Research Article Vol. 16, No. 8 / 1 Aug 2025 / Biomedical Optics Express 3047
Biomedical Optics EXPRESS

Deep learning-based keratoconus detection from Scheimpflug images

JUAN CASADO-MORENO,[1,*] BELEN MASIA,[1] NANJI LU,[2] LELE CUI,[3] AND ALEJANDRA CONSEJO[1]

[1]Aragon Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain
[2]Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China
[3]Eye Hospital, Wenzhou Medical University, Wenzhou, China
[*]jcasado.moreno@unizar.es

Abstract: This study evaluates the effectiveness of deep learning techniques applied to raw Scheimpflug corneal images for keratoconus detection, with a particular focus on forme fruste (FF) keratoconus, which refers to preclinical cases. Using an original dataset of 22,750 images from 910 eyes, a deep learning model based on transfer learning with a pre-trained VGG16 architecture was trained, incorporating specific preprocessing steps and data augmentation strategies. The proposed approach achieved an overall accuracy of 90.70%, with a sensitivity of 80.57%, and a specificity of 80.56% for FF keratoconus classification, and an AUC of 0.89. For clinical keratoconus, the model demonstrated a sensitivity of 93.28%, a specificity of 99.40%, and an AUC of 1.00. These findings highlight the potential of leveraging raw Scheimpflug images in deep learning-based keratoconus detection, particularly for identifying early-stage structural changes that may not be apparent in conventional topographic assessments.

1. Introduction

Keratoconus is a progressive, non-inflammatory corneal disorder characterized by thinning and conical protrusion of the cornea, leading to visual impairment [1,2]. This condition primarily affects adolescents and young adults [3], with a reported prevalence ranging from 1:375 [4] to 1:1,000 [5], depending on the population and diagnostic criteria used. Keratoconus is a bilateral but often asymmetrical disease, typically developing at different rates in each eye. This asymmetry offers a unique opportunity to study eyes that have not yet exhibited clinical signs but are at high risk of progression—a condition known as forme fruste (FF) keratoconus [6]. FF keratoconus can be used as a surrogate for the preclinical stage of keratoconus, as it enables the identification of structurally predisposed eyes while maintaining normal topography. Understanding and detecting these preclinical stages is key to enhancing patient outcomes and preventing irreversible vision loss. It is worth noting that, while FF keratoconus is widely recognized in the literature as a preclinical stage, the term "subclinical keratoconus" lacks a consistent definition. Some authors use it to describe cases with incipient topographic changes, while others apply it interchangeably to preclinical cases as well. In this work, FF keratoconus is specifically used to denote the preclinical stage, while the term "subclinical" appears only when referencing previous studies, regardless of how it was defined in those works.

While early detection of keratoconus is important, it is not sufficient to fully mitigate the risks associated with the disease. There is a pressing need to improve diagnostic strategies to identify keratoconus before it shows clinical signs. Detecting the disease at this stage is essential to prevent iatrogenic postoperative corneal ectasia, a serious complication that can occur after refractive surgery when underlying keratoconus is undiagnosed [7–9]. As the prevalence of myopia continues to rise globally, the demand for refractive surgery increases accordingly, amplifying the importance of accurate preoperative screening [10].

#559663 https://doi.org/10.1364/BOE.559663
Journal © 2025 Received 18 Feb 2025; revised 20 Apr 2025; accepted 30 Apr 2025; published 7 Jul 2025

Biomedical Optics EXPRESS

Scheimpflug-based corneal tomography is considered the gold standard in keratoconus screening due to its ability to provide detailed three-dimensional imaging of the cornea. However, despite its diagnostic capabilities, the proprietary software integrated into these devices for keratoconus characterization often relies on predefined indices and parameters that lack the sensitivity required to detect subclinical cases [11,12]. Consequently, there is a critical need for advanced analytical approaches capable of enhancing the detection of keratoconus before the manifestation of clinically detectable signs with currently available methods.

Due to the complexity of subclinical keratoconus detection, numerous mathematical models [13–15] and traditional machine learning techniques have been proposed [16–19], often relying on the combination of proprietary-specific parameters to automate keratoconus detection. More recently, deep learning algorithms have been introduced, utilizing corneal tomography maps —such as curvature, elevation, and thickness maps— as input data [20–32]. These approaches leverage the ability of deep learning to learn and model complex, non-linear patterns, making it the state-of-the-art technique for image classification tasks [33]. However, the limitation of these approaches in detecting preclinical cases lies in the input data itself—since FF keratoconus does not produce noticeable changes in corneal *shape*, even the most advanced algorithms struggle, particularly when attempting to distinguish them from control cases.

A more recent and promising area of research explores corneal densitometry, which quantifies light backscatter to objectively analyze corneal *tissue* as a reliable marker for detecting subclinical cases [34,35]. Building on this premise, this work introduces an innovative approach by focusing directly on the raw images themselves, minimizing dependence on proprietary, unknown processes. By utilizing raw images, a deep learning-based model can simultaneously leverage both corneal shape (Fig. 1, yellow)—traditionally the primary focus in keratoconus detection—and corneal tissue characteristics (Fig. 1, green), a novel yet promising approach. This dual-focus strategy allows the model to extract and integrate information from both structural and tissue-based features, potentially enhancing its ability to identify subtle changes associated with preclinical keratoconus.

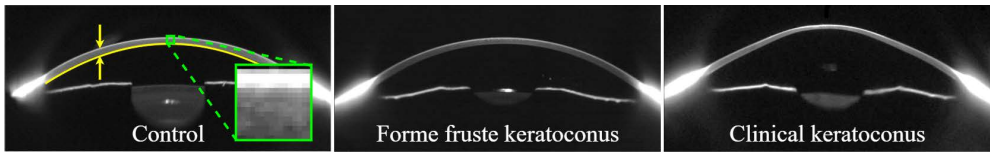

Fig. 1. Representative raw Scheimpflug images from each diagnostic class (control, forme fruste keratoconus, and clinical keratoconus). A deep learning-based model can process these images directly, allowing it to simultaneously extract features related to both corneal shape and tissue. For illustrative purposes, the first image highlights the corneal contour in yellow to indicate shape-related characteristics, while a zoomed-in region in green shows the pixel intensity distribution, which may contain information related to tissue characteristics.

The aim of this work is to develop a deep learning-based approach that leverages raw Scheimpflug images to enhance the detection of preclinical keratoconus, specifically FF keratoconus, by simultaneously analyzing corneal shape and tissue characteristics, thereby overcoming the limitations of traditional topographic-based methods.

