tvst

Article

A Hybrid Deep Learning Construct for Detecting Keratoconus From Corneal Maps

Ali H. Al-Timemy[1,2], Zahraa M. Mosa[3], Zaid Alyasseri[4,5], Alexandru Lavric[6], Marcelo M. Lui[7], Rossen M. Hazarbassanov[7,8], and Siamak Yousefi[9,10]

[1] Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq
[2] Centre for Robotics and Neural Systems, Cognitive Institute, School of Engineering, Computing and Mathematics, Plymouth University, Plymouth, UK
[3] College of Pharmacy, Uruk University, Baghdad, Iraq
[4] Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
[5] ECE Department–Faculty of Engineering, University of Kufa, Najaf, Iraq
[6] Computers, Electronics and Automation Department, Stefan cel Mare University of Suceava, Suceava, Bukovina, Romania
[7] Hospital de Olhos–CRO, Guarulhos, São Paulo, São Paulo, Brazil
[8] Department of Ophthalmology and Visual Sciences, Paulista Medical School, Federal University of São Paulo, São Paulo, Brazil
[9] Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, TN, USA
[10] Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA

Correspondence: Siamak Yousefi, UT Hamilton Eye Institute, 930 Madison Avenue, Suite 726, Memphis, TN 38163, USA.
e-mail: siamak.yousefi@uthsc.edu
Rossen M. Hazarbassanov, Rua Botucatu, 806–Vila Clementino, São Paulo, São Paulo CEP 04023-062, Brazil.
e-mail: hazarbassanov@gmail.com

Received: May 18, 2021
Accepted: November 26, 2021
Published: December 16, 2021

Keywords: artificial intelligence; deep learning; corneal imaging; keratoconus; automated diagnosis

Citation: Al-Timemy AH, Mosa ZM, Alyasseri Z, Lavric A, Lui MM, Hazarbassanov RM, Yousefi S. A hybrid deep learning construct for detecting keratoconus from corneal maps. Transl Vis Sci Technol. 2021;10(14):16, https://doi.org/10.1167/tvst.10.14.16

Purpose: To develop and assess the accuracy of a hybrid deep learning construct for detecting keratoconus (KCN) based on corneal topographic maps.

Methods: We collected 3794 corneal images from 542 eyes of 280 subjects and developed seven deep learning models based on anterior and posterior eccentricity, anterior and posterior elevation, anterior and posterior sagittal curvature, and corneal thickness maps to extract deep corneal features. An independent subset with 1050 images collected from 150 eyes of 85 subjects from a separate center was used to validate models. We developed a hybrid deep learning model to detect KCN. We visualized deep features of corneal parameters to assess the quality of learning subjectively and computed area under the receiver operating characteristic curve (AUC), confusion matrices, accuracy, and F1 score to evaluate models objectively.

Results: In the development dataset, 204 eyes were normal, 123 eyes were suspected KCN, and 215 eyes had KCN. In the independent validation dataset, 50 eyes were normal, 50 eyes were suspected KCN, and 50 eyes were KCN. Images were annotated by three corneal specialists. The AUC of the models for the two-class and three-class problems based on the development set were 0.99 and 0.93, respectively.

Conclusions: The hybrid deep learning model achieved high accuracy in identifying KCN based on corneal maps and provided a time-efficient framework with low computational complexity.

Translational Relevance: Deep learning can detect KCN from non-invasive corneal images with high accuracy, suggesting potential application in research and clinical practice to identify KCN.

Introduction

Keratoconus (KCN) is a noninflammatory ectatic corneal disorder characterized by progressive thinning of the cornea that may lead to reduced vision or even vision loss.[1] Although detecting subclinical KCN is challenging, more advanced cases are easily diagnosed due to the presence of more obvious retinoscopic and biomicroscopic signs such as Munson's sign, Vogt's

striae, or Fleischer's ring[2,3] Some of the early methods for KCN diagnosis rely on subjective evaluation of topographical maps[4]; however, automated models may provide a more accurate and objective evaluation of KCN.

Conventional machine learning models, including neural networks, decision trees, and discriminant analysis, have been applied to corneal topography parameters for detecting KCN.[5–8] Some of the models have used data from the anterior topographic maps of cornea only, whereas follow-up models have used data from posterior cornea to evaluate KCN.[9–11] Machine learning models, however, can combine anterior and posterior corneal data in order to complement information regarding KCN and may lead to improved detection.

Current artificial intelligence (AI) models for the detection of KCN are mainly supervised[6–15] and have achieved area under the receiver operating characteristic curves (AUCs) generally in the range of 0.90 to 1.0 based on Pentacam (OCULUS, Wetzlar, Germany) indices or combined Pentacam and optical coherence tomography (OCT) parameters.[13,14] However, these studies have used datasets with varying number of samples and patients at different KCN stages, making generalization of the results challenging. Unsupervised machine learning models, on the other hand, require no prelabeled data and have been applied to corneal data to identify different severity levels of KCN[16,17] and to predict those individuals who may require more invasive corneal surgery by using topography, elevation, and pachymetry parameters of the cornea.[18]

Significantly fewer deep learning (DL) studies exist for the cornea[19–24] compared with other ocular domains.[25–42] Deep learning models, particularly deep convolutional neural networks (CNNs), have been applied to color-coded corneal maps such as elevation, curvature, and thickness to identify KCN. These studies have typically utilized small subsets of images with fewer than 400 images in spite of the fact that DL models usually require a greater number of samples.[20] Other studies have used simulated corneal images rather than real-world clinical corneal images.[21] As such, it is challenging to generalize their results. Zeboulon et al.[22] used a relatively large dataset with 3000 corneal maps to detect KCN and a history of refractive surgery. They achieved a high accuracy for discriminating KCN from normal, but the ability of the model to detect suspected KCN has not been examined. Additionally, as training and optimizing deep CNN models is typically computationally expensive, models that run faster, such as our currently proposed model, have greater potential to be integrated into clinical practice.

