

Keratoconus detection of changes using deep learning of colour-coded maps

Xu Chen,¹ Jiaxin Zhao,¹ Katja C Iselin,² Davide Borroni ,² Davide Romano,² Akilesh Gokul,³ Charles N J McGhee,³ Yitian Zhao,⁴ Mohammad-Reza Sedaghat,^{5,6} Hamed Momeni-Moghaddam,^{5,6} Mohammed Ziaei,³ Stephen Kaye,^{1,2} Vito Romano ,^{1,2} Yalin Zheng¹

To cite: Chen X, Zhao J, Iselin KC, *et al.* Keratoconus detection of changes using deep learning of colour-coded maps. *BMJ Open Ophthalmology* 2021;**6**:e000824. doi:10.1136/bmjophth-2021-000824

► Additional online supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjophth-2021-000824>).

Received 7 June 2021
Accepted 5 July 2021

ABSTRACT

Objective To evaluate the accuracy of convolutional neural networks technique (CNN) in detecting keratoconus using colour-coded corneal maps obtained by a Scheimpflug camera.

Design Multicentre retrospective study.

Methods and analysis We included the images of keratoconic and healthy volunteers' eyes provided by three centres: Royal Liverpool University Hospital (Liverpool, UK), Sedaghat Eye Clinic (Mashhad, Iran) and The New Zealand National Eye Center (New Zealand). Corneal tomography scans were used to train and test CNN models, which included healthy controls. Keratoconic scans were classified according to the Amsler-Krumeich classification. Keratoconic scans from Iran were used as an independent testing set. Four maps were considered for each scan: axial map, anterior and posterior elevation map, and pachymetry map.

Results A CNN model detected keratoconus versus health eyes with an accuracy of 0.9785 on the testing set, considering all four maps concatenated. Considering each map independently, the accuracy was 0.9283 for axial map, 0.9642 for thickness map, 0.9642 for the front elevation map and 0.9749 for the back elevation map. The accuracy of models in recognising between healthy controls and stage 1 was 0.90, between stages 1 and 2 was 0.9032, and between stages 2 and 3 was 0.8537 using the concatenated map.

Conclusion CNN provides excellent detection performance for keratoconus and accurately grades different severities of disease using the colour-coded maps obtained by the Scheimpflug camera. CNN has the potential to be further developed, validated and adopted for screening and management of keratoconus.

INTRODUCTION

Keratoconus is a non-inflammatory degeneration that leads to progressive corneal thinning, myopia, irregular astigmatism and scarring, resulting in debilitating disease that may significantly affect a patients' quality of life.¹ The early and accurate detection of keratoconus provides opportunities to reduce risk factors and offer treatments to potentially slow its progression.²

Detecting keratoconus in its early stages, however, can be difficult, as patients may have

Key message

What is already known about this subject?

- Early and accurate detection of keratoconus provides opportunities to address risk factors and offer treatments to potentially slow its progression.

What are the new findings?

- The proposed method can automatically analyse Scheimpflug tomography scans and accurately stage the severity of keratoconus.

How might these results change the focus of research or clinical practice?

- The results show that automatic analysis has the potential to allow for a faster and more accurate evaluation of keratoconus.

normal visual acuity and an unremarkable slit-lamp examination. The diagnosis often requires an assessment of the topography and/or tomography of the cornea to reveal subtle changes in corneal morphology.^{3–6} The evaluation of corneal topography and tomography may itself be challenging and open to misinterpretation.⁵ Several indices have been proposed to facilitate differentiation between keratoconus and normal eyes, such as the zone of increasing corneal power, inferior–superior corneal power asymmetry, steepest radial axes, posterior and anterior ectasia and corneal pachymetry.^{7–9}

Utilisation of artificial intelligence (AI) to evaluate and detect early stages of various diseases^{10–14} at much faster rates than physicians has seen a sharp rise in recent years.^{15–17} In particular, AI techniques developed from artificial neural networks, deep learning (DL)-based algorithms show great promise in extracting features and learning patterns from complex data.¹⁸ Indeed, in eyecare, DL techniques have helped physicians in detection of anatomical structures and/or lesions,^{19–22} diagnosis of various diseases²³ and provided prognostic insights.^{24 25} Most studies are, however, primarily focused on



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Vito Romano; Vito.Romano@liverpool.ac.uk

relatively common diseases such as diabetic retinopathy, age-related macular degeneration and glaucoma.²⁶

The sensitivity and specificity of machine learning for the detection of keratoconus has been evaluated in several studies.^{27–40} Among current neural networks, the one which may be more suitable for the evaluation of the keratoconus is the convolutional neural networks technique (CNN), which is one of the main methods of recognising and classifying images through their colour-code pattern.⁴¹ CNN, therefore, can be applied to corneal topographic colour maps. Four studies have applied CNNs to patients with keratoconus,^{27–29 39} but with relatively small numbers, limiting its potential application. We, therefore, aimed to develop a CNN model to automatically detect keratoconus using standard colour-coded corneal maps and to provide a code that could be used in the clinic. This would be beneficial as colour-coded maps provide large amounts of information and clinicians are intrinsically more familiar with interpreting colour-coded maps in comparison to complex topographic indexes.

