments used were Pentacam, Eyesys, Sirius, Orbscan IIz, CSO, RTVue, Envisu R2210, Galilei, Topographic Modeling System (TMS), and CASIA. All devices belonged to one of the groups described in the Reference standards section. Consejo 2020 used the Corvis-ST. In the protocol for this review, we specified that we would not include any devices measuring biomechanical properties (Vandevenne 2021); however, only the first image of each measurement was used for analysis in Consejo 2020. This image is taken before the air stimulus and is the same as a Scheimpflug-based measurement. The studies described 12 different AI algorithms.

The most frequently used algorithm was the neural network; 11 studies used a simple neural network, and 13 studies used the convolutional neural network. The second most common algorithm was logistic regression (N = 10). Seven studies each used decision tree, discriminant analysis, and support vector machine. The right-hand column of Table 1 shows the number of cornea specialists. Thirty studies did not provide information about who made the keratoconus diagnosis or how they made it.

Detection of manifest keratoconus

Figure 5 shows the summary ROC (SROC) plot of sensitivity and specificity of the AI algorithms for detecting manifest keratoconus (54 studies, 50,519 eyes/images). Sensitivities range from 76% to 100%, and the summary sensitivity is 98.6% (95% CI 97.6% to 99.1%). Specificities range from 82% to 100%, and the summary specificity is 98.3% (95% CI 97.4% to 98.9%). Most studies are clustered in the upper-left corner of the graph, indicating a high accuracy of the tested algorithms to diagnose keratoconus. There appears to be little heterogeneity between the studies. The confidence ellipse of the summary point lies close to the upper-left angle of the ROC plane and is hidden behind the symbols of most studies. The prediction interval is larger, which seems to be attributable mainly to three large studies with lower accuracy.

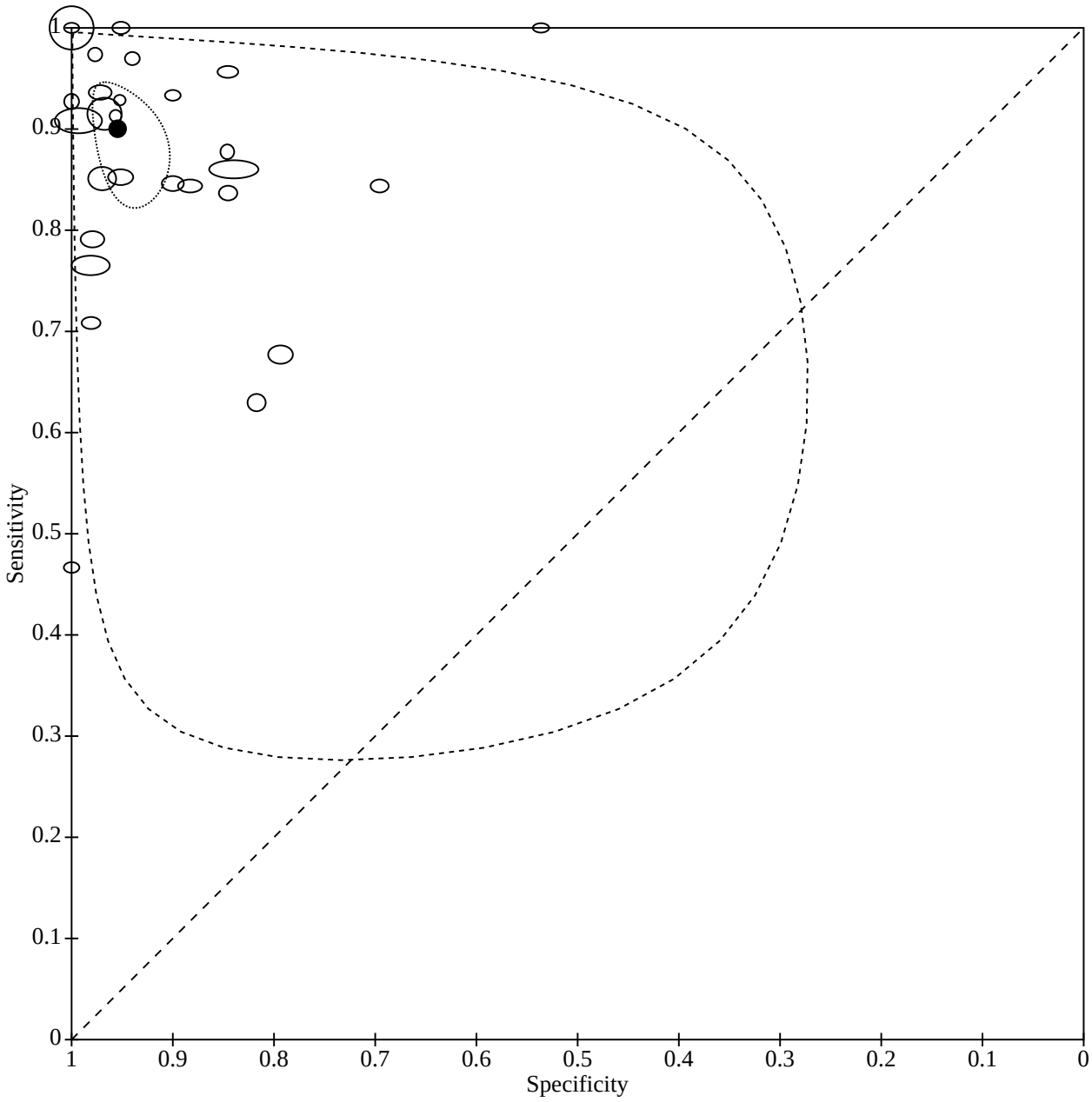
Figure 5. Summary receiver operating characteristics (SROC) plot of accuracy of AI for detecting manifest keratoconus.



Detection of subclinical keratoconus

Figure 6 shows the SROC plot of sensitivity and specificity of the AI algorithms for detecting subclinical keratoconus (28 studies, 9508 eyes/images). Sensitivities range from 47% to 100%, and the summary sensitivity is 90.0% (95% CI 84.5% to 93.8%). Specificities range from 54% to 100%, and the summary specificity is 95.5% (95% CI 91.9% to 97.5%). More than half of the studies are near the y-axis of the SROC plot, indicating high specificity. A few studies are located around the upper left corner. However, the distribution of the dots is fairly wide, indicating high heterogeneity between the studies, particularly for sensitivity.

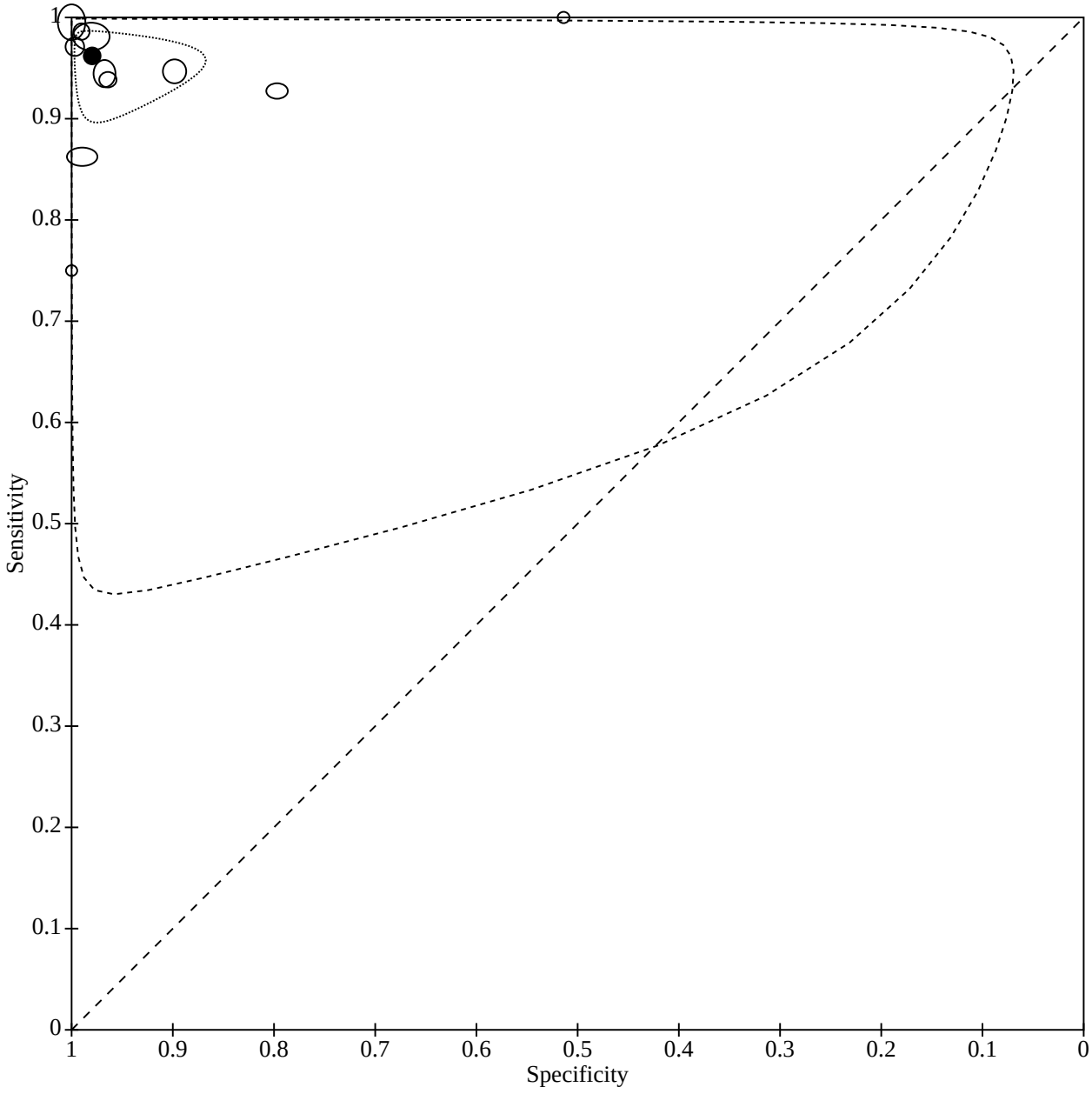
Figure 6. Summary receiver operating characteristics (SROC) plot of accuracy of AI for detecting subclinical keratoconus.



Detection of mixed keratoconus

We analysed the studies that developed and trained algorithms to diagnose and distinguish both manifest and subclinical keratoconus. The corresponding SROC plot includes 11 studies (Figure 7). Sensitivities range from 75% to 100%, and the summary sensitivity is 96.2% (95% CI 92.5% to 98.1%). Specificities range from 51% to 100%, and the summary specificity is 98.0% (95% CI 92.6% to 99.5%). There is a wide distribution of the studies in the curve. Accuracy is almost perfect in five studies, and sensitivity or specificity is high in two other studies.

Figure 7. Summary receiver operating characteristics (SROC) plot of accuracy of AI for detecting subclinical and manifest keratoconus (mixed).



Subgroup analyses

We conducted subgroup analyses, restricted to manifest keratoconus with adequate numerosity, by study design (clinical series versus registry), AI algorithm (as listed in the secondary objectives), imaging technique (tomography, topography, OCT), and data source (parameters, images). None of these covariates were associated with accuracy (Table 2).

Sensitivity analyses

We conducted sensitivity analyses by excluding the 14 studies that used images as the unit of analyses (as this inflated the number of observations several times), and we found that sensitivity and specificity remained very high for both manifest keratoconus (98.5%, 95% CI 97.4% to 99.1%) and subclinical keratoconus (98.5%, 95% CI 97.5% to 99.1%). As most studies were unclear on the unit of analysis, and only 16 of the remaining 49 studies clearly stated that they analysed one eye per participant, we did not run additional sensitivity analyses.

DISCUSSION

Summary of main results

In this DTA review, we investigated the accuracy of AI algorithms for diagnosing keratoconus. These automated tools could help ophthalmologists, optometrists, and other eye care professionals

to recognize the disease and refer people in time. An early diagnosis of keratoconus is beneficial for the patient as it leads to regular follow-up and thus detection and treatment of progression before visual acuity has decreased. AI could be used in a primary or secondary eye care setting as a triage test for people seeking refractive surgery or people whose visual acuity cannot be fully corrected with glasses. Studies investigating these tools should aim for high sensitivity so the AI algorithm can correctly detect as many people with keratoconus as possible.

We included al large number of studies (63), although 58 had a case-control design, and the population of most was not as described in our protocol (Vandevenne 2021), leading to high risk of bias and low applicability in the domain of 'Patient selection'.

We found the diagnostic accuracy achieved with AI algorithms was almost perfect for detecting manifest keratoconus; however, the certainty of evidence was low for sensitivity and specificity. The main issue was the case-control design for all but five studies, which may have led to an overestimation of accuracy. Moreover, while the confidence region is very narrow for detection of manifest keratoconus, the prediction area in the SROC curve is rather large because some studies (including two large studies) had lower sensitivity and specificity (Elsawy 2021, Zeboulon 2020b). The methodological quality of these studies was low, and we could not investigate whether they differed in the population or index test due to poor reporting. However, a possible explanation for the lower sensitivity and specificity in Elsawy 2021 could be that it developed a multidisease algorithm.

The diagnostic accuracy of AI algorithms for subclinical keratoconus may be suboptimal. The summary sensitivity of the 28 studies was 90.0% (95% CI 84.5% to 93.8%). The evidence was of very low certainty due to indirectness (the population did not match our definition) and high risk of selection and other biases (case-control design with poor generalizability). Because a missed diagnosis of subclinical keratoconus in people seeking refractive surgery could lead to iatrogenic ectasia, future studies should maximize sensitivity rather than specificity.

Few studies tested AI algorithms for detection of mixed (both subclinical and manifest) keratoconus, and the distribution of accuracy measures in the SROC plot was very heterogeneous. We were unable to draw any firm conclusions due to the low number of studies.

Subgroup analyses

As stated in our protocol, we aimed to investigate different causes of heterogeneity (Vandevenne 2021); however, it was not feasible to examine all the predefined factors because of poor reporting in the included studies and because the subgroups were too small. We were only able to perform the following subgroup analysis of manifest keratoconus.

- Study design (clinical series and registries): there were no differences between the two subgroups.
- Different AI algorithms (of which there were 10): we found no evidence of a difference in accuracy between the different AI algorithms. This could indicate that different algorithms are suitable for detecting keratoconus based on cornea imaging. However, four subgroups included fewer than five studies, and the number of participants in two subgroups was small.
- Imaging technique: we found no differences between topography-, tomography-, or OCT-based AI algorithms.
- Data source (images versus parameters): there were no subgroup differences.

Strengths and weaknesses of the review

We conducted a comprehensive search; because AI in keratoconus is a very focused topic, it is unlikely that we missed any studies. However, we only included articles published in English. We did not include ongoing studies because it was unclear whether they were still in process. When the two review authors (MV and EF) did not agree on the eligibility of an article, they asked the opinion of the rest of the review team.

The extracted data were sometimes limited due to poor reporting in the articles, especially with regard to the reference standard and comparison between AI algorithms within a study. Almost half of the studies were at unclear risk of bias with regard to who made the initial diagnosis and how they made it.

Methods for quality assessment of directly comparative test accuracy studies have been developed, but this type of analysis was not feasible in our review (Yang 2021). Few articles presented accuracy estimates for multiple AI algorithms, and we selected the algorithm with the highest Youden's index. We are aware that this selection may inflate accuracy, especially in smaller studies. The fact that all AI algorithms yielded very high and similar accuracy in subgroup analyses may support our approach.

We could not manage unit of analysis issues appropriately because data for participants (e.g. only one eye) were rarely available, and many studies analysed both eyes or multiple images of the same participant, or the unit of analysis was unclear. However, accuracy remained high even after the exclusion of studies with important unit of analysis issues (i.e. those that analysed multiple images per eye instead of eyes or participants).

The sensitivity analyses suggested that the covariates listed in Secondary objectives had no effect on accuracy. Unfortunately, we were unable to determine the sources of heterogeneity. Subgroup analyses showed no differences between the studies based on study design, AI algorithm, imaging technique, or data source. We were unable to carry out HSROC models to assess specificity at high sensitivity (as necessary for a triage role) due to clustering of the data at the upper left corner of the SROC plot.

Finally, a few studies had zero counts, especially for false positives and false negatives, and this may have introduced estimation bias.

Comparison with previous research

There are published reviews on AI in ophthalmology, AI in anterior segment diseases, and AI in keratoconus. Hogarty 2019 provides a general overview of relevant AI algorithms and new developments in ophthalmology. Lin 2019 specifically investigates AI for keratoconus and refractive surgery screening, but only provides an overview of the literature. These two reviews concluded that AI is a promising technique to process corneal imaging data and detect keratoconus, but as they were narrative reviews with no meta-analysis, their results are not comparable with ours. Cao 2022 evaluated the accuracy of different AI algorithms in a meta-analysis. It found 30 articles detecting manifest keratoconus with a summary sensitivity of 97% (95% CI

94.9% to 98.2%) and a summary specificity of 98.5% (95% CI 97.1% to 99.3%). Our results were very similar; however, our summary sensitivity was slightly higher. Cao 2022 found 15 articles detecting subclinical keratoconus with a summary sensitivity of 88.2% (95% CI 82.2% to 92.3%) and a summary specificity of 94.7% (95% CI 91.4% to 96.7%). The summary sensitivity in our review was slightly higher, but the sensitivity was similar. Based on their results, the authors of Cao 2022 concluded that AI algorithms are not yet applicable in a clinical setting, and more research is needed.

Applicability of findings to the review question

We defined the eligible population as people with refractive errors:

- whose vision could not be fully corrected with glasses; or
- who were seeking refractive surgery, or
- who had suspected keratoconus (for whom a decision was to be made on referral to cornea specialists).

Figure 3 shows that 'Patient selection' is the most problematic domain for applicability. This is due to the case-control design of most studies, which included people already diagnosed with keratoconus. According to the clinical pathway described in our protocol, the AI algorithm could serve as a triage test for eye care professionals, such as ophthalmologists, optometrists, and others, to detect keratoconus and refer patients to a cornea specialist for appropriate diagnosis, follow-up, and treatment (Vandevenne 2021). However, the applicability of AI is questionable in this population; further research is needed to confirm the reliability of this approach.

AUTHORS' CONCLUSIONS

Implications for practice

We found that artificial intelligence (AI) may be very accurate for detecting keratoconus, but there was some heterogeneity. A local assessment of AI accuracy is advisable wherever this technique is implemented, particularly for detection of subclinical keratoconus. It is important to interpret the results of this review with caution: there is a high risk for overestimation of accuracy due to the case-control design of most included studies.

The goal of AI algorithms is to help eye care professionals to make a medical decision for individuals. If found to be accurate, AI algorithms could be used in a triage setting to detect subclinical keratoconus before refractive surgery. When subclinical keratoconus is detected by the AI algorithm based on corneal imaging, the patient can be referred to a cornea specialist to receive proper follow-up and treatment. If a diagnosis is missed and the patient receives refractive surgery, there is a high chance they will develop iatrogenic ectasia. People whose vision cannot be fully corrected with glasses should receive corneal imaging. The AI algorithm could help the eye care professional to detect whether keratoconus is present. Positive cases would be referred to a cornea specialist. If the diagnosis is missed, there is a chance of progression and eventually the need for corneal transplantation. The impact of a missed diagnosis is much higher than that of a false diagnosis. While a missed diagnosis would lead to visual impairment, a false diagnosis would entail referral to a cornea specialist, where the patient would undergo some additional non-invasive examinations.

Implications for research

AI could be a solution for consistent and unbiased diagnoses; however, the accuracy of the tests for detecting keratoconus – especially subclinical keratoconus – needs to be evaluated in higher-quality studies. The ideal study would include a consecutive cohort of people seeking ophthalmic care or refractive surgery. The following changes would improve the evidence in this area.

- Future studies should have access to larger and more diverse data sets that provide a more representative picture of clinical practice, to validate their results. This is a hurdle every researcher encounters when developing an AI-based algorithm in medicine. An important problem is patient privacy: there is a need for a large, international, anonymized database with corneal imaging data.
- Researchers should establish standardized methodology to develop and validate AI algorithms for detecting keratoconus (or any other disease). This should make it easier to compare studies and draw clear conclusions about the accuracy and applicability of the AI algorithms.
- AI algorithms should be platform-independent so they can analyse data from different imaging devices.
- A clear global consensus on the definition of (subclinical) keratoconus and progression would facilitate test development and comparison between studies.

ACKNOWLEDGEMENTS

We would like to thank:

- Cochrane Eyes and Vision (CEV) for creating and running the electronic searches;
- Roy Schwartz for his comments on this protocol;
- Vito Romano for his comments on the protocol and review;
- the Diagnostic Test Accuracy team for their comments on the protocol and review;
- Anupa Shah, Managing Editor of Cochrane Eyes and Vision, for her assistance throughout the review process;
- the Cochrane Central Editorial Service for completing the editorial processing upon transfer from the Eyes and Visions review group; and
- Julia Turner, copy editor, Cochrane Central Production Service, for her comments and editing of this review.

REFERENCES

References to studies included in this review

Abdelmotaal 2020 *{published data only}*

Abdelmotaal H, Mostafa MM, Mostafa AN, Mohamed AA, Abdelazeem K. Classification of color-coded scheimpflug camera corneal tomography images using deep learning. *Translational Vision Science and Technology* 2020;**9**(13):30.

Accardo 2002 *{published data only}*

Accardo PA, Pensiero S. Neural network-based system for early keratoconus detection from corneal topography. *Journal of Biomedical Informatics* 2002;**35**(3):151-9.

Almeida 2022 *{published data only}*

Almeida GC, Guido R, Silva H, Brandão C, de Mattos LC, Lopes BT, et al. New artificial intelligence index based on Scheimpflug corneal tomography to distinguish subclinical keratoconus from healthy corneas. *Journal of Cataract & Refractive Surgery* 2022;**48**(10):1168-74.

Al-Timemy 2021 *{published data only}*

Al-Timemy AH, Mosa ZM, Alyasseri Z, Lavric A, Lui MM, Hazarbassanov RM, et al. A hybrid deep learning construct for detecting keratoconus from corneal maps. *Translational Vision Science and Technology* 2021;**10**(14):16.

Arbelaez 2012 *{published data only}*

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**(11):2231-8.

Bessho 2006 *{published data only}*

Bessho K, Maeda N, Kuroda T, Fujikado T, Tano Y, Oshika T. Automated keratoconus detection using height data of anterior and posterior corneal surfaces. *Japanese Ophthalmological Society* 2006;**50**(5):409-16.

Cao 2020 *{published data only}*

Cao K, Verspoor K, Sahebjada S, Baird PN. Evaluating the performance of various machine learning algorithms to detect subclinical keratoconus. *Translational Vision Science and Technology* 2020;**9**(2):24.

Cao 2021a *{published data only}*

Cao K, Verspoor K, Chan E, Daniell M, Sahebjada S, Baird PN. Machine learning with a reduced dimensionality representation of comprehensive Pentacam tomography parameters to identify subclinical keratoconus. *Computer in Biology and Medicine* 2021;**138**:104884.

Carvalho 2005 *{published data only}*

Carvalho LA. Preliminary results of neural networks and zernike polynomials for classification of videokeratography maps. *Optometry and Vision Science* 2005;**82**(2):151-8.

Castro-Luna 2020 *{published data only}*

Castro-Luna GM, Martínez-Finkelshtein A, Ramos-López D. Robust keratoconus detection with Bayesian network classifier for Placido-based corneal indices. *Contact Lens and Anterior Eye* 2020;**43**(4):366-72.

Cavas-Martinez 2017 *{published data only}*

Cavas-Martinez F, Bataille L, Fernandez-Pacheco DG, Cañavate FJF, Alio JL. A new approach to keratoconus detection based on corneal morphogeometric analysis. *PLoS One* 2017;**12**(9):e0184569.

Chan 2015 *{published data only}*

Chan C, Ang M, Saad A, Chua D, Mejia M, Lim L, et al. Validation of an objective scoring system for forme fruste keratoconus detection and post-LASIK ectasia risk assessment in Asian eyes. *Cornea* 2015;**34**(9):996-1004.

Chandapura 2019 *{published data only}*

Chandapura R, Salomão MQ, Ambrósio Jr R, Swarup R, Shetty R, Sinha Roy A. Bowman's topography for improved detection of early ectasia. *Journal of Biophotonics* 2019;**12**(10):e201900126.

Chastang 2000 *{published data only}*

Chastang PJ, Borderie VM, Carvajal-Gonzalez S, Rostene W, Laroche L. Automated keratoconus detection using the EyeSys videokeratoscope. *Journal of Refractive Surgery and Cataract* 2000;**26**(5):675-83.

Chen 2021 *{published data only}*

Chen X, Zhao J, Iselin KC, Borroni D, Romano D, Gokul A, et al. Keratoconus detection of changes using deep learning of colour- coded maps. *BMJ Open Ophthalmology* 2021;**6**(1):e000824.

Cohen 2022 *{published data only}*

Cohen E, Bank D, Sorkin N, Giryes R, Varssano D. Use of machine learning to achieve keratoconus detection skills of a corneal expert. *International Ophthalmology* 2022;**42**(12):3837-47.

Consejo 2020 *{published data only}*

Consejo A, Solarski J, Karnowski K, Rozema JJ, Wojtkowski M, Iskander DR. Keratoconus detection based on a single scheimpflug image. *Translational Vision Science and Technology* 2020;**9**(7):36.

De Almeida Jr 2021 *{published data only}*

de Almeida Jr GC, Capobianco Guidod R, Netod JS, Rosad JM, Castiglionia L, de Mattosa LC, et al. Corneal Tomography Multivariate Index (CTMVI) effectively distinguishes healthy corneas from those susceptible to ectasia. *Biomedical Signal Processing and Control* 2021;**10**:102995.

Elsawy 2021 *{published data only}*

Elsawy A, Eleiwa T, Chase C, Ozcan E, Tolba M, Feuer W, et al. Multidisease deep learning neural network for the diagnosis of corneal diseases. *American Journal of Ophthalmology* 2021;**226**:252-61.

Feizi 2016 *{published data only}*

Feizi S, Yaseri M, Kheiri B. Predictive ability of galilei to distinguish subclinical keratoconus and keratoconus from

normal corneas. *Journal of Ophthalmic and Vision Research* 2016;**11**(1):8-16.

Gairola 2022 *{published data only}*

Gairola S, Joshi P, Balasubramaniam A, Murali K, Kwatra N, Jain M. Keratoconus classifier for smartphone-based corneal topographer. *International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* 2022;**2022**:1875-78.

Gao 2022 *{published data only}*

Gao HB, Pan ZG, Shen MX, Lu F, Li H, Zhang XQ. KeratoScreen: early keratoconus classification with zernike polynomial using deep learning. *Cornea* 2022;**41**(9):1158-65.

Ghaderi 2021 *{published data only}*

Ghaderi M, Sharifi A, Jafarzadeh Pour E. Proposing an ensemble learning model based on neural network and fuzzy system for keratoconus diagnosis based on Pentacam measurements. *International Ophthalmology* 2021;**41**(12):3935-48.

Issarti 2019 *{published data only}*

Issarti I, Consejo A, Jimenez-Garcia M, Hershko S, Koppen C, Rozema JJ. Computer aided diagnosis for suspect keratoconus detection. *Computers in Biology and Medicine* 2019;**109**:33-42.

Issarti 2020 *{published data only}*

Issarti I, Consejo A, Jimenez-Garcia M, Kreps EO, Koppen C, Rozema JJ. Logistic index for keratoconus detection and severity scoring (Logik). *Computers in Biology and Medicine* 2020;**122**:103809.

Kalin 1996 *{published data only}*

Kalin NS, Maeda N, Klyce SD, Hargrave S, Wilson SE. Automated topographic screening for keratoconus in refractive surgery candidates. *CLAO Journal* 1996;**22**(3):164-7.

Kamiya 2019 *{published data only}*

Kamiya K, Ayatsuka Y, Kato Y, Fujimura F, Takahashi M, Shoji N, et al. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study. *BMJ Open* 2019;**9**(9):e031313.

