• Intelligent Ophthalmology •

Artificial intelligence for diagnosis of keratoconus using Scheimpflug based corneal tomography

Sadaf Qayyum, Memoona Arshad, Humaima Saeed

Department of Optometry, The University of Faisalabad, Faisalabad 38000, Punjab, Pakistan
Correspondence to: Sadaf Qayyum. Flat 16, Block 2 Al-Shifa Trust Eye Hospital Rawalpindi, Rawalpindi 46000, Punjab, Pakistan. 2023-phd-op-001@tuf.edu.pk
Received: 2025-12-28 Accepted: 2026-02-06

Abstract

● **AIM:** To develop and evaluate the diagnostic accuracy of deep learning (DL) models in differentiating keratoconus (KC) from normal eyes with regular astigmatism.

● **METHODS:** A comparative cross-sectional study was conducted at the Cornea and Diagnostic Department of Al-Shifa Trust Eye Hospital, Pakistan. Galilei dual Scheimpflug-based corneal topography was performed to obtain four corneal maps: anterior axial curvature, posterior axial curvature, corneal thickness, and posterior elevation. Four convolutional neural network models were developed and trained on corneal maps to classify eyes as KC and normal. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity.

● **RESULTS:** A total of 5602 corneal maps were extracted from 1411 eyes (790 KC and 621 normal) of 827 participants, including KC (472) and normal (355) groups, aged 10 to 40y. The DL models achieved the highest accuracy with DenseNet-121 (99.2%), ResNet-50 (99.0%), Inception-V3 (98.6%), and EfficientNet-B0 (98.1%). DenseNet-121 and ResNet-50 achieved an AUC of 1.00. External validation on an independent dataset of 85 participants (150 eyes with 1050 extracted corneal maps) confirmed excellent accuracies for EfficientNet-B0 (98.1%), DenseNet-121 (98.3%), and ResNet-50 (97.1%).

● **CONCLUSION:** All DL models demonstrate excellent diagnostic accuracy for KC detection, highlighting the potential for clinical implementation and optimized KC management with greater precision.

● **KEYWORDS:** artificial intelligence; convolutional neural network; corneal topography; keratoconus
DOI:10.18240/ijo.2026.07.02

Citation: Qayyum S, Arshad M, Saeed H. Artificial intelligence for diagnosis of keratoconus using Scheimpflug based corneal tomography. *Int J Ophthalmol* 2026;19(7):1221-1234

INTRODUCTION

Keratoconus (KC) is characterized by localized thinning of the central or paracentral cornea, which causes irregular astigmatism and a progressive decrease in visual acuity. If left untreated, this condition may lead to significant vision loss, profoundly affecting the patient's quality of life[1]. KC affects approximately 1 in 2000 individuals globally, with its prevalence increasing significantly in recent years, particularly in Asian and Middle Eastern populations[2]. In Southeast Asia, notably higher KC rates have been documented. In India, the prevalence of KC is 2.3%, while 1% of the population in China is affected[3], and a prevalence of 4.79% has been reported in the Saudi Arabian pediatric population[4]. The development of KC is influenced by a combination of genetic and environmental factors, indicating its multifactorial nature[5].

Certain risk factors have been associated with disease progression. Patients with a previous history of allergic eye disease or systemic atopic disease are at increased risk of developing aggressive presentations of KC[6]. Early recognition of KC is critical, particularly to enable timely intervention with treatments such as corneal collagen cross-linking, which can halt disease progression and preserve visual function. However, detecting KC in its earliest form, especially the subclinical variant known as forme fruste KC, remains challenging. Traditional Placido disk-based corneal topography has long served as a valuable diagnostic tool; however, it has key limitations. Most notably, it lacks the capacity to directly assess the posterior corneal surface, where early ectatic changes often first appear[7].