2. Methodology

2.1. Overview

Our model takes as input five raw Scheimpflug images corresponding to five angularly equidistant eye meridians, and outputs a classification of such eye into one of three classes: control (C),

Biomedical Optics EXPRESS

forme fruste (FF) keratoconus or clinical keratoconus (KC); the specific clinical definition of each class can be found in Section 3.1.

This classification is done employing transfer learning, a common strategy in deep learning approaches which involves leveraging pre-trained models—typically trained on large, generic datasets—and fine-tuning them for specific tasks such as keratoconus detection. In our case, a VGG16 backbone is trained to extract features from the input images, and these features are then input to a *feature fusion module* before producing the final classification. Section 2.3 describes the model architecture in detail, while the pre-processing of the images before being fed to the model is explained in Section 2.2. As is common in diagnosis-related problems, the classes in the training dataset are significantly imbalanced, with a low prevalence of keratoconus and especially FF keratoconus, hinderinglearning;wethusexplorestrategiestoaddressclassimbalanceinSection 2.4.

2.2. Data pre-processing

Input data to our model are raw Scheimpflug images (without gamma correction or contrast enhancement), cropped from the original (2400 $\times$ 1782 pixels in size), which included patient identification details, machine specifications, and blank margins. For each raw image (Fig. 2(A)), we conduct an objective corneal segmentation process to accurately identify the anterior and posterior boundaries of the cornea, and select only the area in-between these boundaries. The first step involves automatically cropping the raw image using the corneal apex as the center, with a bandwidth of 1200 pixels, corresponding approximately to the central 7 mm of the cornea. This approach facilitates the exclusion of irrelevant elements such as eyelashes, or lateral scleral brightness artifacts (Fig. 2(B)). The next step is border detection (Fig. 2(C)). In this stage, the anterior and posterior corneal borders are automatically identified by applying Gaussian filtering followed by the Canny edge detection method, a widely used technique for corneal segmentation. Further, the cornea is segmented (Fig. 2(D)) to eliminate the potential influence of the iris, crystalline lens, and background, enhancing the effectiveness of the learning model. A second cropping is then applied (Fig. 2(E)) to remove excess background. In this step, a fixed vertical dimension is selected for the ROI to ensure all corneas fit within the second cropped area. To determine this dimension, the maximum distance between the apex and the posterior border at the corneal periphery is measured across the dataset, and this value is used as a reference.

This segmentation process optimizes memory usage, reduces training time to one-quarter of the original, and improves model performance, as demonstrated in Section 3.3. In our dataset, the resulting cropped images are 1200 $\times$ 300 pixels in size, and are then downsampled by half to optimize memory usage and subsequently normalized, resulting in input images with a spatial resolution of 600 $\times$ 150 pixels.

2.3. Model architecture

The architecture of our model is shown in Fig. 3. Given five input images from equidistant meridians, we use a convolutional neural network (CNN) to extract features from them. This CNN is VGG16 [36], pre-trained on ImageNet [37], which has shown great performance for classification tasks in natural images. Since our Scheimpflug images are substantially different from the images that VGG16 is trained on, we fine-tune the final layers of the network on our own dataset (details on this dataset are given in Section 3.1), so that it learns to extract meaningful features for our classification problem, in a transfer learning approach. We experiment with different degrees of transfer, varying the number of layers that are trained on our data (see Supplement 1).

The features extracted from each of the images are then fed into our *feature fusion module*. This module comprises a global average pooling (GAP) step to reduce dimensionality, followed by feature concatenation from the five branches, and two dense (fully-connected) layers. Two dropout

Biomedical Optics EXPRESS

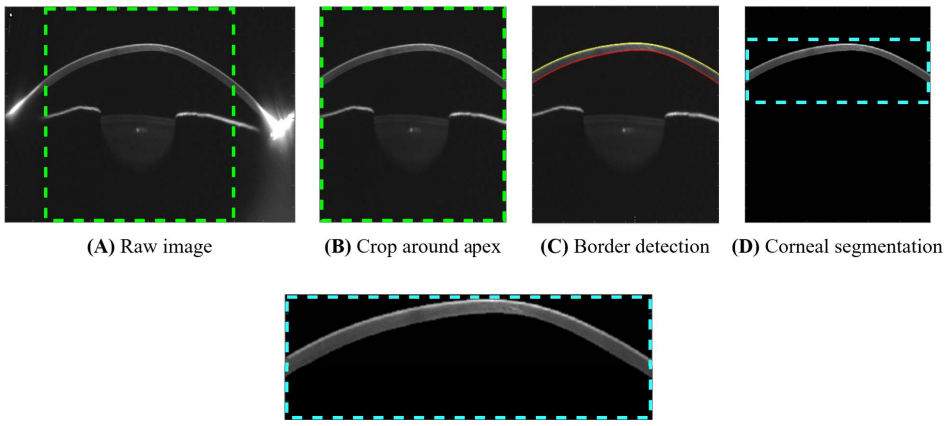

Fig. 2. Main steps of our data pre-processing stage. (A) Sample raw Scheimpflug image with the initial cropped region centered on the apex (highlighted in green), to remove irrelevant elements such as eyelashes and lateral scleral brightness artifacts (2020 × 966 pixels). (B) Resulting image after cropping (1200 × 966 pixels). (C) Border detection step, identifying the anterior (yellow line) and posterior (red) borders of the cornea (1200 × 966 pixels). (D) Corneal segmentation to eliminate any remaining influence from the iris, crystalline lens, and background (1200 × 966 pixels). (E) Final post-processed image after removing excess background (1200 × 300 pixels, then downsampled to 600 × 150).

layers are included after each dense layer to regularize the model during training, enhancing its generalization and preventing overfitting by randomly deactivating neurons. Additionally, after each dropout layer a batch normalization layer follows, normalizing activations to stabilize and accelerate the training process. Finally, a softmax activation is applied to the output layer to obtain, for each of the three classes, the probability that the input sample belongs to that class.

The model is trained using sparse categorical cross-entropy loss function, a more efficient alternative to categorical cross-entropy that works with integer-encoded labels instead of one-hot encoding, reducing computational complexity in multiclass classification [38].