Kamiya 2021 *{published data only}*

Kamiya K, Ayatsuka Y, Kato Y, Shoji N, Mori Y, Miyata K. Diagnosability of keratoconus using deep learning with Placido disk-based corneal topography. *Frontiers in Medicine* 2021;**8**:724902.

Kojima 2020 *{published data only}*

Kojima T, Nishida T, Nakamura T, Tamaoki A, Hasegawa A, Takagi Y, et al. Keratoconus screening using values derived from auto-keratometer measurements: a multicenter study. *American Journal of Ophthalmology* 2020;**215**:127-34.

Kojima 2021 *{published data only}*

Kojima T, Isogai N, Nishida T, Nakamura T, Ichikawa K. Screening of keratoconus using Autokeratometer and Keratometer Keratoconus Index. *Diagnostics* 2021;**11**(11):2120.

Kovacs 2016 *{published data only}*

Kovacs I, Mihaltz K, Kranitz K, Juhasz E, Takacs A, Dienes L, et al. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with preclinical signs of keratoconus. *Journal of Cataract and Refractive Surgery* 2016;**42**(2):275-83.

Kuo 2020 *{published data only}*

Kuo BI, Chang WY, Liao T S, Liu FY, Liu HY, Chu HS, et al. Keratoconus screening based on deep learning approach of corneal topography. *Translational Vision Science and Technology* 2020;**9**(2):53.

Lavric 2021 *{published data only}*

Lavric A, Anchidin L, Popa V, Al-Timemy AH, Alyasseri Z, Takahashi H, et al. Keratoconus severity detection from elevation, topography and pachymetry raw data using a machine learning approach. *IEEE Access* 2021;**9**:84344-55.

Lopes 2018 *{published data only}*

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. *American Journal of Ophthalmology* 2018;**195**:223-32.

Lucena 2021 *{published data only}*

Lucena AR, Araujo MO, Carneiro RF, Cavalcante TD, Ribeiro AB, Anselmo FJ. Development of an application for providing corneal topography reports based on artificial intelligence. *Arquivos Brasileiros de Oftalmologia* 2021;**85**(4):351-8.

Maeda 1994 *{published data only}*

Maeda N, Klyce SD, Smolek MK, Thompson HW. Automated keratoconus screening with corneal topography analysis. *Investigative Ophthalmology and Visual Science* 1994;**35**(6):2749-57.

Maeda 1995a *{published data only}*

Maeda N, Klyce SD, Smolek MK. Comparison of methods for detecting keratoconus using videokeratography. *Archives of Ophthalmology* 1995;**113**(7):870-4.

Maeda 1995b *{published data only}*

Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. *Investigative Ophthalmology and Visual Science* 1995;**36**(7):1327-35.

Mahmoud 2013 *{published data only}*

Mahmoud AM, Nunez MX, Blanco C, Koch DD, Wang L, Weikert MP, et al. Expanding the cone location and magnitude index to include corneal thickness and posterior surface information for the detection of keratoconus. *American Journal of Ophthalmology* 2013;**156**(6):1102-11.

Mahmoud 2021 *{published data only}*

Mahmoud HA, Mengash HA. Automated keratoconus detection by 3D corneal images reconstruction. *Sensors* 2021;**21**(7):2326.

Mohammadpour 2022 *{published data only}*

Mohammadpour M, Heidari Z, Hashemi H, Yaseri M, Fotouhi A. Comparison of artificial intelligence-based machine learning

classifiers for early detection of keratoconus. *European Journal of Ophthalmology* 2022;**32**(3):1352-60.

Pavlatos 2020 *{published data only}*

Pavlatos E, Chen S, Yang Y, Wang Q, Huang D, Li Y. A coincident thinning index for keratoconus identification using OCT pachymetry and epithelial thickness maps. *Journal of Refractive Surgery* 2020;**36**(11):757-65.

Rabinowitz 1999 *{published data only}*

Rabinowitz Y S, Rasheed K. KISA% index: a quantitative videokeratography algorithm embodying minimal topographic criteria for diagnosing keratoconus. *Journal of Cataract and Refractive Surgery* 1999;**25**(10):1327-35.

Ruiz 2016 *{published data only}*

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**(6):827-32.

Ruiz 2017 *{published data only}*

Ruiz Hidalgo I, Rozema JJ, Saad A, Gatinel D, Rodriguez P, Zakaria N, et al. Validation of an objective keratoconus detection system implemented in a scheimpflug tomographer and comparison with other methods. *Cornea* 2017;**36**(6):689-95.

Saad 2014 *{published data only}*

Saad A, Guilbert E, Gatinel D. Corneal enantiomorphism in normal and keratoconic eyes. *Journal of Refractive Surgery* 2014;**30**(8):542-7.

Saad 2016 *{published data only}*

Saad A, Gatinel D. Combining placido and corneal wavefront data for the detection of forme fruste keratoconus. *Journal of Refractive Surgery* 2016;**32**(8):510-6.

Saika 2013 *{published data only}*

Saika M, Maeda N, Hirohara Y, Mihashi T, Fujikado T, Nishida K. Four discriminant models for detecting keratoconus pattern using Zernike coefficients of corneal aberrations. *Japanese Journal of Ophthalmology* 2013;**57**(6):503-9.

Shetty 2015 *{published data only}*

Shetty R, Matalia H, Srivatsa P, Ghosh A, Dupps W J Jr, Sinha Roy A. A novel zernike application to differentiate between three-dimensional corneal thickness of normal corneas and corneas with keratoconus. *American Journal of Ophthalmology* 2015;**160**(3):453-62.

Shi 2020 *{published data only}*

Shi C, Wang M, Zhu T, Zhang Y, Ye Y, Jiang J, et al. Machine learning helps improve diagnostic ability of subclinical keratoconus using Scheimpflug and OCT imaging modalities. *Eye and Vision* 2020;**7**:48.

Sideroudi 2017 *{published data only}*

Sideroudi H, Labiris G, Georgantzoglou K, Ntonti P, Siganos C, Kozobolis V. Fourier analysis algorithm for the posterior corneal keratometric data: clinical usefulness in keratoconus. *Ophthalmic and Physiological Optics* 2017;**37**(4):460-6.

Smadja 2013 *{published data only}*

Smadja D, Touboul D, Cohen A, Doveh E, Santhiago MR, Mello GR, et al. Detection of subclinical keratoconus using an automated decision tree classification. *American Journal of Ophthalmology* 2013;**156**(2):237-46.

Smolek 1997 *{published data only}*

Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Investigative Ophthalmology and Visual Science* 1997;**38**(11):2290-9.

Souza 2010 *{published data only}*

Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from orbscan II examinations. *Clinics* 2010;**65**(12):1223-8.

Subramaniam 2022 *{published data only}*

Subramaniam P, Ramesh GP. Keratoconus classification with convolutional neural networks using segmentation and index quantification of eye topography images by particle swarm optimisation. *BioMed Research International* 2022;**2022**:Article ID 8119685. [DOI: 10.1155/2022/8119685]

Twa 2005 *{published data only}*

Twa MD, Parthasarathy S, Roberts C, Mahmoud AM, Raasch TW, Bullimore MA. Automated decision tree classification of corneal shape. *Optometry and Vision Science* 2005;**82**(12):1038-46.

Xie 2020 *{published data only}*

Xie Y, Zhao L, Yang X, Wu X, Yang Y, Huang X, et al. Screening candidates for refractive surgery with corneal tomographic-based deep learning. *JAMA Ophthalmology* 2020;**138**(5):519-26.

Xu 2017 *{published data only}*

Xu Z, Li W, Jiang J, Zhuang X, Chen W, Peng M, et al. Characteristic of entire corneal topography and tomography for the detection of sub-clinical keratoconus with Zernike polynomials using Pentacam. *Scientific Reports* 2017;**7**(1):16486.

Xu 2022a *{published data only}*

Xu Y, Ren YR, Zhaung XY, Zhang XF. Predictive index based on minimum corneal thickness and symmetry index back of Sirius for early diagnosis of keratoconus. *International Eye Science* 2022;**22**(9):1426-35.

Yang 2021 *{published data only}*

Yang Y, Pavlatos E, Chamberlain W, Huang D, Li Y. Keratoconus detection using OCT corneal and epithelial thickness map parameters and patterns. *Journal of Cataract and Refractive Surgery* 2021;**47**(6):759-66.

Yousefi 2018 *{published data only}*

Yousefi S, Yousefi E, Takahashi H, Hayashi T, Tampo H, Inoda S, et al. Keratoconus severity identification using unsupervised machine learning. *PLoS One* 2018;**13**(11):e0205998.

Zeboulon 2020a *{published data only}*

Zeboulon P, Debellemaniere G, Bouvet M, Gatinel D. Corneal topography raw data classification using a convolutional neural network. American Journal of Ophthalmology 2020;**219**:33-9.

Zeboulon 2020b *{published data only}*

Zeboulon P, Debellemaniere G, Gatinel D. Unsupervised learning for large-scale corneal topography clustering. *Science Reports* 2020;**10**(1):16973.

References to studies excluded from this review

Aatila 2021 *{published data only}*

Aatila M, Lachgar M, Hamid H, Kartit A. Keratoconus severity classification using features selection and machine learning algorithms. *Computational and Mathematical Methods in Medicine* 2021;**2021**:9979560.

Al-Timemy 2022 *{published data only}*

Al-Timemy A, Al-Zubaidi L, Ghaeb N, Takahashi H, Lavric A, Mosa Z, et al. A device-agnostic deep learning model for detecting keratoconus based on anterior elevation corneal maps. In: Investigative Ophthalmology and Visual Science. 7 edition. Vol. 63. 2022:2101.

Buehren 2018 *{published data only}*

Buehren J, Kleinhans S, Herrmann E, Kohnen T. Comparison of metrics obtained with discriminant analysis and decision trees for the detection of subclinical keratoconus. Investigative Ophthalmology and Visual Science 2018;**59**(9):ARVO E-abstract 5724.

Cao 2021b *{published data only}*

Cao K, Verspoor K, Chan E, Daniell M, Sahebjada S, Baird PN. Novel, high-performance machine learning model for detection of subclinical keratoconus. Investigative Ophthalmology and Visual Science 2021;**62**(8):ARVO E-abstract 2157.

Castro-Luna 2021 *{published data only}*

Castro-Luna G Jimenez-Rodriguez D, Castano-Fernandez AB, Perez-Rueda A. Diagnosis of subclinical keratoconus based on machine learning techniques. *Journal of Clinical Medicine* 2021;**10**(18):21.

ChiCTR2000037484 *{published data only}*

ChiCTR2000037484. Study on artificial intelligence-assisted diagnosis of keratoconus diseases. trialsearch.who.int/Trial2.aspx?TrialID=ChiCTR2000037484 (first received 28 August 2020).

ChiCTR2000039070 *{published data only}*

ChiCTR2000039070. Research on early warning diagnosis and prognosis model of keratoconus based on Artificial Intelligence. trialsearch.who.int/Trial2.aspx?TrialID=ChiCTR2000039070 (first received 15 October 2020).

DosSantos 2019 *{published data only}*

Dos Santos VA, Schmetterer L, Stegmann H, Pfister M, Messner A, Schmidinger G, et al. CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning. *Biomedical Optics Express* 2019;**10**(2):622-41.

Elsawy 2022 *{published data only}*

Elsawy A, Abdel-Mottaleb M. PIPE-Net: A pyramidal-input-parallel-encoding network for the segmentation of corneal layer interfaces in OCT images. *Computers in Biology and Medicine* 2022;**147**:105595.

Feng 2021 *{published data only}*

Feng R, Xu Z, Zheng X, Hu H, Jin X, Chen DZ, et al. KerNet: a novel deep learning approach for keratoconus and sub-clinical keratoconus detection based on raw data of the Pentacam HR system. *IEEE Journal of Biomedical and Health Informatics* 2021;**25**(10):3898-910.

Hazarbessanov 2022 *{published data only}*

Hazarbassanov RM, Alyasseri ZA, Al-Timemy A, Lavric A, Abasid AK, Takahashi H, et al. Detecting keratoconus on two different populations using an unsupervised hybrid artificial intelligence model. In: Investigative Ophthalmology and Visual Science. 7 edition. Vol. 63. 2022:2088.

Hernandez 2020 *{published data only}*

Hernandez LA, Sanchez-Huerta V, Ramirez-Fernandez M, Hernandez-Quintela E. Combinatorial approach to determine top performing keratometric features and machine learning algorithms for keratoconus detection. Investigative Ophthalmology and Visual Science 2020;**61**(7):ARVO E-abstract 4750.

Hidalgo 2014 *{published data only}*

Hidalgo IR, Perez PR, Rozema JJ, Tassignon MJB. Comparison of machine learning methods to automatically classify keratoconus. Investigative Ophthalmology and Visual Science 2014;**55**(13):ARVO E-abstract 4206.

Hjordtal 1995 *{published data only}*

Hjordtal JO, Erdmann L, Bek T. Fourier analysis of video-keratographic data. A tool for separation of spherical, regular astigmatic and irregular astigmatic corneal power components. *Ophthalmic and Physiological Optics* 1995;**15**(3):171-85.

Issarti 2018 *{published data only}*

Issarti I, Consejo A, Rozema J. Elevation-based detection of keratoconus. Investigative Ophthalmology and Visual Science 2018;**59**(9):ARVO E-abstract 5810.

JPRN-UMIN000034587 *{published data only}*

JPRN-UMIN000034587. Diagnostic evaluation of keratoconus using anterior segment optical coherence tomography and machine learning. trialsearch.who.int/Trial2.aspx?TrialID=JPRN-UMIN000034587 (first received 1 November 2018).

JPRN-UMIN000040128 *{published data only}*

JPRN-UMIN000040128. Diagnostic evaluation of keratoconus using corneal topography and machine learning. trialsearch.who.int/Trial2.aspx?TrialID=JPRN-UMIN000040128 (first received 10 April 2020).

JPRN-UMIN000040308 *{published data only}*

JPRN-UMIN000040308. Prediction of keratoconus progression using anterior segment optical coherence tomography and deep learning. trialsearch.who.int/Trial2.aspx?TrialID=JPRN-UMIN000040308 (first received 1 October 2020).

JPRN-UMIN000040321 *{published data only}*

JPRN-UMIN000040321. Development of diagnostic artificial intelligence in the ophthalmology. trialsearch.who.int/Trial2.aspx?TrialID=JPRN-UMIN000040321 (first received 7 May 2020).

JPRN-UMIN000043831 *{published data only}*

JPRN-UMIN000043831. Screening for keratoconus using Smartphone and artificial intelligence. trialsearch.who.int/Trial2.aspx?TrialID=JPRN-UMIN000043831 (first received 3 April 2021).

Kleinhans 2019 *{published data only}*

Kleinhans S, Herrmann E, Kohnen T, Buhren J. Comparison of discriminant analysis and decision trees for the detection of subclinical keratoconus. *Klinische Monatsblatter fur Augenheilkunde* 2019;**236**(6):798-805.

Klyce 2005 *{published data only}*

Klyce SD, Karon MD, Smolek MK. Screening patients with the corneal navigator. *Journal of Refractive Surgery* 2005;**21**(5):S617-22.

Kundu 2021 *{published data only}*

Kundu G, Shetty R, Khamar P, Mullick R, Gupta S, Nuijts R, et al. Universal architecture of corneal segmental tomography biomarkers for artificial intelligence-driven diagnosis of early keratoconus. *British Journal of Ophthalmology* 2021;**16**:319309.

Lavric 2019 *{published data only}*

Lavric A, Valentin P. Keratodetect: keratoconus detection algorithm using convolutional neural networks. *Computational Intelligence and Neuroscience* 2019;**2019**:8162567.

Li 2009 *{published data only}*

Li X, Yang H, Rabinowitz YS. Keratoconus: classification scheme based on videokeratography and clinical signs. *Journal of Cataract and Refractive Surgery* 2009;**35**(9):1597-603.

Li 2021 *{published data only}*

Li DF, Dong YL, Xie, S, Guo Z, Li SX, Guo Y, et al. Deep learning based lesion detection from anterior segment optical coherence tomography images and its application in the diagnosis of keratoconus. *Chinese Journal of Ophthalmology* June 2021;**57**(6):447-53.

Liu 2021 *{published data only}*

Liu H, Anwar M, Koaik M, Taylor S, Karanjia R, Mintsioulis G, et al. Deep learning for detection of keratoconus and prediction of crosslinking efficacy. Investigative Ophthalmology and Visual Science 2021;**62**(8):ARVO E-abstract 2044.

Malyugin 2021 *{published data only}*

Malyugin B, Sakhnov S, Izmailova S, Boiko E, Pozdeyeva N, Axenova L, et al. Keratoconus diagnostic and treatment algorithms based on machine-learning methods. *Diagnostics* 2021;**11**(10):1933.

Matalia 2020 *{published data only}*

Matalia H, Matalia J, Pisharody A, Patel Y, Chinnappaiah N, Salomao M, et al. Unique corneal tomography features of allergic eye disease identified by OCT imaging and artificial intelligence. *Journal of Biophotonics* 2020;**13**(10):e202000156.

Nasrin 2018 *{published data only}*

Nasrin F, Iyer RV, Mathews SM. Simultaneous estimation of corneal topography, pachymetry, and curvature. *IEEE Transactions on Medical Imaging* 2018;**37**(11):2463-73.

NCT01746823 *{unpublished data only}*

NCT01746823. Identification and validation of functional biomarkers for keratoconus. clinicaltrials.gov/ct2/show/NCT01746823 (first received 11 December 2012).

NCT04313387 *{published data only}*

NCT04313387. Efficiency of an algorithm derived from corneal tomography parameters to distinguish highly susceptible corneas to ectasia from healthy. clinicaltrials.gov/ct2/show/NCT04313387 (first received 18 March 2020).

NCT04763785 *{published data only}*

NCT04763785. Development of a keratoconus detection algorithm by deep learning analysis and Its validation on eyestar images. clinicaltrials.gov/ct2/show/NCT04763785 (first received 21 February 2021).

Omidi 2022 *{published data only}*

Omidi P, Cayless A, Langenbucher A. Evaluation of optimal Zernike radial degree for representing corneal surfaces. *PLOS One* 2022;**17**(5):e0269119.

Pavlatos 2022 *{published data only}*

Pavlatos E, Huang D, Li Y. Combining OCT corneal topography and thickness maps to diagnose keratoconus using a convolutional neural network. In: Investigative Ophthalmology and Visual Science. 7 edition. Vol. 63. 2022:2109.

Ramos-Lopez 2011 *{published data only}*

Ramos-Lopez D, Martinez-Finkelshtein A, Castro-Luna GM, Pinero D, Alio JL. Placido-based indices of corneal irregularity. *Optometry and Vision Science* 2011;**88**(10):1220-31.

Rozema 2017 *{published data only}*

Rozema JJ, Rodriguez P, Navarro R, Koppen C. Bigaussian wavefront model for normal and keratoconic eyes. *Optometry and Vision Science* 2017;**94**(6):680-87.

Saad 2010 *{published data only}*

Saad A, Gatinel D. Topographic and tomographic properties of forme fruste keratoconus corneas. *Investigative Ophthalmology and Visual Science* 2010;**51**(11):5546-55.

Saad 2012 *{published data only}*

Saad A, Gatinel D. Evaluation of total and corneal wavefront high order aberrations for the detection of forme fruste keratoconus. *Investigative Ophthalmology and Visual Science* 2012;**53**(6):2978-92.

Schatteburg 2022 *{published data only}*

Schatteburg J, Langenbucher A. Protocol for the diagnosis of keratoconus using convolutional neural networks. *PLOS One* 2022;**17**(2):e0264219.

Souza 2008 *{published data only}*

Souza MB, Medeiros FW, Souza DB, Alves MR. Detection of keratoconus based on a neural network with Orbscan. *Arquivos Brasileiros de Oftalmologia* 2008;**71**(6):65-8.

Steinberg 2015a *{published data only}*

Steinberg J, Aubke-Schultz S, Frings A, Hulle J, Druchkiv V, Richard G, et al. Correlation of the KISA% index and Scheimpflug tomography in 'normal', 'subclinical', 'keratoconus-suspect' and 'clinically manifest' keratoconus eyes. *Acta Opthalmologica* 2015;**93**(3):e199-207.

Steinberg 2015b *{published data only}*

Steinberg J, Casagrande MK, Frings A, Katz T, Druchkiv V, Richard G, et al. Screening for subclinical keratoconus using swept-source Fourier domain anterior segment optical coherence tomography. *Cornea* 2015;**34**(11):1413-9.

Takahashi 2021 *{published data only}*

Takahashi H, Al-Timemy AH, Mosa ZM, Alyasseri Z, Lavric A, Filho JA, et al. Detecting keratoconus severity from corneal data of different populations with machine learning. Investigative Ophthalmology and Visual Science 2021;**62**(8):ARVO E-abstract 2145.

Tan 2019 *{published data only}*

Tan A, Yu M, Chen X, Hu L. Application of deep learning in early diagnosis assistant system of keratoconus. *Zhongguo Yiliao Qixie Zazhi* 2019;**43**(2):83-5.

Tas 2021 *{published data only}*

Tas AY, Hasanreisoglu M, Balim H, Gonen M, Sahin A. Automated diagnosis of keratoconus from corneal topography. Investigative Ophthalmology and Visual Science 2021;**62**(8):ARVO E-abstract 2021.

Toprak 2021 *{published data only}*

Toprak I, Cavas F, Velazquez JS, Alio del Barrio JL, Alio JL. Three-dimensional morphogeometric and volumetric characterization of cornea in pediatric patients with early keratoconus. *American Journal of Ophthalmology* 2021;**222**:102-11.

Velazquez-Blazquez 2020 *{published data only}*

Velazquez-Blazquez JS, Bolarin JM, Cavas-Martinez F, Alio JL. Emklas: A new automatic scoring system for early and mild keratoconus detection. *Translational Vision Science and Technology* 2020;**9**(2):1-18.

Vieira de Carvalho 2008 *{published data only}*

Vieira de Carvalho LA, Barbosa MS. Neural networks and statistical analysis for classification of corneal videokeratography maps based on Zernike coefficients: a quantitative comparison. *Arquivos Brasileiros de Oftalmologia* 2008;**71**(3):337-41.

Wang 2022 *{published data only}*

Wang L, Shen M, Shi C, Zhou Y, Chen Y, Pu J, et al. EE-Net: An edge-enhanced deep learning network for jointly identifying corneal micro-layers from optical coherence tomography. *Biomedical Signal Processing and Control* 2022;**71**(Pt B):103213.

Xu 2022b *{published data only}*

Xu Z, Feng R, Jin X, Hu H, Ni S, Xu W, et al. Evaluation of artificial intelligence models for the detection of asymmetric keratoconus eyes using Scheimpflug tomography. In: Clinical & Experimental Ophthalmology. 7 edition. Vol. 50. 2022:714-23.

Yucekul 2022 *{published data only}*

Yücekul B, Dick HB, Taneri S. Systematic Detection of Keratoconus in Optical Coherence Tomography: Corneal and Epithelial Thickness Maps. *Journal of Cataract & Refractive Surgery* 2022;**10**:1097.

Zghal 1997 *{published data only}*

Zghal I, Saragoussi JJ, Cotinat J, Renard G, Pouliquen Y. Automated keratoconus detection in fellow eyes of unilateral clinically keratoconus. *Journal Francais d'Ophtalmologie* 1997;**20**(4):284-91.

Zou 2019 *{published data only}*

Zou HH, Xu JH, Zhang L, Ji SF, Wang Y. Assistant diagnose for subclinical keratoconus by artificial intelligence. *Chinese Journal of Ophthalmology* 2019;**55**(12):911-5.

Additional references

Abràmoff 2016

Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology and Visual Science* 2016;**57**(13):5200-6.

Brown 2018

Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RV, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmology* 2018;**136**(7):803-10.

Brunner 2018

Brunner M, Czanner G, Vinciguerra R, Romano V, Ahmad S, Batterbury M, et al. Improving precision for detecting change in the shape of the cornea in patients with keratoconus. *Scientific Reports* 2018;**8**(1):1-7.

Cao 2022

Cao K, Verspoor K, Sahebjada S, Baird PN. Accuracy of machine learning assisted detection of keratoconus: a systematic review and meta-analysis. *Journal of Clinical Medicine* 2022;**11**(3):478.

Covidence [Computer program]

Covidence. Melbourne: Veritas Health Innovation, accessed 30 November 2022. Available at www.covidence.org.

Fan 2018

Fan R, Chan TC, Prakash G, Jhanji V. Applications of corneal topography and tomography: a review. *Clinical and Experimental Ophthalmology* 2018;**42**(2):133-46.

Ferdi 2019

Ferdi AC, Nguyen V, Gore DM, Allan BD, Rozema JJ, Watson SL. Keratoconus natural progression: a systematic review and meta-analysis of 11 529 eyes. *Ophthalmology* 2019;**126**(7):935-45.

Flynn 2016

Flynn TH, Sharma DP, Bunce C, Wilkins MR. Differential precision of corneal Pentacam HR measurements in early and advanced keratoconus. *British Journal of Ophthalmology* 2016;**100**(9):1183-7.

Gargeya 2017

Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;**124**(7):962-9.

Giri 2017

Giri P, Azar DT. Risk profiles of ectasia after keratorefractive surgery. *Current Opinion in Ophthalmology* 2017;**28**(4):337-42.

Gomes 2015

Gomes JA, Tan D, Rapuano CJ, Belin MW, Ambrósio RJ, Guell JL, et al. Global consensus on keratoconus and ectatic diseases. *Cornea* 2015;**34**(4):359-69.

GRADEpro GDT [Computer program]

GRADEpro GDT. Hamilton (ON): McMaster University (developed by Evidence Prime), accessed 27 February 2023. Available from gradepro.org.

Grassmann 2018

Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology* 2018;**125**(9):1410-20.

Gulshan 2016

Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;**316**(22):2402-10.

Harbord 2007

A unification of models for meta-analysis of diagnostic accuracy studies. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. *Biostatistics* 2007;**8**:239-51.

Hashemi 2020

Hashemi H, Heydarian S, Hooshmand E, Saatchi M, Yekta A, Aghamirsalim M, et al. The prevalence and risk factors for keratoconus: a systematic review and meta-analysis. *Cornea* 2020;**39**(2):263-70.

Hayes 2012

Hayes S, Khan S, Boote C, Kamma-Lorger CS, Dooley E, Lewis J, et al. Depth profile study of abnormal collagen orientation in keratoconus corneas. *Archives of Ophthalmology* 2012;**130**(2):251-2.

Hogarty 2019

Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. *Clinical and Experimental Ophthalmology* 2019;**47**(1):128-39.

Kanellopoulos 2013a

Kanellopoulos AJ, Asimellis G. Revisiting keratoconus diagnosis and progression classification based on evaluation of corneal asymmetry indices, derived from Scheimpflug imaging in keratoconic and suspect cases. *Clinical Ophtholmology* 2013;**7**:1539-48.

Kanellopoulos 2013b

Kanellopoulos AJ, Moustou V, Asimellis G. Evaluation of visual acuity, pachymetry and anterior-surface irregularity in keratoconus and crosslinking intervention follow-up in 737 cases. *International Journal of Keratoconus and Ectatic Corneal Diseases* 2013;**2**(3):95.

Kelly 2011

Kelly T, Williams KA, Coster DJ. Corneal transplantation for keratoconus: a registry study. *Archives of Ophthalmology* 2011;**129**(6):691-7.

LeCun 2015

LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**(7553):436-44.

Leeflang 2022

Leeflang MM, Steingart KR, Scholten RJ, Davenport C. Chapter 12: Drawing conclusions. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 2.0 (updated July 2023). Cochrane, 2023. Available from training.cochrane.org/handbook-diagnostic-test-accuracy/current.

Lin 2019

Lin SR, Ladas JG, Bahadur GG, Al-Hashimi S, Pineda R. A review of machine learning techniques for keratoconus detection and refractive surgery screening. *Seminars in Ophthalmology* 2019;**34**(4):317-26.