MATERIAL AND METHODS

Study population

We included tomographic images of keratoconic and healthy eyes provided by three centres: The Royal Liverpool University Hospital (UK); Sedaghat Eye Clinic, Mashhad (Iran) and The New Zealand National Eye Center (New Zealand). Patients with all stages of keratoconus were included. Images were obtained between January 2013 and January 2020. Scans were obtained using the standard 25 scans setting without pupil dilation, under scotopic conditions by trained technicians using the Pentacam HR (Oculus Optikgeräte, Wetzlar, Germany). Patients were asked to fixate on the target light and asked to blink completely just before each measurement to allow for adequate tear film coverage over the corneal surface. The examiner checked each scan and its quality before recording it and only scans of acceptable quality, reporting OK at quality score section, were included. For the control group subjects with a Belin/Ambrósio Enhanced Ectasia total deviation index (BAD-D) of less than 1.6 SD from normative values (indicating the absence of ectasia) were included.^{42 43}

Four different maps were investigated: axial, anterior elevation, posterior elevation and pachymetry. All of the four different maps used the absolute or standard colour scale, meaning that the maps display a fixed range of value selected in the settings of the tomographer, regardless of the map selected. All scans were categorised into four stages (online supplemental figure 1) according to Amsler-Krumeich grading.^{44 45}

Five different classification tasks were considered: (1) healthy and keratoconus, (2) healthy versus stage 1, (3) stage 1 vs 2, (4) Stage 2 vs 3 and (5) 5-class classification between healthy and each stage of keratoconus. In order to tackle the problem of an unbalanced label, and accurately classify healthy and keratoconic eyes, the weights

of each class were balanced by using the inverse class frequency for the training. For example, the weights for healthy eyes (n=82) and keratoconus eyes (n=1033) were set as $(82+1033)/1033$ and $(82+1033)/82$, respectively.

Classification models

In this work, different classification strategies were developed. Considering that each of the four maps contained numerous parameters that could be used to detect keratoconus and in order to preserve the maximum possible level of information for the prediction, we trained four models with one for each type of map images mentioned above, and a fifth model which used a concatenation of 4 map images (axial, corneal thickness, front and back elevation) as input. During the prediction, the output of each CNN model was the predicted class label of the input map images so that each of the trained models would give a predicted class label based on its own opinion.

We also made a sixth model by using the majority voting strategy from the above mentioned five models for each given task. For this model, a final prediction was made on the class of the input samples through a voting strategy as each of the five models will produce a prediction. More specifically, for each image sample the categorical labels from the above five different models were pooled together the argument of the maxima (abbreviated argmax) operation was applied to determine the best prediction (online supplemental figure 2). Without first performing the argmax operation, the predicted results would be directly influenced by the addition of the probability distribution.

Data preparation

In this work, the whole Liverpool (UK) and New Zealand (NZ) datasets were randomly split into training (80%) and testing (20%) images (table 1). Twenty per cent of the training set were randomly selected as the validation set during the model training process. The entirety of the Mashhad (IRAN) dataset was kept as an independent testing set and not involved at any stage of the training phase.

Maps from the right eye were flipped along the centre vertical axis. Maps from the left eye remained unchanged. Four individual maps were cropped using an in-house programme in Matlab version 2019b (Mathworks,

Table 1 Summary of the training and testing datasets made from the whole Liverpool (UK)+New Zealand (NZ) dataset and the IRAN dataset, only used as validation set

Class name	Training	Testing	Validation
Healthy (n=134)	82	20	32
Stage 1 (n=282)	159	40	83
Stage 2 (n=425)	211	53	161
Stage 3 (n=208)	115	29	64
Stage 4 (n=877)	548	137	192
Total (n=1926)	1115	279	532

Natick, USA). In brief, all of the white pads around the four individual maps and the colour bars were removed respectively and then the four cropped individual maps were resized into 224 by 224 as the final inputs for the model learning. For the purpose of learning, the four individual maps together with the four resized individual maps were concatenated back in the original order.

Neural network architecture

We adopted a VGG16 model⁴⁶ for this study, but in order to prevent overfitting, the number of connected weights of the top layer were reduced. Everything before the flattened layer was the same as with the standard VGG16. After the flattened layer for the 2-class task and in order to fully connect the (FC) layer with 128 outputs with a rectified linear unit (ReLU), an FC layer with 64 outputs with ReLU and an FC layer with two outputs (the final output layer representing two classes) was used with a SoftMax activation function. After the flattened layer for the 5-class task, the order was an FC layer with 128 outputs, a ReLU dropout layer with probability 0.5, FC layer with 64 outputs with ReLU dropout layer with probability 0.5, FC layer with five outputs (the final output layer representing five classes) were used with a softmax activation function. Please see online supplemental figure 2 for more information.

Visualisation

Saliency maps⁴⁶ and Gradient-weighted Class Activation Mapping (Grad-CAM)⁴⁷ were developed for visualisation of the learning, because the outputs of saliency maps and Grad-CAM were the same shape as the input image and could provide some intuition of attention. For saliency maps, the gradient of output with respect to input image from our models were computed. These gradients can highlight certain regions which contribute the most. Higher gradient values mean more contribution towards the output. In Grad-CAM, the computed gradients of output with respect to input image, produce a coarse localisation map or heat map highlighting the important regions in the image to show where the network has been focusing on. This provides a bit more explanation on the black-box characteristic of DL.