2.3.1. Training hyperparameters and implementation details

We trained the model using the Adam optimizer [39], with an early stopping callback after 15 epochs, and a ReduceLROnPlateau callback that reduced the learning rate by a factor of 0.5 if the validation loss did not improve for five consecutive epochs, with a minimum learning rate of 1e-6. Furthermore, during validation, class weights were adjusted such that the forme fruste class was assigned twice the weight of the other two classes, to alleviate its reduced size in terms of training samples.

Our final VGG16-based transfer learning scheme features six unfrozen layers. The dense layers have 512 and 256 channels, applying the ReLU activation function, a dropout rate of 0.4, and a batch size of 32. The initial learning rate is 1e-4, and training runs for a maximum of 100 epochs. These hyperparameters are the result of a comprehensive exploration of the space: for results using other hyperparameters, as well as other pre-trained models for transfer learning, please refer to the Supplement 1.

The model was developed using Python 3.10.12 and the Tensorflow 2.14.0 open-source framework. Models were trained on a NVIDIA GeForce RTX A6000 GPU.

Biomedical Optics EXPRESS

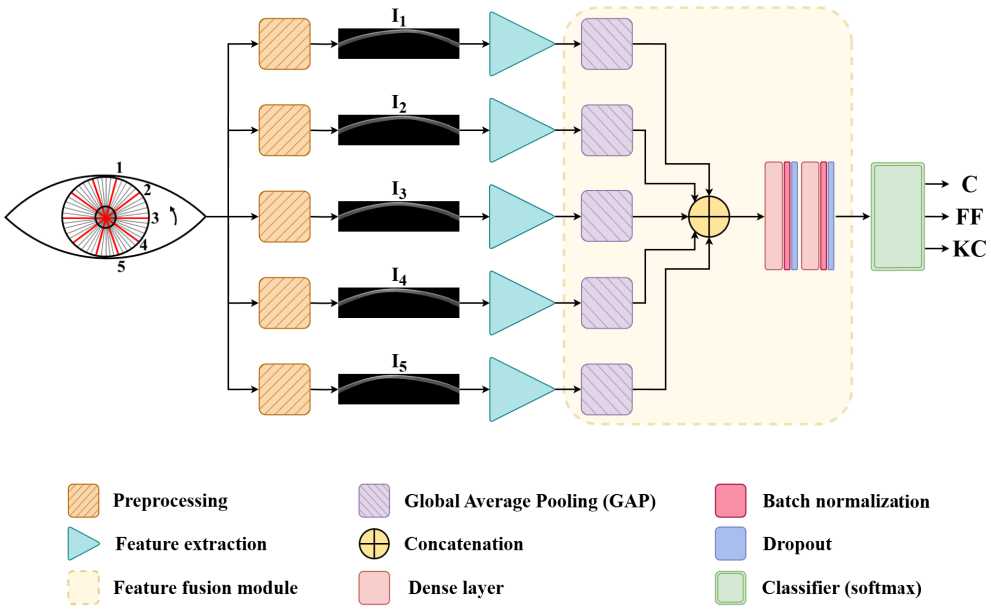

Fig. 3. Workflow of our proposed deep learning-based model. The input consists of raw images of five equidistant meridians out of the 25 meridians of the same eye imaged by the Pentacam HR rotating camera. These raw images undergo a data pre-processing stage (described in Section 2.2) that yields the images I_i, $i \in 1..5$, that are input to our model. Features are extracted from each image I_i with a VGG16-based CNN trained on our specialized dataset. Then, the outputs of each branch are concatenated, and two additional dense layers are applied to extract more complex features, together with batch normalization and dropout layers. Finally, the output is obtained after a Softmax classifier. The output labels are C: control, FF: forme fruste keratoconus, and KC: clinical keratoconus.

2.4. *Addressing class imbalance*

In diagnosis-related problems, it is common to have datasets where the classes are imbalanced. Fortunately, the prevalence of pathological cases in the population is typically lower than that of the normal ones, resulting in the normal class usually having a larger number of samples than the others. This imbalance severely affects the learning process and results in inferior performance, particularly when detecting the underrepresented class, which tends to be the most interesting one. In our dataset (described in detail in Section 3.1) this is also the case: the number of samples belonging to the control class is largest, and the number of FF samples is significantly lower than that of the other two classes. Therefore, we implement a number of strategies to mitigate class imbalance.

We balance class size by applying a combination of random oversampling (ROS) and data augmentation (DA) to our dataset. ROS is a widely used technique that randomly duplicates samples from the minority classes to decrease class imbalance [40]. To increase the variability of the dataset and avoid excessive repetition, this oversampling is combined with DA [41]. DA involves generating new training samples by applying certain transformations to the original samples; the nature of these transformations can vary depending on the particular scenario. Augmentation increases the dataset's diversity and helps prevent overfitting, particularly relevant when there are an important number of duplicate images due to ROS. Inspired by the work of Schaudt et al. [42], we implemented and evaluated two levels of DA, termed *weak* and *strong*. Our DA procedure is illustrated in Fig. 4. In our case, augmentation is conducted using three

Biomedical Optics EXPRESS

different transformations: image rotations, image flips, and image sharpening. Although based on their methodology, we specifically avoided transformations that affect image contrast to preserve the original pixel intensity of the corneal tissue. To further minimize any unintended alterations, sharpening was applied with the lowest probability, as indicated in Fig. 4, and no additional transformations that could compromise the integrity of corneal texture were used.