Lopes 2012

Lopes BT, Ramos IC, Faria-Correia F, Luz A, Freitas Valbon B, Belin MW, et al. Correlation of topometric and tomographic indices with visual acuity in patients with keratoconus. *International Journal of Keratoconus and Ectatic Corneal Diseases* 2012;**1**(3):167-72.

Lopes 2019

Lopes BT, Eliasy A, Ambrosio R. Artificial Intelligence in corneal diagnosis: where are we? *Current Ophthalmology Reports* 2019;**7**(3):204-11.

Macaskill 2010

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editor(s), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from

training.cochrane.org/handbook-diagnostic-test-accuracy/archive/v1.

Martínez-Abad 2017

Martínez-Abad A, Pinero DP. New perspectives on the detection and progression of keratoconus. *Journal of Cataract and Refractive Surgery* 2017;**43**(9):1213-27.

Meek 2005

Meek KM, Tuft SJ, Huang Y, Gill PS, Hayes S, Newton RH, et al. Changes in collagen orientation and distribution in keratoconus corneas. *Investigative Ophthalmology and Visual Science* 2005;**46**(6):1948-56.

Rabinowitz 2021

Rabinowitz YS, Galvis V, Tello A, Rueda D, García JD. Genetics vs chronic corneal mechanical trauma in the etiology of keratoconus. *Experimental Eye Research* 2021;**202**:108328.

Reitsma 2009

Reitsma JB, Rutjes AW, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editor(s), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2009. Available from training.cochrane.org/handbook-diagnostic-test-accuracy/archive/v1.

Review Manager 2020 [Computer program]

Review Manager 5 (RevMan 5). Version 5.4. Copenhagen: Nordic Cochrane Centre: Cochrane Collaboration, accessed 20 March 2023.

Röck 2018

Röck T, Bartz-Schmidt KU, Röck D. Trends in corneal transplantation at the University Eye Hospital in Tübingen, Germany over the last 12 years: 2004–2015. *PLOS One* 2018;**13**(6):e0198793.

SAS software [Computer program]

SAS software. Version 9.2. North Carolina, USA: SAS Institute Inc, accessed 15 October 2022. www.sas.com.

Sedghipour 2012

Sedghipour MR, Sadigh AL, Motlagh BF. Revisiting corneal topography for the diagnosis of keratoconus: use of Rabinowitz's KISA% index. *Clinical Ophthalmology* 2012;**6**:181-4.

Shibata 2018

Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Scientific Reports* 2018;**8**(1):1-9.

Stata software [Computer program]

StataCorp. 2021. Version Release 17. College Station, TX: StataCorp LLC, accessed 10 October 2022. www.stata.com.

Subhash 2013

Subhash HM, Wang RK. Optical coherence tomography: technical aspects. In: Liang R, editors(s). Biomedical Optical Imaging Technologies. Biological and Medical Physics, Biomedical Engineering. Berlin, Heidelberg: Springer, 2013:163-212. [DOI: 10.1007/978-3-642-28391-8_5]

Sykakis 2015

Sykakis E, Karim R, Evans JR, Bunce C, Amissah-Arthur KN, Patwary S, et al. Corneal collagen cross-linking for treating keratoconus. *Cochrane Database of Systematic Reviews* 2015, Issue 3. Art. No: CD010621. [DOI: 10.1002/14651858.CD010621.pub2]

Takwoingi 2022

Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C, et al. Chapter 10: Undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y, editor(s). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 2.0 (updated July 2023). Cochrane, 2023. Available from training.cochrane.org/handbook-diagnostic-test-accuracy/current.

Ting 2017

Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;**318**(22):2211-33.

Ting 2019

Ting DS, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Wong TY. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology* 2019;**103**(2):167-75.

Ting 2020

Ting DS, Foo VH, Yang LW, Sia ST, Ang M, Lin H, et al. Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology. *British Journal of Ophthalmology* 2020;**105**(2):158-68.

Wojtkowski 2010

Wojtkowski M. High-speed optical coherence tomography: basics and applications. *Applied Optics* 2010;**49**(16):D30-61.

Yang 2021

Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, et al, QUADAS-C Group. QUADAS-C: a tool for assessing risk of bias in comparative diagnostic accuracy studies. *Annals of Internal Medicine* 2021;**174**(11):1592-9.

Zadnik 1996

Zadnik K, Barr JT, Gordon MO, Edrington TB. Biomicroscopic signs and disease severity in keratoconus. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study Group. *Cornea* 1996;**15**(2):139-46.

References to other published versions of this review

Vandevenne 2021

Vandevenne MM, Favuzza E, Veta M, Lucenteforte E, Berendschot T, Mencucci R, et al. Artificial intelligence for detecting keratoconus. *Cochrane Database of*

Systematic Reviews 2021, Issue 12. Art. No: CD014911. [DOI: 10.1002/14651858.CD014911]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Abdelmotaal 2020

Study characteristics	
Patient Sampling	Single-centre, retrospective, case-control study. Scheimpflug tomographic (Pentacam (Oculus GmbH, Wetzlar, Germany)) images obtained from non-consecutive refractive surgery candidates, people with unilateral or bilateral keratoconus, or subclinical keratoconus in Egypt (3218 images from 3218 eyes of 1669 people).
Patient characteristics and setting	• 1108 healthy eyes selected from non-consecutive refractive surgery candidates • 1038 keratoconus eyes • 1072 subclinical keratoconus eyes
Index tests	Convolutional neural network (CNN) for 4-map selectable display images. The CNN was trained with corneal colour-coded maps of whole Scheimpflug images.
Target condition and reference standard(s)	The keratoconus class included those with a clinical diagnosis of keratoconus or an irregular cornea (as determined by distorted keratometry mires or distortion of retinoscopic red reflex, or both) and the following topographic findings. • Focal steepening located in a zone of protrusion surrounded by concentrically decreasing power zones • Focal areas with dioptric values > 47.0 D • I-S asymmetry measured as > 1.4 D, or angling of the hemi-meridians in an asymmetric or broken bowtie pattern with skewing of the steepest radial axis. The subclinical keratoconus class included subtle corneal topographic changes in the aforementioned keratoconus abnormalities in the absence of slit-lamp or visual acuity changes typical of keratoconus. The cases were labelled before the analysis with the algorithm by 2 experienced corneal specialists and any disagreements were reviewed by a 3rd-specialist.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			

Abdelmotaal 2020 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Accardo 2002

Study characteristics	
Patient Sampling	Single-centre, retrospective, case-control study including 396 corneal topographic maps (396 eyes, 198 participants), obtained with a videokeratoscope (EyeSys, EyeSys Vision, Houston, Texas), selected from the cases of keratoconus or other conditions recorded over a 3-year period at the centre, in Italy.
Patient characteristics and setting	• Keratoconus cases (110 images, 110 eyes, 55 participants): mild and moderate keratoconus severity, with a sagittal cone apex power < 53 D, where no clinical sign was present (early keratoconus) or only the Vogt's striae were detected (mild or moderate keratoconus) • Others (166 images of 166 eyes, 83 participants): various bilateral non-keratoconus conditions, congenital astigmatism, contact lens corneal warpage
Index tests	A neural network using as input the parameters of both eyes of the same subject and as output the 3 categories of clinical classification (normal, keratoconus, other alterations) for each subject, a low number of neurons in the hidden layer (< 10), and a learning rate of 0.1.
Target condition and reference standard(s)	Keratoconus group comprised cases of mild and moderate severity with a sagittal cone apex power < 53 D, with no clinical sign (early keratoconus) or only the Vogt's striae, and keratoconus suspects that met one of the following criteria. • The pathology was already present in the other eye with the same topographic pattern • A family history of keratoconus was present • The apex progression of the corneal protrusion was >1 D after 1 or 2 years of follow-up The maps were classified before the analysis with the algorithm, the number of observers is unclear.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable.
Notes	This work was partially supported by the University of Trieste (MURST60%) and by Burlo Garofolo Hospital in Trieste, Italy.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Accardo 2002 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Al-Timemy 2021

Study characteristics	
Patient Sampling	The study design is unclear, it seems to be case-control study. 3794 Pentacam (Oculus GmbH, Wetzlar, Germany) corneal images from University of Sao Paulo were included and an independent validation subset with 1050 images was collected from 150 eyes of 85 subjects from a separate centre in Brazil.

Al-Timemy 2021 *(Continued)*

Patient characteristics and setting	The criteria for keratoconus diagnosis are unclear. People with manifest keratoconus, suspected keratoconus and normal eyes were included.
Index tests	A hybrid deep learning model which integrates multiple convolutional neural network (CNN) models for detecting keratoconus based on corneal topographic maps.
Target condition and reference standard(s)	The criteria for keratoconus diagnosis are unclear. Eyes were labelled as keratoconus suspects if corneal topography included atypical, localized steepening or an asymmetrical bowtie pattern; the keratometric curvature was > 47.00 D, the oblique cylinder was > 1.50 D, the central corneal thickness was < 500 µm, BAD-D was between 1.6 and 3.0. 3 corneal specialists performed the eye classification, before the analysis with the algorithm.
Flow and timing	All cases were included in reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported by National Institute of Health (NIH), National Eye Institute (NEI), and Bright Focus Foundation.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Al-Timemy 2021 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Almeida 2022

Study characteristics	
Patient Sampling	Multicentre, case-control study. All participants were examined at the Visum Eye Center and Rio Claro Eye Institute between January 2012 and January 2019. Exclusion criteria were a history of ocular trauma, corneal scarring, and neurotrophic keratopathy. • Normal Eyes: 2296 people who underwent LASIK or photorefractive keratectomy and were stable after at least 18 months of follow-up. • Very Asymmetric Eyes With Normal Topography and Tomography Group: 187 eyes of 187 people with very asymmetric eyes with normal topography (VAE-NT) in 1 eye and frank ectasia (VAE-E) in the fellow eye • Ectasia Group: 410 people (1 eye each) with bilateral clinical keratoconus
Patient characteristics and setting	• 410 eyes with keratoconus • 2296 healthy corneas • 187 very asymmetric eyes with a normal topography of the cornea.
Index tests	Multiple logistic regression analysis (MLRA) is based on the logistic function that bounds its output within the range of 0 to 1. To build the algorithm extracted from MLRA, 22 variables were used.
Target condition and reference standard(s)	All eyes were examined by rotating Scheimpflug corneal and anterior segment tomography (Pentacam HR, Oculus Optikgerate GmbH). Image quality was checked to ensure that only cases with acceptable-quality images were included. All cases were reviewed by an experienced fellowship-trained corneal specialist (G.C.A.J.) for correct classification into keratoconus and VAE-NT groups. Objective criteria for considering normal topography included objective front surface curvature metrics derived from Pentacam. Normal topography criteria were rigorously considered based on the objective criteria of a maximum keratometry curvature (Kmax; steepest front keratometry) of < 47.2 D, a

Almeida 2022 *(Continued)*

	paracentral I-S asymmetry value at 6 mm (3 mm radii) of < 1.45, and a keratoconus percentage index score of < 60. An objective criterion for normal tomography criteria was adopted for the control group and the VAE-NT group, and the maximum values were 3.8 mm for anterior chamber depth, 4 mm for front apical elevation, 5 mm for front corneal elevation at the thinnest point, and 12 mm for front corneal elevation in the central 4.0 mm. The corresponding posterior elevation values were 7 mm, 13 mm, and 25 mm.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This work was supported by the Sao Paulo State Research Support Foundation (FAPESP, grant nos: 2015/17226-7 and 2019/04475-0) and the National Council for Scientific and Technological Development (CNPq, grant no: 306808/2018-8.). The funding organizations had no role in design or conduct of this research, and they have no related commercial interests.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Almeida 2022 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Arbelaez 2012

Study characteristics	
Patient Sampling	Retrospective case series. Clinical data and corneal examinations were retrieved from clinical records from 2 centres (Oman and Italy). 3502 eyes were enroled.
Patient characteristics and setting	According to the clinical diagnosis, participants were classified into the following 4 groups. • Keratoconus • Subclinical keratoconus (early keratoconus, forme fruste and suspected) • Other conditions (history of refractive surgery, penetrating keratoplasty, or ocular trauma)

Arbelaez 2012 *(Continued)*

	• Normal eyes, enroled among subjects undergoing a routine ophthalmological examination for minor refractive defects. Each group was divided into a training set (including 200 eyes) to be used to develop the keratoconus detection program and a validation set (including the remaining eyes). Participants were excluded if tomography scans had poor quality.
Index tests	Classification algorithm based on support vector machine (SVM), a supervised learning technique that can be used for pattern classification. It analysed symmetry index of front and back corneal curvature, best fit radius of the front corneal surface, Baiocchi Calossi Versaci front index (BCVf) and BCV back index (BCVb), root-mean-square of front and back corneal surface higher order aberrations, and thinnest corneal point provided by a Scheimpflug camera combined with Placido corneal topography (Sirius, CSO, Italy).
Target condition and reference standard(s)	Unclear who performed the classification of the eyes, which was done before the analysis with the algorithm.
Flow and timing	It is unclear if all cases received the same reference standard. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	No		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Arbelaez 2012 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Bessho 2006

Study characteristics	
Patient Sampling	Retrospective, multicentre, case-control study. 165 eyes of 120 subjects were included at 2 centres in Japan.
Patient characteristics and setting	• People with keratoconus or keratoconus suspect comprised the keratoconus group. • People with post-photorefractive keratectomy or with-the-rule astigmatism and people without disease comprised the non-keratoconus group.
Index tests	Fourier-incorporated keratoconus detection Index (FKI) created performing a logistic regression analysis with a training set to differentiate the keratoconus group from the non-keratoconus group. The index is based on information obtained by Fourier analysis from not only the anterior corneal surface but also from the posterior corneal surface and corneal thickness.
Target condition and reference standard(s)	Corneal topographic analysis was performed with a slit-scanning corneal topographer (Orbscan II, Bausch & Lomb). It is unclear how the diagnosis was made; however, cases were classified before the inclusion.
Flow and timing	It is unclear if all cases received the same reference standard. All data were included in a 2 × 2 table.

Bessho 2006 *(Continued)*

Comparative	Not applicable
Notes	This study was supported in part by Grant-in-Aid No.15591854 for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (N. Maeda), and by a research grant from the Osaka Eye Bank Foundation, Suita, Japan (N. Maeda).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		

Bessho 2006 *(Continued)*

Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		

Cao 2020

Study characteristics	
Patient Sampling	Retrospective, case-control study including the following groups. • People with subclinical keratoconus recruited from public and private clinics at the Royal Victorian Eye and Ear Hospital and private consulting rooms and optometry clinics in Melbourne, Australia, as part of the Australian Study of Keratoconus • Controls recruited from the GEnes in Myopia study
Patient characteristics and setting	• Subclinical keratoconus was defined as eyes with abnormal corneal topography, including I-S localized steepening or asymmetric bowtie pattern and no detectable clinical signs. • The control group consisted of refractive error subjects with no ocular disease that may affect refraction in the eyes.
Index tests	Random forest method using 11 tomographic parameters (Pentacam) for the diagnosis of subclinical keratoconus.
Target condition and reference standard(s)	Subclinical keratoconus was defined as those eyes with abnormal corneal topography, including I-S localized steepening or asymmetric bowtie pattern and no detectable clinical signs. It is unclear how the diagnosis was made; however, cases were classified before the inclusion.
Flow and timing	All cases were included in reference standard and index test. All data were included in a 2 × 2 table.
Comparative	It is unclear whether different AI tests were developed and interpreted blind or independently and without knowledge of the results of each other, and whether missing data and their causes were similar for each AI test.
Notes	This study was supported by the Australian National Health and Medical Research Council (NHMRC) project Ideas grant APP1187763 and Senior Research Fellowship (1138585 to PNB), the Louisa Jean De Bretteville Bequest Trust Account, University of Melbourne, the Angior Family Foundation, Keratoconus Australia, Perpetual Impact Philanthropy grant (SS), and a Lions Eye Foundation Fellowship (SS). The Centre for Eye Research Australia (CERA) receives Operational Infrastructure Support from the Victorian Government.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			

Cao 2020 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		

Cao 2020 *(Continued)*

Are the proportions and reasons for missing data similar for all index tests?	Unclear	
		Unclear risk

Cao 2021a

Study characteristics			
Patient Sampling	Single-centre, retrospective, case-control study. The data collection was conducted at the Royal Victorian Eye and Ear Hospital in Australia from between 2007 and 2019.		
Patient characteristics and setting	• Subclinical keratoconus group: subjects were defined as those eyes with abnormal corneal tomography, including I-S localized steepening or an asymmetric bowtie pattern, but without detectable clinical signs on slit-lamp biomicroscopy and retinoscopy examination. • The control group consisted of eyes that had no known history of any corneal disorder, but may have presented with another ocular condition that did not introduce a corneal change.		
Index tests	Random forest-based model trained using a modest number (15) of components derived from a reduced dimensionality representation of complete Pentacam system parameters.		
Target condition and reference standard(s)	The diagnosis was made by an experienced optometrist together with > 1 cornea specialist, using Pentacam corneal tomography system; cases were classified before the inclusion.		
Flow and timing	All cases were included in reference standard and index. All data were included in a 2 × 2 table.		
Comparative	Not applicable		
Notes	This study was supported by the Australian National Health and Medical Research Council (NHMRC) project Ideas grant APP1187763 and Senior Research Fellowship (1138585 to PNB), Lions Eye Donation Service (SS), Angior Family Foundation (SS), Perpetual Impact Philanthropy grant (SS), and Keratoconus Australia Funding (SS). The Centre for Eye Research Australia (CERA) receives Operational Infrastructure Support from the Victorian Government. The sponsor or funding organizations had no role in the design or conduct of this research.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		

Cao 2021a *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Carvalho 2005

Study characteristics			

Carvalho 2005 *(Continued)*

Patient Sampling	Retrospective, single-centre case-control study. 80 corneal maps were selected from the database of the EyeSys System 2000 (EyeSys Vision, Houston, TX) topographer in Brazil.
Patient characteristics and setting	80 corneal maps of different people were selected according to the following 5 categories (16 corneas for each category). • Regular cornea • With-the-rule astigmatisms • Against-the-rule astigmatisms • Keratoconus • Post-LASIK Criteria for diagnosis of keratoconus are unclear. Corneal maps had few or no nose or eyelid shadows; only the right eye of each person was allowed. Right and left eyes of the same person were not used. The investigators excluded corneas with incipient keratoconus, keratoconus with high degrees of symmetrical astigmatism, and other cases for which a single prevailing diagnosis could not be issued. In the case of regular profiles, given that even the most symmetrical corneas have some degree of with-the-rule astigmatism, only corneas with simulated keratometry 0.25 D were considered "regular" or "normal."
Index tests	A neural network which used the first 15 Zernike coefficients
Target condition and reference standard(s)	The selection and classification were performed by 2 eye care specialists, with unclear criteria. However, cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported in part by FAPESP (São Paulo Research Foundation) process #03132–8.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			

Carvalho 2005 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Castro-Luna 2020

Study characteristics	
Patient Sampling	Retrospective single-centre case-control study in 60 eyes from 60 people from the Department of keratoconus of INVISION Ophthalmology clinic in Almería, Spain
Patient characteristics and setting	Participants were divided into the following 2 groups depending on their preliminary diagnosis based on the classical topographic criteria. • Control group without topographic alteration (30 eyes) • Keratoconus group (30 eyes). The keratoconus group included people with asymmetric bow tie in the topographic image and ≥ 1 sign of keratoconus in the examination with the slit lamp, such as stromal thinning, conical protrusion of the cornea at the apex, Fleischer's ring, Vogt's striae, or anterior stromal scar.

Castro-Luna 2020 *(Continued)*

	Grade 4 keratoconus with excessively distorted corneal topography was excluded. All cases were examined using the CSO topography system (CSO, Firenze, Italy).
Index tests	Bayesian network classifier for keratoconus identification that uses previously developed topographic indices, calculated directly from the digital analysis of the Placido ring images.
Target condition and reference standard(s)	It is unclear who performed the selection and classification. However, cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This research was partially supported by the Andalucian regional government (grant PIN-0530-2017). Research of A.M.-F. and D.R.-L. was also supported in part by the Spanish Government – European Regional Development Fund (grant MTM2017-89941-P), the Andalucian regional government (research group FQM-229), and the University of Almería (Campus de Excelencia Internacional del Mar CEIMAR). A.M.-F. acknowledges an additional support from the Carlos I Institute of Theoretical and Computational Physics, while D.R.-L. thanks the support from CDTIME (Center for Development and Transfer of Mathematical Research to Industry, University of Almería).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Castro-Luna 2020 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Cavas-Martinez 2017

Study characteristics	
Patient Sampling	Single-centre case-control study in Spain, including 464 eyes of 464 participants
Patient characteristics and setting	Participants were divided into the following 2 groups. • Control group (143 healthy eyes) • Keratoconus group (321 keratoconus eyes)
Index tests	A model of detection of early keratoconus (only grade 1) obtained by logistic regression considering the new parameters defined according to a new geometric approach
Target condition and reference standard(s)	The standard criteria for keratoconus diagnosis were the presence of an asymmetric bowtie pattern in corneal topography, KISA% index ≥ 100%, a central keratometry with different cut-off values to keratoconus suspect (> 47.2 D), an I-S asymmetry with a cut-off value of 1.4 D difference between average inferior and superior corneal powers at 3 mm from the centre of the cornea, as well as other topographic indices and ≥ 1 keratoconus sign on slit-lamp examination, such as stromal thinning, conical protrusion on the cornea at the apex, Fleischer's ring, Vogt's striae, or anterior stromal scar. Corneal analysis was performed by the Sirius system (CSO, Italy). It seems that a single experienced ex-

Cavas-Martinez 2017 *(Continued)*

	aminer was involved in selection and classification. However, cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	The study was carried out in the framework of the Thematic Network for Co-Operative Research in Health (RETICS) reference number RD12/0034/0007 and RD16/0008/0012, financed by the Carlos III Health Institute ± General Subdirection of Networks and Cooperative Investigation Centers (R&D&I National Plan 2008±2011) and the European Regional Development Fund (FEDER). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		

Cavas-Martinez 2017 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Chan 2015

Study characteristics	
Patient Sampling	Retrospective case-control study; 128 topographic images of 128 people were selected at the Singapore National Eye Center
Patient characteristics and setting	• The forme fruste keratoconus group (24 images) involved clinically and topographically normal eyes with the contralateral eye showing frank keratoconus. These cases were obtained from the database of people with keratoconus from the Singapore National Eye Center. The diagnosis of keratoconus in the contralateral eye was reconfirmed by clinical examination and evaluation of topographies from the Orbscan IIz corneal topography system (Bausch + Lomb TechnoLas, Munich, Germany) and Tomey keratoconus screening system (TMS, software version 2.4.2J, Tomey TMS-2N; Tomey Corp, Nagoya, Japan) by a corneal sub-specialist. • The control group (104 images) involved normal preoperative topographies of people who had myopic LASIK (with or without astigmatism) performed at least 4 years before with no resultant ectasia.
Index tests	The SCORE Analyzer is based on a linear regression analysis that constructs a set of linear functions of variables known as discriminant functions. It combines 12 Placido and tomographic indices in a weighted fashion to classify corneas as suspicious for keratoconus or normal.
Target condition and reference standard(s)	Clinically evident keratoconus was defined by evidence of ≥ 1 slit-lamp biomicroscopic findings including conical protrusion of the cornea at the apex, Fleischer's ring, Vogt's striae, and corneal stromal thinning. The classification was performed by 1 corneal specialist. Cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.

Chan 2015 *(Continued)*

Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		

Chan 2015 *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk
DOMAIN 5: Comparative	

Chandapura 2019

Study characteristics	
Patient Sampling	Retrospective case-control study, involving 439 eyes from 2 centres (India and Brazil). Comparison between 4 AI models.
Patient characteristics and setting	• Keratoconus • Form fruste keratoconus • Very asymmetric eyes with normal topography (VAE-NT)
Index tests	Random forest models based on Pentacam (Oculus GmbH, Wetzlar, Germany) or OCT parameters (OCT topography of the Bowman's layer)
Target condition and reference standard(s)	• Keratoconus group: corneas with significant inferior steepening, asymmetric astigmatism, and corneal thinning on both OCT (RTVue, Optovue Inc., Irvine) and Scheimpflug (Pentacam, OCULUS Optikgerate). • Form fruste keratoconus: very mild localized steepening and suspicious anterior corneal surface topography on both devices. • Very asymmetric eyes with normal topography (VAE-NT): fellow eyes of people with highly asymmetric keratoconus. Examination of topographies of the anterior surface was performed by only 1 experienced refractive surgeon, who was masked to the information (disease present in 1 or both eyes) about the participants and the eyes. Classification was performed before the index tests.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	It is unclear whether different AI tests were developed and interpreted blind or independently and without knowledge of the results of each other. Missing data and their causes were similar for each AI test.
Notes	Indo-German Science and Technology Center, Grant/Award Number: SIBAC

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		

Chandapura 2019 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Yes		
		Unclear risk	

Chastang 2000

Study characteristics	
Patient Sampling	Retrospective case-control study, single-centre (France) involving 208 corneal topographies (EyeSys System 2000) of 208 corneas from 8 groups of participants.
Patient characteristics and setting	Participants were classified by the following diagnoses. • Normal • Regular astigmatism • Cataract surgery • Radial keratotomy • Excimer laser photorefractive keratectomy • Non-freeze myopic keratomileusis • Penetrating keratoplasty • Keratoconus
Index tests	Binary decision tree. In the first step, the distribution of keratoconic and non-keratoconic patterns was studied based on the value of each index in the training set. For each index, corneas with an index value higher than the threshold (or cut-off value) were classified as keratoconic corneas (positive test), whereas corneas with an index value less than the threshold were classified as non-keratoconic (negative test). In the second step, binary decision trees were built by combining 2 indices to improve the classification method. The 6 indices with the highest sensitivity and specificity were used in these models. The first index was used to divide the training set into 2 populations (i.e. population with a positive test and population with a negative test) based on the previously calculated optimum threshold. In each of these populations, the distribution of keratoconic and non-keratoconic patterns was studied based on the value of the second index. In each population, sensitivity and specificity curves as a function of the second index threshold were generated to evaluate the optimum cut-off value. This resulted in 2 thresholds according to the response to the first test. In fact, the second index's most efficient threshold (i.e. the threshold with maximum sensitivity and specificity) in the population with a positive test was different from that in the population with a negative test. A cornea was classified as keratoconic when the second test was positive.
Target condition and reference standard(s)	Maps were classified by 2 cornea specialists based on clinical records and topographic appearances, before the index test.
Flow and timing	All cases were included in reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Unclear whether different AI tests were developed and interpreted blind or independently and without knowledge of the results of each other, and whether missing data and their causes were similar for each AI test.
Notes	Supported in part by the Fondation Claude Bernard, Paris, France.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			

Chastang 2000 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Chastang 2000 *(Continued)*

DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Unclear		
		Unclear risk	

Chen 2021

Study characteristics			
Patient Sampling	Multicentre retrospective case-control study comparing 4 convolutional neural network methods. It included 1926 images from the Pentacam (Oculus GmbH, Wetzlar, Germany) of keratoconic and healthy volunteers' eyes provided by 3 centres (UK, Iran, New Zealand).		
Patient characteristics and setting	• Keratoconic scans were classified according to the Amsler-Krumeich classification. Only scans of acceptable quality were included. • Control group: subjects with a BAD-D < 1.6 SDs from normative values.		
Index tests	Convolutional neural network method that uses 4 colour-coded corneal maps obtained by a Scheimpflug camera (Pentacam)		
Target condition and reference standard(s)	The definition of keratoconus is unclear. Unclear who performed the classification. However, cases were classified before inclusion.		
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.		
Comparative	Unclear whether different AI tests were developed and interpreted blind or independently and without knowledge of the results of each other. Missing data and their causes were similar for each AI test.		
Notes	The study authors have not declared a specific grant for this research from any funding agency in the public, commercial, or not-for-profit sectors.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		

Chen 2021 *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Yes		
		Unclear risk	

Cohen 2022

Study characteristics	
Patient Sampling	Single centre, retrospective, case-control study that evaluated 8526 corneal tomography examinations of 2525 participants obtained between November 2010 and July 2017 with the Galilei dual Scheimpflug/Placido disc analyzer system (software version 5.2.1; Ziemer Ophthalmic Systems, Port, Switzerland). Low-quality samples were excluded.
Patient characteristics and setting	Of the 7104 included samples: • 4088 were labelled as normal; • 1299 were labelled as suspect irregular cornea; and • 2614 were labelled as keratoconus. Label distribution was similar in train and test sets.
Index tests	Random forest; the model integrated keratoconus prediction indexes of the device in addition to the 94 instrument-derived output parameters. The model was first trained and tested, then validated with a separate validation set.
Target condition and reference standard(s)	All images were graded by a single cornea specialist (D.V.). A normal cornea would have a regular spherical or spherocylindrical curvature, thinning toward the centre without epicentral posterior or anterior elevation, with relatively normal numerical values. A suspected irregular cornea describes an at-risk cornea. Such a cornea may have subtle inconclusive signs like I-S values outside the reference range or aberrant C-shaped or mild posterior surface elevations. Alternatively, a suspected irregular cornea may have an unusual corneal thinning.
Flow and timing	All cases were included in reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	All study authors declared that they received no grant support or research funding for the study. All study authors certified that they had no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the manuscript.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		

Cohen 2022 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Cohen 2022 *(Continued)*

Consejo 2020

Study characteristics	
Patient Sampling	Prospective case-control study, involving 50 eyes selected in a single centre in Belgium
Patient characteristics and setting	Scheimpflug single-image snapshots obtained with Corvis-ST (Oculus, Germany) were analysed and grouped as follows. • Normal (25 eyes): intraocular pressure of 15–17 mmHg and corneal astigmatism < 0.75 D • Keratoconus (25 eyes): various stages of keratoconus, including clear cornea, non-severe corneal thinning, Fleischer's ring at the apex, and anterior or posterior corneal steepening
Index tests	The combination of central corneal thickness with microscopic parameters extracted from statistical modelling of light intensity distribution
Target condition and reference standard(s)	The definition of keratoconus is unclear. The classification was performed by only 1 experienced ophthalmologist, before the index test.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable.
Notes	This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779960 and support from the Statutory Funds of Wroclaw University of Science and Technology. MW received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 666295 and from the financial resources for science in the years 2016 to 2019 awarded by the Polish Ministry of Science and Higher Education for the implementation of an international co-financed project. JJR received a grant from the Flemish Fund for Scientific Research (FWO-TBM T000416N).