In this study, we utilized a relatively large dataset with 4844 corneal images to train and validate a hybrid deep learning model for detecting KCN. We developed seven deep CNN models based on the EfficientNet-b0 architecture such that each model becomes an expert in identifying KCN-induced features from that particular corneal map. We then developed a hybrid model that integrates deep corneal features extracted from each corneal map based on a support vector machine (SVM) to provide a DL model for the identification of KCN. To explain the outcome of the model and to subjectively evaluate the relevance of deep features and DL decisions, we developed *t*-distributed stochastic neighbor embedding (*t*-SNE) and visualized the outcome.

Methods

Datasets and Preprocessing

The protocol of the study (0094/2020) was approved by the Institutional Review Board of Federal University of São Paulo–UNIFESP/EPM as the coordinator center and Hospital de Olhos–CRO, Guarulhos, as the side center. Corresponding data use agreements were signed among parties to use the data. The study was conducted in accordance with the tenets of the Declaration of Helsinki. If required, respective informed consent was obtained from participants, and the data were de-identified in Brazil before any further processing.

Three corneal specialists (including RMH) conducted vision tests and ophthalmic examinations under standard conditions and collected corneal images using Scheimpflug imaging systems (Pentacam). Three corneal trained specialists performed the eye classification. We dealt with disagreements favoring two versus one vote. The clinicians were instructed to grade each eye as normal, suspected KCN, or KCN. Eyes were labeled as KCN suspect based on standard criteria in earlier studies.[43,44] More specifically, eyes were labeled as suspected KCN if the corneal topography included atypical, localized steepening or an asymmetrical bowtie pattern. Eyes were labeled as suspected KCN if the keratometric curvature was greater than 47.00 diopters (D), the oblique cylinder was more than 1.50 D, or the central corneal thickness (CT) was below 500 µm. Each eye of each patient was evaluated independently.

Furthermore, raw data of Belin/Ambrosio Enhanced Ectasia Display (BAD) values based on elevation and thickness maps including Pentacam-generated percentage thickness increase (PTI) and corneal-thickness spatial profile were used in this

study. Eyes were labeled as suspected KCN if the best fit sphere with the optical zone of 9.0 mm front elevation map was between 12 and 15 µm and the back elevation map was between 17 and 20 µm. Then, BAD indices, including standard deviations of front and back elevation maps, standard deviations of the PTI, and standard deviations of the minimal thickness, were evaluated. These indices compose the overall standard deviation, or what is referred to as the final D (BAD-D). Eyes were labeled as suspected KCN when BAD-D was between 1.6 and 3.0.

The development (training) dataset included corneal images collected using different Pentacam instruments with different settings (different color-scale steps of the maps compared with the previous subset; see Supplementary Material). All color scales were based on decimal-scale grading using microns for CT and elevation (EL) maps and diopters for axial/sagittal curvature maps. An additional independent dataset, collected from a different clinic in Brazil, was also used to validate the proposed hybrid DL approach. The image dataset of this study for seven corneal maps is being made available to the research community (https://drive.google.com/drive/folders/1GR9T-p7GWGY_0nI5sm8GdJ4V6qlV4vZ2?usp=sharing).

Input Image Preprocessing

Seven corneal maps including anterior and posterior eccentricity (EC), anterior and posterior EL, anterior and posterior sagittal (SAG) curvature, and CT were extracted from a Pentacam instrument and saved in JPEG format. These maps are typically examined in routine clinical settings for KCN assessment. Development set images (7 maps × 542 eyes = 3794 images), in addition to the independent test images (7 maps × 150 eyes), were resized to 224 × 224 × 3 pixels to match the input of the standard DL models.

Deep Learning Model

Figure 1 shows the block diagram of the AI construct that was developed to detect KCN from corneal images. We first developed seven models based on EfficientNet-b0,[45] a DL architecture, pretrained on ImageNet, to identify KCN-induced signs in each corneal map separately. EfficientNet-b0[45] is a newly developed efficient DL architecture with 290 layers. It has fewer parameters compared with AlexNet,[46] GoogleNet,[47] VGG19,[48] and Resnet-50,[49] and it has achieved greater accuracy when tested on ImageNet[50] datasets.

Each EfficientNet-b0 DL model extracted deep features from the last convolutional layers of the

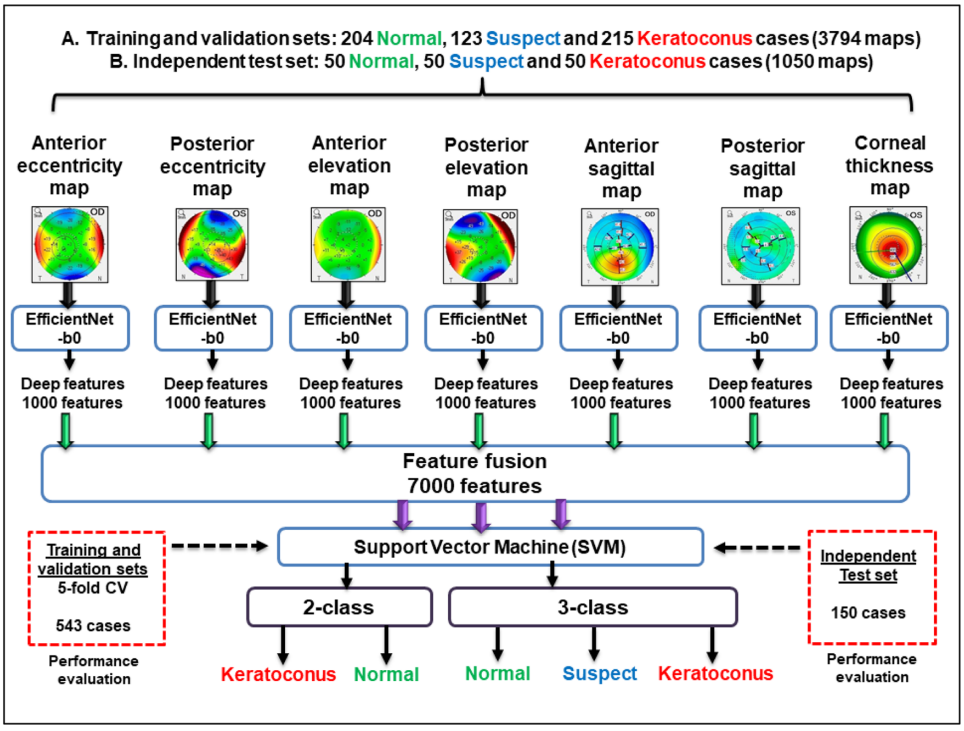

Figure 1. Diagram of the proposed hybrid DL construct for detecting eyes with suspected KCN, normal eyes, and eyes with KCN.

respective models of anterior and posterior EC, anterior and posterior EL, anterior and posterior SAG, and CT maps (Fig. 1).