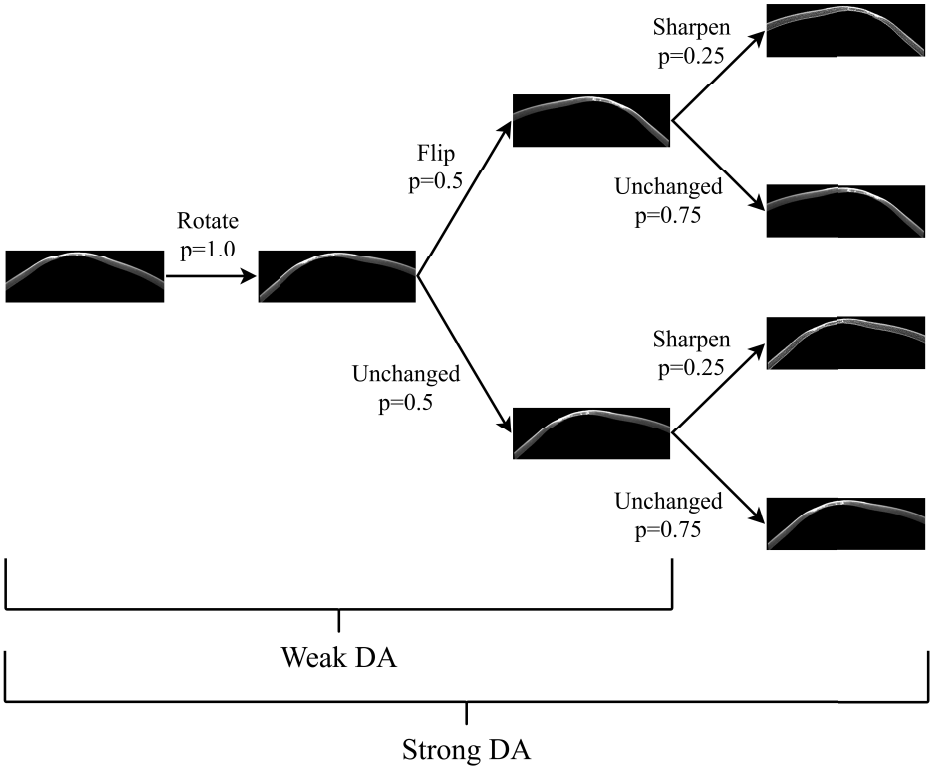

Fig. 4. Diagram of the data augmentation (DA) process with its stacked probabilistic transformations.

Weak DA involves first rotating images between −5 and +5 degrees with a 100% probability (i.e., all images are rotated a random amount within the specified range), and then flipping images horizontally with a 50% probability. Strong DA implies adding image sharpening after the weak DA, applying it to the image with a 25% probability. Since images are input to our model in groups of five (five equidistant meridians from the same eye), all five images from an eye undergo the same DA. This synchronization is achieved by implementing DA as custom layers that, at the beginning of each training step, apply the same random transformation to all five images, using shared random seeds. Section 3.3 evaluates the performance of the model with the different augmentation strategies.

3. Evaluation

This section presents the evaluation of this work. First, we describe in detail the dataset used (Section 3.1), and the evaluation metrics employed (Section 3.2). Finally, Section 3.3 shifts focus to the results of the proposed model.

Biomedical Optics EXPRESS

3.1. Dataset

Our dataset consists of corneal tomographies that were retrospectively collected from one prior investigation carried out at the Eye Hospital of Wenzhou Medical University [19]. Approval for the study was granted by the Hospital Research Ethics Committee, and they adhered to the principles outlined in the Declaration of Helsinki. Written informed consent was obtained from all participants after they were briefed on the nature and potential consequences of the research.

The dataset includes 910 eyes of patients with an average age of 22.3 ± 6.3 years, and a male/female ratio of 59/41, respectively. Participants are divided into three groups according to labels provided by clinical experts: controls (492 eyes), clinical keratoconus (339 eyes), and FF keratoconus (79 eyes), considered to be the non-pathological fellow eye of a keratoconus patient [6]. Specifically, the three groups are defined as follows:

- *Control (C)*: No slit-lamp findings suggestive of corneal ectasia and normal tomography.
- *Keratoconus (KC)*: Clinical and tomographic signs consistent with keratoconus (e.g., anterior and/or posterior corneal steepening, corneal thinning, stromal thinning, Fleischer ring at the cone base, Vogt striae). All of the patients were diagnosed by a cornea specialist in a tertiary center [43].
- *Forme fruste keratoconus (FF)*: Contralateral, asymptomatic eye showing no clinical signs of ectasia of a subject with clinical ectasia in the other eye as a very or highly asymmetric ectasia [43]. These eyes had central average keratometry (K_{max}) ≤ 46.3 diopter, normal scores for topographical keratoconus classification and ABCD (A0B0C0) [44].

To avoid short-term induced corneal tissue changes by lens wear [45], patients were asked to discontinue soft contact lenses for at least two weeks before the examination, or at least four weeks in cases of rigid gaspermeable contact lenses wear. Patients who could not comply were eliminated from the dataset. A comprehensive ophthalmologic examination was conducted for all patients, including corneal Scheimpflug tomography images using a Pentacam HR (Oculus Optikgeräte GmbH, Wetzlar, Germany), which provided the dataset used in this study.

For each explored eye, the Pentacam HR captures images for 25 corneal meridians. Thus, for each eye, our dataset includes the 25 raw Scheimpflug images from the Pentacam HR, and a label (one of three classes, C, KC or FF) assigned by cornea specialists according to the criteria above. Images corresponding to the 25 meridians of an eye share the same label. This results in a total of 22,750 labeled images. These grayscale images have a spatial resolution of 2400 × 1782 pixels before pre-processing (see Section 2.2).

This dataset is then split into training set (~85%) and a test set (~15%). To ensure reliable assessment of our model's performance, we do the split such that there is no leakage between both sets, i.e., data from a certain patient is either in one or in the other set, but never in both. This test set, which is completely separate from the training one and never used during training, is the one used for the results shown here (Section 3.3). Given the size of our training set, when validating different design choices to select the best model we conduct a 5-fold cross-validation to improve generalization.

3.2. Evaluation metrics

To assess the performance of the developed model and provide a meaningful discussion, different metrics commonly used in the literature for health diseases detection (specifically keratoconus detection) were chosen.

The first one is *accuracy*, which measures the ratio of correct classifications to the total number of classifications. Additionally, *sensitivity* represents the proportion of correctly classified pathological cases among all individuals who actually have the disease (true positives divided

Biomedical Optics EXPRESS

by the total number of positives), while *specificity* refers to the ratio of non-pathological cases correctly identified as such (true negatives divided by the total number of negatives).

Besides, the Receiver Operating Characteristic (ROC) curve is a plot that displays the relationship between the proportion of correct positive predictions and the proportion of incorrect positive predictions for a model, calculated across different thresholds for the model's score. Each point on the curve represents a different trade-off between false positives and false negatives. By adjusting the threshold, it is possible to prioritize sensitivity over specificity, or vice versa, allowing the model to be optimized for different clinical or operational needs. Additionally, ROC curves are used to derive the Area Under the ROC curve (*AUC* or AUROC), a popular metric that measures the probability that a positive case (pathological) will have a higher score than a negative case (healthy).

To delve deeper into differences in crucial classification areas for the models, and to get a visual insight into how the model is working internally, GradCAM [46] attributions are utilized. GradCAM is a technique that visualizes the gradients of the classification score in relation to the final convolutional feature map, thereby highlighting important regions of an image.