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Consejo 2020 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

De Almeida Jr 2021

Study characteristics	
Patient Sampling	Prospective, single-centre, case-control study, including a data set of 777 images of 777 participants
Patient characteristics and setting	Data set included data from a 777 people, distributed as follows. • 411 healthy eyes (people undergoing LASIK or photorefractive keratectomy)

De Almeida Jr 2021 *(Continued)*

	• 302 people with bilateral clinical keratoconus (KISA% index > 100% and ≥ 1 of the following biomicroscopic signs: Fleischer's ring, focal stromal thinning, Vogt's striae) • 64 people with very asymmetric ectasia and with 1 healthy cornea and frank ectasia in the fellow eye.
Index tests	AI model based on Paraconsistent Feature Engineering (PFE) and Support Vector Machine (SVM), that received subsets of the 52 original Pentacam tomographic descriptors as input and produced, as output, the scalar value called Corneal Tomography Multivariate Index (CTMVI)
Target condition and reference standard(s)	Classification was performed by only 1 cornea specialist, before the index test.
Flow and timing	All cases were included in the reference standard and index. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Funding: CAPES (Coordination for Improvement of Higher-education Personnel), FAPESP (São Paulo Research Foundation) process no. 2015/17226-7 and process no. 2019/04475-0, and CNPq (National Council for Scientific and Technological Development) process no. 306808/20018-8. The study was supported by FAPESP [2015/17226- 7] to GCAJr and [2019/04475-0] to RCG; CNPq (National Council for Scientific and Technological Development) [306808/20018-8] to RCG; PIBIC-CNPq to JSN.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

De Almeida Jr 2021 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Elsawy 2021

Study characteristics	
Patient Sampling	Prospective, single-centre, case-control study. People with keratoconus, dry eye, Fuchs' dystrophy, or normal corneas were consecutively recruited from a single centre in the USA.
Patient characteristics and setting	The investigators prospectively recruited 478 participants (875 eyes, 158,220 AS-OCT images). The images were grouped as follows according to the clinical diagnosis. • 45,900 healthy • 16,740 Fuchs' endothelial dystrophy • 64,800 keratoconus • 30,780 dry eye syndrome
Index tests	A multidisease deep learning diagnostic network of common corneal diseases, using AS-OCT images (Envisu R2210, LEICA, USA)
Target condition and reference standard(s)	Classification was performed by 1 cornea specialist, who was masked to the automated diagnosis given by the algorithm.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable

Elsawy 2021 *(Continued)*

Notes	The study was supported by an NEI K23 award (K23EY026118), NEI core center grant to the University of Miami (P30 EY014801), and Research to Prevent Blindness (RPB). Financial Disclosures: United States Non-Provisional Patent (application no. 14/247903) and United States Provisional Patent (application no. 62/445,106) (to M.A.); United States Non-Provisional Patents (application no. 8992023 and 61809518), and PCT/US2018/013409 (to M.A. and A.E.). The patents and Patent Cooperation Treaty are owned by University of Miami and licenced to Resolve Ophthalmics, LLC. M.A. is an equity holder and sits on the Board of Directors for Resolve Ophthalmics, LLC. The funding organization had no role in the design or conduct of the research.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Elsawy 2021 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Feizi 2016

Study characteristics	
Patient Sampling	Prospective single-centre case-control study. 210 eyes of 210 people were included in 1 centre in Iran.
Patient characteristics and setting	• Refractive surgery candidates (normal subjects) • People affected by keratoconus • People affected by subclinical keratoconus
Index tests	Logistic regression analysis of sets of parameters obtained with Galilei dual Scheimpflug system (Ziemer Ophthalmic System AG, Port, Switzerland)
Target condition and reference standard(s)	The diagnosis of subclinical keratoconus and keratoconus was based on clinical slit-lamp findings (stromal thinning, conical protrusion, Fleischer's ring, and Vogt's striae) and characteristic patterns based on Placido disc corneal topography (Tomey, EM-3000, version 4.20, Nagoya, Japan). Participants who had 1 abnormal biomicroscopic finding and 1 major or 2 minor criteria were diagnosed with keratoconus. Participants with a normal appearing cornea and 1 major or 2 minor topographic criteria were diagnosed with subclinical keratoconus. It seems that a single experienced examiner was involved in classification. However, cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funds, grants, or other support were received.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			

Feizi 2016 *(Continued)*

Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Gairola 2022

Study characteristics

Gairola 2022 *(Continued)*

Patient Sampling	Single-centre, retrospective, case-control study. First data set (64 participants, 114 eye samples) obtained from topography device, SmartKC at the Sankara Eye Hospital in Bengaluru, India. Second data set (2110 samples; 1637 non-keratoconus and 473 keratoconus) consisted of downloaded anonymized data from the Keratron device database for all the people who took the corneal topography examination at the hospital from April 2008 to May 2010.
Patient characteristics and setting	• First data set included 68 non-keratoconus eyes and 46 keratoconus eyes. • Second data set included 2110 samples, of which 1673 were non-keratoconus and 473 were keratoconus.
Index tests	Convolutional neural network. The network is organized into 2 branches – 1 each for axial and tangential heatmaps – with a shared convolutional backbone, followed by 2 distinct feed-forward classifiers, 1 for each branch. The shared backbone comprises the convolutional layers from a ResNet34 model. The model has a 2-class (keratoconus versus non-keratoconus) classification task. The article provides a clear explanation of the model and training procedure.
Target condition and reference standard(s)	The first data set was diagnosed by 1 senior ophthalmologist at the hospital. The diagnoses in the second data set were obtained based on the PPK-based classification.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		

Gairola 2022 *(Continued)*

If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Gao 2022

Study characteristics	
Patient Sampling	Single-centre, retrospective, case-control study, in the Affiliated Eye Hospital of Wenzhou Medical University, China. A total of 208 participants (1040 corneal topography images) were evaluated and divided into the following 3 groups. • Keratoconus • Subclinical keratoconus • Healthy controls

Gao 2022 *(Continued)*

	Data were collected between 2012 and 2018 using the Pentacam system and analysed from February 2019 to December 2021. The corneal data of pachymetry and elevation were exported from the Pentacam HR system (Oculus, Optikgeräte GmbH, Wetzlar, Germany).
Patient characteristics and setting	Each image was previously assigned to the following 3 groups. • Normal (70 eyes; mean age 28.7 (SD 2.6) years) • Subclinical keratoconus (48 eyes; mean age 24.6 (SD 5.7) years) • Keratoconus (90 eyes; mean age, 25.9 (SD 5.4) years). The data set was randomly split into 70% for training and 30% for testing.
Index tests	Neural network, called Keratoscreen. The preprocessing model separately established Zernike coefficients (ZC) data sets of 5 corneal surface maps: the anterior corneal curvature (DA), posterior corneal curvature (DP), anterior corneal elevation (DAE), posterior corneal elevation (DPE), and corneal thickness data (pachymetry DPAC). The first N (N #15) ZC terms (5 orders, when N = 15) obtained from each map were used to form ZC data sets. KeratoScreen, with L = 4 layers (1 input layer, 2 hidden layers, and 1 output layer), was used in this study. The neurons on the input layer have N nodes that take an input vector ZC and outputs through 2 hidden layers. To train the algorithm they used 20 repeats with the randomly generated training and test sets, and the results were averaged. However, it is unclear whether they reused the same data for training and testing.
Target condition and reference standard(s)	• For the keratoconus group, 1 of the following on slit-lamp biomicroscopy: Vogt's striae, stromal thinning, Fleischer's ring, or corneal scarring. The central average keratometry was > 47.0 D. Asymmetric topographical features included I-S values >2.0 D of the vertical power gradient across the central 6-mm region. The subjects had no history of wearing contact lenses, ocular surgery, or extensive scarring. • The normal eye group was selected using the following criteria: no clinical symptoms of keratoconus or subclinical keratoconus in either eye, central average keratometry < 45.0 D, I-S value of corneal topographical features < 1.4 D at the vertical power gradient across the 6-mm region, the spherical equivalent was within 66.0 D, and astigmatism was within 2 D. The subjects also had no ocular disease, no prior ophthalmic surgery, and no cornea ectasia on a slit lamp. Only 1 eye was randomly selected from normal subjects for analysis. • The subclinical keratoconus group was selected from the participants with undiagnosed eyes of unilateral keratoconus, based on the following criteria: 1 eye was diagnosed with keratoconus; in the contra, there were no clinical symptoms of keratoconus during a slit-lamp examination, retinoscopy, or ophthalmoscopy; the central average keratometry in this eye group was 45.0 D; the I-S value of the corneal topographical features was > 1.4 D at the vertical power gradient through the central 6-mm region; the spherical equivalent was within 66.0 D; astigmatism was within 2 D; and these subjects had no history of contact lens wear or ocular surgery.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	The research received funding through 21NDJC309YBM from the Zhejiang Philosophy and Social Science Planning Project (China), LZ21F020008 from the Natural Science Foundation of Zhejiang Province (China), 2019C03045 from the National Key Project of Research and Development Program of Zhejiang Province (China), and LY21A040001 from the Natural Science Foundation of Zhejiang Province (China).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			

Gao 2022 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Gao 2022 *(Continued)*

Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Ghaderi 2021

Study characteristics			
Patient Sampling	Retrospective, single-centre (Iran), case-control study. A data set of 450 people with keratoconus and healthy controls was analysed.		
Patient characteristics and setting	Data set of people with keratoconus and healthy controls. Scheimpflug tomographic images (Pentacam HR (Oculus GmbH, Wetzlar, Germany)) were analysed.		
Index tests	Ensemble learning system, based on combining multiple initial classifiers as experts for primary classification and a combination rule for combining results of classifiers.		
Target condition and reference standard(s)	Unclear definition of keratoconus. Unclear who performed the classification; it seems 1 cornea specialist. Unclear if reference standard results were interpreted without knowledge of the results of the index test.		
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.		
Comparative	Not applicable		
Notes	No funds, grants, or other support were received.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**

Ghaderi 2021 *(Continued)*

DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Issarti 2019

Study characteristics	
Patient Sampling	Retrospective, single-centre, case-control study that used a previously collected database containing the Pentacam Scheimpflug measurements of 851 eyes (Pentacam (Oculus GmbH, Wetzlar, Germany)).

Patient characteristics and setting	• Normal eyes (n = 312): bilateral normal topography, with no systemic or ocular disease, without any prior ocular surgery, and no slit-lamp findings suggestive of cornea ectasia, recruited during a previous epidemiology study
	• Keratoconic eyes (n = 539), divided into the following 3 groups
	  ○ Mild keratoconus
	  ○ Moderate keratoconus
	  ○ Keratoconus suspect

Index tests	A CAD (computer-aided diagnosis) system, which combines a feedforward neural network (FFN) and a Grossberg-Runge Kutta architecture to detect clinical and suspect keratoconus.
Target condition and reference standard(s)	• Mild keratoconus: clear cornea, tomography maps compatible with Keratoconus, a Fleischer's ring at the apex base, slight thinning, and anterior or posterior corneal steepening • Moderate keratoconus: corneas with slit-lamp findings compatible with keratoconus, corneal thinning at the apex, Vogt's striae, a clearly visible Fleischer's ring, and corneal tomography compatible with keratoconus • Keratoconus suspect: unilateral keratoconus, 1 eye normal tomography and no biomicroscopic signs. An ophthalmologist and an optometrist performed the classification. Cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	
Notes	This work was supported in part by a research grant from the Flemish government agency for Innovation by Science and Technology (grant no. TBM-T000416N).

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	

Issarti 2019 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Issarti 2020

Study characteristics	
Patient Sampling	Retrospective, multicentre, case-control study. Scheimpflug Pentacam measurements (Oculus GmbH, Wetzlar, Germany) of 812 eyes were retrospectively collected from 2 centres in Belgium.
Patient characteristics and setting	• Normal eyes: eyes without slit-lamp findings suggestive of corneal ectasia and normal tomography • Keratoconic eyes were divided into the following 4 groups. ○ Suspected keratoconus

Issarti 2020 *(Continued)*

	○ Early keratoconus ○ Mild keratoconus ○ Moderate to advanced keratoconus
Index tests	A score-based machine learning system based on feedforward neural network (Logistic Index for Keratoconus, Logik), capable of classifying keratoconus according to its severity, to objectively discriminate suspect keratoconus from healthy eyes, and to provide a consistent, time-continuous scoring system for keratoconus progression
Target condition and reference standard(s)	• Normal: eyes without slit-lamp findings suggestive of corneal ectasia and normal tomography • Suspect keratoconus (forme fruste keratoconus): the contralateral, asymptomatic eye of a person with clinical keratoconus in 1 eye without clinical or tomographic signs of ectasia • Early keratoconus: eyes with a subtle sign of keratoconus, such as localized steeping in anterior or posterior surface, without significant changes of the cornea in the slit lamp • Mild keratoconus: tomographic changes consistent with keratoconus (anterior or posterior corneal steepening, corneal thinning, stromal thinning), Fleischer's rings at the cone base, partial or circular Fleischer's rings, but no visible Vogt's striae • Moderate to advanced keratoconus: clear cornea, corneal thinning at the apex, visible Vogt's striae, clearly visible circular Fleischer's ring, and corneal tomography findings An ophthalmologist and an optometrist performed the classification. Cases were classified before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This work was supported in part by a research grant from the Flemish government agency for Innovation by Science and Technology (grant no. TBM-T000416N).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Issarti 2020 *(Continued)*

DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Kalin 1996

Study characteristics	
Patient Sampling	Prospective, consecutive, cross-sectional study, including 106 eyes of 53 consecutive refractive surgery candidates in a single centre (USA)

Kalin 1996 *(Continued)*

Patient characteristics and setting	53 consecutive refractive surgery candidates for myopia correction with no history of ophthalmic diseases or ocular surgery were enroled during a 2-year interval.
Index tests	Expert system classification: an algorithm that incorporates 8 indices. Discriminant analysis was used to produce keratoconus prediction index (KPI) and in a binary decision tree.
Target condition and reference standard(s)	Keratoconus was diagnosed if clinical signs were present and the topography performed with TMS-1 (Computer anatomy 1, New York) was abnormal (irregular astigmatism, loss of radial symmetry, or absence of the normal progressive flattening from the centre to the periphery). The classification was performed by an experienced ophthalmologist, before the index test.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported in part by US Public Health Service grants EY10056 and EY0311 from the National Eye Institute, National Institutes of Health, Bethesda Maryland.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Kalin 1996 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Kamiya 2019

Study characteristics	
Patient Sampling	Single-centre (Japan), retrospective case-control study, which included a total of 304 keratoconic eyes and 239 healthy eyes (refractive surgery candidates and contact lens fitting candidates)
Patient characteristics and setting	The data of people with keratoconus who underwent corneal tomography obtained by a swept-source anterior segment OCT (CASIA SS-1000, Tomey, Aichi, Japan) between March 2013 and April 2018 at Miyata Eye Hospital were retrospectively reviewed. 304 keratoconic eyes with good quality scans of corneal tomography were enroled and divided according to the Amsler-Krumeich classification, as follows. • Grade 1 (108 eyes) • Grade 2 (75 eyes) • Grade 3 (42 eyes) • Grade 4 (79 eyes) The control group comprised 239 eyes in subjects with normal corneal and ocular findings applying for a contact lens fitting or a refractive surgery consultation.
Index tests	Deep learning (convolutional neural network) of the arithmetical mean output data of 6 colour-coded maps of an anterior segment OCT.
Target condition and reference standard(s)	Diagnosis of keratoconus was performed based on evident findings characteristic of keratoconus (e.g. corneal tomography with asymmetric bowtie pattern with or without skewed axes), and ≥ 1 keratoconus sign (e.g. stromal thinning, conical protrusion of the cornea at the apex, Fleischer's ring, Vogt's striae, or anterior stromal scar) on slit-lamp examination. It is unclear how

Kamiya 2019 *(Continued)*

	many corneal specialists classified the cases. However, classification was performed before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	The study authors declared no specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Kamiya 2019 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Kamiya 2021

Study characteristics	
Patient Sampling	Retrospective, single-centre (Japan), case-control study involving 349 keratoconus eyes and 170 normal eyes (refractive surgery candidates, contact lens fitting candidates).
Patient characteristics and setting	A total of 349 eyes with good-quality images of corneal topography measured with a Placido disc corneal topographer (TMS-4 TM, Tomey, Aichi, Japan) were included. The disease was graded according to the Amsler-Krumeich classification, as follows. • Grade 1 (54 eyes) • Grade 2 (52 eyes) • Grade 3 (23 eyes) • Grade 4 (50 eyes) Control group: 170 eyes in people with normal ocular findings applying for a contact lens fitting or for a refractive surgery consultation, who had a refractive error of < 6 D as well as astigmatism of < 3 D.
Index tests	Deep learning (convolutional neural network) of a single colour-coded topography map.
Target condition and reference standard(s)	Multiple corneal specialists diagnosed keratoconus with distinctive features (e.g. corneal colour-coded map with asymmetric bowtie pattern with or without skewed axes), and ≥ 1 keratoconus sign (e.g. stromal thinning, conical bulging, Fleischer's ring, Vogt's striae, or apical scar). It is unclear how many corneal specialists classified the cases; however, classification was performed before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This work was in part supported by Grants-in-Aid for Scientific Research (Grant Number 21K09706).

Kamiya 2021 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Kojima 2020

Study characteristics	
Patient Sampling	Multicentre (Japan), retrospective, case-control study which included 329 eyes (healthy controls and people with keratoconus).
Patient characteristics and setting	• People with keratoconus were aged < 50 years, visited 1 of the facilities between January 2015 and December 2018, were diagnosed with keratoconus, Amsler-Krumeich classification stage 1 (average K value < 48 D and corneal astigmatism < 5 D), had no ocular diseases other than keratoconus and refractive error, and had no history of ocular surgery. • Healthy subjects were people aged < 50 years who had undergone an ophthalmic screening.
Index tests	Multivariate logistic regression analysis to create an equation that predicts early keratoconus (keratometer keratoconus index) using auto-keratometer parameters.
Target condition and reference standard(s)	Keratoconus diagnosis was based on corneal topography or tomography results (corneal steepening and asymmetric astigmatism, protrusion of the posterior cornea, and thinning of the cornea at the area of protrusion) and slit-lamp findings. 2 corneal specialists classified the cases before inclusion.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Kojima 2020 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Kojima 2021

Study characteristics	
Patient Sampling	Single centre, retrospective, case-control study that included 179 eyes of 99 consecutive people with keratoconus and suspected keratoconus (68 males and 31 females, mean age 33.48 (SD 15.41) years), who visited the Nagoya Eye Clinic from January 2019 to December 2020 and were tested with an auto-keratometer (ARK-1s, Gamagori, Japan, NIDEK). During the same period, 468 eyes of 235 consecutive people (125 men and 110 women; mean age 37.55 (SD 22.70) years) examined for refractive correction were included as normal controls. The control group included people who had no abnormalities on slit-lamp biomicroscopy examination and corneal topography.
Patient characteristics and setting	People already diagnosed with keratoconus or suspected keratoconus were included. The exact reason why healthy controls visited the centre is unclear.
Index tests	Regression algorithm, modified keratoconus keratometer index (KKI) using 3 variables: steep K-value, flat K-value, and astigmatism. logit = 1.284 × steep K (dioptre) − 0.618 × flat K (dioptre) − 3.163 × (0: non-with-the-rule

Kojima 2021 *(Continued)*

	astigmatism; 1: with-the-rule astigmatism) − 28.662, KKI = exp(logit)/(1 + exp[logit]). The cut-off value of 0.461 had been determined previously.
Target condition and reference standard(s)	2 cornea specialists diagnosed keratoconus through slit-lamp microscopy and corneal topography. Keratoconus signs found in both slit-lamp microscopy and corneal topography were classified as keratoconus, while signs found only in corneal topography were classified as suspected keratoconus. Forme fruste keratoconus was defined as an eye with normal corneal topography in the contralateral eye of the keratoconus. The severity of keratoconus was based on the Amsler–Krumeich classification.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This research received no external funding.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			

Kojima 2021 *(Continued)*

Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Kovacs 2016

Study characteristics	
Patient Sampling	Retrospective, single-centre (Hungary), case-control study, which involved 135 eyes of people with bilateral keratoconus (keratoconus group), normal fellow eyes of people with unilateral keratoconus (fellow-eye group), and eyes of refractive surgery candidates (control group).
Patient characteristics and setting	• Keratoconus group: 60 eyes of 30 people with bilateral keratoconus • Fellow-eye group: 15 normal fellow eyes of people with unilateral keratoconus • Control group: 60 eyes of 30 refractive surgery candidates
Index tests	Multilayer perception classifier (neural network) trained on bilateral data of index of height decentration.
Target condition and reference standard(s)	Keratoconus was diagnosed according to classic corneal biomicroscopic and topographic findings using the criteria of Rabinowitz: the existence of central protrusion of the cornea with Fleischer's ring, Vogt's striae, or both, by slit-lamp examination in addition to the following topographic findings: a central keratometry (K) value > 47.2 D or an I-S value > 1.4 D, or KISA% >

Kovacs 2016 *(Continued)*

	100%. Both eyes in the keratoconus group and the affected eye in the unilateral keratoconus group had abnormal keratoconus indices measured by a Scheimpflug camera (Pentacam HR). It is unclear who classified the cases. Classification was performed before the index test.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Unclear whether different AI tests were developed and interpreted blind or independently and without knowledge of the results of each other. However, missing data and their causes were similar for each AI test.
Notes	Supported by OTKA NN106649 from the Hungarian Scientific Research Fund. The funder had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	No		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Kovacs 2016 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Yes		
		Unclear risk	

Kuo 2020

Study characteristics	
Patient Sampling	Retrospective, single-centre (Taiwan), case-control study. The investigators retrospectively collected corneal topographies (TMS-4; Tomey Corporation, Nagoya, Japan) of the study group with clinically manifested keratoconus, the subclinical keratoconus group, and the control group with regular astigmatism (354 images of 206 participants).
Patient characteristics and setting	• 170 keratoconus pictures • 28 subclinical keratoconus pictures (criteria based on topographic pattern and no slit-lamp keratoconus findings) • 156 normal topographic pictures (from candidates for refractive surgery without any previous manifestations and with regular astigmatism)
Index tests	3 convolutional neural network models
Target condition and reference standard(s)	The diagnosis of keratoconus was based on clinical signs (the existence of central protrusion of the cornea, Fleischer's ring, Vogt's striae, and focal corneal thinning on slit-lamp examination) and topographic criteria (central K value > 47 D, I-S value > 1.4 D, KISA% >100%, and asymmetric bowtie presentation). 4 corneas specialists classified the cases before the index test.

Kuo 2020 *(Continued)*

Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Unclear whether different AI tests were developed and interpreted blind or independently and without knowledge of the results of each other. However, missing data and their causes were similar for each AI test.
Notes	Supported by Grants 107L891002 and 108L891002 from National Taiwan University.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Kuo 2020 *(Continued)*

DOMAIN 4: Flow and timing		
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear	
Are the proportions and reasons for missing data similar for all index tests?	Yes	
		Unclear risk

Lavric 2021

Study characteristics	
Patient Sampling	Retrospective case-control study. Pentacam data (Oculus GmbH, Wetzlar, Germany) obtained from people screened for keratoconus disease in Brazil. Elevation, topography, and pachymetry parameters were obtained from 5881 eyes of 2800 participants.
Patient characteristics and setting	Participant characteristics and setting are not clearly described in the study. Data seems to originate from a larger data set which included people with keratoconus and healthy controls.
Index tests	Support vector machine that uses elevation, topography or pachymetry parameters obtained from the raw data of the Pentacam to detect keratoconus. The workflow for the development of the algorithm was as follows: splitting the initial data set in elevation, topography and pachymetry data sets; data cleaning and elimination; feature selection; machine learning validation; and performance evaluation. It is unclear whether different data were used for testing and validating the model.
Target condition and reference standard(s)	The target condition was keratoconus; however, the article provided no definition. Tomography images of the Pentacam were used in this study. It is unclear who interpreted the images and made the diagnosis; however, the diagnosis was made before the algorithms analysed the images.
Flow and timing	The article did not describe the reference standard, nor did it describe whether all participants received the same reference standard. All data were included in a 2 × 2 table.