The deep features were extracted from the fully connected layer (efficientnet-b0|model|head|dense|MatMul), layer number 288, which provides an output with 1000 features, without retraining the DL model, which was originally trained on ImageNet. The first and last layers of the EfficientNet-b0 are shown in Supplementary Figure S4.

To develop the hybrid model for integrating the information from each individual deep CNN model, we fused 1000 deep features from each map to generate a concatenated vector with 7000 deep features. This hybrid model was developed on the representation level rather than on the decision level of each model, as doing so generated higher performance. We then used an SVM classifier with quadratic kernel function and set the box constraint level to 1. To make multiclass feasible, we employed a one-versus-one classification in the three-class scenario. We estimated the classification metrics based on the SVM classifier.

Subjective Evaluations

To identify the effectiveness of the generated deep features, we applied *t*-SNE on deep features and visualized the results using two-dimensional (2D) plots for the two-class (normal vs. KCN) and three-class (normal vs. KCN vs. suspected) KCN detection problems. Visualization of deep features through 2D *t*-SNE plots improves our understanding of how deep feature extraction is relevant in identifying KCN.

Objective Evaluations

We employed AUCs to evaluate the model performance. We also computed the confusion matrices along with accuracy and F1 scores to compare the performance and quality of learning objectively. To evaluate models, we split the dataset into 80% for model development and 20% for validation. To reduce possible data selection bias, we repeated this process five times each time the model was evaluated based on randomly selecting the split of 80%/20%. An independent test set with 150 eyes (50 eyes from each group) was also utilized to validate the DL models. We also estimated the time for training and validation. We computed the time for the entire pipeline to extract features and train the SVM classifier on a machine with a Core i5 central processing unit (CPU; Intel Corporation, Santa Clara, CA) and 16 GB of RAM equipped with MATLAB 2020b (MathWorks, Natick, MA). We estimated the time of testing a new sample on this framework, as well. Classification metrics such as accuracy or F1 score were estimated with SVM classifiers for all classes and networks investigated in this study. We also included another testing scheme in which we merged 542 eyes in the development subset with 150 eyes in the independent subset, then randomly divided the entire set into a 50% training set, a 25% validation set, and a 25% test set. We re-evaluated the accuracy based on these sets to ensure that the model is stable.

Results

A total of 204 eyes of 104 patients were normal, 215 eyes of 113 patients had KCN, and 123 eyes of 63 patients were suspected KCN. The mean ages ± SD of the subjects in the normal, KCN, and suspected KCN groups were 33.4 ±10.1 years, 29.0 ± 9.3 years, and 28.6 ± 9.4 years, respectively. Images from 56 normal eyes and 58 eyes with KCN were collected using a Pentacam instrument with settings different from others (Fig. 2; samples in the dashed green lines); see the Supplementary Materials for more examples of different color scales.

The independent validation subset included 150 eyes of 85 patients collected from Hospital de Olhos–CRO, a private hospital located in São Paulo, Brazil. This dataset included 50 normal eyes from 29 subjects, 50 KCN eyes from 31 patients, and 50 suspect KCN eyes from 25 patients. The mean ages ± SD of the subjects in the normal, KCN, and suspected KCN groups were 29.5 ± 4.7 years, 26.3 ± 6.8 years, and 29.1 ± 5.3 years, respectively. The *t*-SNE plot of the entire dataset of 692 eyes (development and test sets) for the three-class case is shown in the lower panel of Figure 2.

We extracted 1000 deep features from the last fully connected layer of each of the seven deep learning models (Fig. 1). Figure 2, top left and top right, shows the *t*-SNE plots of 7000 fused deep features extracted from seven maps of eyes in the development dataset for the two-class (419 eyes) and three-class (542 eyes) scenarios, respectively. Figure 3, left and right, presents the *t*-SNE plots of 7000 fused deep features extracted from seven maps of eyes in the independent validation dataset for the two-class (100 eyes) and three-class (150 eyes) problems, respectively.

The receiver operating characteristic (ROC) curves of the proposed model for the two-class and three-class problems based on the development dataset are presented in Figure 4. The accuracy of the proposed model in detecting KCN from the seven corneal maps