3.3. Results

The proposed model achieved an overall accuracy of 90.70% and an AUC of 0.96. Table 1 presents class-specific sensitivity, specificity, and AUC in the case in which, for each test sample, the predicted class is the one for which the model outputs the highest probability. Additionally, Fig. 5 displays the ROC curves for individual classes.

Table 1. Evaluation metrics (specificity, sensitivity and AUC) for each class. For each test sample, the predicted class is the one with the highest probability.

	Specificity	Sensitivity	AUC
C	86.12%	96.89%	0.97
FF	97.34%	49.71%	0.89
KC	99.40%	93.28%	1.00

AUC, area under the curve; C, control; FF, forme fruste keratoconus; KC, clinical keratoconus.

While the model performs well overall in this scenario, the sensitivity for FF keratoconus (50%) is low, meaning half of FF cases are misclassified as normal. To address this, instead of using the class with the highest output probability as the predicted class, we implemented a probability-based threshold adjustment for FF classification. If an eye's predicted probability for FF exceeded an established threshold, it was assigned to the FF class, even if another class had a higher probability. After an exhaustive loop-based search, we set the threshold to 0.20. This adjustment significantly improved FF sensitivity from 49.71% to 80.57%, as shown in Fig. 6, which may be highly valuable depending on the clinical scenario (e.g., patients considering undergoing refractive surgery, who could incur in serious complications if they have FF keratoconus).

Table 2 shows, for various approaches to address class imbalance: global accuracy, sensitivity, specificity, and AUC for each class independently. The first row highlights the overall top-performing model discussed earlier. The combination of Strong DA and ROS yielded the best overall performance, significantly enhancing FF sensitivity while maintaining high accuracy and

Biomedical Optics EXPRESS

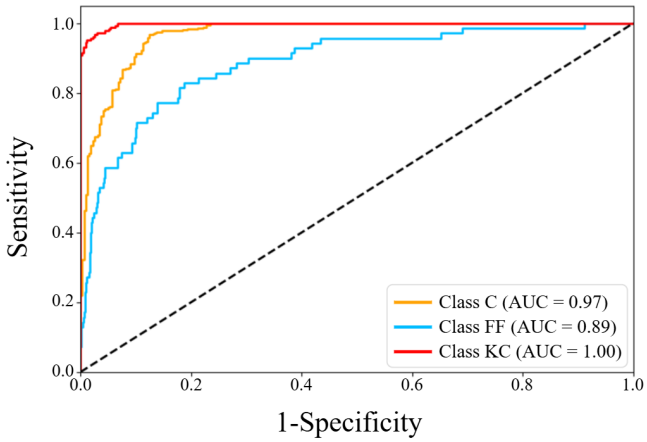

Fig. 5. ROC curve of our model for each of the three classes. AUC, area under the curve; C, control; FF, forme fruste keratoconus; KC, clinical keratoconus.

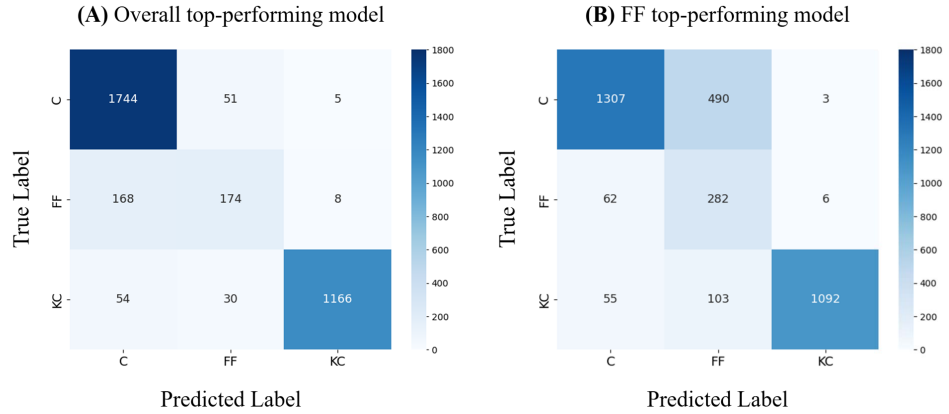

Fig. 6. Comparison of the confusion matrices between the overall top-performing model (A) and the best-performing model for the FF keratoconus (B). C, control; FF, forme fruste keratoconus; KC, clinical keratoconus.

specificity. Given its robustness, this configuration was subsequently used as the foundation for further analyses and testing.

Table 3 shows the results obtained when evaluating different input data schemes for the transfer learning model. Specifically, we evaluate whether the segmentation done during the pre-processing stage indeed leads to improved results, as well as the number of images per eye required by the model during inference. The best performance was achieved using segmented images, in groups of five (5 images/eye). Note that, during training, we typically use all available images per eye, e.g., for the case of a model that has 5 images/eye as input, the 25 meridians of one eye yield five groups of five images, leading to five training samples; this is somewhat akin to a form of data augmentation. Importantly, and as mentioned before, the whole data for one patient is either in the training or in the test set, but never in both. The table shows how, given the same number of images in inference, our strategy of leveraging the full dataset during training results in improved performance (row 2 vs. row 3). We also see that, for the same amount of training data, inference with 5 images/eye leads to better results than with 1 image/eye (row 2 vs. row 4), justifying our choice.

Biomedical Optics EXPRESS

Table 2. Performance metrics for different techniques used to address class imbalance in the dataset. We vary, in each row, whether we apply ROS or not (first column), and whether we conduct no DA, weak DA or strong DA (second column). Our proposed scheme is marked in boldface.

Options		Metrics					
ROS	DA	Accuracy	Sensitivity	Specificity	AUC C	AUC FF	AUC KC
✓	✓**(strong)**	**90.70%**	**79.96%**	**94.29%**	**0.97**	**0.89**	**1.00**
✓	✓(weak)	88.68%	74.17%	92.66%	0.96	0.86	0.99
✓	×	79.85%	77.58%	90.77%	0.94	0.83	0.99
×	✓(weak)	88.53%	67.35%	92.27%	0.96	0.88	0.99
×	×	80.74%	62.89%	89.22%	0.95	0.76	0.99

ROS, random oversampling; DA, data augmentation; AUC, area under the curve; C: control; FF: forme fruste keratoconus; KC: clinical keratoconus.