Lavric 2021 *(Continued)*

Comparative	In total, 6 algorithms were developed, tested, and compared: decision tree, discriminant naïve Bayes, support vector machine, k-nearest neighbour, and ensemble.
Notes	This work was supported in part by a grant from the Romanian Ministry of Research and Innovation, CCCDI-UEFISCDI, within PNCDI III, under Project PN-III-P2-2.1-PTE-2019-0642, and in part by the Romania National Council for Higher Education Funding, CNFIS, under Project CNFIS-FDI-2021-0357.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Lavric 2021 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Yes		
		Unclear risk	

Lopes 2018

Study characteristics	
Patient Sampling	Retrospective, multicentre, case-control study. A total of 3693 people were enroled from 5 different clinics. Participants were divided into 2 data sets: 1 training set and 1 validation set. The training set included the preoperative data of the following 3 groups. • Stable LASIK cases • Ectasia susceptibility • Clinical keratoconus The algorithm was independently tested in a different set of stable LASIK cases and people with very asymmetric ectasia; these people had clinically diagnosed ectasia in 1 eye and normal topography in the fellow eye.
Patient characteristics and setting	The participants were grouped as follows. • LASIK cases (2980 participants with minimum follow-up of 7 years) • Ectasia susceptibility (71 eyes of 45 participants who developed post-LASIK ectasia) • Clinical keratoconus (182 participants) • Participants with very asymmetric ectasia (188 participants); these people had clinically diagnosed ectasia in 1 eye and normal topography in the fellow eye.
Index tests	Random forest: multiple decision trees were built and merged to improve accuracy of the prediction. 2 steps of validation were used to assess the generalization and clinical validity of the models and their ability to correctly classify new data. The first was a holdout validation: the training set was randomly split into 2 data sets: the first comprised 70% of the total data set and was used to actually train the models; the other 30% was used to test the model accuracy. The second validation step was an independent test with cases that were not part of the training set. The algorithm

Lopes 2018 *(Continued)*

	analysed the raw tomographic data to identify the different patterns and detect keratoconus.
Target condition and reference standard(s)	All eyes were examined by rotating Scheimpflug corneal and anterior segment tomography (Pentacam HR; Oculus GmbH, Wetzlar, Germany). Image quality was checked, so that only cases with acceptable-quality images were included in the study. 1 experienced fellowship-trained corneal specialist reviewed all the cases so that they were correctly classified in the keratoconus and very asymmetric ectasia groups. All cases were diagnosed before the algorithm analysed the images.
Flow and timing	All eyes received the reference standard and were included in the 2 × 2 table of the index test.
Comparative	5 models were developed and compared: regularized discriminant analysis (RDA), support vector machine (SVM), naïve Bayes (NB), neural networks (NN), and random forest (RF). It is unclear if all tests were developed and interpreted without knowledge of each other and if all data were used for each test.
Notes	No funding or grant support.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Lopes 2018 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Unclear		
		Unclear risk	

Lucena 2021

Study characteristics	
Patient Sampling	Retrospective, case-control study. A diagnostic pattern bank generated by a specialist physician. The database consisted of 1172 examples of corneal topography, divided into 275 spherical patterns, 302 regular symmetrical astigmatism patterns, 295 regular asymmetrical astigmatism patterns, and 300 irregular astigmatism patterns (keratoconus).
Patient characteristics and setting	This study is a registry-based study; the registry contained healthy controls and people with keratoconus.
Index tests	Convolutional neural network. The algorithm is a semi-automatic, manual interference model. It uses a hierarchical system that tries to represent the structure in relation to the recognition of an image, where pixels form

Lucena 2021 *(Continued)*

	edges, edges form patterns, patterns form objects, which in turn describe the scenes. The algorithm analyses the topographic images and decides whether it is regular or irregular astigmatism (keratoconus). The algorithm was developed with a training phase and a validation phase.
Target condition and reference standard(s)	A specialist physician developed the diagnostic pattern bank of images made by topographers. He divided the included topographies into the following 4 groups. • Spherical patterns • Regular symmetrical astigmatism patterns • Regular asymmetrical astigmatism patterns • Irregular astigmatism patterns (keratoconus) The cases were divided before the algorithm analysed the images.
Flow and timing	All topographies included in the diagnostic pattern bank were judged by the specialist physician and all images were included in the analysis.
Comparative	Not applicable
Notes	This study was supported by Government of the State of Ceará (Foundation for the Support to Scientific and Technological Development of Ceará), and Ophthalmology School of Ceará.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Lucena 2021 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			High
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Maeda 1994

Study characteristics	
Patient Sampling	Single-centre, case-control study. Videokeratographs were drawn from the Louisiana State University Eye Center patient population and were divided randomly by category into 2 sets. Each set comprised 8 categories: normal, keratoconus, keratoplasty, epikeratophakia, excimer laser photorefractive keratectomy, radial keratotomy, contact lens-induced warpage, and other.
Patient characteristics and setting	The study included the following categories of participants. • Healthy controls • People with keratoconus • People who underwent refractive surgery or who had undergone a keratoplasty The study did not include people seeking refractive surgery.
Index tests	Combined discriminant analysis and classification tree analysing images from the TMS-1. The keratoconus detection programme was developed using a training set of 100 corneas and evaluated with a validation set of an additional 100 corneas. Maps were first classified as either keratoconus, borderline, or non-keratoconus. The borderline maps were then divided into keratoconus or non-keratoconus by certain indices. Next, all keratoconus patterns were classified into either peripheral or central keratoconus using a threshold combination

Maeda 1994 *(Continued)*

	of these indices. Final output of the system was the display of the certainty of keratoconus.
Target condition and reference standard(s)	The included cases were diagnosed by 3 corneal topography researchers based on clinical records and topography. All images were made by a TMS-1. The diagnosis was made before the algorithm analysed the images.
Flow and timing	All participants received the index and reference test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported in part by National Institutes of Health grants EY03311 and EYO2377 and by Computed Anatomy, Inc. and Menicon, Co., Ltd.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Maeda 1994 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and timing		
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		

Maeda 1995a

Study characteristics	
Patient Sampling	Single-centre, case-control study. Corneal topographic maps of the TMS-1 were drawn from the Louisiana State University Eye Center patient population.
Patient characteristics and setting	Maps from 176 eyes of 125 people were selected and grouped as follows. • 44 with topographic features typical of keratoconus • 132 with topographic features typical of a variety of non-keratoconus conditions (normal, with-the-rule-astigmatism, contact lens-induced warpage, excimer laser photorefractive keratectomy, penetrating keratoplasty, and pellucid marginal degeneration). Maps of eyes with keratoconus were selected from the charts of people previously diagnosed as having keratoconus in our clinic.
Index tests	Combined discriminant analysis and classification tree, based on topographic images. The algorithm determined whether a keratoconus-like pattern was seen in a particular map in the binary classification tree and, if so, reported a value between 5% and 95% in proportion to the linear discriminant function to quantify the severity of the keratoconus pattern.
Target condition and reference standard(s)	Topography images were made with the TMS-1. It is unclear who diagnosed the cases; however, the diagnosis was made before the algorithm analysed the images.
Flow and timing	It is unclear whether all eyes were diagnosed with the same reference standard. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This study was supported in part by US Public Health Service grants EY03311 and 02377 from the National Eye Institute, National Institutes of Health, Bethesda, Md; an unrestricted departmental grant from Research to Prevent Blindness Ine, New York, NY; and funds from Computed Anatomy Ine, New York, NY, and Menicon Co, Ltd, Nagoya, Japan. The Conecare data analysis software used in this study was provided courtesy of Yaron S. Rabinowitz, MD.

Maeda 1995a *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		

Maeda 1995a *(Continued)*

Could the patient flow have introduced bias?	Unclear risk
DOMAIN 5: Comparative	

Maeda 1995b

Study characteristics	
Patient Sampling	Single-centre, case-control study. Corneal topographic maps of the TMS-1 were drawn from the Louisiana State University Eye Center patient population.
Patient characteristics and setting	Participants were divided into the following 7 categories. • Normal • With-the-rule astigmatism • Mild keratoconus • Moderate keratoconus • Advanced keratoconus • Post-photorefractive keratectomy • Post-keratoplasty The study excluded maps in which focusing or alignment were not properly achieved or that contained atypical topographic appearances. For this project, 183 eyes were selected. The maps were randomly divided into a training set (108 maps) and a test set (75 maps).
Index tests	Neural network constructed in 3 layers. The input layer consisted of 11 neurons equal to 11 topographic indices. There was 1 hidden layer of 18 neurons. The output layer consisted of 7 neurons, 1 for each topographic category.
Target condition and reference standard(s)	Topography images were made with the TMS-1. All eyes were diagnosed by the 3 study authors and classified according to the topographic image and clinical records. The diagnosis was made before the images were analysed by the algorithm.
Flow and timing	All eyes received the same reference standard, and all data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported in part by US Public Health Service grant EY03311 and EY01377 from the National Eye Institute, National Institutes of Health (Bethesda, MD); Computed Anatomy (New York, NY); Menicon Co. Ltd. (Nagoya, Japan).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		

Maeda 1995b *(Continued)*

Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Mahmoud 2013

Study characteristics	
Patient Sampling	Participants were selected from 4 different hospitals: Department of Ophthalmology of the Ohio State University, Columbus, Ohio; Clinica de Oftalmologia de Cali, Pontificia Universidad Javeriana, Cali, Colom-

Mahmoud 2013 *(Continued)*

	bia; Cullen Eye Institute, Department of Ophthalmology, Baylor College of Medicine, Houston, Texas; and Department of Ophthalmology, University Hospital of Bern, Inpelspital, Bern, Switzerland.
Patient characteristics and setting	• People with keratoconus were clinically identified by characteristic refractive and slit-lamp signs (e.g. unstable refraction; oblique astigmatism; irregular retinoscopic and keratometry mires; and biomicroscopic signs such as Vogt's striae or Fleischer's ring) and no history of corneal surgery. • Normal subjects had no documented history of corneal disease or corneal surgery.
Index tests	Logistic regression. The algorithm included both anterior and posterior curvature maps; results were divided into 3 categories: normal (0–0.25), suspect (0.25–0.8), and keratoconus (0.8–1.0).
Target condition and reference standard(s)	Tomography images from the Galilei Dual Scheimpflug-Placido tomographer were used. It was unclear how the reference standard made the diagnosis; however, all cases were diagnosed before the index test analysed them.
Flow and timing	It was unclear whether all participants received the same reference standard, but all cases were diagnosed before inclusion.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Mahmoud 2013 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Mahmoud 2021

Study characteristics	
Patient Sampling	Case-control study. 250 cases were extracted and a physician made the medical diagnoses. Medical evaluations were found in the data set Dataverse.
Patient characteristics and setting	The study included people with keratoconus and healthy controls. Further participant characteristics are described in Dataverse.
Index tests	Convolutional neural network using a 3D reconstruction of corneal images as input. The output was healthy or keratoconus. The algorithm also defined severity of keratoconus.
Target condition and reference standard(s)	The OCT SS-1000 (CASIA) was used to make the corneal images. The diagnosis was made by a physician. All cases were diagnosed before the algorithm analysed the images.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fast-track Research Funding Program.

Mahmoud 2021 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Mohammadpour 2022

Study characteristics	
Patient Sampling	Prospective, diagnostic test accuracy study including 217 eyes of 212 people aged 17–49 years who were referred to the Keratoconus Clinic or were refractive surgery candidates at the Refractive Surgery Unit. Exclusion criteria: a history of ocular surgery, corneal cross-linking, or ring implantation; corneal hydrops or scarring; signs and symptoms of dry eye or ocular diseases other than keratoconus; connective tissue diseases; systemic diseases affecting the eyes; corneal haze; pregnancy; and contact lens use in the previous month.
Patient characteristics and setting	The study included people already diagnosed with keratoconus or suspected keratoconus.
Index tests	The algorithm combines Placido and Scheimpflug technologies to provide complete information on the anterior and posterior corneal surfaces. Sirius (Costruzione Strumenti Oftalmici, Florence, Italy) takes 25 Scheimpflug images and 1 Placido image in < 1 second. Height, slope, and curvature data are then calculated with an arc-step method. This system provides comprehensive information on the entire cornea and classifies keratoconus via the Phoenix software through a neural network process. The study performed a comparison of existing algorithms, which are already validated.
Target condition and reference standard(s)	Participants were grouped based on the clinical diagnosis of 2 independent experienced corneal specialists (M. Mohammadpour, K. Amanzadeh), through slit-lamp biomicroscopy, retinoscopy, corrected distance visual acuity (CDVA) measurement with a Snellen chart, and evaluation of the Pentacam Refractive 4 Maps. The specialists were blinded to classification reports. Diagnostic discrepancies were resolved by a third expert examiner (A. Moghaddasi) for a definitive diagnosis.
Flow and timing	All cases were included in the reference standard and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	The study authors received no financial support for the research, authorship, or publication of the article.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Mohammadpour 2022 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Pavlatos 2020

Study characteristics	
Patient Sampling	Case-control study that recruited participants from the Casey Eye Institute at Oregon Health & Science University (Portland, Oregon) and the Affiliated Eye Hospital of Wenzhou Medical College (Wenzhou, China)
Patient characteristics and setting	The study grouped eyes as follows. • Normal eyes: normal slit-lamp examination, corrected distance visual acuity (CDVA) ≥ 20/20, normal topography appearance, and KISA% index < 100% • Keratoconus eyes: manifest, subclinical, and forme fruste
Index tests	Custom-made MATLAB algorithms generated pattern deviation maps of pachymetry and epithelial thickness, captured using Fourier-domain OCT images of the cornea. The co-localized thinning of the 2 maps was quantified using a novel coincident thinning index, which was calculated from Gaussian fits of the regions of maximum relative thinning.
Target condition and reference standard(s)	OCT scans were obtained using commercial Fourier-domain OCT systems (RTVue or Avanti; Optovue, Inc) with a corneal adaptor module for imaging of the anterior eye. It is unclear how the diagnosis was made. It seems the cases were divided into normal and keratoconus groups before they were analysed with the algorithm.
Flow and timing	It is unclear if all cases received the same reference standard. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported by the National Institutes of Health, Bethesda, MD (Grant Nos. R01EY028755, R01EY029023, T32EY023211, and P30EY010572); a research grant and equipment support from Optovue, Inc, Fremont, California; and unrestricted grants to Casey Eye Institute and Bascom Palmer Eye Institute from Research to Prevent Blindness, Inc, New York, New York.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Pavlatos 2020 *(Continued)*

DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Rabinowitz 1999

Study characteristics	
Patient Sampling	Case-control study. The cases were selected from a database of more than 400 people with keratoconus recruited for longitudinal videokeratography and genetic studies of keratoconus at the Cedars-Sinai Medical Center (CSMC), Los Angeles, California.
Patient characteristics and setting	• Keratoconus: inclusion criteria included slit-lamp or retroillumination signs of keratoconus in 1 or both eyes, absence of corneal scarring, no history of contact lens wear of any type, no signs or history of other corneal disease, and no previous ocular surgery. • 195 normal control participants were also studied.

Rabinowitz 1999 *(Continued)*

Index tests	Regression algorithm. The KISA% index quantifies the topographic features seen in people with clinical keratoconus.
Target condition and reference standard(s)	Topography images were taken of both eyes of each study participant with the TMS-1. It was unclear how the diagnosis was made. The cases were labelled before the analysis with the algorithm.
Flow and timing	It is unclear if all cases received the same reference standard. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported by NIH Grant EY09052 and The Eye Birth Defects Research Foundation, Inc.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Rabinowitz 1999 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and timing		
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		

Ruiz 2016

Study characteristics	
Patient Sampling	Retrospective case-control study; unclear how the data were collected.
Patient characteristics and setting	Participants were grouped as follows. • Normal: eyes that had no systemic or ocular diseases, prior ocular surgery, or regular astigmatism ≤ 1.5 D • Astigmatism: symmetric and straight bowtie pattern on the anterior keratometry with AST < 1.5 D • After refractive surgery: people who had undergone laser-assisted subepithelial keratectomy for myopia, hyperopia, or astigmatism correction. The spherical corrections ranged from –1 D to –6 D, and cylindrical corrections ranged from –1 D to–3 D. • Forme fruste: non-symptomatic Pentacam tomography and elevation, but fellow eye has manifest keratoconus • Keratoconus: diagnosis of mild or moderate keratoconus taking into account the presence of a zone of increased power with abnormal thinning and elevation in the anterior and posterior surfaces combined with patient and family history, changes in refraction, and slit-lamp examination Exclusion criteria: any systemic disease, a history of ocular surgery (except for laser refractive surgery in the refractive surgery group), cross-linking, ametropia > 610 D (for the normal group), and very advanced keratoconus with corneal scarring. Note that keratoconus suspect cases (i.e. eyes that the investigators could not reliably categorize as either normal or keratoconus based on Pentacam or slit-lamp examination) were removed from the analysis, as they could not be properly categorized for training purposes.
Index tests	Support vector machine First, the data were preprocessed to ensure there were no impossible values or missing data. The second step was a correlation-based hierarchical clustering. The output of this analysis is dendrogram, in which variables are organized along branches according to their degree of correlation with each other. Next, the dendrogram was revised to select only 1 variable from each branch, which was typically the variable with the highest clinical relevance. This procedure led to a data set of 22 variables. All classifications were

Ruiz 2016 *(Continued)*

	performed using a linear kernel support vector machine; this method provides a linear decision boundary for binary classification problems.
Target condition and reference standard(s)	An experienced keratoconus specialist and an experienced optometrist classified the eyes into 5 groups using criteria based on patient history and corneal tomography maps from the Pentacam.
Flow and timing	All participants received the same reference standard and were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Study supported by an Agency for Innovation by Science and Technology grant.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		

Ruiz 2016 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Ruiz 2017

Study characteristics	
Patient Sampling	Retrospective, case-control study including 131 eyes from 102 people during routine clinical practice at the Rothschild Foundation in the period between October 2015 and January 2016.
Patient characteristics and setting	Participants were grouped as follows. • Normal • Keratoconus • Forme fruste and suspect keratoconus • Postrefractive surgery for myopia All participants had a complete ophthalmic evaluation that consisted of manifest refraction, slit-lamp examination, and corneal tomography evaluation by both Pentacam HR and Orbscan IIz. Exclusion criteria: a history of ocular surgery (apart from laser correction in the refractive surgery group), corneal cross-linking, and very advanced keratoconus with corneal scarring. Rigid contact lens wearers were asked to remove their lenses at least 1 week before testing.
Index tests	The support vector machine uses 25 topography and tomography parameters of the anterior and posterior corneal surfaces from the Pentacam. The algorithm automatically classifies corneal patterns and shows the probability that the cornea has keratoconus, forme fruste or suspect keratoconus, photorefractive surgery, normal, or regular astigmatism.
Target condition and reference standard(s)	Participants were clinically diagnosed at the Rothschild Foundation based on slit-lamp examination and corneal topography and classified into 4 groups.

Ruiz 2017 *(Continued)*

	Topography and tomography were measured with the Pentacam HR and the Orbscan IIz. Participants were diagnosed before the corneal images were analysed with the algorithm.
Flow and timing	It was unclear whether all participants received the same reference standard. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	Supported by a research grant of the Flemish government agency for Innovation by Science and Technology (grant nr. IWT/110684).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Ruiz 2017 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Saad 2014

Study characteristics	
Patient Sampling	Case-control study; part of a prospective study evaluating clinical, topographic, tomographic, and biomechanical characteristics of people with keratoconus and keratoconus suspect at the Rothschild Foundation, Paris, France.
Patient characteristics and setting	Participants were grouped as follows. • Normal group • Preoperative LASIK with 4-year follow-up • Clinically evident bilateral keratoconus Exclusion criteria: corneal scarring in the anterior or posterior segment.
Index tests	Discriminant analysis composed of 3 variables: the difference between steep and flat keratometry, the 3-mm irregularity, and the anterior elevation of the thinnest point. The algorithm was developed to discriminate keratoconus eyes from normal eyes.
Target condition and reference standard(s)	All keratoconus eyes were diagnosed by 1 corneal specialist based on clinical and topographic signs.
Flow and timing	All participants received the same reference and index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Saad 2014 *(Continued)*

DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Saad 2016

Study characteristics	
Patient Sampling	Case-control study; part of a prospective study evaluating clinical, topographic, tomographic, and biomechanical characteristics of people with keratoconus and keratoconus suspect at the Rothschild Foundation, Paris, France. The study included healthy people and people with keratoconus suspect.
Patient characteristics and setting	119 eyes of 176 people from the Department of Ophthalmology of the Rothschild Foundation were included and separated into 2 groups: normal and keratoconus suspect. The 2 groups were divided based on the results of the Nidek Corneal Navigator (NCN) automated corneal classification software in the OPD-Scan (Nidek Co. Ltd., Gamagori, Japan). • The keratoconus suspect group was composed of 62 topographically normal eyes of people with keratoconus in the fellow eye. • The normal group was composed of 114 eyes of 57 people with post-myopic LASIK who had 4-year follow-up, with no postoperative complications such as ectasia. Only the preoperative topographies were considered in the normal group.
Index tests	Discriminant analysis, combining Placido (topography) and corneal wavefront data to detect early forms of keratoconus and to classify corneas as healthy, keratoconus suspect, or keratoconus.
Target condition and reference standard(s)	The article did not mention who made the diagnosis of keratoconus or keratoconus suspect; however, all cases were separated in the 2 groups before the discriminant analysis.
Flow and timing	It was unclear whether all participants received the same reference standard. All cases did receive the same index test, and all data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High

Saad 2016 *(Continued)*

DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Saika 2013

Study characteristics	
Patient Sampling	Single-centre case-control study including 212 eyes: 51 eyes of 37 people (24 men, 13 women) with keratoconus; 46 eyes of 35 people (20 men, 15 women) with keratoconus suspect; 50 eyes of 28 people (13 men, 15 women) who had undergone LASIK for myopia; and 65 healthy control eyes of 65 people (38 men, 27 women). The corneal surface of each was measured with a Placido-based corneal topographer.
Patient characteristics and setting	Included eyes were divided into the following groups. • Keratoconus: the presence of central thinning of the cornea, a Fleischer's ring, or Vogt's striae observed by slit-lamp examination

Saika 2013 *(Continued)*

	• Keratoconus suspect: a typical topographic pattern of keratoconus obtained by TMS-4 Advance Corneal Topographer (Tomey Corporation, Nagoya, Japan) without any findings on slit-lamp examination • LASIK for myopia • Healthy control
Index tests	Linear discriminant analysis is a statistical method of finding a linear combination of several explanatory variables that characterizes or separates categories of objects. The variables used were: Zernike expansion coefficients of the 2nd to 4th terms of the 4th-order approximation for a 4-mm-diameter pupil; the 2nd to 6th terms of the 6th-order approximation for a 6-mm-diameter pupil; and the Simulated Keratometry. The categories were: keratoconus, keratoconus suspect, LASIK, and healthy control. The model was trained with 2 different sets of participants.
Target condition and reference standard(s)	It was unclear how the diagnosis of keratoconus was made by the reference standard. All cases were labelled before analysis with the AI algorithm.
Flow and timing	It is unclear whether all participants received the same reference standard. All participants were included in the analysis.
Comparative	Not applicable
Notes	Supported in part by the Japan Ministry of Education, Science, Sports, and Culture, Tokyo, Japan (No. 24592669).

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	

Saika 2013 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Shetty 2015

Study characteristics	
Patient Sampling	Retrospective single-centre case-control study in people with clinically diagnosed keratoconus before they received corneal crosslinking (85 eyes) and normal controls before they received an ablation (43 eyes).
Patient characteristics and setting	• The diagnosis of keratoconus was based on evidence of stromal thinning, focal protrusion, or increase in corneal curvature, Fleischer's ring, Vogt's striae, scissoring of the red reflex, an abnormal retinoscopy, and curvature asymmetry leading to abnormal corneal astigmatism. • For normal eyes, manifest spherical error and astigmatism were limited to ±2 D. Exclusion criteria were ocular hypertension, corneal inflammation, prior eye surgery, and current topical medication use.
Index tests	Logistic regression using Zernike coefficients, curvature, corneal volume, and corneal anterior wavefront analyses to calculate if keratoconus is present.
Target condition and reference standard(s)	The diagnosis of keratoconus was based on evidence of stromal thinning, curvature asymmetry leading to abnormal corneal astigmatism, or increase in corneal curvature measured by the Pentacam, and clinical signs (e.g. Fleischer ring's, Vogt's striae, scissoring of the red reflex,

Shetty 2015 *(Continued)*

	an abnormal retinoscopy, and focal protrusion). However, it was unclear who made the diagnosis.
Flow and timing	It was unclear whether all participants received the same reference standard. All participants were included in the analysis.
Comparative	Multiple logistic regression equations were compared; however, it was unclear if the development and interpretation was done without the knowledge of each other and if all cases were analysed by all the equations.
Notes	Dr Dupps is a recipient of National Eye Institute, USA R01 grant (# NIH R01 EY023381).

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	No		

Shetty 2015 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Unclear		
		Unclear risk	

Shi 2020

Study characteristics	
Patient Sampling	Prospective, single-centre, case-control study. People with keratoconus and subclinical keratoconus were recruited from Affiliated Eye Hospital of Wenzhou Medical University. Normal subjects were recruited from the hospital's working staff and students.

Patient characteristics and setting	• Normal eyes had to meet the following criteria.
	  ◦ No clinical signs or suspected subclinical keratoconus or keratoconus patterns from corneal topography images
	  ◦ Central average keratometry < 45.0 D
	  ◦ I-S values < 1.4 D of the vertical power gradient across the 6-mm region
	  ◦ Myopia < −6.0 and astigmatism < −2.0 D
	  ◦ No history of contact lens wear, ocular surgery, or trauma
	• Keratoconus eyes had to meet ≥ 1 of the following criteria.
	  ◦ Vogt's striae
	  ◦ Stromal thinning
	  ◦ Fleischer's ring > 2 mm arc
	  ◦ Central average keratometry > 47.0 D
	  ◦ Asymmetric topographical features with I-S values ≥ 2.0 D of the vertical power gradient across the 6-mm region
	  ◦ No history of contact lens use, ocular surgery, or extensive scarring
	• Subclinical keratoconus eyes were identified from the other eyes of people with unilateral keratoconus. They had to meet all the following criteria.

Shi 2020 *(Continued)*

- No clinical signs of keratoconus during slit-lamp biomicroscope examination, retinoscopy, and ophthalmoscopy
- Diagnosis of keratoconus in the contralateral eye
- Central average keratometry < 45.0 D
- Corneal topographical features with I-S values < 1.4 D of the vertical power gradient across the 6-mm region
- Myopia < –6.0 D with astigmatism < −2.0 D
- No history of contact lens wear or ocular surgery

Index tests	An automated classification system using a machine learning classifier to distinguish clinically unaffected eyes in people with keratoconus from a normal control population based on a combination of Scheimpflug camera images and ultra-high-resolution optical coherence tomography (UHR-OCT) imaging data. A neural network was used.
Target condition and reference standard(s)	2 experienced doctors (YY and JJ) performed a comprehensive ocular exam, including a review of family and medical history, corrected-distance visual acuity, slit-lamp biomicroscope examination, fundus examination, and corneal topography.
Flow and timing	All participants received the same reference standard and were included in the analysis.
Comparative	Not applicable
Notes	This study was supported by research grants from Key R&D Program Projects in Zhejiang Province (2019C03045), the National Major Equipment Program of China (2012YQ12008004), the National Key Research and Development Program of China (2016YFE0107000, the National Nature Science Foundation of China (Grant No. 81570880).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Shi 2020 *(Continued)*

If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Sideroudi 2017

Study characteristics	
Patient Sampling	A cross-sectional, observational study. Study participants were recruited from the cornea outpatient service in a consecutive if eligible basis; controls were refractive surgery candidates. 80 eyes formed the keratoconus group, 55 eyes formed the subclinical keratoconus group, and 50 normal eyes populated the control group.
Patient characteristics and setting	• The keratoconus group included people diagnosed with stage I keratoconus according to the Amsler–Krumeich classification system. Inclusion criteria were as follows.