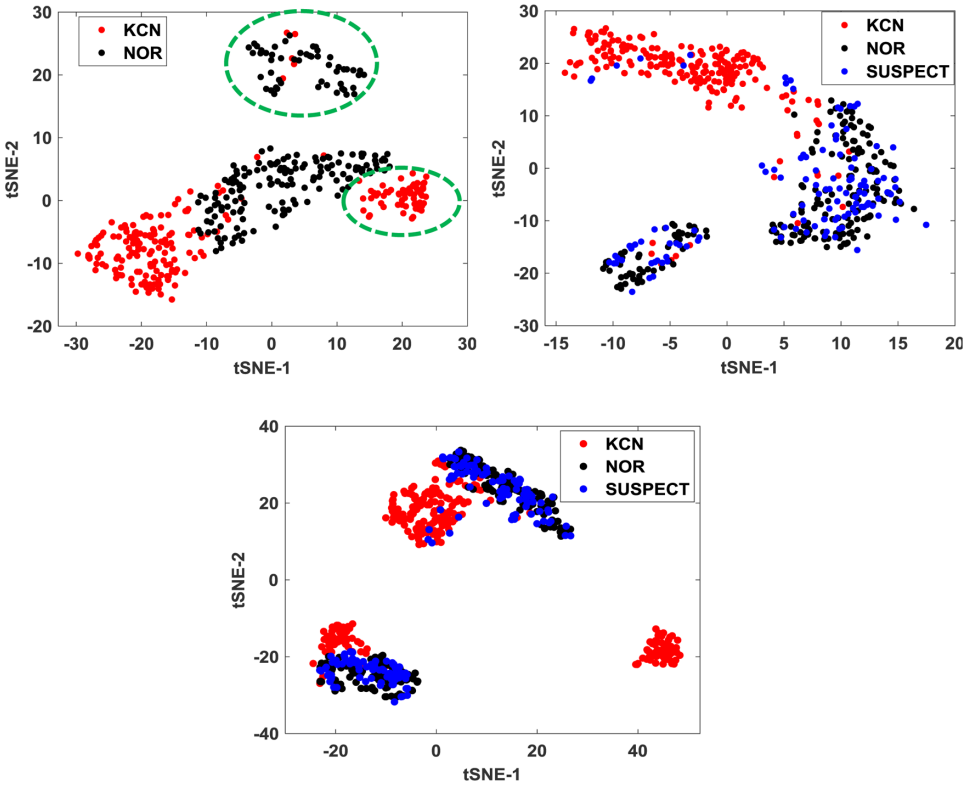

Figure 2. Visualization of 7000 deep features extracted by the EfficientNet-b0 DL architecture from seven different corneal maps using *t*-SNE. (*Top left*) The *t*-SNE plot of features of the two-class problem (eyes from the Pentacam instrument with a different setting are separated with *dashed green lines*). (*Top right*) The *t*-SNE plot of features of the three-class problem. (*Lower panel*) A *t*-SNE visualization of the features of the whole dataset (692 eyes) for the three-class problem based on the development and independent test sets.

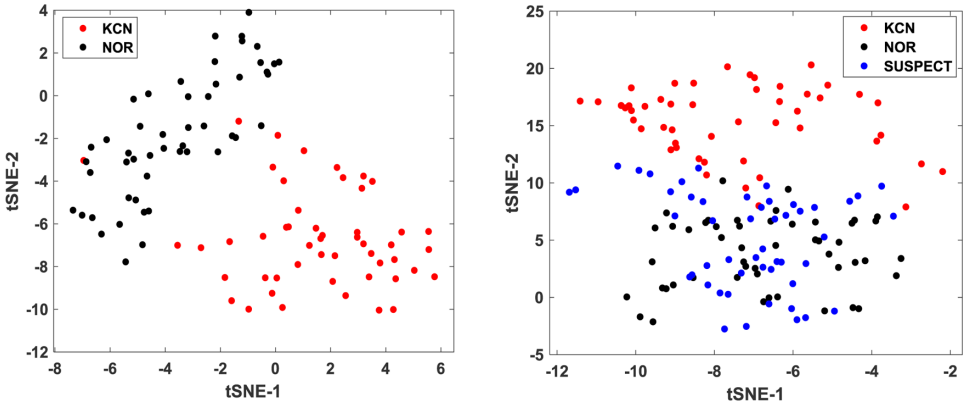

Figure 3. Visualization of 7000 deep features for the independent test set, extracted by the EfficientNet-b0 DL architecture from seven corneal maps using *t*-SNE. (*Left*) A *t*-SNE plot of features of the two-class problem. (*Right*) A *t*-SNE plot of features of the three-class problem.

was 98.8% for the two-class problem (normal vs. KCN) and 81.5% for the three-class problem (normal vs. KCN vs. suspected KCN). The AUC and F1 score of the model based on the development dataset were 0.99 and 0.99 for the two-class problem, respectively, and the AUC and F1 score for the three-class problem were 0.93 and 0.81, respectively. The confusion matrices of the two-class and three-class scenarios are presented in Figure 5.

The accuracies of the seven individual CNN models for the two-class problem, based on corneal anterior and posterior EC, anterior and posterior elevation EL, anterior and posterior SAG, and CT maps for detecting KCN were 97.6%, 96.9%, 95.7%, 97.9%,

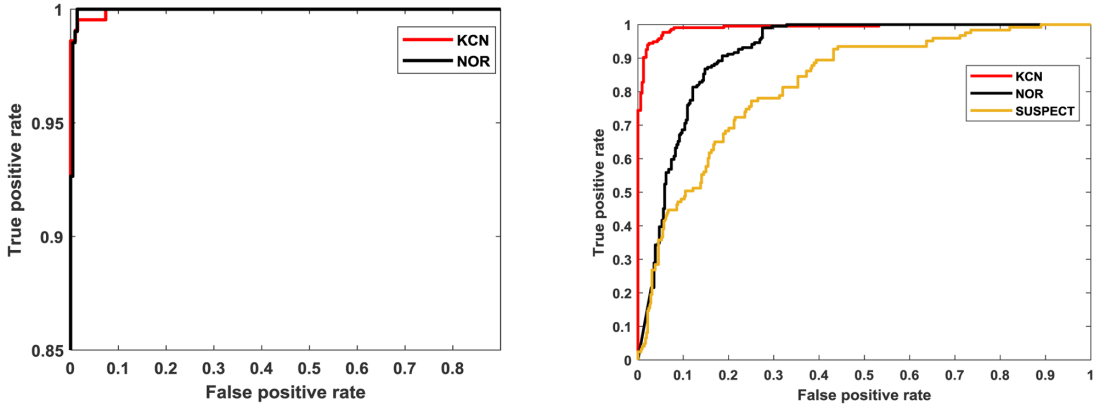

Figure 4. ROC curve of the hybrid model based on the development subset. (*Left*) ROC curves of the two-class problem for discriminating between normal and KCN cases. (*Right*) ROC curves of the three-class problem for discriminating among normal, suspected KCN, and KCN.

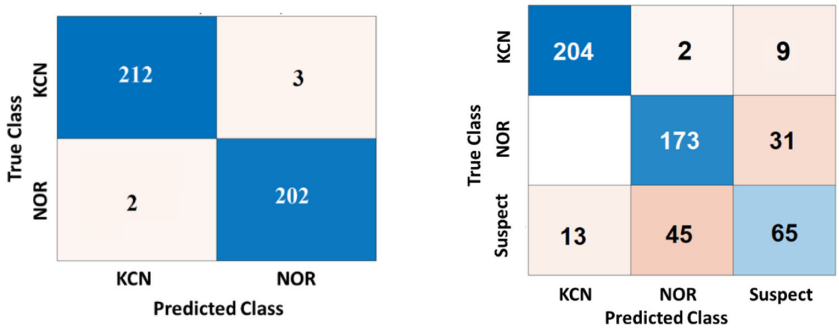

Figure 5. Confusion matrix of the hybrid model for KCN diagnosis obtained based on fivefold cross-validation on the development dataset. (Left) Confusion matrix of the two-class problem. (*Right*) Confusion matrix of the three-class problem. NOR, normal; SUSPECT, suspected KCN.