Implementation

We implemented our classification models by using Python V.3.7, Keras V.2.3.1 and Tensorflow V.1.14 as back-ends. All of the models were trained by using the Adam optimiser. We searched the optimal learning rate from the set of (10⁻³, 10⁻⁴, 10⁻⁵, 10⁻⁶) for each model based on the validation set. Binary cross-entropy was the loss function used for the 2-class classification models, and the categorical cross-entropy loss function for the 5-class classification models. The batch size was set to 20. All the training processes were performed on a TESLA V100 GPUs and all the testing experiments were conducted using a 2080Ti GPU.

Evaluation metrics

In this work, accuracy, sensitivity, specificity and area under the receiver operating characteristic curve (AUC) were used. In brief, Accuracy (ACC)=(TP+TN)/(TP+TN + FP+FN); Sensitivity: TP/(TP+FN); Specificity: TN/(TN+FP), where TP is a true positive, FP a false positive, TN a true negative and FN a false negative. AUC is valued between 0 and 1, where 1 means perfect classification. AUC will be more informative when the number of samples in different classes are imbalanced. De Long's method was introduced to construct confidence intervals for accuracy, sensitivity, specificity and AUC, by bootstrapping with 2000 samples to calculate 95% CI.⁴⁸

RESULTS

A total of 1926 corneal tomography scans were obtained comprising of 134 healthy controls and 1702 patients with keratoconus. A total of 1394 scans were collected from UK/NZ including 102 healthy, 199 stage 1, 264 stage 2, 144 stage 3 and 685 stage 4. An additional 532 keratoconus scans (32 healthy, 83 stage 1, 161 stage 2, 64 stage 3, and 192 stage 4) collected from IRAN between February 2017 to January 2020 were used as the validation set.

We will first present the results on the testing datasets of the UK/NZ datasets for the five different classification tasks, and then present the results on the IRAN dataset as the external validation set. Due to the imbalance of the classification problems, we will focus on the AUC results.

Healthy (n=20) vs keratoconus (n=259)

All the models using the thickness map, front elevation map, the back elevation map, the concatenated four maps and majority voting performed well, all with AUC higher than 0.80 (table 2). The model using the concatenated four maps as the input achieved the best AUC of 0.9423 (95% CI 0.8773 to 0.9942), followed by the model using the back elevation map (0.9173, 95% CI 0.8453 to 0.9793) and the majority voting model (0.8942, 95% CI 0.8115 to 0.9641).

Healthy (n=20) vs stage 1 (n=40)

The majority voting model achieved the best AUC of 0.90 (95% CI 0.8242 to 0.9653), followed by the model using concatenated 4 maps 0.8875 (95% CI 0.8103 to 0.9556) (table 3).

Stage 1 (n=40) vs stage 2 (n=53)

All models except the axial and front elevation maps performed well with AUCs higher than 0.88. The model using the back elevation map had the best performance with an AUC of 0.9153 (95% CI 0.8618 to 0.9592), followed by the majority voting model 0.9092 (95% CI 0.8557 to 0.9569). The difference, however, was marginal (table 4).

**Table 2** Classification results of health volunteers (n=20) and patients (n=259) with various stages of keratoconus on the testing sets

Healthy versus rest	Accuracy	Sensitivity	Specificity	AUC
Axial map	0.9283 (0.9032 to 0.9534)	1.0 (1.0 to 1.0)	0.0 (0.0 to 0.0)	0.5 (0.5 to 0.5)
Thickness map	0.9642 (0.9462 to 0.9821)	0.9768 (0.9612 to 0.9922)	0.8 (0.6316 to 0.9412)	0.8884 (0.8046 to 0.9583)
Front elevation map	0.9642 (0.9427 to 0.9821)	0.9884 (0.9766 to 1.0)	0.65 (0.4706 to 0.8261)	0.8192 (0.7292 to 0.9080)
Back elevation map	0.9749 (0.9570 to 0.9892)	0.9846 (0.9695 to 0.9962)	0.85 (0.7083 to 0.9667)	0.9173 (0.8453 to 0.9793)
Concatenated	0.9785 (0.9642 to 0.9928)	0.9846 (0.9699 to 0.9962)	0.9 (0.7692 to 1.0)	0.9423 (0.8773 to 0.9942)
Majority voting	0.9749 (0.9570 to 0.9892)	0.9884 (0.9767 to 1.0)	0.8 (0.6364 to 0.9375)	0.8942 (0.8115 to 0.9641)

95% CI in brackets.

AUC, area under the curve.

Stage 2 (n=53) vs stage 3 (n=29)

The model using the concatenated four maps achieved the best AUC of 0.8165 (95% CI 0.7398 to 0.8908) (table 5).

Five-class classification

The majority voting model achieved the best AUC of 0.888 (95% CI 0.8677 to 0.907), followed by the concatenated 4 maps model 0.8835 (95% CI 0.8641 to 0.9029) (online supplemental table 1). The confusion matrices of the above models are shown in online supplemental table 2.