Table 3. Comparison of different input data schemes (rows) for the model. We vary three parameters: whether segmentation is applied or not during the image preprocessing stage (first column, Seg); the number of images (meridians) per eye required as input to the model to conduct inference (second column); and the number of images per eye included in the training dataset (third column, grouping in parentheses, see text for details). For each combination of parameters, the table compiles the corresponding evaluation metrics. Our proposed model is highlighted in boldface.

Options			Metrics					
Seg	Input	Training images per eye	Accuracy	Sensitivity	Specificity	AUC C	AUC FF	AUC KC
×	5 images/eye	25 (5 groups × 5 images)	86.44%	74.67%	93.43%	0.96	0.88	0.98
✓	**5 images/eye**	**25 (5 groups × 5 images)**	**90.70%**	**79.96%**	**94.29%**	**0.97**	**0.89**	**1.00**
✓	5 images/eye	5 (1 group × 5 images)	86.03%	63.33%	90.37%	0.96	0.78	0.97
✓	1 image/eye	25 (25 groups × 1 image)	88.09%	71.03%	92.82%	0.96	0.88	0.98

Seg, segmentation; AUC, area under the curve; C, control; FF, forme fruste keratoconus; KC, clinical keratoconus.

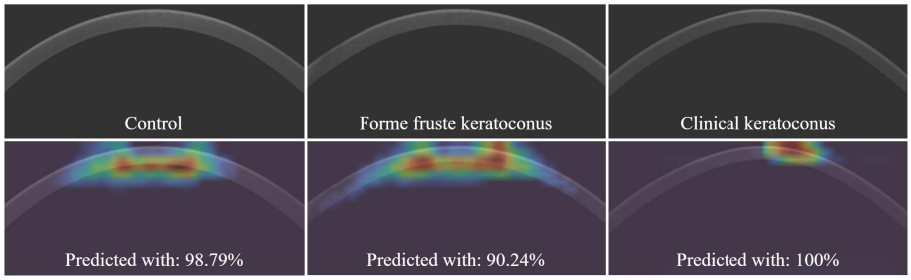

Fig. 7. GradCAM visualizations for the three classes. Top row: Input images. Bottom row: Corresponding GradCAM heatmaps showing areas of high predictive influence, together with prediction accuracies (in all three cases the correct class is predicted).

Biomedical Optics EXPRESS

We also designed an additional experiment to assess whether the model relies on tissue-related features in addition to shape information. For this purpose, the input images were binarized to preserve only the corneal contour, thereby isolating shape information. Table S3 in the Supplement 1 presents the model's performance on these binarized images compared to its performance on the original input images. A marked drop in performance when using the binarized data supports the hypothesis that tissue-related features are a key component in the model's predictive ability, especially in FF cases.

GradCAM heatmaps [46] were also utilized, in order to enhance model interpretability by visually indicating which areas of the input image were most influential in the model's decision-making process, offering insights into how the model was working internally. Figure 7 shows GradCAM visualizations for sample images from the three classes: C, FF, and KC. In both control and FF keratoconus, which are similar in shape between them, the model focuses on a large central portion of the cornea. In the case of clinical keratoconus, the area of influence is much smaller, confined, and located in a region of high curvature.

4. Discussion

This study presents the, to our knowledge, first application of deep learning to raw Scheimpflug images for keratoconus detection, a significant departure from previous approaches that relied on preprocessed tomographic maps. The complexity of the problem is heightened by our focus on FF keratoconus, a condition that lacks distinct topographic alterations when compared to control eyes, making its detection inherently challenging. Accurate identification of keratoconus, particularly in this preclinical stage, remains a major challenge in ophthalmology. This detection is crucial for preventing iatrogenic corneal ectasia following refractive surgery. FF keratoconus serves as a surrogate for the preclinical stage, since it represents structurally predisposed eyes despite normal topography. Our model demonstrates the feasibility of using Scheimpflug images to detect subtle corneal changes, achieving a sensitivity of 80.57%, a specificity of 80.56% at a 0.20 cut-off, and an AUC of 0.89 for FF classification (see Figs. 5 and 6(B)). Notably, this threshold is in line with values reported in the literature for similar early-stage keratoconus detection tasks [19,24]. For clinical keratoconus detection, the model achieves higher performance, reaching a sensitivity of 93.28%, a specificity of 99.40%, and an AUC of 1.00 for KC classification (Table 1). This superior performance is expected, as KC cases typically exhibit more noticeable and distinct changes in corneal shape and alterations in corneal tissue appearance, making them easier to identify than C or FF cases, which tend to share similar morphological characteristics.

Our proposed model achieves an overall accuracy of 90.70%, which falls within the range of accuracy reported in the literature (81%–99%) [20–22,25,27–32,47–51]. While some studies have reported higher values within this range, many of these focused on classification tasks that were inherently less challenging, often distinguishing only between clinical keratoconus and normal eyes [22,29–31,47,48]. Additionally, in cases where subclinical keratoconus was considered, the definition was often imprecise, with many so-called subclinical cases likely representing early keratoconus already exhibiting slight topographic alterations [20,21,25,27,28,32]. This distinction is crucial, as our study specifically targets FF keratoconus, a stage where conventional topographic methods fail to detect structural abnormalities, making the classification task significantly more complex. The GradCAM heatmaps (see Fig. 7) support the model's interpretability. Interestingly, control and FF cases often lead to activations in a large central portion of the cornea, while clinical keratoconus cases the focus is on more specific areas of high curvature. This could be consistent with the model focusing also on corneal tissue in the former cases (C and FF), and mainly on shape in the latter (KC), but more investigation would be needed to derive conclusions in this direction. This provides additional indication that the model is not relying on spurious correlations but rather learning meaningful diagnostic features.

Biomedical Optics EXPRESS

The use of raw Scheimpflug images provides a unique opportunity to simultaneously infer both corneal shape and tissue characteristics, extending beyond the limitations of prior studies that predominantly focused on topographical parameters, i.e., shape analysis. For results of an additional experiment using binarized images—designed to isolate shape information and remove tissue-related features—please refer to the Supplement 1. This perspective is supported by previous research that has explored corneal tissue properties (e.g., corneal densitometry) as an alternative approach, recognizing its potential in detecting subclinical keratoconus. In this context, Koc et al. [35] and Consejo et al. [34] pioneered corneal densitometry-based methods for subclinical keratoconus detection; however, they employed traditional statistical techniques rather than artificial intelligence based approaches. Koc et al. reported a sensitivity of 75.0%, a specificity of 90.0%, and an AUC of 0.883 using densitometric analysis from built-in Pentacam HR software [35], while Consejo et al. improved upon this with their densitometry distribution analysis (DDA), achieving a sensitivity of 90.0%, a specificity of 95.0%, and an AUC of 0.97 [34]. Despite these promising results, the latter study was based on a limited sample of 40 eyes (20 C *vs* 20 FF), which may have inflated performance metrics due to reduced variability and more controlled conditions. In contrast, our approach includes 571 eyes (492 C *vs* 79 FF) and leverages deep learning to extract both structural and tissue-based features directly from raw images, eliminating dependence on predefined parameters and enhancing the model's generalizability. Additionally, those previous studies were conducted on Caucasian eyes and two classes, whereas our dataset consists primarily of Asian eyes and three classes, making direct comparisons challenging.