Table 1. Performance Metrics Including AUC, F1 Score, and Accuracy Based on Different Datasets

Dataset	Classes, *n*	AUC	F1 Score	Accuracy (%)
Development dataset 1 (542 eyes)	2	0.99	0.99	98.5
	3	0.93	0.81	81.5
Independent test dataset 2 (150 eyes)	2	0.99	0.92	92
	3	0.81	0.69	68.7
Merged datasets (692 eyes)	2	0.99	0.98	97.7
	3	0.96	0.85	84.4

98.3%, 95.2%, and 95.5%, respectively. Accordingly, for the three-class problem, the obtained accuracies were 75.5%, 72.1%,76.6%, 75.3%, 80.8%, 72%, and 73.8%, respectively. Based on the independent validation subset, for the two-class problem, the AUC, F1 score, and accuracy were 0.99, 0.92, and 92%, respectively; for the three-class problem, the AUC, F1 score, and accuracy were 0.81, 0.69, and 68.7%, respectively. Based on the merged development subset (542 cases) and independent validation subset (150 eyes), and randomly splitting the whole set into 50%/25%/25% training/validation/testing, for the two-class problem the AUC, F1 score, and accuracy were 0.99, 0.98, and 97.7%, respectively; for the three-class problem, the AUC, F1 score, and accuracy were 0.96, 0.85, and 84.4%, respectively. Table 1 illustrates a summary of the AUC, F1 score, and accuracy metrics.

Figure 6 shows sample images of eyes that were misclassified by the DL model for the two-class problem (Figs. 6A, 6B) and three-class problem (Figs. 6C, 6D). More examples of suspect eyes are provided in the Supplementary Materials. The times required to extract features from the seven maps and for training the SVM classifier for the two-class and three-class problems were about 9.4 minutes and 11.9 minutes, respectively, and the time for testing a new sample

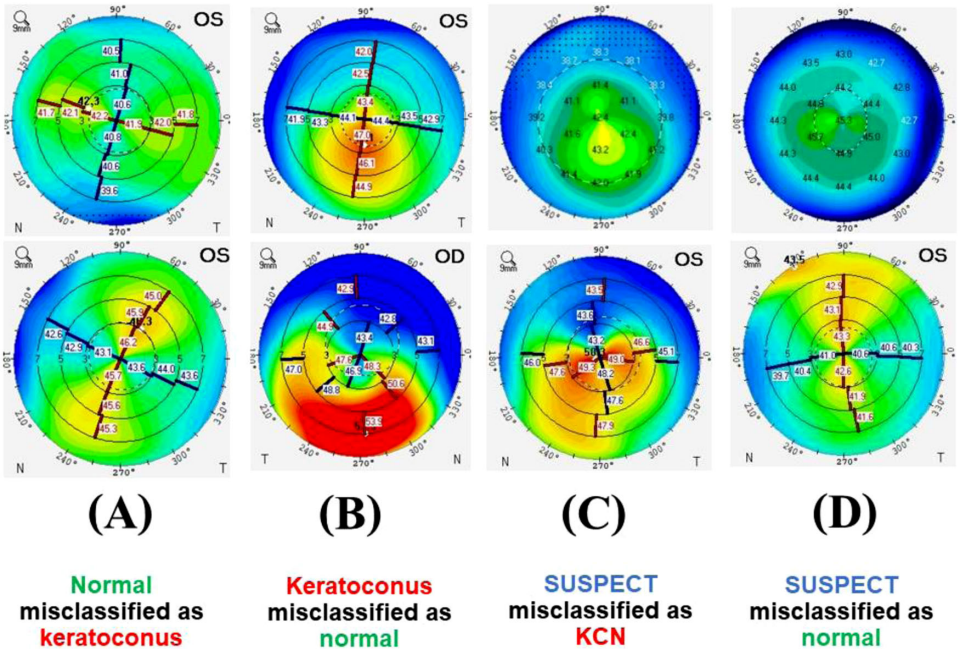

Figure 6. Anterior sagittal curvature maps of eight eyes that were misclassified by the hybrid DL framework. (A) Two normal eyes that were misclassified as KCN. (B) Two KCN eyes that were misclassified as normal. (C) Two suspected KCN eyes that were misclassified as KCN. (D) Two suspected KCN eyes that were misclassified as normal.

for the two-class or three-class problem was about 1.2 seconds.

Discussion

We developed a hybrid DL construct to diagnose KCN from non-invasive corneal topographic images. Our strategy in developing the new construct was to address some of the limitations in previous models. We used a relatively large development dataset with 3794 corneal images to develop and validate the DL models. We then tested the developed models on an independent test set of 1050 images. We first developed multiple DL models, each of which was trained to extract relevant deep features from a particular corneal map for identification of KCN-induced lesions related to that map only. We then fused deep features from different corneal maps in the hybrid construct and compared the performance of single-map and hybrid models. Because single CNN models are known to be sometimes highly sensitive to slight perturbations in the pixels of input images and thus may be error prone,[51] hybrid CNN models could be an appropriate alternative that provides a more solid platform.

Several conventional models have been developed to assess KCN.[9–11] However, there is a significant overlap between parameters of normal eyes versus eyes with KCN based on these models. To address this weakness and to provide automated and more objective methods, a variety of machine learning classifiers have been used to detect KCN, as reported in the previous literature, using decision trees,[52,53] SVM,[54,55] and neural networks.[56] These models rely heavily on handcrafted features or indices extracted from the machine. In contrast, DL models can provide an end-to-end solution that learns to extract features without supervision and with no requirement for handcrafted features or machine-generated parameters.[46]

Multiple DL models have been proposed to diagnose KCN previously[20,21]; however, the models either have utilized small sample sizes of fewer than 400 images or have used simulated images to train and test the models.[21] Because DL models generally require large representative datasets to successfully learn different features corresponding to the underlying condition. In contrast to the previous study,[22] we trained the model to identify both KCN and suspected KCN, which is critical as eyes with suspected KCN will more likely progress to KCN.