From the above results of different classification tasks on the testing sets, it is apparent that the top three models are always concatenated map model, major voting model and the model using back elevation maps. The differences in performance between the aforementioned models, however, were marginal. The concatenated map model performed consistently across all classification tasks when compared with the voting model but importantly

required almost 80% less computational resources. We, therefore, selected only the model using concatenated four maps for the validation testing of the IRAN dataset.

Results on the external IRAN validation dataset

In the IRAN dataset, there were 532 images comprising: Healthy (n=32); stage 1 (n=83); stage 2 (n=161); stage 3 (n=64) and stage 4 (n=192). As all the images were from patients classified as keratoconus, we were not able to evaluate the sensitivity and specificity of this model on this dataset.

The AUC was 0.9737 (95% CI 0.9653 to 0.9821) for the classification of healthy (n=32) vs keratoconus (n=500). The AUC was 0.7304 (95% CI 0.6817 to 0.7783) for the classification of healthy (n=32) vs stage 1 (n=83). The AUC was 0.8988 (95% CI 0.8649 to 0.9305) for the classification of stage 1 (n=83) vs stage 2 (n=161). The AUC was 0.8285 (95% CI 0.7793 to 0.8754) for the classification of stage 2 (n=161) vs stage 3 (n=64). The AUC was 0.859 (95% CI 0.8438 to 0.8749) for the 5-class classification.

Table 3 Classification results of health volunteers (n=20) and patients with stages 1 keratoconus (n=40) on the testing sets

Healthy versus stage 1	Accuracy	Sensitivity	Specificity	AUC
Axial map	0.6667 (0.5667 to 0.7667)	1.0 (1.0 to 1.0)	0.0 (0.0 to 0.0)	0.5 (0.5 to 0.5)
Thickness map	0.8667 (0.8 to 0.9333)	0.875 (0.7857 to 0.9524)	0.8500 (0.7143 to 0.9630)	0.8625 (0.7792 to 0.9375)
Front elevation map	0.8000 (0.7167 to 0.8833)	0.8750 (0.7838 to 0.9535)	0.6500 (0.4583 to 0.8235)	0.7625 (0.6584 to 0.8562)
Back elevation map	0.8500 (0.7667 to 0.9167)	0.8750 (0.7838 to 0.9524)	0.8 (0.6471 to 0.9412)	0.8375 (0.7457 to 0.9167)
Concatenated	0.9 (0.8333 to 0.9500)	0.9245 (0.8537 to 0.9783)	0.85 (0.7143 to 0.96)	0.8875 (0.8103 to 0.9556)
Majority voting	0.9167 (0.8500 to 0.9667)	0.9500 (0.8864 to 1.0)	0.8500 (0.7143 to 0.9600)	0.9000 (0.8242 to 0.9653)

95% CI in brackets.

AUC, area under the curve.

Table 4 Classification results of patients with stages 1 (n=40) and 2 keratoconus (n=53) on the testing sets

Stage 1 versus stage 2	Accuracy	Sensitivity	Specificity	AUC
Axial map	0.5699 (0.4839 to 0.6559)	1.0 (1.0 to 1.0)	0.0 (0.0 to 0.0)	5.0 (5.0 to 5.0)
Thickness map	0.8925 (0.8387 to 0.9462)	0.9245 (0.8571 to 0.9808)	0.85 (0.75 to 0.9429)	0.8873 (0.8281 to 0.9383)
Front elevation map	0.7957 (0.7204 to 0.8602)	0.8302 (0.74 to 0.913)	0.75 (0.6316 to 0.8571)	0.7901 (0.7178 to 0.8592)
Back elevation map	0.914 (0.8602 to 0.9570)	0.9057 (0.8276 to 0.9636)	0.925 (0.8462 to 0.9796)	0.9153 (0.8618 to 0.9592)
Concatenated	0.9032 (0.8495 to 0.9462)	0.9245 (0.86 to 0.9811)	0.875 (0.7805 to 0.9535)	0.8998 (0.8423 to 0.9505)
Majority voting	0.914 (0.8602 to 0.957)	0.9434 (0.8837 to 1.0)	0.875 (0.7812 to 0.9535)	0.9092 (0.8557 to 0.9569)

95% CI in brackets.
AUC, area under the curve.

The confusion matrices of the above models are shown in online supplemental table 2.

Visualisation of the learnt features

We also studied the area of interest (focus) of the models during its learning process. Saliency maps and Grad-CAM were developed to obtain heat maps to show where the network highlighted the important regions in the image. The following are some good examples (online supplemental figure 3).

DISCUSSION

Our findings suggest that accurate automated detection of keratoconus and its evolution are possible using a CNN. When provided with all four maps (axial, corneal thickness, front and back elevation) the model is automatically able to discern between keratoconus and healthy eyes with an accuracy of 99.07% (97.57%–99.43%). Moreover, it can detect different stages of keratoconus with an accuracy of 93.12% (86.75%–93.98%). We postulate that the CNN model has better accuracy than the current gold

standard of human interpretation. The use of Amsler-Krumeich classification of keratoconus in four stages is related to its worldwide use in daily practice.

We believe that DL will aid in screening and in staging of keratoconus in a clinical setting, because the precise detection of early keratoconus is still challenging in daily practice.