Sideroudi 2017 *(Continued)*

	◦ Eccentric steepening ◦ Myopia or induced astigmatism < 5.00 D ◦ Mean central keratometry readings < 48.00 D • The subclinical keratoconus group included people diagnosed with subclinical keratoconus. Inclusion criteria were as follows. ◦ Diagnosis of keratoconus in the fellow eye according to the Amsler-Krumleich criteria ◦ KISA% index between 60% and 100% in the subclinical keratoconus eye ◦ Lack of any keratoconus-related findings/signs in the slit-lamp biomicroscopy • Refractive surgery candidates with no evidence of ectasia populated the control group. Exclusion criteria: previous incisional eye surgery, corneal scars and opacities, history of herpetic keratitis, severe eye dryness, current corneal infection, glaucoma, suspicion of glaucoma, intraocular pressure-lowering treatment, pregnancy or nursing, contact lens use, and underlying autoimmune disease.
Index tests	A self-developed algorithm based on logistic regression in Visual Basic for Microsoft Excel performed a Fourier series harmonic analysis for the posterior corneal sagittal curvature data and evaluated the derived parameters in the diagnosis of Subclinical Keratoconus and Keratoconus.
Target condition and reference standard(s)	The diagnosis of keratoconus was based on the Amsler-Krumeich classification and tomography images; however, it was unclear who made the diagnosis. The diagnosis was made before the analysis with the AI algorithm.
Flow and timing	It was unclear whether all included participants were diagnosed by the same person(s). All participants were included in the analyses.
Comparative	Different algorithms were developed; it was unclear whether the results were interpreted separately. All data were included in the different analyses.
Notes	No funding source mentioned.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			

Sideroudi 2017 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Yes		
		Unclear risk	

Smadja 2013

Study characteristics	
Patient Sampling	Retrospective, single centre, case-control study. Normal eyes were selected from suitable candidates undergoing a screening examination for refractive surgery and among the general population undergoing a routine ophthalmologic examination.
Patient characteristics and setting	• Normal eyes: discontinued daily-wear soft contact lens use at least 1 week before evaluation. Eyes were considered normal when no clinical signs of keratoconus and no suggestive topographic or tomographic patterns of suspect keratoconus were found, such as asymmetric bowtie with a skewed radial axes, focal or inferior steepening, central keratometry > 47.0 D, or corneas thinner than 500 mm. Exclusion criteria: previous ocular surgery, ocular pathology, familial history of keratoconus, and contact lens wearing in the past week. • Forme-fruste keratoconus: evident keratoconus in the fellow eye. These eyes had no clinical signs of keratoconus and a normal topographic aspect with no asymmetric bowtie and no focal or inferior steepening pattern. • Keratoconus: corneal topography with asymmetric bowtie pattern or localized steepening, irregular cornea determined by distortion of the retinoscopic or ophthalmoscopic red reflex, and ≥ 1 slit-lamp finding (stromal thinning, Fleischer's ring > 2 mm arc, Vogt's striae, or corneal scarring). Exclusion criteria: specific treatment for keratoconus (e.g. collagen cross-linking, intracorneal rings, keratoplasty) and marginal pellucid degeneration.
Index tests	A new screening program for the detection of forme fruste keratoconus using the GALILEI Dual Scheimpflug Analyzer (Ziemer Ophthalmic Systems AG, Port, Switzerland). The method is based on an automated decision tree classification that helps to discriminate between normal corneas, forme fruste keratoconus eyes, and keratoconus eyes.
Target condition and reference standard(s)	Participants were divided into 3 groups, but it was unclear who made the diagnosis. The diagnoses were made before the analysis with the classification tree.
Flow and timing	It was unclear if the diagnoses were made by the same person(s). All participants were included in the analyses.
Comparative	Different algorithms were developed. It was unclear whether the results were interpreted separately. All data were included in the different analyses.
Notes	No funding source mentioned.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		

Smadja 2013 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Unclear		
Are the proportions and reasons for missing data similar for all index tests?	Yes		

Smadja 2013 *(Continued)*

	Unclear risk

Smolek 1997

Study characteristics	
Patient Sampling	Retrospective, single-centre, case-control study. 300 TMS-1 examinations (Tomey USA, Cambridge, MA) were collected from medical records at the LSU Eye Center
Patient characteristics and setting	• Normal: < 1.5D cylinder, with-the-rule astigmatism >1.5 D cylinder, contact lens-induced warpage, pellucid marginal degeneration, penetrating keratoplasty, radial keratectomy, and photorefractive keratectomy • Keratoconus suspect: absence of any clinical signs, symptoms, or medical history that might point to the presence of keratoconus or any other condition in that eye, even though the videokeratographic appearance bore a striking resemblance to that of an eye with a minor conelike ectasia • Keratoconus Exclusion criteria were unclear.
Index tests	The study reports the development of a pair of neural networks. One network detects and classifies clinical keratoconus and keratoconus suspects from among a variety of potentially confounding topographic patterns. A second network quantifies the severity of any conelike feature that matches the topographic pattern of clinical keratoconus or keratoconus suspect.
Target condition and reference standard(s)	Unclear who made the diagnosis. All cases were diagnosed before the analyses.
Flow and timing	It was unclear if the diagnoses were made by the same person(s). All participants were included in the analyses.
Comparative	Not applicable
Notes	Supported by National Eye Institute grants EY03311 and EY02377.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		

Smolek 1997 *(Continued)*

Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Unclear
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Souza 2010

Study characteristics	
Patient Sampling	Retrospective, single-centre, case-control study, in which data was collected from medical records: normal (n = 172), astigmatism (n = 89), keratoconus (n = 46), and photorefractive keratectomy (n = 11).
Patient characteristics and setting	There were 4 groups: normal, astigmatism, keratoconus, and photorefractive keratectomy.

Souza 2010 *(Continued)*

	• Clinically diagnosed normal eyes, with no abnormal flattening or steepening on tangential map and absence of irregular astigmatism, were included in the normal (< 1.5 D cylinder) or astigmatism (> 1.5 D cylinder) groups. • Keratoconus: central corneal power > 48.7 D, an I-S asymmetry > 1.913 or Vogt's striae or Fleischer's ring. Exclusion criteria: Orbscan II maps with poor corneal coverage, missing data points, poor fixation, or lid artefacts
Index tests	The performance of support vector machine was evaluated to detect keratoconus apart from all other corneal patterns, using Orbscan II data.
Target condition and reference standard(s)	It was unclear if the diagnoses were made by the same person(s). All participants were included in the analyses.
Flow and timing	Participants were divided into 4 groups, but it is unclear who made the diagnosis. The diagnoses were made before the analysis with the classification tree.
Comparative	Different algorithms were developed; it is unclear whether the results were interpreted separately. All data were included in the different analyses.
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Souza 2010 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.	Yes		
Are the proportions and reasons for missing data similar for all index tests?	Unclear		
		Unclear risk	

Subramaniam 2022

Study characteristics	
Patient Sampling	The SyntEye KTC model (Rozema et al.) has been used to generate the data set that is to be used for the training of the convolutional neural network. The data set consists of topography images of healthy normal eyes, developing keratoconus eyes, and keratoconus eyes.
Patient characteristics and setting	Data set consists of subclinical keratoconus and keratoconus eyes.
Index tests	Convolutional neural network. It analyses topography images and classifies them into 3 categories: normal, subclinical and keratoconus.

Subramaniam 2022 *(Continued)*

	The article provides a clear explanation of the model and training procedure.
Target condition and reference standard(s)	The topography images were artificially synthesized by a program called Synteye; it made 300 images of each classification: normal, subclinical keratoconus, and keratoconus. No human observation is mentioned.
Flow and timing	All cases were included in the reference standard and index test. All data were presented in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Subramaniam 2022 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Twa 2005

Study characteristics	
Patient Sampling	Retrospective, single-centre, case-control study that included eyes diagnosed with keratoconus and a reference group of normal eyes of corneal refractive surgery candidates
Patient characteristics and setting	• Normal group included both eyes of people who had myopia with or without astigmatism who were examined for the surgical correction of refractive error and had videokeratography examination data. Exclusion criteria: no other documented ocular disease. The normal participants were sequentially selected from all people who were examined for the surgical correction of refractive error. • Keratoconus: keratoconus diagnosis in 1 or both eyes by International Classification of Diseases, 9th Revision diagnosis codes between August 1998 and January 2000. Their diagnostic classification was confirmed by chart review, applying criteria established by the Collaborative Longitudinal Evaluation of Keratoconus (CLEK) study. Exclusion criteria: lack of full videokeratography data over the region of study (central 7 mm), or obvious artifacts in videokeratography data resulting from obstruction of the corneal surface by the eyelids.
Index tests	An automated decision tree analysis that analyses videokeratography and classifies the cases into normal or keratoconus.
Target condition and reference standard(s)	Unclear if the diagnoses were made by the same person(s). All participants were included in the analyses.

Twa 2005 *(Continued)*

Flow and timing	Participants were divided into 2 groups, but it was unclear who made the diagnosis. The diagnoses were made before the analysis with the classification tree.
Comparative	Not applicable
Notes	This study was supported by National Institutes of Health grants EY16225 and EY13359 (MDT), American Optometric Foundation Ocular Sciences Ezell Fellowship, Ameritech faculty fellowship (SP), and NIH-EY12952 (MAB).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Twa 2005 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Did all patients receive the same reference standard?	Unclear	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk
DOMAIN 5: Comparative		

Xie 2020

Study characteristics	
Patient Sampling	Retrospective, case-control study, including people throughout China who wanted to undergo refractive surgery, had a primary diagnosis of keratoconus, and had stable postoperative refractive states. In total, 6465 corneal tomographic images from 1385 people were collected to develop the AI model.
Patient characteristics and setting	The following groups were included: normal cornea, suspected irregular cornea, early-stage keratoconus, keratoconus, and myopic postoperative cornea. • Normal cornea has a natural shape with all the indices within a normal range. Normal corneas included with-the-rule astigmatism and normally thin corneas. • A suspected irregular cornea describes an at-risk cornea. Such a cornea may have I-S values outside the reference range or aberrant C-shaped or round posterior surface elevations. Alternatively, a suspected irregular cornea may have an unusual pachymetric progression. • Keratoconus and early-stage keratoconus: comprehensive analysis of all the morphologic and characteristic indices according to an Asian database by 3 cornea specialists
Index tests	InceptionResNetV2 architecture in a convolutional neural network on the TensorFlow platform to create the AI model with transfer learning technique. The algorithm uses a deep learning algorithm with corneal tomographic imaging and divides the images into the previously mentioned groups. This model may aid in identifying at-risk corneas and determining which people are unsuited to corneal refractive surgery, thereby assisting in surgery decision-making.
Target condition and reference standard(s)	The expert team included 3 senior ophthalmologists with at least 5 years of practical experience in the refractive surgery centre of the study clinic. Each image was independently labelled by the 3 experts, none of whom knew the labels selected by the others. When the labels differed, that chosen by 2 of the 3 experts was selected as the standard.
Flow and timing	The data from all participants were checked by 3 ophthalmologists. The diagnosis was made before the analysis with the AI algorithm.
Comparative	Not applicable

Xie 2020 *(Continued)*

Notes	The research received funding through grants 2018YFC0116500 from the National Key R&D Program of China, 31671000 from the Natural Science Foundation of China, 201804020007 from the Guangzhou Science and Technology Planning Project, 81822010 from the National Natural Science Foundation of China, 2018B010109008 from the Science and Technology Planning Projects of Guangdong Province, and 2017TX04R031 from the Guangdong Science and Technology Innovation Leading Talents.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Xie 2020 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Unclear
DOMAIN 4: Flow and timing		
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk
DOMAIN 5: Comparative		

Xu 2017

Study characteristics	
Patient Sampling	Prospective, single-centre, cross-sectional study. People with keratoconus were enroled at the Affiliated Eye Hospital of Wenzhou Medical University in China. Complete ocular examinations were performed by 2 experienced doctors, including a review of medical and family history, corrected distance visual acuity, slit-lamp biomicroscopy, fundus examination, and corneal topography using Medmont E300 (Medmont, Inc., Nunawading Melbourne, Australia).

Patient characteristics and setting	Participants were divided into the following 3 groups.
	
	• Normal group (healthy eyes from normal subjects) were enroled from the hospital staff and university students if they met the following screening criteria.
	  ◦ Central average keratometry < 45.0D
	  ◦ Myopia < −6.00 D and astigmatism < −2.00 D
	  ◦ No clinical signs or suggestive topographic patterns for suspicious subclinical keratoconus, keratoconus, or pellucid marginal degeneration
	  ◦ No history of ocular surgery or trauma
	  ◦ Stopped contact lens wear for ≥ 8 weeks for rigid gas-permeable contact lenses and ≥ 4 week for soft contact lenses.
	• Keratoconus group (mild or moderate keratoconus eyes) had to meet the following criteria.
	  ◦ Central average keratometry > 47.0D
	  ◦ ≥ 1 slit-lamp sign (stromal thinning, Vogt's striae, or Fleischer's ring > 2-mm arc)
	  ◦ Asymmetric topographical features with I-S ≥ 1.9 D of the vertical gradient power across the 6-mm region
	  ◦ No history of contact lens wear, ocular surgery, or extensive scarring
	• Subclinical keratoconus group (fellow eye of unilateral keratoconus) had to meet the following criteria.
	  ◦ Central average keratometry < 45.0D
	  ◦ Diagnosis of keratoconus in the contralateral eye
	  ◦ No clinical signs of keratoconus on slit-lamp biomicroscopy, retinoscopy, and ophthalmoscopy

Xu 2017 *(Continued)*

- Corneal topographical features with I-S values < 1.4D of the vertical gradient power across the 6-mm region
- No history of contact lens wear, ocular surgery, or trauma

Index tests	Participants were divided into a training set (normal, subclinical keratoconus group, and keratoconus group) used to build the discrimination function, and a validation set (normal and subclinical group) used to test the diagnostic power. The goal of the present study was to apply the Zernike fitting method to describe the 3D varying complexity of corneal shapes and the 3D distribution of corneal thickness, and to characterize the entire corneal topography and tomography data in subclinical eyes, keratoconus eyes, and normal eyes using Pentacam tomography. Furthermore, the metrics constructed from Zernike polynomials were compared to improve the diagnostic sensitivity and specificity for the detection of subclinical keratoconus corneas.
Target condition and reference standard(s)	2 experienced doctors performed complete ocular examinations, including a review of medical and family history, corrected distance visual acuity, slit-lamp biomicroscopy, fundus examination, and corneal topography using Medmont E300 (Medmont, Inc., Nunawading Melbourne, Australia).
Flow and timing	The data from all participants were checked by 3 ophthalmologists. The diagnosis was made before the analysis with the AI algorithm.
Comparative	Not applicable
Notes	This study was supported by the National Natural Science Foundation of China (81400441 to Shen), the National Key Research and Development Program of China (2016YFC0102500 to Wang), and the Zhejiang provincial Natural Science Foundation of China (LQ17H120008 to Xu).

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Unclear		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Xu 2017 *(Continued)*

If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Xu 2022a

Study characteristics	
Patient Sampling	Single centre, retrospective, case-control study. The control group consisted of refractive surgery candidates with normal clinical and topographic features. Early keratoconus was defined as local corneal steepening and asymmetric astigmatism, or the contralat-

Xu 2022a *(Continued)*

	eral eye was diagnosed with keratoconus. It also included people with keratoconus.
Patient characteristics and setting	Data set consisted of subclinical keratoconus and keratoconus eyes.
Index tests	A predictive index, Sirius Keratoconus Index (SKI), was constructed using LASSO and Logistic regression analyses based on topographic, pachymetric, and aberrometry variables of the Sirius. The cut-off value of the SKI was set at 0.44.
Target condition and reference standard(s)	Unclear how the cases were diagnosed.
Flow and timing	Unclear whether all cases received the same reference standard. All cases were included in the index test. All data were included in a 2 × 2 table.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Unclear risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			

Xu 2022a *(Continued)*

Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			
Were different AI tests were developed and interpreted without knowledge of each other.			
Are the proportions and reasons for missing data similar for all index tests?			

Yang 2021

Study characteristics	
Patient Sampling	Cross-sectional, observational study that recruited participants at the Casey Eye Institute at Oregon Health and Science University (OHSU), Portland, Oregon. Age-matched normal participants were recruited from volunteers and people seeking refractive surgery consultation. All participants were aged 18 years or older.
Patient characteristics and setting	A clinical diagnosis of keratoconus was established using a combination of corrected distance visual acuity (CDVA), slit-lamp physical findings, topographic patterns, and the quantitative topography KISA%. Each keratoconic eye was assigned to 1 of the 3 keratoconic subgroups according to following classification scheme. • Manifest keratoconus: slit-lamp findings associated with keratoconus (Vogt's striae, Fleischer's ring, Munson's sign, iron ring, Rizzuti's sign, and apparent focal corneal bulging and thinning) or CDVA < 20/20; topography characteristic of keratoconus (asymmetric bowtie with a skewed radial axis, central or inferior steep zone, and claw shape); and KISA% > 100%. • Subclinical keratoconus: CDVA ≥ 20/20; no slit-lamp findings of keratoconus; topography characteristic of keratoconus or pellucid marginal degeneration; and KISA% > 100%. • Forme fruste keratoconus: the better eye of people with asymmetric keratoconus (AKC) with CDVA ≥ 20/20; no slit-lamp findings of keratoconus; and KISA% < 100%.

Yang 2021 *(Continued)*

	Age-matched normal participants were recruited from volunteers and people seeking refractive surgery consultation. All normal eyes had CDVA ≥ 20/20, no signs of keratoconus on slit-lamp examination, regular axial power map topography pattern (round, oval, symmetric bowtie, etc.), KISA% < 100%, and no ocular pathology other than myopia or hyperopia. Exclusion criteria: previous corneal surgeries, recent contact lens usage (soft contact lens within 1 week or rigid gas-permeable lens within 3 weeks), inability to give informed consent, or inability to maintain stable fixation for imaging. Severe keratoconus with corneal scarring has unpredictable corneal and epithelial thickness patterns and does not pose a challenge for clinical diagnosis.
Index tests	A 2-step decision tree. Step 1 uses quantitative OCT pachymetric and epithelial thickness map parameters. If any of the 4 parameters listed in the previous section exceeds the cut-off, the eye is suspicious for keratoconus and proceeds to step 2. If none of the 4 parameters exceeds the cut-off, then the eye is considered normal and does not require step 2 examination. Step 2 requires a human grader to visually inspect the corneal and epithelial thickness maps and search for characteristic keratoconic map patterns of coincident thinning and concentric epithelial thinning.
Target condition and reference standard(s)	Unclear who made the diagnosis. All cases were diagnosed before the analysis with the 2-step decision tree.
Flow and timing	Unclear if all cases were diagnosed by the same cornea specialists. All cases were included in the analysis.
Comparative	Not applicable
Notes	Supported by the National Institutes of Health, Bethesda, Maryland, USA (R01EY028755, R01EY029023, T32EY023211, and P30EY010572; E. Pavlatos, D. Huang, and Y. Li); a research grant and equipment support from OptoVue, Inc., Fremont, California (D. Huang and Y. Li); unrestricted grants to Casey Eye Institute from Research to Prevent Blindness, Inc., New York, New York (E. Pavlatos, W. Chamberlain, D. Huang, and Y. Li); National Natural Science Foundation of China, Beijing, China (81900830; Y. Yang). The sponsors did not participate in the data collection, data management, or data analysis in the study.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			

Yang 2021 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Yousefi 2018

Study characteristics	
Patient Sampling	Multicentre, retrospective, case-control study. Corneal OCT images were collected from 12,242 eyes of 3162 people using SS-1000 CASIA OCT Imaging Systems (Tomey, Japan) and other parameters from the electronic health record (EHR) system.
Patient characteristics and setting	All available data were collected without any preconditions. The investigators then selected a single visit from each eye and excluded eyes with missing ectasia status index (ESI). Eyes were grouped as follows according to ESI. • Normal: ESI 0–4 • Forme fruste keratoconus (or keratoconus suspect): ESI 5–29 • Keratoconus: ESI ≥ 30 Using Casia labels, the data set included 1970 healthy eyes, 796 eyes with forme fruste keratoconus, and 390 eyes with keratoconus.
Index tests	The algorithm included 3 major steps, as follows. • Principal component analysis (PCA) was used to linearly reduce the dimensionality of the input data from 420 to 8 significant principal components. • Manifold learning was used to further reduce the selected principal components non-linearly to 2 eigen parameters. • Finally, a density-based clustering was applied to the eigen parameters to identify eyes with keratoconus.
Target condition and reference standard(s)	The article does not state who made the diagnosis. All cases were diagnosed before the analysis.
Flow and timing	Unclear whether all participants were diagnosed by the same cornea specialists. Unclear if all cases included in the analysis.
Comparative	Not applicable
Notes	The study authors were funded by an unrestricted grant from Research to Prevent Blindness (RPB), New York, NY. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Unclear		

Yousefi 2018 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Unclear
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Unclear
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Unclear		
Could the patient flow have introduced bias?		Unclear risk	
DOMAIN 5: Comparative			

Zeboulon 2020a

Study characteristics	
Patient Sampling	Retrospective, machine-learning, experimental study. The Orbscan (Bausch & Lomb, Bridgewater, NJ) database was exported using the batch export functionality both as image files and the underlying numeric data matrixes represented by each colour map. They selected 3000 examinations in total, 1000 per class (normal, keratoconus, history of refractive surgery). All 3000 ex-

Zeboulon 2020a *(Continued)*

	aminations were obtained from different people, and only 1 eye per person was selected. They balanced the examinations to have exactly 500 left eyes and 500 right eyes in each class. The selection process was as follows: consecutive examinations were preselected by a resident and reviewed by a corneal tomography expert.
Patient characteristics and setting	• Keratoconus group: an anterior curvature map showing 1 of the classic keratoconus patterns described by Rabinowitz et al. associated with corneal thinning. • Refractive surgery group: examinations were an oblate anterior surface (flat in its centre), a prolate posterior surface (steep in its centre), central corneal thinning, and lower central curvature values compared with the periphery (cases of myopic laser surgery). • Normal group: examinations were preselected if no corneal condition could be detected.
Index tests	The possibility of using numeric data matrixes instead of colour maps to train a convolutional neural network (CNN) for a classification task. Specifically, the investigators used 4 maps that are frequently used in clinical practice, stacked together as if they were 4 colour channels of a single image to classify examinations into 3 categories: normal, keratoconus, and history of refractive surgery. The training set was trained during 15 epochs with a learning rate of 0.0001 and a batch size of 2.
Target condition and reference standard(s)	The diagnosis was made by a resident and corneal tomography specialist with at least 5 years of experience. The diagnosis was made before the convolutional neural network analysis.
Flow and timing	All participants were diagnosed by 2 cornea specialists. All cases were included in the analysis.
Comparative	Not applicable
Notes	The study authors received no funds or support for the study.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			

Zeboulon 2020a *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Unclear		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

Zeboulon 2020b

Study characteristics	
Patient Sampling	Retrospective, machine-learning, experimental study. 22,066 Orbscan (Bausch&Lomb, USA) examinations were randomly extracted from the Orbscan database using the batch export functionality. The last examination of the first visit for each eye of each person was selected. This process reduced the number of examinations to 13,705. The cases were divided into the following groups: normal, keratoconus, history of myopic refractive surgery, Fuchs' corneal dystrophy, and other.

Zeboulon 2020b *(Continued)*

Patient characteristics and setting	• Normal • Keratoconus group: if the anterior curvature map showed 1 of the classic keratoconus patterns described by Rabinowitz et al. associated with corneal thinning • Refractive surgery group: cases that underwent myopic laser surgery and had an oblate anterior surface (flat in its centre), a prolate posterior surface (steep in its centre), a central corneal thinning and lower central curvature values compared to the periphery. • Fuchs' examinations had a central pachymetry > 600 microns with reduced peripheral thickness in the corneal periphery and an oblate posterior surface. • Other: all examinations that could not be assigned to a specific class with a good level of confidence, including bad quality examinations
Index tests	The efficiency of unsupervised algorithms was tested to extract and sort usable examinations from a large unlabelled corneal topography database into different diagnostic clusters, with little human intervention, data cleaning or feature selection. Convolutional neural network (CNN) was used.
Target condition and reference standard(s)	All 13,705 examinations were manually labelled and checked by 2 corneal topography experts (with at least 5 years of practice in a corneal and refractive surgery department) in a random order. The cases were diagnosed before the convolutional neural network analysis.
Flow and timing	All participants were diagnosed by 2 cornea specialists. All cases were included in the analysis.
Comparative	Not applicable
Notes	No funding source mentioned.

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (All tests)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Zeboulon 2020b *(Continued)*

If a threshold was used, was it pre-specified?	Unclear		
Was the model designed in an appropriate manner?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	
DOMAIN 5: Comparative			

AI: artificial intelligence; AS-OCT: anterior segment optical coherence tomography; AST: astigmatism index; BAD-D: Belin-Ambrósio Enhanced Ectasia Display total deviation; D: dioptre; I-S: inferior-superior; KISA% index: keratoconus percentage index, derived from central keratometry, the inferior-superior value, the astigmatism index, and the SRAX index, an expression of irregular astigmatism occurring in keratoconus; LASIK: laser-assisted in situ keratomileusis; OCT: optical coherence tomography; PPK: percent probability of keratoconus; SD: standard deviation; TMS: Topographic Modeling System.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Aatila 2021	No 2 × 2 data available.
Al-Timemy 2022	No article available, conference proceedings.
Buehren 2018	Conference proceedings.
Cao 2021b	Conference proceedings.

Study	Reason for exclusion
Castro-Luna 2021	Ineligible index test/reference test.
ChiCTR2000037484	Clinical trial protocol/ongoing study.
ChiCTR2000039070	Clinical trial protocol/ongoing study.
DosSantos 2019	Ineligible outcomes.
Elsawy 2022	Ineligible outcomes.
Feng 2021	No 2 × 2 data available.
Hazarbessanov 2022	No article available; conference proceedings.
Hernandez 2020	Conference proceedings.
Hidalgo 2014	Conference proceedings.
Hjordtal 1995	Ineligible population.
Issarti 2018	Conference proceedings.
JPRN-UMIN000034587	Clinical trial protocol.
JPRN-UMIN000040128	Clinical trial protocol.
JPRN-UMIN000040308	Clinical trial protocol.
JPRN-UMIN000040321	Clinical trial protocol.
JPRN-UMIN000043831	Clinical trial protocol/ongoing study.
Kleinhans 2019	Language.
Klyce 2005	Ineligible outcomes.
Kundu 2021	No 2 × 2 data available.
Lavric 2019	Ineligible population.
Li 2009	Ineligible outcomes.
Li 2021	Language.
Liu 2021	Conference proceedings.
Malyugin 2021	No 2 × 2 data available.
Matalia 2020	Ineligible population.
Nasrin 2018	Clinical trial protocol/ongoing study.
NCT01746823	Clinical trial protocol.
NCT04313387	Clinical trial protocol/ongoing study.

Study	Reason for exclusion
NCT04763785	Clinical trial protocol/ongoing study.
Omidi 2022	Ineligible outcomes.
Pavlatos 2022	No article available, conference proceeding.
Ramos-Lopez 2011	No 2 × 2 data available.
Rozema 2017	Ineligible outcomes.
Saad 2010	Ineligible index test/reference test.
Saad 2012	Ineligible index test/reference test.
Schatteburg 2022	No article available, clinical trial protocol.
Souza 2008	Language.
Steinberg 2015a	Wrong index test/reference test.
Steinberg 2015b	Wrong index test/reference test.
Takahashi 2021	Conference proceedings.
Tan 2019	Language.
Tas 2021	Conference proceedings.
Toprak 2021	Ineligible index test/reference test.
Velazquez-Blazquez 2020	No 2 × 2 data available.
Vieira de Carvalho 2008	No 2 × 2 data available.
Wang 2022	Ineligible population.
Xu 2022b	No 2 × 2 data available.
Yucekul 2022	Wrong index test.
Zghal 1997	Language.
Zou 2019	Language.