We used the EfficientNet-b0[45] architecture, which has been shown to provide superior performance in some applications compared with previous DL architectures such as AlexNet,[46] GoogleNet,[47] VGG19,[48] and Resnet-50.[49] The EfficientNet-b0

architecture includes a fewer number of parameters compared with most of the current DL architectures, making it an ideal candidate for most of the problems, including those with relatively small numbers of samples. Moreover, the computational complexity of the proposed hybrid model is relatively low, allowing it to be run on machines without graphics processing units (GPUs) and only on CPUs, thus providing a suitable platform for many settings with limited resources.

The proposed hybrid construct showed a high AUC of 0.99, F1 score of 0.99, and accuracy of 98.8% for the two-class problem in discriminating eyes with KCN from normal eyes. There were only five misclassifications out of 419 eyes (Fig. 4, left), which included three of the eyes with KCN that were misclassified as normal. Although the model was highly accurate in detecting KCN cases versus normal, there was an obvious overlap between features of suspected KCN and normal eyes in the three-class scenario, which is supported from *t*-SNE of the deep features presented in Figure 2 (top right), where features from normal and suspected KCN eyes show a noticeable overlap. This can be explained by the fact that, although the Pentacam tomographic maps are reasonably distinguishable for normal and established KCN eyes, there is still a significant overlap between maps of normal eyes and eyes suspected of KCN. Nevertheless, this is not a problem associated with this study, but it exists in many biological studies, as the spectrum of normal and disease is a continuum and it can be highly challenging to find a single threshold that separates normal from suspected or early disease.

To assess the generalizability of the model to unseen data, we evaluated the accuracy of the model based on an independent validation subset with 150 eyes. We obtained an AUC of 0.99 (F1 score, 0.92; accuracy, 92%) for the two-class problem and an AUC of 0.81 (F1 score, 0.69; accuracy, 68.7%) for the three-class problem. Although the accuracy levels for the two-class problem were similar, based on the development and independent validation subsets, the accuracy of the three-class problem declined (AUC of 0.93 vs. 0.81). This is not surprising, as features of normal eyes and KCN eyes are more separated compared with features of normal eyes and suspect eyes, particularly when dealing with different datasets. This observation could be partially explained by the following. The independent validation subset was collected from instruments in a private hospital in which the color scale was substantially different from our development subset. Also, fine-tuning of models based on greater numbers of classes is more challenging compared with models that work on only two classes. It should also be noted that the results on the independent set are lower than those of the development set in both the two-class and three-class problems (based on F1 score). This shows that generalizability to different corneal maps that were collected under different instrument color scales is limited. This is expected, as AI models typically will be generalizable on unseen datasets with similar distributions and not on datasets with dissimilar distributions (which is the case here). One solution could be merging the datasets, as was shown in a recent study on retinopathy of prematurity (ROP).[57] We also observed that, if we merge both datasets, then the F1 scores will be improved (F1 score = 0.98 and 0.85 for the two-class and three-class problems, respectively).

For DL models that work with multiple numbers of classes, we typically require greater numbers of samples compared with two classes. Nevertheless, the model was highly generalizable to two classes. Also, based on merged development and independent validation subsets, we allowed the DL model to learn color scale, as well, and reproduce the accuracy level that we obtained based on the development dataset alone. In this study, images were collected from different clinics based on different color scales, which poses a major challenge due to the distributions of the datasets being dissimilar. This issue may be addressed by standardization of the scales of the acquired maps (if the same color palate is used). However, follow-up studies are desirable to investigate whether standardization can fully address this problem.

Future studies with larger numbers of normal and suspected KCN are warranted to further elucidate slight differences between corneal maps of normal eyes and eyes with suspected KCN. To understand why the hybrid DL model made incorrect decisions, corneal maps of several misclassified eyes were evaluated (Fig. 6). In fact, these eyes were challenging for the corneal specialists to make clinical diagnoses, as the inferior inclination or high regular astigmatism (astigmatism against the rule) was presented on the axial/sagittal front maps. In such cases, clinicians typically require additional information, including maps of the BAD-D enhanced ectasia screen and corneal thickness PTI, to make clinical diagnosis.

There is a compromise between a single end-to-end CNN model and a hybrid model that integrates multiple CNN models. Although the former provides a simpler approach, the latter generates a more robust and less error-prone model (less likely to be fooled by misleading images). The model is more robust because we know that voting overall increases the robustness and, in many cases, the accuracy, which we showed empirically. In addition to improved robustness, the likelihood of seven CNN models being fooled is substantially lower than the likelihood of a

Table 2. Previous Literature Investigating the Detection of KCN From Corneal Topographic Images

Study	KCN Classes	Device Used	Dataset/Number of Maps	Evaluation Method	Network Used	Accuracy
Kamiya et al.[19]	Normal and 4 grades of KCN	Tomey CASIA	543 cases/6 maps	Fivefold CV	ResNet-18	99%
Kuo et al.[20]	Normal, KCN	Tomey TMS-4 Corneal Topographer	354 cases/1 map	Training, testing, and subclinical testing	VGG16 InceptionV3 ResNet152	93.1% 93.1% 95.8%
Lavric and Valentin[21]	Normal, KCN	Synthetic maps	SyntEyes and SyntEyes KTC models[58]/1 map	Training, validation, and testing	KeratoDetect	99.3%
Zeboulon et al.[22]	Normal/KCN and history of refractive surgery	Bausch + Lomb Orbscan	3000 cases/4 maps	Tenfold CV	CNN	99.3%
Al-Timemy et al.[23]	Normal, KCN	OCULUS Pentacam	534 cases/4 maps	Training, validation, and testing	EDTL with AlexNet and product fusion	98.3%
Current study	Normal, KCN, suspected KCN	OCULUS Pentacam	692 eyes/7 maps	Training, validation, and independent testing	EfficientNet-b0 DL with SVM	Two-class, 98% Three-class, 81.6%

CV, cross-validation; EDTL, ensemble deep transfer learning.

single CNN model to be fooled, which further improves the dependability of the construct. It should be noted that corneal experts typically use these seven corneal maps to assess KCN; thus, this architecture mimics clinical evaluation. The accuracies of the seven individual CNN models, based on corneal anterior and posterior EC, anterior and posterior elevation EL, anterior and posterior SAG, and CT maps for detecting KCN were 97.6%, 96.9%, 95.7%, 97.9%, 98.3%, 95.2%, and 95.5%, respectively. In comparison, the hybrid model provided slightly higher accuracy while generating a more solid framework consistent with clinical KCN diagnosis.