The current gold standard of keratoconus diagnosis and staging relies on human interpretation of biomicroscopy features and tomography scans, and has previously been shown to be limited by poor reproducibility.^{5 49} The most commonly used parameter to determine keratoconus progression has been maximum keratometry (Kmax).^{50–52} Kmax is a single point reading representing the maximum curvature typically taken from the axial or sagittal anterior corneal curvature map. Kmax has numerous limitations as a single point reading is a poor descriptor of the cone morphology, a change in cone morphology (eg, a nipple cone progressing to a globular cone) can sometimes be associated with a reduction in

Table 5 Classification results of patients with stages 2 (n=53) and 3 keratoconus (n=29) on the testing sets

Stage 2 versus stage 3	Accuracy	Sensitivity	Specificity	AUC
Axial map	0.6463 (0.561 to 0.7317)	1.0 (1.0 to 1.0)	0.0 (0.0 to 0.0)	0.5 (0.5 to 0.5)
Thickness map	0.7683 (0.6951 to 0.8415)	0.4138 (0.2727 to 0.5769)	0.9623 (0.9184 to 1.0)	0.6880 (0.6151 to 0.7708)
Front elevation map	0.7561 (0.6707 to 0.8293)	0.4483 (0.2903 to 0.6071)	0.9245 (0.8571 to 0.9808)	0.6864 (0.5977 to 0.7712)
Back elevation map	0.7927 (0.7195 to 0.8659)	0.5172 (0.3571 to 0.6774)	0.9434 (0.8846 to 0.9825)	0.7303 (0.6463 to 0.8128)
Concatenated	0.8537 (0.7927 to 0.9146)	0.6897 (0.5455 to 0.8333)	0.9434 (0.8846 to 0.9836)	0.8165 (0.7398 to 0.8908)
Majority voting	0.7317 (0.6585 to 0.8049)	0.2759 (0.1389 to 0.4194)	0.9811 (0.9444 to 1.0)	0.6285 (0.5577 to 0.7012)

95% CI in brackets.
AUC, area under the curve.

Kmax, single point readings tend to have poor reproducibility, changes in Kmax do not correlate to changes in visual function and Kmax is limited to the anterior corneal surface, ignoring the posterior cornea, thereby having no ability to detect early or subclinical disease or early disease progression.^{53–57}

The ability of an algorithm to detect keratoconus is based on an operationalisable distinction between a normal and ectatic cornea. A Global Consensus Panel in 2015 was able to agree on a definition of keratoconus in terms of abnormal posterior ectasia, abnormal corneal thickness distribution and clinical non-inflammatory corneal thinning.⁴⁵ Various indices have been evaluated with regard to their ability to discriminate an ectatic from a normal cornea. Among these, the Smolek/Klyce and the Klyce/Maeda (KCI) have been shown to possess a good specificity and sensitivity in distinguishing between keratoconus and healthy eyes.⁴² The Tomographic and Biomechanical Index uses AI to combine Scheimpflug tomography and corneal biomechanical parameters to optimise ectasia detection with good sensitivity and specificity.⁵⁸ This index has been shown to be even more accurate than Corneal Biomechanical Index.⁵⁹

Regarding the detection of keratoconus progression, however, the Global Consensus Panel noted that specific quantitative data were lacking and, moreover, would most likely be device specific. Determinants for assessing keratoconus progression have been reviewed by Duncan *et al.*⁶⁰ They concluded that this multitude of suggested progression parameters highlights the need for a new or standardised method to document progression.^{61–65}

The use of colour-coded maps for DL provides more complete information with the global status of the cornea, instead of using topographic and/or tomographic numeric indices as done in the past.^{32 34 35 37 38 40 66–69} Numeric values can easily exemplify corneal shape, but they fail to represent the spatial gradients and distributions of corneal curvature, elevation, refractive power and thickness.

However, is not possible to demine the superiority of the CNN over a single numerical index, as the Kmax, in view of the overall learning process, which required four maps: axial, anterior elevation, posterior elevation and pachymetry. Furthermore, for the evaluation of keratoconus, a single parameter is not sufficient.⁴⁵

In this study, the images of four colour-coded maps (axial, corneal thickness, front and back elevation) were used for DL, instead of topographic and tomographic numeric indices. The reason of our choice is based on the capability of colour-coded maps to hold a larger amount of corneal information than these numeric values for this learning. The maps were obtained via tomography Scheimpflug imaging which has an advantage over Placido disk-based corneal topography as it is able to evaluate both the anterior and posterior surfaces of the cornea. Evaluation of the posterior corneal surface is essential as both curvature and elevation of the posterior corneal surface have to be considered for the detection of early keratoconus.^{70–72}

A multiplicity of machine-learning techniques such as neural network, support vector machine, decision tree, unsupervised machine learning, custom neural network, feedforward neural network and CNN have been used in previous studies but only in four studies has a combination of colour-coded maps and CNN been used used.^{27–29 39} The reason for opting for CNN over other machine learning methods in this study was based on the ability of the CNN to directly extract the morphological characteristics from the obtained images without preliminary learning, subsequently providing higher classification precision, especially in the field of image recognition. This study is different from the aforementioned ones as Lavric and Valentin²⁸ did not use real clinical data, Zéboulon *et al.*²⁷ focused on refractive surgery; Kamiya *et al.*³⁹ used a no tomography device (AS-OCT), and Kuo *et al.*²⁹ had a smaller sample size.