To address the limitations of conventional topography in detecting early ectatic changes, Scheimpflug-based corneal tomography offers a more advanced imaging modality. It operates on the principle that tilting the camera plane relative to the lens and image planes allows for an extended depth of focus, facilitating sharp imaging of the entire corneal cross-section[8]. Modern rotating Scheimpflug systems, such

as Pentacam, Galilei, and Orbscan capture multiple radial portions, typically 25 to 50, within seconds. These images are then used to reconstruct a detailed three-dimensional model of the cornea, significantly enhancing diagnostic precision in KC assessment[9-10].

The zone of increasing corneal power, inferior-superior (IS) corneal power asymmetry, skewing of the steepest radial axis, and other combined indices are among the indicators identified to distinguish between KC and normal eyes. However, ophthalmologists face a clinical challenge due to the complex indices provided by corneal topography[11].

Current real-world applications of artificial intelligence (AI) in medicine encompass a wide range of functions. These include disease diagnosis, risk stratification, risk prediction, AI-assisted clinical decision-making, and personalized treatment planning[12]. Deep learning (DL) algorithms in ophthalmology are typically developed using a disease-based learning approach. In this method, clinicians annotate specific, well-defined disease features on the image data, which the algorithm subsequently learns. Currently, the most promising AI tools in ophthalmology are applied in the field of retinal diseases, particularly for the detection of diabetic retinopathy[13], age-related macular degeneration[14], and retinopathy of prematurity[15]. However, AI models are also being developed for a variety of anterior segment conditions, such as glaucoma[16], KC, cataract, and oculoplastic disorders[17].

Machine learning (ML) algorithms build predictive models, such as logistic regression, artificial neural networks, and decision trees, by learning from sample training datasets. These models make data-driven decisions without requiring explicit programming for each scenario. Depending on the availability and nature of labeled data in the training set, either supervised or unsupervised learning can be applied[18]. Several ML approaches have been employed for the detection of KC using Placido-based corneal topography, tomographic numerical indices using slit-scanning tomography, and dual rotating Scheimpflug cameras[19-23]. AI models based on Pentacam indices or combined parameters from Pentacam and optical coherence tomography are mainly supervised and have achieved area under curve (AUC) values ranging from 0.90 to 1.00[24-25].

To identify KC, several DL models, such as convolutional neural networks (CNNs), have been applied to color-coded corneal maps, including elevation, thickness, and curvature maps[26-27]. These studies utilized a smaller dataset, usually fewer than 400 images, despite the fact that DL required a larger data set[12]. Synthetic corneal images were utilized instead of real-world images, which improved model accuracy but made generalization of the results more challenging[28]. Agharezaei *et al*[29] demonstrated the effectiveness of DL models applied to corneal topographic maps for detecting KC. Their results showed that the DL model achieved high diagnostic performance when distinguishing between KC and normal eyes. These findings underscore the potential of DL to improve the diagnostic precision of KC.

Recently, DL models use whole images of color-coded corneal maps based on Scheimpflug image tomography, which determine thickness, curvature, and elevation maps of the anterior and posterior corneal surfaces to assess the diagnostic accuracy of KC. This study involves DL models that use pattern recognition to detect subtle topographic changes in KC and normal eyes. In this study, we utilized a total 5602 corneal images. Four DL CNN models based on transfer learning strategies were employed and fine-tuned on a substantial dataset to enhance the diagnostic accuracy of KC and normal group. Instead of using extensive topographic indices, each model learned to detect KC characteristics from individual corneal maps and a range of topographic patterns.

PARTICIPANTS AND METHODS

Ethical Approval This comparative cross-sectional study was conducted at the Cornea and Diagnostic Department of Al-Shifa Trust Eye Hospital, Rawalpindi, between February 2025 and August 2025, following approval from the Institutional Ethical Review Committee (ERC-15/AST-2025). This study was conducted in accordance with the principles outlined in the Declaration of Helsinki. Informed consent was obtained from all participants.