Previous published studies have investigated different DL models for KCN detection, where different number of maps were investigated. Our work is different in terms of the number of maps utilized and the use of seven DL models for each map without the need for training. It should be noted that Kamiya et al.[19] trained six separate models, and the output from all networks was averaged to obtain a decision. In our work, we directly extracted features from the seven models by the process of convolution of the network with all input maps, without the need for training of the seven networks. Then, we fused the features from all networks and fed them to the SVM classifier to obtain a decision. This approach saved significant amounts of computational time and time spent tuning the parameters of the DL models, as our method does not require training. Table 2 provides a comparison of previous studies investigating KCN detection that includes details regarding the device used, number of eyes/maps, DL models used, and evaluation methods.

Although we used a relatively large dataset and employed a robust hybrid platform, our study has potential limitations, as well. First, greater numbers of suspected eyes will be required to enhance the learning process and improve the accuracy of the model with regard to identifying suspected KCN. Second, although we performed development and testing five times to ensure that the accuracy of the model is consistent, independent subsets would be desirable to confirm our findings. Third, the data were collected from two centers in Brazil, in addition to the independent test set, which was collected from a different clinical setting; thus, obtaining data for other populations with different races is warranted to validate models independently and to ensure generalizability. Finally, future studies should evaluate the impact of the fusion of different corneal maps and their combinations on accuracy and generalizability.

Conclusions

We developed a hybrid DL model composed of seven individual models to extract deep features from corneal maps using a relatively large dataset with 4844 images. The model achieved close to perfect accuracy for discriminating eyes with KCN from normal eyes and reasonable accuracy in distinguishing among normal, suspected KCN, and KCN eyes. The proposed framework provides a robust platform with low computational complexity. Successful validation and deployment of this model may assist clinicians in managing patients with KCN.

Acknowledgments

The authors thank Claudio Alan Oliveira da Rosa and Camila Palmeira Griz for their assistance in the data collection, clinical evaluation, and diagnosis.

Supported by National Institute of Health (NIH), National Eye Institute (NEI), and Bright Focus Foundation.

The preliminary version of this manuscript was accepted for oral presentation at the ARVO Imaging in the Eye Conference, May 1, 2021.

Disclosure: **A.H. Al-Timemy**; None, **Z.M. Mosa**; None, **Z. Alyasseri**; None, **A. Lavric**; None, **M.M. Lui**; None, **R.M. Hazarbassanov**; None, **S. Yousefi**; None

References

1. Rabinowitz YS. Keratoconus. *Surv Ophthalmol.* 1998;42(4):297–319.
2. de Sanctis U, Loiacono C, Richiardi L, Turco D, Mutani B, Grignolo FM. Sensitivity and specificity of posterior corneal elevation measured by Pentacam in discriminating keratoconus/subclinical keratoconus. *Ophthalmology.* 2008;115(9):1534–1539.
3. Gordon-Shaag A, Millodot M, Ifrah R, Shneor E. Aberrations and topography in normal, keratoconus-suspect, and keratoconic eyes. *Optom Vis Sci.* 2012;89(4):411–418.
4. Maguire LJ, Bourne WM. Corneal topography of early keratoconus. *Am J Ophthalmol.* 1989;108(2):107–112.
5. Maeda N, Klyce SD, Smolek MK, Thompson HW. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci.* 1994;35(6):2749–2757.
6. Smolek MK, Klyce SD. Current keratoconus detection methods compared with a neural network approach. *Invest Ophthalmol Vis Sci.* 1997;38(11):2290–2299.
7. Chastang PJ, Borderie VM, Carvajal-Gonzalez S, Rostene W, Laroche L. Automated keratoconus detection using the EyeSys videokeratoscope. *J Cataract Refract Surg.* 2000;26(5):675–683.
8. Twa MD, Parthasarathy S, Roberts C, Mahmoud AM, Raasch TW, Bullimore MA. Automated decision tree classification of corneal shape. *Optom Vis Sci.* 2005;82(12):1038–1046.
9. Ambrosio R, Jr, Alonso RS, Luz A, Coca Velarde LG. Corneal-thickness spatial profile and corneal-volume distribution: tomographic indices to detect keratoconus. *J Cataract Refract Surg.* 2006;32(11):1851–1859.
10. Pinero DP, Alio JL, Aleson A, Escaf Vergara M, Miranda M. Corneal volume, pachymetry, and correlation of anterior and posterior corneal shape in subclinical and different stages of clinical keratoconus. *J Cataract Refract Surg.* 2010;36(5):814–825.
11. Fernandez Perez J, Valero Marcos A, Martinez Pena FJ. Early diagnosis of keratoconus: what difference is it making? *Br J Ophthalmol.* 2014;98(11):1465–1466.
12. Lavric A, Popa V, Takahashi H, Yousefi S. Detecting keratoconus from corneal imaging data using machine learning. *IEEE Access.* 2020;8:149113–149121.
13. Hwang ES, Perez-Straziota CE, Kim SW, Santhiago MR, Randleman JB. Distinguishing highly asymmetric keratoconus eyes using combined Scheimpflug and spectral-domain OCT analysis. *Ophthalmology.* 2018;125(12):1862–1871.
14. Lopes BT, Ramos IC, Salomao MQ, et al. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. *Am J Ophthalmol.* 2018;195:223–232.
15. Saad A, Gatinel D. Evaluation of total and corneal wavefront high order aberrations for the detection of forme fruste keratoconus. *Invest Ophthalmol Vis Sci.* 2012;53(6):2978–2992.
16. Yousefi S, Yousefi E, Takahashi H, et al. Keratoconus severity identification using unsupervised machine learning. *PLoS One.* 2018;13(11):e0205998.
17. Zéboulon P, Debellemanière G, Gatinel D. Unsupervised learning for large-scale corneal topography clustering. *Sci Rep.* 2020;10(1):16973.
18. Yousefi S, Takahashi H, Hayashi T, et al. Predicting the likelihood of need for future keratoplasty intervention using artificial intelligence. *Ocul Surf.* 2020;18(2):320–325.
19. 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(9):e031313.
20. Kuo BI, Chang WY, Liao TS, et al. Keratoconus screening based on deep learning approach of corneal topography. *Transl Vis Sci Technol.* 2020;9(2):53.
21. Lavric A, Valentin P. KeratoDetect: keratoconus detection algorithm using convolutional neural networks. *Comput Intell Neurosci.* 2019;2019:8162567.
22. Zeboulon P, Debellemaniere G, Bouvet M, Gatinel D. Corneal topography raw data classification using a convolutional neural network. *Am J Ophthalmol.* 2020;219:33–39.
23. Al-Timemy AH, Ghaeb NH, Mosa ZM, Escudero J. Deep transfer learning for improved