Our study has a number of limitations. First, the number of eyes is still modest in nature and there are only a small number of eyes in some specific groups and this may create a classification model bias. A potential strategy for this is to use generative adversarial networks to synthesise images from a small number of real images but such models would require further external validation. Second, other risk factors of keratoconus were not included in the used prediction models and such factors will be useful for further refinement of the prediction performance. Future studies incorporating such risk factors such as family history, atopy or ethnicity,^{73–75} may improve the overall function of the model.

Moreover, given the multicentre nature of the study a number of technicians were used to obtain the scans but the Scheimpflug tomographer has previously been reported to have good intra and interobserver repeatability in healthy patients⁷⁶ and those with keratoconus.⁷⁷ Finally, the adoption of more recent CNN models and tricks (eg, attention and customised loss functions) can potentially further enhance the performance of the model.

In summary, our results demonstrate that AI models provide excellent detection performance for keratoconus and can accurately grade different severities of disease and therefore have the potential to be further developed, validated and adopted for screening and management of keratoconus. Clinical implication of automated detection and screening are of considerable importance in view to their ability to provide diagnosis in shorter time, increasing in this way the patient care. Indeed, it could be deployed particularly in regions with a high burden of disease so that the CNN model may have the potential to provide earlier diagnosis of keratoconus, improve access to treatments such as corneal cross-linking and potentially reduce preventable visual loss. A larger external validation study with another study population including healthy controls is required to confirm this study's preliminary findings.

Author affiliations

¹Department of Eye and Vision Science, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK

²Department of Ophthalmology, St Paul's Eye Unit, Royal Liverpool University Hospital, Liverpool, UK

³Department of Ophthalmology, New Zealand National Eye Centre, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

⁴Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China

⁵Eye Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

⁶Health Promotion Research Center, Zahedan University of Medical Sciences, Zahedan, Iran

Twitter Vito Romano @DrVitoRomano

Contributors VR, YZ and SK: conceptualisation, methodology, investigation, visualisation, review and editing. XC and JZ: analysis, data curation, writing. KCI, DB, DR, AG, CNJM, YZ, M-RS, HM-M and MZ: data collection, writing

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study received Institutional Review Board approval (A02719) and was conducted accordingly to the ethical standards set in the 1964 Declaration of Helsinki, as revised in 2000.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Davide Borroni <http://orcid.org/0000-0001-6952-5647>

Vito Romano <http://orcid.org/0000-0002-5148-7643>

REFERENCES

- Rabinowitz YS. Keratoconus. *Surv Ophthalmol* 1998;42:297–319.
- Sorkin N, Varssano D. Corneal collagen crosslinking: a systematic review. *Ophthalmologica* 2014;232:10–27.
- Mohammadpour M, Heidari Z, Hashemi H. Updates on managements for keratoconus. *J Curr Ophthalmol* 2018;30:110–24.
- Rocha-de-Lossada C, Prieto-Godoy M, Sánchez-González José-María, et al. Tomographic and aberrometric assessment of first-time diagnosed paediatric keratoconus based on age ranges: a multicentre study. *Acta Ophthalmol* 2020;91.
- Brunner M, Czanner G, Vinciguerra R, et al. Improving precision for detecting change in the shape of the cornea in patients with keratoconus. *Sci Rep* 2018;8.
- Ambrósio R, Randleman JB. Screening for ectasia risk: what are we screening for and how should we screen for it? *J Refract Surg* 2013;29:230–2.
- Holladay JT. Keratoconus detection using corneal topography. *J Refract Surg* 2009;25:S958–62.