All participants underwent detailed ophthalmic examinations, including visual acuity assessment using a logMAR chart, subjective refraction using an auto refractometer (Topcon Corporation, Tokyo, Japan), objective refraction using a retinoscope (Heine), and anterior segment examination using slit-lamp bio microscopy (Topcon Corporation, Tokyo, Japan). Corneal tomography based on a Scheimpflug imaging system (Galilei, Ziemer, Port, Switzerland) was performed to collect corneal images. Topography images in the dataset were carefully reviewed by a team of corneal specialists to assess, diagnose, and label KC and normal eyes. A double-labeling technique was used, in which each image was labeled by two independent reviewers. Any discrepancies between the reviewers were resolved by a third reviewer. Reviewers were instructed to label each image as KC and normal.

The inclusion criteria for the KC group were a diagnosis based on clinical findings and corneal topographic parameters[12,30-31]. Clinical signs included central corneal protrusion, Fleischer's ring, Vogt's striae, and focal corneal thinning on slit-lamp examination[31]. Topographic criteria comprised a central K value >47 diopters (D), an IS value >1.4 D, and an asymmetric bow-tie pattern[12]. The normal group consisted of age-matched candidates for refractive surgery or contact lens fitting with

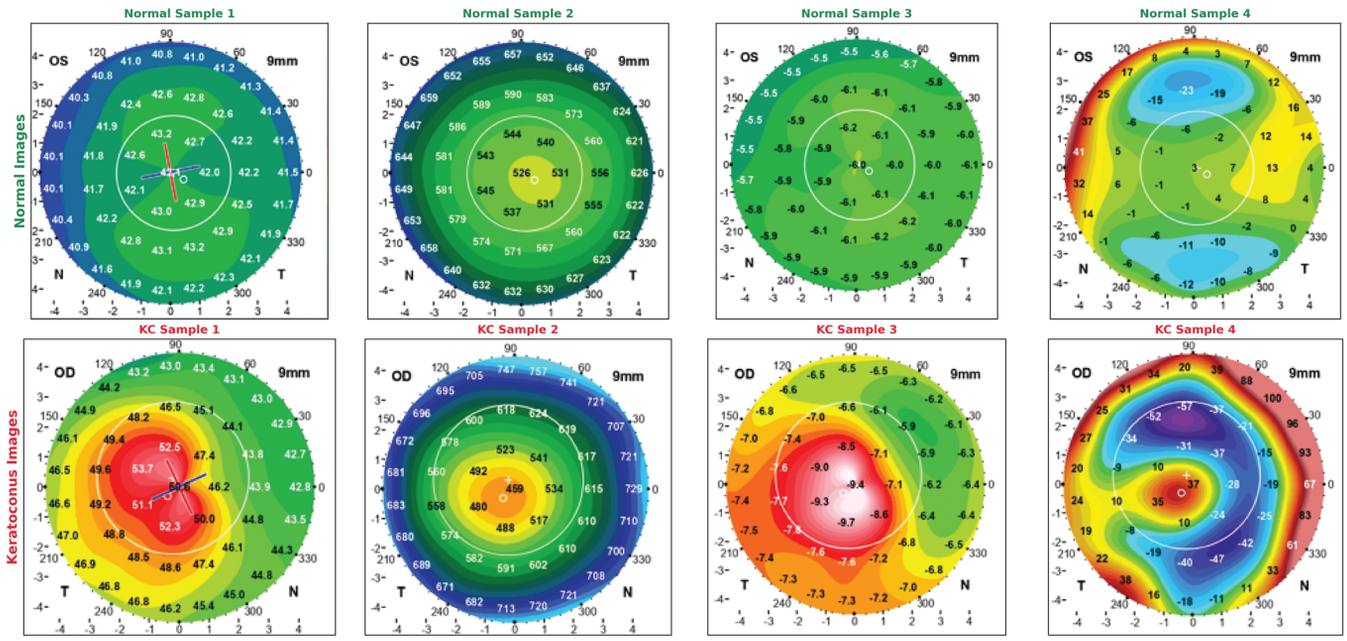

Figure 1 Example of four corneal topography maps Anterior axial curvature map, pachymetry map, posterior axial curvature, and posterior elevation map of keratoconus (KC) and normal patient.