DATA

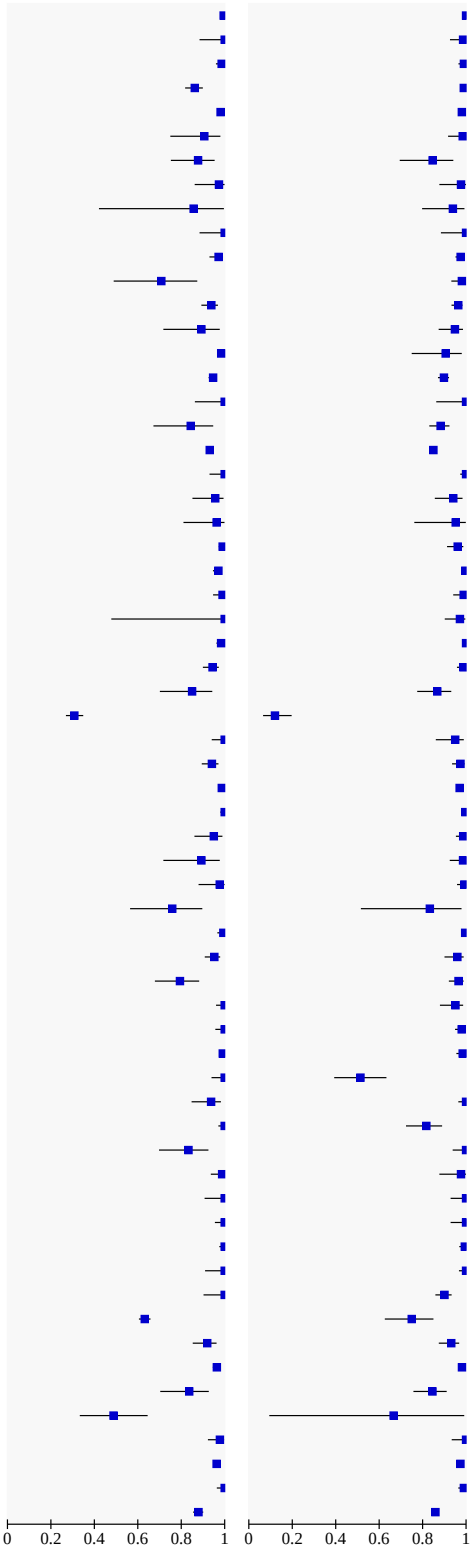
Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 Artificial intelligence (all studies)	63	56364
2 Artificial intelligence (manifest keratoconus)	54	50519
3 Artificial intelligence (subclinical keratoconus)	28	9508
4 Artificial intelligence (mixed)	11	11644

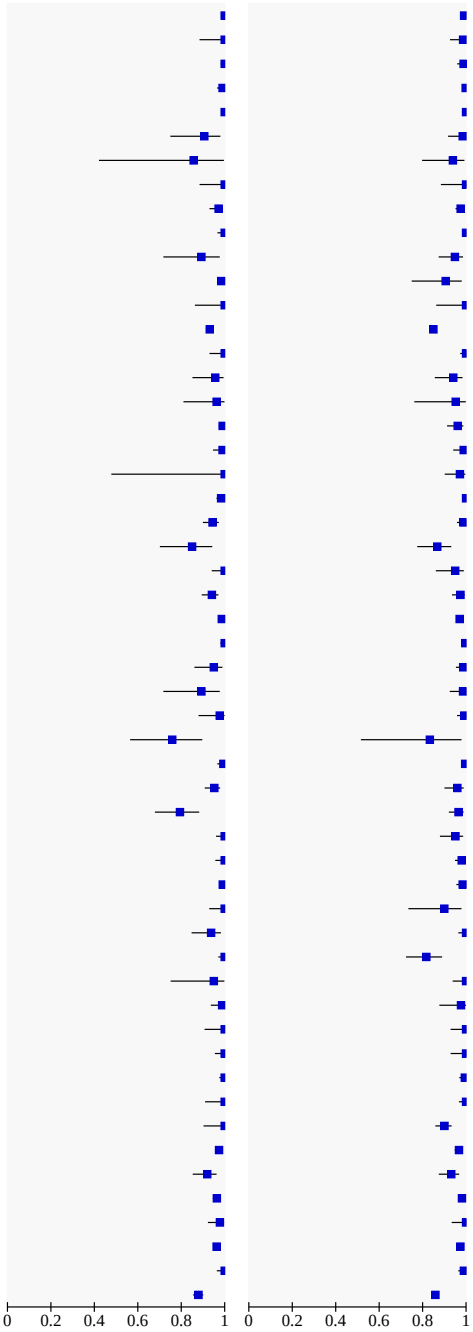
Test 1. Artificial intelligence (all studies)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdelmotaal 2020	1679	0	8	887	1.00 [0.99, 1.00]	1.00 [1.00, 1.00]
Accardo 2002	30	1	0	72	1.00 [0.88, 1.00]	0.99 [0.93, 1.00]
Al-Timemy 2021	212	2	3	202	0.99 [0.96, 1.00]	0.99 [0.97, 1.00]
Almeida 2022	257	12	41	1136	0.86 [0.82, 0.90]	0.99 [0.98, 0.99]
Arbelaez 2012	886	35	17	1771	0.98 [0.97, 0.99]	0.98 [0.97, 0.99]
Bessho 2006	29	1	3	64	0.91 [0.75, 0.98]	0.98 [0.92, 1.00]
Cao 2020	43	6	6	33	0.88 [0.75, 0.95]	0.85 [0.69, 0.94]
Cao 2021a	37	1	1	42	0.97 [0.86, 1.00]	0.98 [0.88, 1.00]
Carvalho 2005	6	2	1	31	0.86 [0.42, 1.00]	0.94 [0.80, 0.99]
Castro-Luna 2020	30	0	0	30	1.00 [0.88, 1.00]	1.00 [0.88, 1.00]
Cavas-Martinez 2017	139	8	4	313	0.97 [0.93, 0.99]	0.98 [0.95, 0.99]
Chan 2015	17	2	7	102	0.71 [0.49, 0.87]	0.98 [0.93, 1.00]
Chandapura 2019	168	9	11	242	0.94 [0.89, 0.97]	0.96 [0.93, 0.98]
Chastang 2000	25	4	3	74	0.89 [0.72, 0.98]	0.95 [0.87, 0.99]
Chen 2021	492	3	8	29	0.98 [0.97, 0.99]	0.91 [0.75, 0.98]
Cohen 2022	587	59	33	521	0.95 [0.93, 0.96]	0.90 [0.87, 0.92]
Consejo 2020	25	0	0	25	1.00 [0.86, 1.00]	1.00 [0.86, 1.00]
De Almeida Jr 2021	27	24	5	181	0.84 [0.67, 0.95]	0.88 [0.83, 0.92]
Elsawy 2021	7208	1136	532	6424	0.93 [0.93, 0.94]	0.85 [0.84, 0.86]
Feizi 2016	51	0	0	136	1.00 [0.93, 1.00]	1.00 [0.97, 1.00]
Gairola 2022	44	4	2	64	0.96 [0.85, 0.99]	0.94 [0.86, 0.98]
Gao 2022	26	1	1	20	0.96 [0.81, 1.00]	0.95 [0.76, 1.00]
Ghaderi 2021	317	5	3	125	0.99 [0.97, 1.00]	0.96 [0.91, 0.99]
Issarti 2019	303	1	9	311	0.97 [0.95, 0.99]	1.00 [0.98, 1.00]
Issarti 2020	101	1	1	91	0.99 [0.95, 1.00]	0.99 [0.94, 1.00]
Kalin 1996	5	2	0	69	1.00 [0.48, 1.00]	0.97 [0.90, 1.00]
Kamiya 2019	299	0	5	239	0.98 [0.96, 0.99]	1.00 [0.98, 1.00]
Kamiya 2021	169	2	10	168	0.94 [0.90, 0.97]	0.99 [0.96, 1.00]
Kojima 2020	34	11	6	72	0.85 [0.70, 0.94]	0.87 [0.78, 0.93]
Kojima 2021	166	95	373	13	0.31 [0.27, 0.35]	0.12 [0.07, 0.20]
Kovacs 2016	60	3	0	57	1.00 [0.94, 1.00]	0.95 [0.86, 0.99]
Kuo 2020	160	4	10	152	0.94 [0.89, 0.97]	0.97 [0.94, 0.99]
Lavric 2021	4100	51	55	1675	0.99 [0.98, 0.99]	0.97 [0.96, 0.98]
Lopes 2018	182	9	0	2971	1.00 [0.98, 1.00]	1.00 [0.99, 1.00]
Lucena 2021	57	2	3	150	0.95 [0.86, 0.99]	0.99 [0.95, 1.00]
Maeda 1994	25	1	3	71	0.89 [0.72, 0.98]	0.99 [0.93, 1.00]
Maeda 1995a	43	1	1	131	0.98 [0.88, 1.00]	0.99 [0.96, 1.00]
Maeda 1995b	22	2	7	10	0.76 [0.56, 0.90]	0.83 [0.52, 0.98]
Mahmoud 2013	162	1	1	243	0.99 [0.97, 1.00]	1.00 [0.98, 1.00]
Mahmoud 2021	160	4	8	96	0.95 [0.91, 0.98]	0.96 [0.90, 0.99]
Mohammadpour 2022	54	5	14	139	0.79 [0.68, 0.88]	0.97 [0.92, 0.99]
Pavlatos 2020	91	4	0	78	1.00 [0.96, 1.00]	0.95 [0.88, 0.99]
Rabinowitz 1999	84	4	0	195	1.00 [0.96, 1.00]	0.98 [0.95, 0.99]
Ruiz 2016	450	3	4	191	0.99 [0.98, 1.00]	0.98 [0.96, 1.00]
Ruiz 2017	59	35	0	37	1.00 [0.94, 1.00]	0.51 [0.39, 0.63]
Saad 2014	60	0	4	102	0.94 [0.85, 0.98]	1.00 [0.96, 1.00]
Saad 2016	125	17	0	76	1.00 [0.97, 1.00]	0.82 [0.72, 0.89]
Saika 2013	40	0	8	58	0.83 [0.70, 0.93]	1.00 [0.94, 1.00]
Shetty 2015	84	1	1	42	0.99 [0.94, 1.00]	0.98 [0.88, 1.00]
Shi 2020	38	0	0	50	1.00 [0.91, 1.00]	1.00 [0.93, 1.00]
Sideroudi 2017	80	0	0	50	1.00 [0.95, 1.00]	1.00 [0.93, 1.00]
Smadja 2013	148	1	0	176	1.00 [0.98, 1.00]	0.99 [0.97, 1.00]
Smolek 1997	39	0	0	111	1.00 [0.91, 1.00]	1.00 [0.97, 1.00]
Souza 2010	36	27	0	245	1.00 [0.90, 1.00]	0.90 [0.86, 0.93]
Subramaniam 2022	818	16	476	48	0.63 [0.61, 0.66]	0.75 [0.63, 0.85]
Twa 2005	103	9	9	123	0.92 [0.85, 0.96]	0.93 [0.87, 0.97]
Xie 2020	562	12	21	625	0.96 [0.95, 0.98]	0.98 [0.97, 0.99]
Xu 2017	41	15	8	82	0.84 [0.70, 0.93]	0.85 [0.76, 0.91]
Xu 2022a	21	1	22	2	0.49 [0.33, 0.65]	0.67 [0.09, 0.99]
Yang 2021	89	0	2	54	0.98 [0.92, 1.00]	1.00 [0.93, 1.00]
Yousefi 2018	1142	51	44	1919	0.96 [0.95, 0.97]	0.97 [0.97, 0.98]
Zeboulon 2020a	100	2	0	198	1.00 [0.96, 1.00]	0.99 [0.96, 1.00]
Zeboulon 2020b	700	814	96	4961	0.88 [0.85, 0.90]	0.86 [0.85, 0.87]



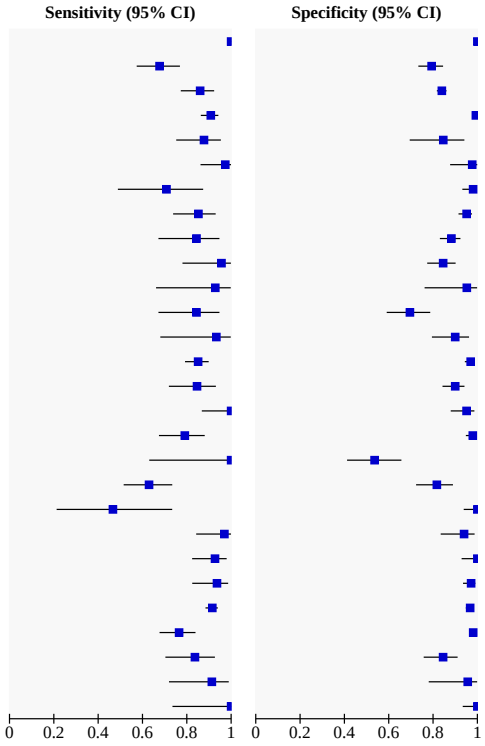
Test 2. Artificial intelligence (manifest keratoconus)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdelmotaal 2020	789	8	0	887	1.00 [1.00, 1.00]	0.99 [0.98, 1.00]
Accardo 2002	30	1	0	72	1.00 [0.88, 1.00]	0.99 [0.93, 1.00]
Al-Timemy 2021	204	2	0	173	1.00 [0.98, 1.00]	0.99 [0.96, 1.00]
Almeida 2022	203	2	2	1146	0.99 [0.97, 1.00]	1.00 [0.99, 1.00]
Arbelaez 2012	643	0	0	1029	1.00 [0.99, 1.00]	1.00 [1.00, 1.00]
Bessho 2006	29	1	3	64	0.91 [0.75, 0.98]	0.98 [0.92, 1.00]
Carvalho 2005	6	2	1	31	0.86 [0.42, 1.00]	0.94 [0.80, 0.99]
Castro-Luna 2020	30	0	0	30	1.00 [0.88, 1.00]	1.00 [0.88, 1.00]
Cavas-Martinez 2017	139	8	4	313	0.97 [0.93, 0.99]	0.98 [0.95, 0.99]
Chandapura 2019	109	0	0	216	1.00 [0.97, 1.00]	1.00 [0.98, 1.00]
Chastang 2000	25	4	3	74	0.89 [0.72, 0.98]	0.95 [0.87, 0.99]
Chen 2021	492	3	8	29	0.98 [0.97, 0.99]	0.91 [0.75, 0.98]
Consejo 2020	25	0	0	25	1.00 [0.86, 1.00]	1.00 [0.86, 1.00]
Elsawy 2021	7208	1136	532	6424	0.93 [0.93, 0.94]	0.85 [0.84, 0.86]
Feizi 2016	51	0	0	136	1.00 [0.93, 1.00]	1.00 [0.97, 1.00]
Gairola 2022	44	4	2	64	0.96 [0.85, 0.99]	0.94 [0.86, 0.98]
Gao 2022	26	1	1	20	0.96 [0.81, 1.00]	0.95 [0.76, 1.00]
Ghaderi 2021	317	5	3	125	0.99 [0.97, 1.00]	0.96 [0.91, 0.99]
Issarti 2020	101	1	1	91	0.99 [0.95, 1.00]	0.99 [0.94, 1.00]
Kalin 1996	5	2	0	69	1.00 [0.48, 1.00]	0.97 [0.90, 1.00]
Kamiya 2019	299	0	5	239	0.98 [0.96, 0.99]	1.00 [0.98, 1.00]
Kamiya 2021	169	2	10	168	0.94 [0.90, 0.97]	0.99 [0.96, 1.00]
Kojima 2020	34	11	6	72	0.85 [0.70, 0.94]	0.87 [0.78, 0.93]
Kovacs 2016	60	3	0	57	1.00 [0.94, 1.00]	0.95 [0.86, 0.99]
Kuo 2020	160	4	10	152	0.94 [0.89, 0.97]	0.97 [0.94, 0.99]
Lavric 2021	4100	51	55	1675	0.99 [0.98, 0.99]	0.97 [0.96, 0.98]
Lopes 2018	182	9	0	2971	1.00 [0.98, 1.00]	1.00 [0.99, 1.00]
Lucena 2021	57	2	3	150	0.95 [0.86, 0.99]	0.99 [0.95, 1.00]
Maeda 1994	25	1	3	71	0.89 [0.72, 0.98]	0.99 [0.93, 1.00]
Maeda 1995a	43	1	1	131	0.98 [0.88, 1.00]	0.99 [0.96, 1.00]
Maeda 1995b	22	2	7	10	0.76 [0.56, 0.90]	0.83 [0.52, 0.98]
Mahmoud 2013	162	1	1	243	0.99 [0.97, 1.00]	1.00 [0.98, 1.00]
Mahmoud 2021	160	4	8	96	0.95 [0.91, 0.98]	0.96 [0.90, 0.99]
Mohammadpour 2022	54	5	14	139	0.79 [0.68, 0.88]	0.97 [0.92, 0.99]
Pavlatos 2020	91	4	0	78	1.00 [0.96, 1.00]	0.95 [0.88, 0.99]
Rabinowitz 1999	84	4	0	195	1.00 [0.96, 1.00]	0.98 [0.95, 0.99]
Ruiz 2016	450	3	4	191	0.99 [0.98, 1.00]	0.98 [0.96, 1.00]
Ruiz 2017	50	3	0	27	1.00 [0.93, 1.00]	0.90 [0.73, 0.98]
Saad 2014	60	0	4	102	0.94 [0.85, 0.98]	1.00 [0.96, 1.00]
Saad 2016	125	17	0	76	1.00 [0.97, 1.00]	0.82 [0.72, 0.89]
Saika 2013	19	0	1	58	0.95 [0.75, 1.00]	1.00 [0.94, 1.00]
Shetty 2015	84	1	1	42	0.99 [0.94, 1.00]	0.98 [0.88, 1.00]
Shi 2020	38	0	0	50	1.00 [0.91, 1.00]	1.00 [0.93, 1.00]
Sideroudi 2017	80	0	0	50	1.00 [0.95, 1.00]	1.00 [0.93, 1.00]
Smadja 2013	148	1	0	176	1.00 [0.98, 1.00]	0.99 [0.97, 1.00]
Smolek 1997	39	0	0	111	1.00 [0.91, 1.00]	1.00 [0.97, 1.00]
Souza 2010	36	27	0	245	1.00 [0.90, 1.00]	0.90 [0.86, 0.93]
Subramaniam 2022	450	16	12	476	0.97 [0.96, 0.99]	0.97 [0.95, 0.98]
Twa 2005	103	9	9	123	0.92 [0.85, 0.96]	0.93 [0.87, 0.97]
Xie 2020	562	12	21	625	0.96 [0.95, 0.98]	0.98 [0.97, 0.99]
Yang 2021	89	0	2	54	0.98 [0.92, 1.00]	1.00 [0.93, 1.00]
Yousefi 2018	1142	51	44	1919	0.96 [0.95, 0.97]	0.97 [0.97, 0.98]
Zeboulon 2020a	100	2	0	198	1.00 [0.96, 1.00]	0.99 [0.96, 1.00]
Zeboulon 2020b	700	814	96	4961	0.88 [0.85, 0.90]	0.86 [0.85, 0.87]



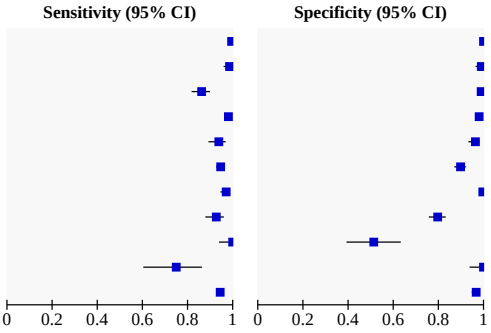
Test 3. Artificial intelligence (subclinical keratoconus)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdelmotaal 2020	857	0	0	887	1.00 [1.00, 1.00]	1.00 [1.00, 1.00]
Al-Timemy 2021	65	45	31	173	0.68 [0.57, 0.77]	0.79 [0.73, 0.85]
Almeida 2022	80	184	13	964	0.86 [0.77, 0.92]	0.84 [0.82, 0.86]
Arbelaez 2012	208	7	21	1029	0.91 [0.86, 0.94]	0.99 [0.99, 1.00]
Cao 2020	43	6	6	33	0.88 [0.75, 0.95]	0.85 [0.69, 0.94]
Cao 2021a	37	1	1	42	0.97 [0.86, 1.00]	0.98 [0.88, 1.00]
Chan 2015	17	2	7	102	0.71 [0.49, 0.87]	0.98 [0.93, 1.00]
Chandapura 2019	52	11	9	216	0.85 [0.74, 0.93]	0.95 [0.91, 0.98]
De Almeida Jr 2021	27	24	5	181	0.84 [0.67, 0.95]	0.88 [0.83, 0.92]
Feizi 2016	22	21	1	115	0.96 [0.78, 1.00]	0.85 [0.77, 0.90]
Gao 2022	13	1	1	20	0.93 [0.66, 1.00]	0.95 [0.76, 1.00]
Issarti 2020	27	28	5	64	0.84 [0.67, 0.95]	0.70 [0.59, 0.79]
Kovacs 2016	14	6	1	54	0.93 [0.68, 1.00]	0.90 [0.79, 0.96]
Lopes 2018	160	9	28	289	0.85 [0.79, 0.90]	0.97 [0.94, 0.99]
Mohammadpour 2022	44	16	8	144	0.85 [0.72, 0.93]	0.90 [0.84, 0.94]
Pavlatos 2020	26	4	0	78	1.00 [0.87, 1.00]	0.95 [0.88, 0.99]
Ruiz 2016	53	4	14	190	0.79 [0.67, 0.88]	0.98 [0.95, 0.99]
Ruiz 2017	8	32	0	37	1.00 [0.63, 1.00]	0.54 [0.41, 0.66]
Saad 2016	51	17	30	76	0.63 [0.52, 0.73]	0.82 [0.72, 0.89]
Saika 2013	7	0	8	58	0.47 [0.21, 0.73]	1.00 [0.94, 1.00]
Shi 2020	32	3	1	47	0.97 [0.84, 1.00]	0.94 [0.83, 0.99]
Sideroudi 2017	51	0	4	50	0.93 [0.82, 0.98]	1.00 [0.93, 1.00]
Smadja 2013	44	5	3	172	0.94 [0.82, 0.99]	0.97 [0.94, 0.99]
Subramaniam 2022	388	16	36	476	0.92 [0.88, 0.94]	0.97 [0.95, 0.98]
Xie 2020	88	12	27	625	0.77 [0.68, 0.84]	0.98 [0.97, 0.99]
Xu 2017	41	15	8	82	0.84 [0.70, 0.93]	0.85 [0.76, 0.91]
Xu 2022a	21	1	2	22	0.91 [0.72, 0.99]	0.96 [0.78, 1.00]
Yang 2021	12	0	0	54	1.00 [0.74, 1.00]	1.00 [0.93, 1.00]



Test 4. Artificial intelligence (mixed)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdelmotaal 2020	1679	0	8	887	1.00 [0.99, 1.00]	1.00 [1.00, 1.00]
Al-Timemy 2021	212	2	3	202	0.99 [0.96, 1.00]	0.99 [0.97, 1.00]
Almeida 2022	257	12	41	1136	0.86 [0.82, 0.90]	0.99 [0.98, 0.99]
Arbelaez 2012	886	35	17	1771	0.98 [0.97, 0.99]	0.98 [0.97, 0.99]
Chandapura 2019	168	9	11	242	0.94 [0.89, 0.97]	0.96 [0.93, 0.98]
Cohen 2022	587	59	33	521	0.95 [0.93, 0.96]	0.90 [0.87, 0.92]
Issarti 2019	303	1	9	311	0.97 [0.95, 0.99]	1.00 [0.98, 1.00]
Kojima 2021	166	95	13	373	0.93 [0.88, 0.96]	0.80 [0.76, 0.83]
Ruiz 2017	59	35	0	37	1.00 [0.94, 1.00]	0.51 [0.39, 0.63]
Saika 2013	36	0	12	58	0.75 [0.60, 0.86]	1.00 [0.94, 1.00]
Subramaniam 2022	818	16	48	476	0.94 [0.93, 0.96]	0.97 [0.95, 0.98]



ADDITIONAL TABLES

Table 1. Study characteristics

Study ID	Study design	Study population	Sample size	Country	Instrument	Index test	Reference standard
Abdelmotaal 2020	Retrospective, single-centre, case-control	Refractive surgery candidates, subclinical KC, and manifest KC	3218 images (3218 eyes of 1669 participants)	Egypt	Pentacam	Convolutional neural network	2 cornea specialists
Accardo 2002	Retrospective, single-centre, case-control	Healthy controls, KC, and other ocular diseases	396 images (396 eyes of 198 participants)	Italy	Eyesys	Neural network	Unclear
Almeida 2022	Retrospective, multicentre, case-control	Healthy controls (who underwent PRK or LASIK), subclinical KC, manifest KC	2893 eyes (2893 participants)	Brazil	Pentacam	Multiple logistic regression	1 cornea specialist
Al-Timemy 2021	Retrospective, multicentre, case-control	Healthy controls, subclinical KC, and manifest KC	1050 images (150 eyes of 85 participants)	Brazil	Pentacam	Convolutional neural network	3 cornea specialists
Arbelaez 2012	Retrospective, multicentre, case series	Healthy controls, subclinical KC, and manifest KC	3502 eyes	Oman, Italy	Sirius	Support vector machine	Unclear
Bessho 2006	Retrospective, multicentre, case-control	Healthy controls and KC	165 eyes (120 participants)	Japan	Orbscan IIz	Logistic regression	Unclear
Cao 2020	Retrospective, case-control	Healthy controls and subclinical KC	88 participants	Australia	Pentacam	Random forest	Unclear
Cao 2021a	Retrospective, case-control	Healthy controls and subclinical KC	267 eyes (226 participants): 186 training, 81 test	Australia	Pentacam	Random forest	2 cornea specialists
Carvalho 2005	Retrospective, single-centre, case-control	Instrument database	80 eyes: 40 training, 40 test set	Brazil	Eyesys	Neural network	2 cornea specialists
Castro-Luna 2020	Retrospective, single-centre, case-control	Control group and KC	60 eyes (60 participants)	Spain	CSO topography system	Bayesian network	Unclear
Cavas-Martinez 2017	Retrospective, single-centre, case-control	Control group and KC	464 eyes (464 participants)	Spain	Sirius	Logistic regression	Unclear

Table 1. Study characteristics *(Continued)*

Chan 2015	Retrospective, single-centre, case-control	Database of KC of the Singapore National Eye Center	128 images (128 participants)	Singapore	Orbsan II	Discriminant analysis	1 cornea specialist
Chandapura 2019	Retrospective, multicentre, case-control	Healthy controls, subclinical KC, and manifest KC	439 eyes	India, Brazil	OCT RTVue + Pentacam	Random forest	1 cornea specialist
Chastang 2000	Retrospective, single-centre, case-control	Control group (e.g. healthy, regular astigmatism, radial keratotomy) and KC	208 eyes: 104 training, 104 validation set	France	Eyesys	Binary decision tree	2 cornea specialists
Chen 2021	Retrospective, multicentre, case-control	Healthy controls and KC	1926 images	UK, Iran, New Zealand	Pentacam	Convolutional neural network	Unclear
Cohen 2022	Retrospective, single-centre, case-control	Healthy controls, KC, and subclinical KC	8526 corneal tomography examinations (2525 participants)	Israel	Galilei Dual Scheimpflug Analyzer	Random forest	1 cornea specialist
Consejo 2020	Prospective, single-centre, case-control	Control group and KC	50 eyes	Belgium	Corvis-ST	Support vector machine	1 ophthalmologist
De Almeida Jr 2021	Prospective, single-centre, case-control	LASIK or PRK candidates and KC	Training 777 eyes, validation 237 eyes	Brazil	Pentacam	Support vector machine	1 cornea specialist
Elsawy 2021	Prospective, single-centre, case-control	Control group (healthy, dry eye, Fuchs' endothelial dystrophy) and KC	158,220 images (879 eyes, 478 participants): training 134,460, validation 23,760	USA	Envisu R2210 (AS-OCT)	Neural network	6 cornea specialists
Feizi 2016	Prospective, single-centre, case-control	Refractive surgery candidates, subclinical KC, and manifest KC	210 eyes (207 participants)	Iran	Galilei Dual Scheimpflug Analyzer	Logistic regression	Unclear
Gairola 2022	Retrospective, single-centre, case-control	Healthy controls and KC	2224 images	India	Topography (Keratron and KC-smart device)	Convolutional neural network	1 ophthalmologist
Gao 2022	Retrospective, single-centre, case-control	Healthy controls, KC, and subclinical KC	1040 images (208 participants)	China	Pentacam	Neural network	Unclear