- Prakash G, Agarwal A, Mazhari AI, et al. A new, pachymetry-based approach for diagnostic cutoffs for normal, suspect and keratoconic cornea. *Eye* 2012;26:650–7.
- Dumitrica DM, Colin J. Indices for the detection of keratoconus. *Ophthalmologia* 2010;54:19–29.
- Katzen J, Dodelzon K. A review of computer aided detection in mammography. *Clin Imaging* 2018;52:305–9.
- Watanabe AT, Lim V, Vu HX, et al. Improved cancer detection using artificial intelligence: a retrospective evaluation of missed cancers on mammography. *J Digit Imaging* 2019;32:625–37.
- Johansson G, Olsson C, Smith F, et al. AI-aided detection of malignant lesions in mammography screening - evaluation of a program in clinical practice. *BJR Open* 2021;3:20200063.
- Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. *Eur Radiol* 2019;29:4825–32.
- Cui M, Zhang DY. Artificial intelligence and computational pathology. *Lab Invest* 2021;101:412–22.
- Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ* 2019;7:e7702.
- Dong H, Yang G, Liu F. Automatic brain tumor detection and segmentation using U-Net based fully Convolutional networks. In: *Communications in computer and information science*. , 2017: 723, 506–17.
- Hu S, Gao Y, Niu Z. Special section on emerging deep learning theories and methods for biomedical engineering weakly supervised deep learning for COVID-19 infection detection and classification from CT images.
- Cao C, Liu F, Tan H, et al. Deep learning and its applications in biomedicine. *Genomics Proteomics Bioinformatics* 2018;16:17–32.
- Williams BM, Borroni D, Liu R, et al. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. *Diabetologia* 2020;63:419–30.
- Ma Y, Hao H, Xie J. Rose: a retinal OCT-Angiography vessel segmentation dataset and new model. *IEEE Transactions on Medical Imaging* 2020;40.
- Yang M, Xiao X, Liu Z, et al. Deep RetinaNet for dynamic left ventricle detection in Multiview echocardiography classification. *Sci Program* 2020;2020:1–6.
- Ali A-R, Li J, Kanwal S, et al. A novel fuzzy multilayer Perceptron (F-MLP) for the detection of irregularity in skin lesion border using Dermoscopic images. *Front Med* 2020;7:297.
- Abramoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
- Bridge J, Harding S, Zheng Y. Development and validation of a novel prognostic model for predicting AMD progression using longitudinal fundus images. *BMJ Open Ophthalmol* 2020;5:e000569.
- Yan Q, Weeks DE, Xin H, et al. Deep-learning-based prediction of late age-related macular degeneration progression. *Nat Mach Intell* 2020;2:141–50.
- Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res* 2019;72:100759.
- Zéboulon P, Debellemanière G, Bouvet M, et al. Corneal topography RAW data classification using a Convolutional neural network. *Am J Ophthalmol* 2020;219:33–9.
- Lavric A, Valentin P, KeratoDetect: keratoconus detection algorithm using Convolutional neural networks. *Comput Intell Neurosci* 2019;2019:1–9.
- Kuo B-I, Chang W-Y, Liao T-S. Special issue keratoconus screening based on deep learning approach of corneal topography 2020.
- Dos Santos VA, Schmetterer L, Stegmann H, et al. CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning. *Biomed Opt Express* 2019;10:622–41.
- Issarti I, Consejo A, Jiménez-García M, et al. Computer aided diagnosis for suspect keratoconus detection. *Comput Biol Med* 2019;109:33–42.
- Maeda N, Klyce SD, Smolek MK, et al. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci* 1994;35:2749–57.
- Ruiz Hidalgo I, Rozema JJ, Saad A, et al. Validation of an objective keratoconus detection system implemented in a scheinpluf Tomographer and comparison with other methods. *Cornea* 2017;36:689–95.
- Smadja D, Touboul D, Cohen A, et al. Detection of subclinical keratoconus using an automated decision tree classification. *Am J Ophthalmol* 2013;156:237–46.
- Kovács I, Miháltz K, Kránitz K, et al. Accuracy of machine learning classifiers using bilateral data from a scheinpluf camera for identifying eyes with preclinical signs of keratoconus. *J Cataract Refract Surg* 2016;42:275–83.
- Ruiz Hidalgo I, Rodriguez P, Rozema JJ, et al. Evaluation of a Machine-Learning classifier for keratoconus detection based on scheinpluf tomography. *Cornea* 2016;35:827–32.
- Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. *Invest Ophthalmol Vis Sci* 1995;36:1327–35.