To optimize keratoconus classification, different combinations of techniques have been explored to address class imbalance, as shown in Table 2. As expected, the absence of data balancing techniques (last row) led to low sensitivity and specificity, confirming the necessity of class balancing techniques. Among the tested approaches, strong DA combined with ROS (first row) yielded the best performance, significantly improving FF sensitivity while maintaining overall accuracy. The reason for strong DA outperforming weak DA (second row) may be that the variations introduced by weak DA were not significant enough to make a considerable difference compared to the data without augmentation.

We also explored different schemes for the input data, including the influence of the segmentation step of our preprocessing stage, and the number of input images to our model (Table 3, the best scheme is shown in row 2). When segmentation was not applied (row 1 vs. row 2), overall performance decreased, highlighting the importance of this preprocessing step in enhancing model performance; besides, processing time was also improved from ten hours to two hours and a half, approximately. The use of several images per eye improved with respect to using one, even when using the same amount of training data (row 2 vs. row 4); this is probably due to having more information to conduct inference, as well as the potential of establishing relationships between different meridian views. Lastly, when not using the full training dataset (row 3 vs. row 2), sensitivity also dropped significantly, reinforcing the well-known importance of dataset size, as well as validating our approach of leveraging the full training dataset.

Despite its strengths, this work is not free of limitations. The dataset, while large relative to prior studies, remains geographically limited and may lead the model to over-rely on locally augmented features; further validation on larger and more diverse populations would help ensure generalizability. Additionally, the reliance on Pentacam HR images restricts direct comparisons with other imaging modalities such as anterior-segment OCT, which provides higher axial resolution. Future studies should explore whether combining Scheimpflug imaging with other modalities could further enhance performance.

In conclusion, this work demonstrates the feasibility of creating a supervised deep learning model that uses raw labeled corneal Scheimpflug tomography images. By considering both

Biomedical Optics EXPRESS

corneal shape and tissue, it becomes a novel contribution to the field and serves as a starting point for further research.

Funding. HORIZON EUROPE European Research Council (101162733); Aragon Institute for Engineering Research (IMPULSO2023); Science and Technology Department of Sichuan Province (No. 2024YFFK0302); Postdoctoral Research Fund of West China Hospital, Sichuan University.

Acknowledgment. The authors would like to thank Maria Romanos-Ibanez for her valuable support in the preprocessing of the data.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. Y. Rabinowitz, "Keratoconus," Surv Ophthalmol. **42**(4), 297–319 (1998).
2. D. Piñero, J. Nieto, and A. Lopez-Miguel, "Characterization of corneal structure in keratoconus," J. Cataract. & Refract. Surg. **38**(12), 2167–2183 (2012).
3. N. Chatzis and F. Hafezi, "Progression of keratoconus and efficacy of pediatric [corrected] corneal collagen cross-linking in children and adolescents," J. Refract. Surg. **28**(11), 753–758 (2012).
4. D. A. Godefrooij, G. A. de Wit, C. S. Uiterwaal, *et al.*, "Age-specific incidence and prevalence of keratoconus: A nationwide registration study," Am. J. Ophthalmol. **175**, 169–172 (2017).
5. H. Hashemi, S. Heydarian, E. Hooshmand, *et al.*, "The prevalence and risk factors for keratoconus: A systematic review and meta-analysis," Cornea **39**(2), 263–270 (2020).
6. A. Saad and D. Gatinel, "Topographic and tomographic properties of forme fruste keratoconus corneas," Invest. Ophthalmol. Vis. Sci. **51**(11), 5546–5555 (2010).
7. R. Ambrósio and J. B. Randleman, "Screening for ectasia risk: what are we screening for and how should we screen for it?" J. Refract. Surg. **29**(4), 230–232 (2013).
8. J. F. Pérez, A. V. Marcos, and F. J. Pe na, "Early diagnosis of keratoconus: what difference is it making?" Br. J. Ophthalmol. **98**(11), 1465–1466 (2014).
9. M. R. Santhiago, N. T. Giacomin, D. Smadja, *et al.*, "Ectasia risk factors in refractive surgery," Clin. Ophthalmol. **10**, 713–720 (2016).
10. B. A. Holden, T. R. Fricke, D. A. Wilson, *et al.*, "Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050," Ophthalmology **123**(5), 1036–1042 (2016).
11. J. Duncan, M. Belin, and M. Borgstrom, "Assessing progression of keratoconus: novel tomographic determinants," Eye and Vis. **3**(1), 6 (2016).
12. M. Jiménez-García, S. Dhubhghaill, C. Koppen, *et al.*, "Baseline findings in the retrospective digital computer analysis of keratoconus evolution (REDCAKE) Project," Cornea **40**(2), 156–167 (2021).
13. N. J. Lu, F. Hafezi, C. Koppen, *et al.*, "New keratoconus staging system based on oct," J. Cataract. & Refract. Surg. **49**(11), 1098–1105 (2023).
14. I. Issarti, A. Consejo, M. Jiménez-García, *et al.*, "Computer aided diagnosis for suspect keratoconus detection," Computers Biol. Med. **109**, 33–42 (2019).
15. R. Ambrósio Jr., A. P. Machado, E. Leão, *et al.*, "Optimized artificial intelligence for enhanced ectasia detection using scheimpflug-based corneal tomography and biomechanical data," Am. J. Ophthalmol. **251**, 126–142 (2023).
16. M. Smolek and S. Klyce, "Current keratoconus detection methods compared with a neural network approach," Invest. Ophthalmol. Vis. Sci. **38**(11), 2290–2299 (1997).
17. P. Chastang, V. Borderie, S. Carvajal-Gonzalez, *et al.*, "Automated keratoconus detection using the EyeSys videokeratoscope," J. Cataract. & Refract. Surg. **26**(5), 675–683 (2000).
18. A. Lavric, V. Popa, H. Takahashi, *et al.*, "Detecting keratoconus from corneal imaging data using machine learning," IEEE Access **8**, 149113–149121 (2020).
19. N. J. Lu, C. Koppen, F. Hafezi, *et al.*, "Combinations of scheimpflug tomography, ocular coherence tomography and air-puff tonometry improve the detection of keratoconus," Cont. Lens Anterior Eye **46**(3), 101840 (2023).
20. R. Feng, Z. Xu, X. Zheng, *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 J. Biomed. Health Inform. **25**(10), 3898–3910 (2021).
21. A. H. Al-Timemy, L. Alzubaidi, Z. M. Mosa, *et al.*, "A deep feature fusion of improved suspected keratoconus detection with deep learning," Diagnostics **13**(10), 1689 (2023).
22. X. Chen, J. Zhao, K. C. Iselin, *et al.*, "Keratoconus detection of changes using deep learning of colour-coded maps," BMJ Open Ophth. **6**(1), e000824 (2021).
23. P. Zéboulon, G. Debellemanière, M. Bouvet, *et al.*, "Corneal topography raw data classification using a convolutional neural network," Am. J. Ophthalmol. **219**, 33–39 (2020).