refractive errors ≤8.0 D and astigmatism ≤3.0 D, without any clinical or topographic abnormalities. Exclusion criteria for all participants included inflammatory eye conditions, ocular surface diseases, glaucoma, systemic diseases affecting the eye, a history of ocular surgery, contact lens wear, or the use of topical medications other than artificial tears. The sample size of this study was 827 participants (472 KC and 355 normal).

Data Loading and Preprocessing Technique The training pipeline included a data-feeding methodology designed for KC and normal image classification. The PyTorch DataLoader framework was utilized with custom dataset classes (KC and normal) that address the requirements of complex imaging processes, quality validation, format standardization, and memory-efficient loading. Data feeding began with comprehensive preprocessing. Corneal images obtained from Galilei topography were saved in JPG format. Four corneal maps (anterior axial curvature, corneal pachymetry, posterior axial curvature, and posterior elevation) were automatically extracted from corneal topography images, as shown in Figure 1, using Python Enhanced Scan Selector V-2.

The custom PyQt5 GUI tool allowed manual selection of up to four circular regions of interest (ROIs) per corneal image, each extracted as a standardized 256×256-pixel scan. Each corneal tomography image contained distinct diagnostic feature maps for KC detection (one region of interest per corneal map, *i.e.*, anterior axial curvature, posterior axial curvature, corneal thickness, and posterior elevation). Four ROIs were extracted per image to ensure that the models could learn all clinically relevant areas of interest rather than missing important diagnostic features. By extracting four ROIs from each corneal image, the final dataset size increased to 5602 images from 1411 eyes. Four ROIs were drawn on each image, with each circle assigned a different color for identification, such as red, green, blue, and yellow. Resizing or repositioning of circles was performed using the adjustment mode. The spatial coordinates of ROIs were saved in JPG format for further computational processing.

Preprocessing Pipeline Using Python Imaging Libraries, images were loaded with automatic format conversion to handle different image formats. A bilateral filter was applied for noise reduction. This filter was applied selectively to reduce noise while preserving edge information critical for corneal structure assessment. The filter was applied only when image noise exceeded a threshold (SD>20), ensuring that high-quality images were not unnecessarily processed. This preserved sharp edges while reducing noise in homogeneous regions by considering both spatial proximity and intensity similarity during image smoothing, thereby maintaining anatomical details relevant to KC diagnosis.

Contrast Limited Adaptive Histogram Equalization (CLAHE) enhancement was applied to the L channel to enhance local contrast without oversaturating the image. The tile approach (8×8) was effective for corneal images with varying illumination, and a clip limit of 1.5 prevented over-enhancement that could cause artifacts. This approach was adopted to enhance localized contrast in topography images with non-uniform illumination while preserving clinically relevant features. All images were resized to 256×256 pixels while preserving the original aspect ratio by maintaining proportional relationships between different corneal regions;

padding was added to achieve uniform input dimensions. The final step included brightness adjustment and quality control, which corrected systematic brightness variations across images and ensured pixel values were within the range of 0–255. Preprocessed images were then converted to float32 tensors with values scaled to the range of 0–1. Figure 2 represents visualization of preprocessing techniques.

Augmentation Strategy During Training Process To address the limitation of small sample sizes in previous studies, which is particularly critical for training DL models, we utilized data augmentation techniques applied only to the training dataset by applying different transformations, while maintaining unmodified images for the testing and validation datasets. This approach is particularly suitable for medical images, where a balance between data diversity and preservation of diagnostic features is critical. For model training, augmentation techniques used conservative parameters that provided data variation while preserving diagnostic features.