detection of keratoconus using corneal topographic maps [published online ahead of print June 16, 2021]. *Cognit Comput*, https://doi.org/10.1007/s12559-021-09880-3.
24. Abdelmotaal H, Mostafa MM, Mostafa ANR, Mohamed AA, Abdelazeem K. Classification of color-coded Scheimpflug camera corneal tomography images using deep learning. *Transl Vis Sci Technol.* 2020;9(13):30.
25. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402–2410.
26. Ting DSW, Cheung CY, Lim G, 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):2211–2223.
27. Li X, Hu X, Yu L, Zhu L, Fu CW, Heng PA. CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE Trans Med Imaging.* 2020;39(5):1483–1493.
28. Wang H, Yuan G, Zhao X, et al. Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening. *Comput Methods Programs Biomed.* 2020;191:105398.
29. Arcadu F, Benmansour F, Maunz A, Willis J, Haskova Z, Prunotto M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med.* 2019;2(1):92.
30. Mao J, Luo Y, Liu L, et al. Automated diagnosis and quantitative analysis of plus disease in retinopathy of prematurity based on deep convolutional neural networks. *Acta Ophthalmol.* 2020;98(3):e339–e345.
31. Campbell JP, Kim SJ, Brown JM, et al. Evaluation of a deep learning-derived quantitative retinopathy of prematurity severity scale. *Ophthalmology.* 2021;128(7):1070–1076.
32. Burlina P, Freund DE, Joshi N, Wolfson Y, Bressler NM. Detection of age-related macular degeneration via deep learning. Paper presented at 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 13–16 April 2016, Prague, Czech Republic.
33. Grassmann F, Mengelkamp J, Brandl C, 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):1410–1420.
34. Rim TH, Lee AY, Ting DS, et al. Detection of features associated with neovascular age-related macular degeneration in ethnically distinct data sets by an optical coherence tomography: trained deep learning algorithm. *Br J Ophthalmol.* 2021;105(8):1133–1139.
35. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med.* 2017;82:80–86.
36. Liu H, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol.* 2019;137(12):1353–1360.
37. Lee J, Kim YK, Park KH, Jeoung JW. Diagnosing glaucoma with spectral-domain optical coherence tomography using deep learning classifier. *J Glaucoma.* 2020;29(4):287–294.
38. Ran AR, Cheung CY, Wang X, et al. Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis. *Lancet Digit Health.* 2019;1(4):e172–e182.
39. Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA. Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans. *JAMA Ophthalmol.* 2020;138(4):333–339.
40. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R. A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS One.* 2019;14(7):e0219126.
41. Thakur A, Goldbaum M, Yousefi S. Predicting glaucoma before onset using deep learning. *Ophthalmol Glaucoma.* 2020;3(4):262–268.
42. Sun J, Huang X, Egwuagu C, et al. Identifying mouse autoimmune uveitis from fundus photographs using deep learning. *Transl Vis Sci Technol.* 2020;9(2):59.
43. Sonmez B, Doan MP, Hamilton DR. Identification of scanning slit-beam topographic parameters important in distinguishing normal from keratoconic corneal morphologic features. *Am J Ophthalmol.* 2007;143(3):401–408.
44. Kawamorita T, Uozato H, Kamiya K, et al. Repeatability, reproducibility, and agreement characteristics of rotating Scheimpflug photography and scanning-slit corneal topography for

corneal power measurement. *J Cataract Refract Surg.* 2009;35(1):127–133.
45. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. In: Chaudhuri K, Salakhutdinov R, eds. *Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PML4 97*, 2019.
46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM.* 2017;60(6):84–90.
47. Szegedy C, Wei L, Yangqing J, et al. Going deeper with convolutions. Paper presented at 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015, Boston, MA.
48. Simonyan A, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*, arXiv:1409.1556v6.
49. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *arXiv*, arXiv:1602.07261v2.
50. Deng J, Dong W, Socher R, Li L, Kai L, Li F-F. ImageNet: a large-scale hierarchical image database. Paper presented at 2009 IEEE Conference on Computer Vision and Pattern Recognition, 20–25 June 2009, Miami, FL.
51. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. Paper presented at 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015, Boston, MA.
52. Marsolo K, Twa M, Bullimore MA, Parthasarathy S. Spatial modeling and classification of corneal shape. *IEEE Trans Inf Technol Biomed.* 2007;11(2):203–212.
53. Smadja D, Touboul D, Cohen A, et al. Detection of subclinical keratoconus using an automated decision tree classification. *Am J Ophthalmol.* 2013;156(2):237–246.e1.
54. 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(6):827–832.
55. Musa ZM, Hussein N, Ali AH. Detecting keratoconus by using SVM and decision tree classifiers with the aid of image processing. *Baghdad Sci J.* 2019;16(4):1022–1029.
56. Maeda N, Klyce SD, Smolek MK. Neural network classification of corneal topography. Preliminary demonstration. *Invest Ophthalmol Vis Sci.* 1995;36(7):1327–1335.
57. Chen JS, Coyner AS, Ostmo S, et al. Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. *Ophthalmol Retina.* 2021;5(10):1027–1035.
58. Rozema JJ, Rodriguez P, Ruiz Hidalgo I, Navarro R, Tassignon MJ, Koppen C. SyntEyes KTC: higher order statistical eye model for developing keratoconus. *Ophthalmic Physiol Opt.* 2017;37(3):358–365.