Biomedical Optics EXPRESS

24. B. Kuo, W. Chang, T. Liao, *et al.*, "Keratoconus screening based on deep learning approach of corneal topography," Trans. Vis. Sci. Tech. **9**(2), 53 (2020).
25. W. Quanchareonsap, N. Kasetsuwan, U. Reinprayoon, *et al.*, "Deep learning algorithm for keratoconus detection from tomographic maps and corneal biomechanics: A diagnostic study," J. Curr. Ophthalmol. **36**(1), 46–53 (2024).
26. K. Kamiya, Y. Ayatsuka, Y. Kato, *et al.*, "Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study," BMJ Open **9**(9), e031313 (2019).
27. H. B. Gao, Z. G. Pan, M. X. Shen, *et al.*, "KeratoScreen: Early keratoconus classification with Zernike polynomial using deep learning," Cornea **41**(9), 1158–1165 (2022).
28. H. Abdelmotaal, M. M. Mostafa, A. N. R. Mostafa, *et al.*, "Classification of color-coded Scheimpflug camera corneal tomography images using deep learning," Translational Vision Science & Technology **9**(13), 30 (2020).
29. Z. Agharezaei, R. Firouzi, S. Hassanzadeh, *et al.*, "Computer-aided diagnosis of keratoconus through VAE-augmented images using deep learning," Sci. Rep. **13**(1), 20586 (2023).
30. A. H. Al-Timemy, N. H. Ghaeb, Z. M. Mosa, *et al.*, "Deep transfer learning for improved detection of keratoconus using corneal topographic maps," Cogn. Comput. **14**(5), 1627–1642 (2022).
31. M. Fırat, C. Çankaya, A. Çınar, *et al.*, "Automatic detection of keratoconus on pentacam images using feature selection based on deep learning," Int. J. Imaging Syst. Tech. **32**(5), 1548–1560 (2022).
32. A. H. Al-Timemy, Z. M. Mosa, Z. Alyasseri, *et al.*, "A hybrid deep learning construct for detecting keratoconus from corneal maps," Trans. Vis. Sci. Tech. **10**(14), 16 (2021).
33. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**(7553), 436–444 (2015).
34. A. Consejo, M. Jiménez-García, I. Issarti, *et al.*, "Detection of subclinical keratoconus with a validated alternative method to corneal densitometry," Transl. Vis. Sci. Technol. **10**(9), 32 (2021).
35. M. Koc, K. Tekin, M. Tekin, *et al.*, "An early finding of keratoconus: increase in corneal densitometry," Cornea **37**(5), 580–586 (2018).
36. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, (2015).
37. J. Deng, W. Dong, R. Socher, *et al.*, "ImageNet: A large-scale hierarchical image database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE, 2009), pp. 248–255.
38. "Accuracy metrics," https://keras.io/api/metrics/accuracy_metrics/. Accessed: 2025-02-12.
39. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, (2015).
40. M. Buda, A. Maki, and M. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Networks **106**, 249–259 (2018).
41. C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data **6**(1), 60 (2019).
42. D. Schaudt, R. von Schwerin, A. Hafner, *et al.*, "Augmentation strategies for an imbalanced learning problem on a novel COVID-19 severity dataset," Sci. Rep. **13**(1), 18299 (2023).
43. M. Q. Salomão, A. L. Hofling-Lima, L. P. Esporcatte, *et al.*, "Ectatic diseases," Exp. Eye Res. **202**, 108347 (2021).
44. M. W. Belin, H. S. Jang, and M. Borgstrom, "Keratoconus: Diagnosis and staging," Cornea **41**(1), 1–11 (2022).
45. A. Consejo, D. Alonso-Caneiro, M. Wojtkowski, *et al.*, "Corneal tissue properties following scleral lens wear using Scheimpflug imaging," Ophthalmic Physiol. Opt. **40**(5), 595–606 (2020).
46. R. Selvaraju, M. Cogswell, A. Das, *et al.*, "Grad-CAM: visual explanations from deep networks via gradient-based localization," Int. J. Comput. Vis. **128**(2), 336–359 (2020).
47. H. Abdelmotaal, R. M. Hazarbassanov, R. Salouti, *et al.*, "Keratoconus detection-based on dynamic corneal deformation videos using deep learning," Ophthalmology Science **4**(2), 100380 (2024).
48. L. M. Hartmann, D. S. Langhans, V. Eggarter, *et al.*, "Keratoconus progression determined at the first visit: A deep learning approach with fusion of imaging and numerical clinical data," Trans. Vis. Sci. Tech. **13**(5), 7 (2024).
49. K. Y. Tey, E. Z. K. Cheong, and M. Ang, "Potential applications of artificial intelligence in image analysis in cornea diseases: A review," Eye and Vis. **11**(1), 10 (2024).
50. N. S. Bodmer, D. G. Christensen, L. M. Bachmann, *et al.*, "Deep learning models used in the diagnostic workup of keratoconus: A systematic review and exploratory meta-analysis," Cornea **43**(7), 916–931 (2024).
51. Z. J. Muhsin, R. Qahwaji, I. Ghafir, *et al.*, "Advances in machine learning for keratoconus diagnosis," Int. Ophthalmol. **45**(1), 128 (2025).