Augmentation was applied only to the training data to mimic realistic image variations while preserving disease-related topographic patterns. The augmentation pipeline included horizontal flipping (P=0.2), applied with a 20% probability, as corneal topography is generally symmetrical, making horizontal flipping clinically reasonable; this low probability preserved the original orientation in approximately 80% of cases. Additionally, small rotations (≤3°, P=0.2) introduced realistic variations in patient positioning while preserving diagnostic integrity. Brightness and contrast adjustments were applied with a limit of 0.05 (probability 0.2). Small variations (5%) represented realistic differences in imaging conditions, and the low probability ensured that most images retained their original appearance, as shown in Figure 3.

Validation and Test Data Augmentation Data were split before any augmentation. A stratified split strategy was used to divide the data into three independent datasets. The training set includes 60% of the data for model training. Augmentation was applied only to the training dataset. The validation and testing sets each included 20% of the data. No augmentation was applied to the validation or testing sets. All corneal images were evaluated in their original form to ensure accurate and unbiased model evaluation. No normalization was applied during image preprocessing to prevent potential data leakage.

Deep Learning Model Architecture and Modifications This study included four DL models based on CNN-based pretrained architectures with transfer learning to identify KC and normal images. To utilize transfer learning, DL models were employed as a backbone feature extractor. These included ResNet-50[32], DenseNet-121[33], EfficientNet-B0[34], and Inception-V3[32]. These models were initialized with ImageNet pretrained weights to incorporate rich feature representations. According to the requirements of each architecture, input images were resized (256×256 for ResNet-50, DenseNet-121, and EfficientNet-B0; 299×299 for Inception-V3). These different resolutions were chosen because of model-specific requirements (default input size).

Global average pooling (GAP) was applied after the convolutional backbone to reduce spatial dimensions while preserving discriminative information. Depending on the architecture, this resulted in fixed-length feature vectors of varying dimensionality (ResNet-50: 2048-D, DenseNet-121: 1024-D, EfficientNet-B0: 1280-D, Inception-V3: 2048-D). For binary classification (KC and normal), to adapt generic representation, the original classification layers of pertained models were replaced with custom classifier head. The architecture was comprised of the following sequence of layers. The GAP layer incorporated spatial map information into fixed-length vector representations. These feature vectors were then projected into a fully connected layer of 512 neurons with rectified linear unit (ReLU) activation, enabling dimensionality reduction. A dropout layer with a rate of 50% was applied to reduce overfitting, followed by a second fully connected layer with 256 neurons and ReLU activation to refine discriminative features. To further improve generalization, another dropout layer with a rate of 30% was added. Through a fully connected layer of two neurons, the final network output was mapped using a softmax activation function, which generated class probabilities for the KC and normal categories, as depicted in Figure 4. This architecture adopted a progressive dropout strategy (0.5–0.3), following a comprehensive regularization approach.

First layer (0.5 dropout): High dropout prevented overfitting in the high-dimensional feature space; Second layer (0.3 dropout): Moderate dropout allowed greater information flow for final classification; Final layer (no dropout): Ensured maximum information flow for the final decision. The combination of regularization techniques achieved optimal performance by preventing overfitting. Batch normalization was not included in the custom classifier head. This follows standard practice of keeping the classifier head unchanged. Since this design is consistent with standard transfer learning strategies. Batch normalization plays a critical role in deep convolutional learning models, which are already pretrained with batch normalization layers, whereas the custom classifier head is shallow and does not face severe internal covariate shift. However, adding batch normalization to the shallow classifier head with a small batch size can affect training stability and reduce model performance.