Table 1. Study characteristics *(Continued)*

Ghaderi 2021	Retrospective, single-centre, case-control	Healthy controls and KC (single-centre database)	450 eyes (separated into training, validation, and test sets)	Iran	Pentacam	Ensemble learning system	Unclear
Issarti 2019	Retrospective, single-centre, case-control	Healthy controls and KC (single-centre database)	851 eyes	Belgium	Pentacam	Feedforward neural network	1 ophthalmologist, 1 optometrist
Issarti 2020	Retrospective, multicentre, case-control	Healthy controls and KC (multicentre database)	812 eyes	Belgium	Pentacam	Feedforward neural network	1 ophthalmologist, 1 optometrist
Kalin 1996	Prospective, consecutive, cross-sectional study	Refractive surgery candidates and KC	106 eyes (53 participants)	USA	TMS-1	Binary decision tree	1 ophthalmologist
Kamiya 2019	Retrospective, single-centre, case-control	Refractive surgery candidates, contact lens fitting candidates, and KC	543 eyes	Japan	CASIA SS-1000	Convolutional neural network	Cornea specialists
Kamiya 2021	Retrospective, single-centre, case-control	Refractive surgery candidates, contact lens fitting candidates, and KC	349 eyes	Japan	TMS-4 topographer	Convolutional neural network	Cornea specialists
Kojima 2020	Retrospective, multicentre, case-control	Healthy controls and KC	329 eyes	Japan	Auto-keratometer	Logistic regression	2 cornea specialists
Kojima 2021	Retrospective, single-centre, case-control	Healthy controls, KC, and subclinical KC	647 eyes (335 participants)	Japan	Auto-keratometer (ARK-1)	Regression algorithm	2 cornea specialists
Kovacs 2016	Retrospective, single-centre, case-control	Refractive surgery candidates, normal eye of unilateral KC, and KC	135 eyes: training 70%, test set 30%	Hungary	Pentacam	Neural network	Unclear
Kuo 2020	Retrospective, single-centre, case-control	Refractive surgery candidates and KC	354 images (206 participants)	Taiwan	TMS-4	Convolutional neural network	4 cornea specialists
Lavric 2021	Retrospective, case-control	Controls and KC	5881 eyes (2800 participants)	Brazil	Pentacam	Support vector machine	Unclear

Table 1. Study characteristics *(Continued)*

Lopes 2018	Retrospective, multicentre, case-control	LASIK cases, post-LASIK ectasia, and KC	3648 eyes	USA, Brazil, UK, Italy	Pentacam	Random forest	1 cornea specialist
Lucena 2021	Retrospective, case-control	Control group and KC	1172 images: training 960, test set 212	Brazil	Topographers	Convolutional neural network	1 cornea specialist
Maeda 1994	Single-centre, case-control	Control group and KC	200 eyes: training 100, test 100	USA	TMS-1	Combined discriminant analysis and classification tree	3 cornea specialists
Maeda 1995a	Single-centre, case-control	Control group and KC	176 eyes (125 participants)	USA	TMS-1	Combined discriminant analysis and classification tree	Unclear
Maeda 1995b	Single-centre, case-control	Control group and KC	183 eyes: training 108, test set 75	USA	TMS-1	Neural network	Unclear
Mohammadpour 2022	Prospective, diagnostic test accuracy study	Healthy controls, subclinical KC, and manifest KC	217 eyes (212 participants)	Iran	Sirius	Neural network	2 cornea specialists
Mahmoud 2013	Retrospective, multicentre, case-control	Healthy controls and KC	407 eyes	Colombia, Switzerland, USA	Galilei Dual Scheimpflug-Placido tomographer	Logistic regression	Unclear
Mahmoud 2021	Case-control	Healthy controls and KC	250 eyes	Unclear	CASIA SS-1000	Convolutional neural network	1 ophthalmologist
Pavlatos 2020	Prospective, multicentre, case-control	Healthy controls, subclinical KCT, and manifest KC	215 eyes	USA, China	OCT RTVue or Avanti	CTN index	Unclear
Rabinowitz 1999	Retrospective, single-centre, case-control	Healthy controls and KC	281 participants	USA	TMS-1	KISA% index	Unclear
Ruiz 2016	Retrospective, single-centre, case-control	Healthy controls, refractive surgery candidates, irregular astigmatism, subclinical KC, and manifest KC	860 eyes	Belgium	Pentacam	Support vector machine	1 cornea specialist, 1 optometrist

Table 1. Study characteristics *(Continued)*

Ruiz 2017	Retrospective, multicentre, case-control	Healthy controls, post-refractive surgery candidates, subclinical KC, and manifest KC	131 eyes (102 participants)	Belgium, France	Topographers	Support vector machine	Unclear
Saad 2014	Prospective, single-centre, case-control	Refractive surgery candidates, subclinical KC, and manifest KC	166 eyes	France	Orbscan IIz	Discriminant analysis	1 cornea specialist
Saad 2016	Prospective, single-centre, case-control	Refractive surgery candidates, subclinical KC, and manifest KC	119 eyes (176 participants)	France	Placido disk topographer	Discriminant analysis	Unclear
Saika 2013	Single-centre case-control	Healthy controls, LASIK candidates, subclinical KC, and manifest KC	212 eyes	Japan	Placido disk topographer	Linear discriminant analysis	Unclear
Shetty 2015	Retrospective, single-centre, case-control	Healthy controls and KC	128 eyes	India	Pentacam	Logistic regression	Unclear
Shi 2020	Prospective, single-centre, case-control	Healthy controls and KC	121 eyes (121 participants)	China	Scheimpflug and UHR-OCT	Neural network	2 cornea specialists
Sideroudi 2017	Prospective, cross-sectional, non-randomized observational study	Refractive surgery candidates, subclinical KC, and manifest KC	185 eyes (185 participants)	Greece	Pentacam	Logistic regression	Unclear
Smadja 2013	Retrospective, single-centre, case-control	Refractive surgery or routine ophthalmic examination, referrals, subclinical KC, and manifest KC	372 eyes (197 participants)	France	Galilei rotating Scheimpflug tomography	Tree classification	Unclear
Smolek 1997	Retrospective, single-centre, case-control	Normal, with-the-rule astigmatism, KC, subclinical KC, contact lens-induced corneal warpage, pellucid marginal degeneration, PRK, radial keratotomy, and keratoplasty	300 examinations (150 training, 150 test)	USA	TMS-1	Neural network	Unclear
Souza 2010	Retrospective, single-centre, case-control	Healthy controls, astigmatism, photorefractive keratectomy, and KC	318 participants	Brazil	Orbscan IIz	Support vector machine	Unclear
Subramaniam 2022	Case-control study	Healthy controls, subclinical KC, and manifest KC	1500 images	India	Topography images syn-	Convolutional neural network	Unclear

Table 1. Study characteristics *(Continued)*

					thesized with SyntEye		
Twa 2005	Retrospective, single-centre, case-control	Refractive surgery candidates and KC	224 eyes	USA	Topography	Decision tree	Unclear
Xie 2020	Retrospective, observational	Refractive surgery candidates, KC	6465 images (1385 participants)	China	Pentacam	Convolutional neural network	3 ophthalmologists
Xu 2017	Prospective, single-centre, cross-sectional	Healthy controls, subclinical KC, and manifest KC	363 eyes (363 participants)	China	Pentacam	Discriminant analysis	2 ophthalmologists
Xu 2022a	Retrospective, single-centre, case-control	Healthy controls and subclinical KC	92 eyes (80 participants)	China	Sirius	Logistic regression	Unclear
Yang 2021	Cross-sectional, observational	Healthy controls, refractive surgery candidates, subclinical KC, and manifest KC	176 eyes (124 participants)	USA	OCT	Decision tree (2-step)	Unclear
Yousefi 2018	Retrospective, multicentre, case-control	Healthy controls and KC	3156 participants	Japan, USA	CASIA OCT	Density-based clustering	Unclear
Zeboulon 2020a	Retrospective, case-control	Healthy controls, refractive surgery candidates, subclinical KC, and manifest KC	3000 examinations	France	Orbscan	Convolutional neural network	1 ophthalmology resident, 1 corneal tomography expert
Zeboulon 2020b	Retrospective, case-control	Healthy controls, history of myopic refractive surgery, Fuchs' corneal dystrophy, and KC	6979 participants	France	Orbscan	Convolutional neural network	1 ophthalmology resident, 1 corneal tomography expert

AS-OCT: anterior segment optical coherence tomography; CTN index: Coincident Thinning Index; KC: keratoconus; KISA% index: keratoconus percentage index, derived from central keratometry, the inferior-superior value, the astigmatism index, and the SRAX index, an expression of irregular astigmatism occurring in keratoconus; LASIK: laser-assisted in situ keratomileusis; OCT: optical coherence tomography; PRK: photorefractive keratectomy; TMS: Topographic Modeling System; UHR-OCT: ultrahigh-resolution optical coherence tomography.

Table 2. Subgroup analyses

Subgroups		No. of studies (participants)	Sensitivity (95% CI)	P-value for relative sensitivity	Specificity (95% CI)	P-value for relative specificity
Study design	**Clinical series**	46 (38,788)	0.987 (0.977, 0.993)	Reference	0.984 (0.975, 0.993)	Reference
	Registries	8 (11,731)	0.975 (0.919, 0.993)	0.458	0.975 (0.936, 0.990)	0.464
AI algorithm	**Logistic regression**	8 (2,889)	0.983 (0.957, 0.993)	Reference	0.992 (0.974, 0.997)	Reference
	Bayesian network	3 (788)	0.994 (0.972-0.999)	0.260	0.982 (0.834, 0.998)	0.666
	Convolutional neural network	13 (13,452)	0.979 (0.945-0.991)	0.734	0.978 (0.960, 0.988)	0.110
	Discriminant analysis	3 (462)	0.977 (0.945, 0.990)	0.628	1.000 (0.814, 1.000)	0.093
	Decision tree	5 (8,96)	0.976 (0.895, 0.995)	0.731	0.978 (0.935, 0.993)	0.299
	Neural network	10 (16,296)	0.973 (0.920, 0.991)	0.561	0.968 (0.931, 0.986)	0.093
	Other	6 (4,338)	0.990 (0.892, 0.999)	0.629	0.968 (0.931, 0.987)	0.068
	Random forest	2 (3,487)	1.000 (0, 1.000)	0.038	0.997 (0.994, 0.999)	0.270
	SVM	4 (7,911)	0.994 (0.982, 0.998)	0.203	0.993 (0.928, 0.999)	0.916
Imaging technique	**OCT**	6 (19,585)	0.971 (0.941, 0.985)	Reference	0.984 (0.885, 0.998)	Reference
	Tomography	26 (27,267)	0.993 (0.985, 0.996)	0.042	0.986 (0.976, 0.992)	0.910
	Topography	21 (3,579)	0.965	0.744	0.978	0.756

Table 2. Subgroup analyses *(Continued)*

			(0.931, 0.983)		(0.958, 0.989)	
Data type	**Images**	13 (27,532)	0.980 (0.950, 0.992)	Reference	0.975 (0.947, 0.988)	Reference
	Parameters	39 (22,792)	0.987 (0.976, 0.947)	0.461	0.984 (0.975, 0.990)	0.342

CI: confidence interval.

APPENDICES

Appendix 1. Glossary of terms

Asphericity: a measure of corneal shape and how it affects the refraction of light
Astigmatism: refractive error due to an abnormal shape of the cornea
BAD-D: Belin-Ambrósio Enhanced Ectasia Display total deviation
Curvature: the rate of change of direction of a curve with respect to distance along the curve
Dioptre (D): a unit of measurement for the strength of a lens (i.e. the light breaking ability of a lens)
Iatrogenic ectasia: weakening of the biomechanical stability of the cornea due to surgery, which leads to the development of a keratoconus-like ectasia
Keratometry: the measurement of the corneal radius of curvature
Kmax: maximum keratometry expressed in dioptres
Pachymetry: corneal thickness
Tomography: imaging by sections, able to describe the anterior and posterior surface of an object
Topography: imaging and description of the features of a surface

Appendix 2. CENTRAL search strategy

#1 MeSH descriptor: [Keratoconus] this term only
#2 keratoconus*
#3 cornea* near/5 ectatic*
#4 cornea* near/5 ectasia
#5 conical near/2 cornea*
#6 cornea* near/2 thinning
#7 #1 OR #2 OR #3 OR #4 OR #5 OR #6
#8 MeSH descriptor: [Artificial Intelligence] this term only
#9 MeSH descriptor: [Deep Learning] this term only
#10 MeSH descriptor: [Machine Learning] explode all trees
#11 MeSH descriptor: [Neural Networks, Computer] this term only
#12 MeSH descriptor: [Algorithms] this term only
#13 MeSH descriptor: [Decision Trees] this term only
#14 MeSH descriptor: [Automation] this term only
#15 MeSH descriptor: [Databases, Factual] this term only
#16 MeSH descriptor: [Electronic Data Processing] this term only
#17 artificial NEAR/1 intelligence
#18 (deep or machine) NEAR/2 learning
#19 vector NEAR/3 machine
#20 AI or DL or DLS
#21 (deep or convolutional or neural) NEAR/3 network*
#22 automat* NEAR/2 (screen* or detect* or diagnos* or algorithm* or identif* or grading or graded or method*)
#23 Bagging
#24 Naive NEAR/1 Bayes
#25 Multilayer NEAR/1 Perceptron
#26 (multi-layer NEAR/1 perceptron) or MLP
#27 Radial NEAR/1 Basis NEAR/1 Function
#28 Random NEAR/1 Forest

#29 Ensemble NEAR/1 Selection
#30 (Ada or gradient) NEAR/1 boost*
#31 LASSO
#32 Elastic NEAR/1 Net
#33 genetic NEAR/1 algorithm*
#34 (decision or classification or regression or probability or model*) NEAR/3 tree*
#35 logistic* NEAR/2 regression NEAR/15 learn*
#36 augment* NEAR/1 clinical NEAR/1 decision* NEAR/1 mak*
#37 nearest NEAR/1 (neighbor or neighbour)
#38 fuzzy NEAR/3 (logit or logic or logistic)
#39 kernel
#40 #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39
#41 #7 AND #40

Appendix 3. MEDLINE Ovid search strategy

1. Keratoconus/
2. keratoconus$.tw.
3. (cornea$ adj5 ectatic$).tw.
4. (cornea$ adj5 ectasia).tw.
5. (conical adj2 cornea$).tw.
6. (cornea$ adj2 thinning).tw.
7. or/1-6
8. artificial intelligence/
9. deep learning/
10. exp machine learning/
11. "neural networks (computer)"/
12. fuzzy logic/
13. algorithms/
14. decision tree/
15. automation/
16. databases, factual/
17. information processing/
18. (artificial adj1 intelligence).tw.
19. ((deep or machine) adj2 learning).tw.
20. (vector adj3 machine).tw.
21. (AI or DL or DLS).tw.
22. ((deep or convolutional or neural) adj3 network$).tw.
23. (automat$ adj2 (screen$ or detect$ or diagnos$ or algorithm$ or identif$ or grading or graded or method$)).tw.
24. Bagging.tw.
25. (Naive adj1 Bayes).tw.
26. (Multilayer adj1 Perceptron).tw.
27. ((multi-layer adj1 perceptron) or MLP).tw.
28. (Radial adj1 Basis adj1 Function).tw.
29. (Random adj1 Forest).tw.
30. (Ensemble adj1 Selection).tw.
31. ((Ada or gradient) adj1 boost$).tw.
32. LASSO.tw.
33. (Elastic adj1 Net).tw.
34. (genetic adj1 algorithm$).tw.
35. ((decision or classification or regression or probability or model$) adj3 tree$).tw.
36. (logistic$ adj2 regression adj15 learn$).tw.
37. (augment$ adj1 clinical adj1 decision$ adj1 mak$).tw.
38. (nearest adj1 (neighbor or neighbour)).tw.
39. (fuzzy adj3 (logit or logic or logistic)).tw.
40. kernel.tw.
41. or/8-40
42. 7 and 41

Appendix 4. Embase Ovid search strategy

1. keratoconus/

2. keratoconus$.tw.
3. (cornea$ adj5 ectatic$).tw.
4. (cornea$ adj5 ectasia).tw.
5. (conical adj2 cornea$).tw.
6. (cornea$ adj2 thinning).tw.
7. or/1-6
8. artificial intelligence/
9. deep learning/
10. machine learning/
11. supervised machine learning/ or support vector machine/ or unsupervised machine learning/
12. perceptron/
13. artificial neural network/
14. convolutional neural network/
15. deep neural network/
16. automated pattern recognition/
17. decision tree/
18. detection algorithm/
19. learning algorithm/
20. classification algorithm/
21. data classification/
22. disease classification/
23. disease simulation/
24. automation/
25. information processing/
26. feature extraction/
27. bayesian learning/
28. fuzzy system/
29. k nearest neighbor/
30. kernel method/
31. random forest/
32. (artificial adj1 intelligence).tw.
33. ((deep or machine) adj2 learning).tw.
34. (vector adj3 machine).tw.
35. (AI or DL or DLS).tw.
36. ((deep or convolutional or neural) adj3 network$).tw.
37. (automat$ adj2 (screen$ or detect$ or diagnos$ or algorithm$ or identif$ or grading or graded or method$)).tw.
38. Bagging.tw.
39. (Naive adj1 Bayes).tw.
40. (Multilayer adj1 Perceptron).tw.
41. ((multi-layer adj1 perceptron) or MLP).tw.
42. (Radial adj1 Basis adj1 Function).tw.
43. (Random adj1 Forest).tw.
44. (Ensemble adj1 Selection).tw.
45. ((Ada or gradient) adj1 boost$).tw.
46. LASSO.tw.
47. (Elastic adj1 Net).tw.
48. (genetic adj1 algorithm$).tw.
49. ((decision or classification or regression or probability or model$) adj3 tree$).tw.
50. (logistic$ adj2 regression adj15 learn$).tw.
51. (augment$ adj1 clinical adj1 decision$ adj1 mak$).tw.
52. (nearest adj1 (neighbor or neighbour)).tw.
53. (fuzzy adj3 (logit or logic or logistic)).tw.
54. kernel.tw.
55. or/8-54
56. 7 and 55

Appendix 5. OpenGrey search strategy

keratoconus AND (Artificial intelligence OR deep learning OR machine learning)

Appendix 6. ISRCTN search strategy

keratoconus AND (Artificial intelligence OR deep learning OR machine learning)

Appendix 7. ClinicalTrials.gov search strategy

keratoconus AND (Artificial intelligence OR deep learning OR machine learning)

Appendix 8. WHO ICTRP search strategy

keratoconus AND Artificial intelligence OR keratoconus AND deep learning OR keratoconus AND machine learning

Appendix 9. QUADAS 2 guidance

DOMAIN	Low risk/concern	Unclear	High risk/concern
PATIENT SELECTION	Describe methods of patient selection; describe included patients (prior testing, presentation, intended use of index test and setting):		
Was a consecutive or random sample of patients enroled?	Consecutive sampling or random sampling seeking refractive error correction or refractive surgery in eye services.	Unclear whether consecutive or random sampling used.	Selection of non-consecutive patients.
Was a case-control design avoided?	No selective recruitment of people with or without keratoconus.	Unclear selection mechanism.	Selection of either cases or control in a predetermined, non-random fashion; or enrichment of the cases from a selected population.
Did the study avoid inappropriate exclusions?	Exclusions are detailed and felt to be appropriate (e.g. people already diagnosed with keratoconus or with other corneal diseases).	Exclusions are not detailed (pending contact with study authors).	Inappropriate exclusions are reported (e.g. of people with borderline index test results).
Risk of bias: could the selection of patients have introduced bias?	'No' for any of the above		
Concerns regarding applicability: are there concerns that the included patients do not match the review question?	Inclusion of patients seeking refractive error correction or refractive surgery in primary or secondary care eye services.	Unclear inclusion criteria.	Inclusion of patients attending cornea services for known disease, population-based studies, registry-based studies.
INDEX TEST	Describe the index test and how it was conducted and interpreted:		
Were the index test results interpreted without knowledge of the results of the reference standard?	Test performed "blind" or "independently and without knowledge of" reference standard results are sufficient and full details of the blinding procedure are not required; or clear temporal pattern to the order of testing that precludes the need for formal blinding.	Unclear whether results are interpreted independently.	Reference standard results available to those who conducted or interpreted the index test.
If a threshold was used, was it prespecified?	The study authors declare that the selected cut-off used to dichotomize data was specified a priori, or a protocol is available with this information.	No information on preselection of index test cut-off values.	A study is classified at higher risk of bias if the authors define the optimal cut-off post hoc based on their own study data.
Risk of bias: could the conduct or interpretation of the	'No' for any of the above.		

(Continued)

index test have introduced bias?			
Concerns regarding applicability: are there concerns that the index test, its conduct, or interpretation differ from the review question?	Tests used and testing procedure clearly reported and tests executed by personnel with sufficient training.	Unclear execution of the tests or unclear study personnel profile, background, and training.	Tests used are not validated, or study personnel is insufficiently trained.
REFERENCE STANDARD	Describe the reference standard and how it was conducted and interpreted:		
Is the reference standard likely to correctly classify the target condition?	Topography and/or tomography interpreted independently by 2 or more cornea specialists.	Topography and/or tomography interpreted by cornea specialists, but not enough details to adjudicate 'yes' or 'no'.	Topography and/or tomography interpreted by only one cornea specialist.
Were the reference standard results interpreted without knowledge of the results of the index test?	Reference standard performed "blind" or "independently and without knowledge of" index test results are sufficient and full details of the blinding procedure are not required; or clear temporal pattern to the order of testing that precludes the need for formal blinding.	Unclear whether results are interpreted independently.	Index test results available to those who conducted the reference standard.
Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias?	'No' for any of the above.		
Concerns regarding applicability: are there concerns that the target condition as defined by the reference standard does not match the review question?	Same or similar definition of the target condition as described in the protocol.	Unclear definition of the target disease diagnosed by the reference standard.	Different definition of the target condition as defined in the protocol.
FLOW AND TIMING	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2 × 2 table (refer to flow diagram): describe the time interval and any interventions between index test(s) and reference standard.		
Was there an appropriate interval between index test(s) and reference standard?	No more than three months between index and reference test execution.	—	More than three months between index and reference test execution.
Did all patients receive a reference standard?	All participants receiving the index test are verified with the reference standard.	—	Not all participants receiving the index test are verified with the reference standard.
Did all patients receive the same reference standard?	Not applicable for this review.		
Were all patients included in the analysis?	The number of participants included in the study matches the number in analyses, or participants with undefined or borderline test results are excluded.	—	The number of participants included in the study does not match the number in analyses, or participants with undefined

(Continued)

			or borderline test results are excluded.
Risk of bias: could the patient flow have introduced bias?	'No' for any of the above,		
ADDITIONAL QUESTIONS	These questions concern the direct comparisons between AI tests,		
Were different AI tests developed and interpreted without knowledge of each other?	Different AI tests were developed and interpreted "blind" or "independently and without knowledge of" the results of each other.	—	Different AI tests were developed or their results interpreted with knowledge of the results of each other.
Are the proportions and reasons for missing data similar for all index tests?	Missing data and their causes were similar for each AI test.	—	The amount of missing data or their causes differed between AI tests.

HISTORY

Protocol first published: Issue 12, 2021

CONTRIBUTIONS OF AUTHORS

MMSV: conception of the review, design of the review, search and selection of studies for inclusion in the review, collection of data for the review, assessment of the risk of bias in the included studies, assessment of the certainty in the body of evidence, interpretation of data, writing of the review
EF: search and selection of studies for inclusion in the review, collection of data for the review, assessment of the risk of bias in the included studies
MV: critical revision of artificial intelligence sections of the review, providing advice on artificial intelligence
EL: analysis of data, critical revision of statistical section
TB: critical revision of the review
RM: critical revision of clinical sections of the review
RMMAN: critical revision of clinical sections of the review
GV: design of the review, analysis of data, assessment of the certainty in the body of evidence, critical revision of all review sections
MMD: conception of the review, critical revision of all review sections

DECLARATIONS OF INTEREST

MMSV, EF, MV, EL, TB, RM: have no conflicts of interest.
RMMAN: Carl Zeiss Meditec AG (Independent Contractor - Consultant), Alcon Laboratories Inc (Independent Contractor - Consultant), Johnson & Johnson Health Care Systems Inc. (Independent Contractor - Consultant)
GV: former Cochrane Editor, has not been involved in the editorial process of this review
MMD: Maastricht University (Employment), Maastricht Universitair Medisch Centrum (MUMC+) (Employment)

SOURCES OF SUPPORT

Internal sources

- University Eye Clinic Maastricht, Maastricht University Medical Center (MUMC+), Maastricht, Netherlands

  Authors' place of employment (MMSV, TB, RMMAN, MMD)
- Department of Neurosciences, Psychology, Drug Research and Child Health (NEUROFARBA), University of Florence, Florence, Italy

  Authors' place of employment (RM, GV)
- Department of Neurosciences, Psychology, Pharmacology and Child Health, University of Florence, Florence, Italy

  Authors' place of employment (EF)
- Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Authors' place of employment (MV)
- Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy

Authors' place of employment (EL)
- Department of Neuroscience, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, Netherlands

Authors' place of employment (MMSV)
- Centre for Public Health, Queen's University Belfast, UK

Authors' place of employment (GV)

External sources

- Public Health Agency, UK

This review was supported by the HSC Research and Development (R&D) Division of the Public Health Agency which funds the Cochrane Eyes and Vision editorial base at Queen's University Belfast.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

1. During the quality assessment process, we noticed that the QUADAS-2 tool was not entirely fitted to our review. In the domain 'Index test', we added the question 'Was the model designed in an appropriate manner?'. We considered a study at low risk of bias if data from a single participant were reserved to only one data partition, parameters were tuned, and the optimal model was selected. We considered a study at high risk of bias if data from a single participant were not reserved to only one data partition, parameters were not tuned, and the optimal model was not selected. When the design of the model was unclear, and we could not determine the above-mentioned properties, we considered the study at unclear risk. In the protocol for this review, we stated that the 'Concerns regarding applicability' question in the domain 'Reference standard' ('Are there concerns that the target condition as defined by the reference standard does not match the review question?') was not applicable to this review (Vandevenne 2021); however, we corrected this during quality assessment. Additionally, in the domain 'Flow and timing', we removed the question 'Was there an appropriate interval between index test(s) and reference standard?', as it was not applicable to this review. The reference test and index test were performed on the same corneal images or parameters, so an interval between index and reference test is irrelevant.
2. We had planned to use a hierarchical summary receiver operating characteristic (HSROC) model and estimate the average sensitivity at fixed specificity values according to cut-offs for terciles of specificity (Macaskill 2010). However, we found accuracy was nearly maximal in the vast majority of studies, which clustered close to the upper-left corner of the ROC plane. Thus, we pooled data using a bivariate model, which is equivalent to an HSROC model in absence of covariates (Harbord 2007).
3. We had planned to conduct direct comparisons between the index tests, (different types or data sources for AI) if sufficient data were available. We conducted these analyses with a test covariate in the bivariate model as suggested in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Takwoingi 2022).
4. We had planned to conduct analyses using the 'metadas' user-written command in SAS software (SAS software) and to make predictions at fixed specificities using NLMIXED procedure postestimation commands. Since we fitted bivariate models, we used Stata software*metandi* and *melogit* commands, as recommended in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (*Takwoingi 2022).
5. The search line (logistic$ adj2 regression$).tw. in the protocol was amended to (logistic$ adj2 regression adj15 learn$).tw. in the review (Vandevenne 2021). The original line retrieved all reports of logistic regression being used as a statistical method in studies relating to keratoconus. The search phrase was retrieving too many false hits and was edited to identify reports where logistic regression was used in conjunction with some form of machine learning.

INDEX TERMS

Medical Subject Headings (MeSH)

*Artificial Intelligence; Case-Control Studies; Cross-Sectional Studies; *Keratoconus [diagnostic imaging]; Physical Examination

MeSH check words

Humans