- 38 Souza MB, Medeiros FW, Souza DB, *et al.* Evaluation of machine learning classifiers in keratoconus detection from Orbscan II examinations. *Clinics*;65:1223–8.
- 39 Kamiya K, Ayatsuka Y, Kato Y, *et al.* Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study. *BMJ Open* 2019;9:31313.
- 40 Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Invest Ophthalmol Vis Sci* 1997;38:2290–9.
- 41 Shin H-C, Roth HR, Gao M, *et al.* Deep Convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35:1285–98.
- 42 Goebels S, Eppig T, Wagenpfeil S, *et al.* Staging of keratoconus indices regarding tomography, topography, and biomechanical measurements. *Am J Ophthalmol* 2015;159:733–8.
- 43 Villavicencio OF, Gilani F, Henriquez MA, *et al.* Independent population validation of the Belin/Ambrósio enhanced ectasia display: implications for keratoconus studies and screening. *Int J Keratoconus Ectatic Corneal Dis* 2014;3:1–8.
- 44 Amsler M. Kératocône classique et kératocône fruste; arguments unitaires. *Ophthalmologica* 1946;111:96–101.
- 45 Gomes JAP, Tan D, Rapuano CJ, *et al.* Global consensus on keratoconus and ectatic diseases. *Cornea* 2015;34:359–69.
- 46 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- 47 Selvaraju RR, Cogswell M, Das A. Grad-CAM: visual explanations from deep networks via Gradient-Based localization. *Proceedings of the IEEE International Conference on Computer Vision. Vol 2017-October. Institute of Electrical and Electronics Engineers Inc*, 2017:618–26.
- 48 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837.
- 49 Ziaei M, Barsam A, Shamie N, *et al.* Reshaping procedures for the surgical management of corneal ectasia. *J Cataract Refract Surg* 2015;41:842–72.
- 50 Wittig-Silva C, Chan E, Islam FMA, *et al.* A randomized, controlled trial of corneal collagen cross-linking in progressive keratoconus: three-year results. *Ophthalmology* 2014;121:812–21.
- 51 O'Brart DPS, Chan E, Samaras K, *et al.* A randomised, prospective study to investigate the efficacy of riboflavin/ultraviolet A (370 nm) corneal collagen cross-linkage to halt the progression of keratoconus. *Br J Ophthalmol* 2011;95:1519–24.
- 52 Sykakis EHS. Cochrane database of systematic reviews corneal collagen cross-linking for treating keratoconus (review) 2015.
- 53 Mahmoud AM, Nuñez MX, Blanco C, *et al.* Expanding the cone location and magnitude index to include corneal thickness and posterior surface information for the detection of keratoconus. *Am J Ophthalmol* 2013;156:1102–11.
- 54 de Sanctis U, Loiacono C, Richiardi L, *et al.* Sensitivity and specificity of posterior corneal elevation measured by Pentacam in discriminating Keratoconus/Subclinical keratoconus. *Ophthalmology* 2008;115:1534–9.
- 55 Tomidokoro A, Oshika T, Amano S, *et al.* Changes in anterior and posterior corneal curvatures in keratoconus. *Ophthalmology* 2000;107:1328–32.
- 56 Lopes BT, Ramos IC, Faria-Correia F. Correlation of Topometric and tomographic indices with visual acuity in patients with keratoconus. *Int J Keratoconus Ectatic Corneal Dis*;1:167–72.
- 57 Guber I, McAlinden C, Majo F, *et al.* Identifying more reliable parameters for the detection of change during the follow-up of mild to moderate keratoconus patients. *Eye Vis* 2017;4:24.
- 58 Ambrósio R, Lopes BT, Faria-Correia F, *et al.* Integration of Scheimpflug-based corneal tomography and biomechanical assessments for enhancing ectasia detection. *J Refract Surg* 2017;33:434–43.
- 59 Ferreira-Mendes J, Lopes BT, Faria-Correia F, *et al.* Enhanced ectasia detection using corneal tomography and biomechanics. *Am J Ophthalmol* 2019;197:7–16.
- 60 Duncan JK, Belin MW, Borgstrom M. Assessing progression of keratoconus: novel tomographic determinants. *Eye Vis* 2016;3.
- 61 Vinciguerra R, Romano V, Arbabi EM, *et al.* In vivo early corneal biomechanical changes after corneal cross-linking in patients with progressive keratoconus. *J Refract Surg* 2017;33:840–6.
- 62 Romano V, Vinciguerra R, Arbabi EM, *et al.* Progression of keratoconus in patients while awaiting corneal cross-linking: a prospective clinical study. *J Refract Surg* 2018;34:177–80.
- 63 Vinciguerra R, Tzamalís A, Romano V, *et al.* Assessment of the association between in vivo corneal biomechanical changes after corneal cross-linking and depth of demarcation line. *J Refract Surg* 2019;35:202–6.
- 64 Pagano L, Gadhvi KA, Borroni D, *et al.* Bilateral keratoconus progression: immediate versus delayed sequential bilateral corneal cross-linking. *J Refract Surg* 2020;36:552–6.
- 65 Shah H, Pagano L, Vakharia A, *et al.* Impact of COVID-19 on keratoconus patients waiting for corneal cross linking. *Eur J Ophthalmol* 2021;112067212110013.
- 66 Accardo PA, Pensiero S. Neural network-based system for early keratoconus detection from corneal topography. *J Biomed Inform* 2002;35:151–9.
- 67 Ruiz Hidalgo I, Rodriguez P, Rozema JJ, *et al.* Evaluation of a machine-learning classifier for keratoconus detection based on scheimpflug tomography. *Cornea* 2016;35:827–32.
- 68 Ruiz Hidalgo I, Rozema JJ, Saad A, *et al.* Validation of an objective keratoconus detection system implemented in a scheimpflug tomographer and comparison with other methods. *Cornea* 2017;36:689–95.
- 69 Arbelaez MC, Versaci F, Vestri G, *et al.* Use of a support vector machine for keratoconus and subclinical keratoconus detection by topographic and tomographic data. *Ophthalmology* 2012;119:2231–8.
- 70 Shi Y. Strategies for improving the early diagnosis of keratoconus. *Clinical Optometry* 2016;8:13–21.
- 71 Ishii R, Kamiya K, Igarashi A, *et al.* Correlation of corneal elevation with severity of keratoconus by means of anterior and posterior topographic analysis. *Cornea* 2012;31:253–8.
- 72 Kamiya K, Ishii R, Shimizu K, *et al.* Evaluation of corneal elevation, pachymetry and keratometry in keratoconic eyes with respect to the stage of Amsler-Krumeich classification. *Br J Ophthalmol* 2014;98:459–63.
- 73 Rahi A, Davies P, Ruben M, *et al.* Keratoconus and coexisting atopic disease. *Br J Ophthalmol* 1977;61:761–4.
- 74 Li X, Rabinowitz YS, Rasheed K, *et al.* Longitudinal study of the normal eyes in unilateral keratoconus patients. *Ophthalmology* 2004;111:440–6.
- 75 Goh YW, Gokul A, Yadegarfar ME, *et al.* Prospective clinical study of keratoconus progression in patients awaiting corneal cross-linking. *Cornea* 2020;39:1256–60.
- 76 McAlinden C, Khadka J, Pesudovs K. A comprehensive evaluation of the precision (repeatability and reproducibility) of the oculus Pentacam HR. *Invest Ophthalmol Vis Sci* 2011;52:7731–7.
- 77 Wonneberger W, Sterner B, MacLean U, *et al.* Repeated same-day versus single tomography measurements of Keratoconic eyes for analysis of disease progression. *Cornea* 2018;37:474–9.