For binary classification (KC and normal), ImageNet pretrained weights were adapted according to architecture-specific modifications. A fine-tuning approach with pretrained weights

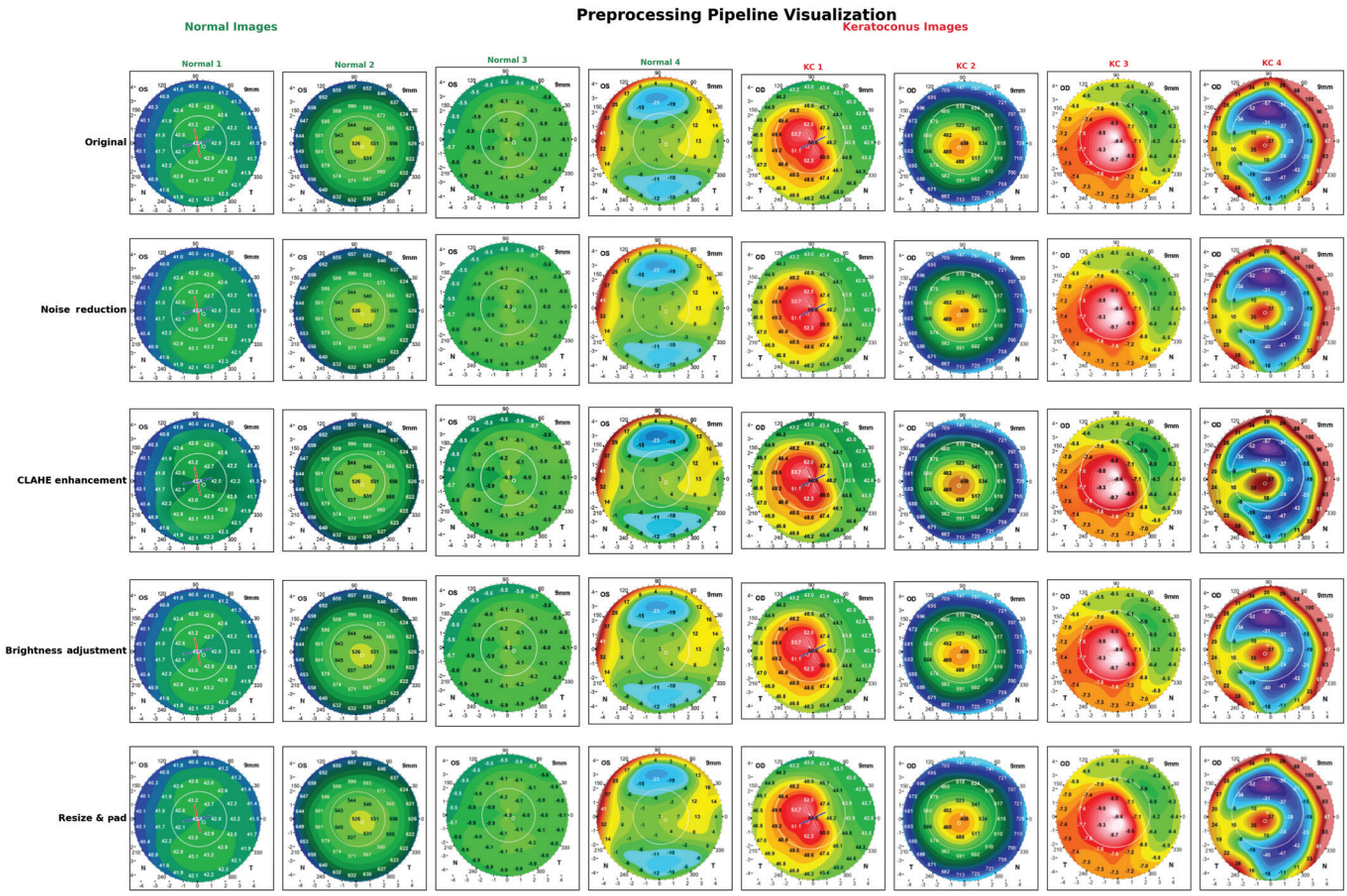

Figure 2 Visualization of preprocessing techniques applied to KC and normal maps KC: Keratoconus.

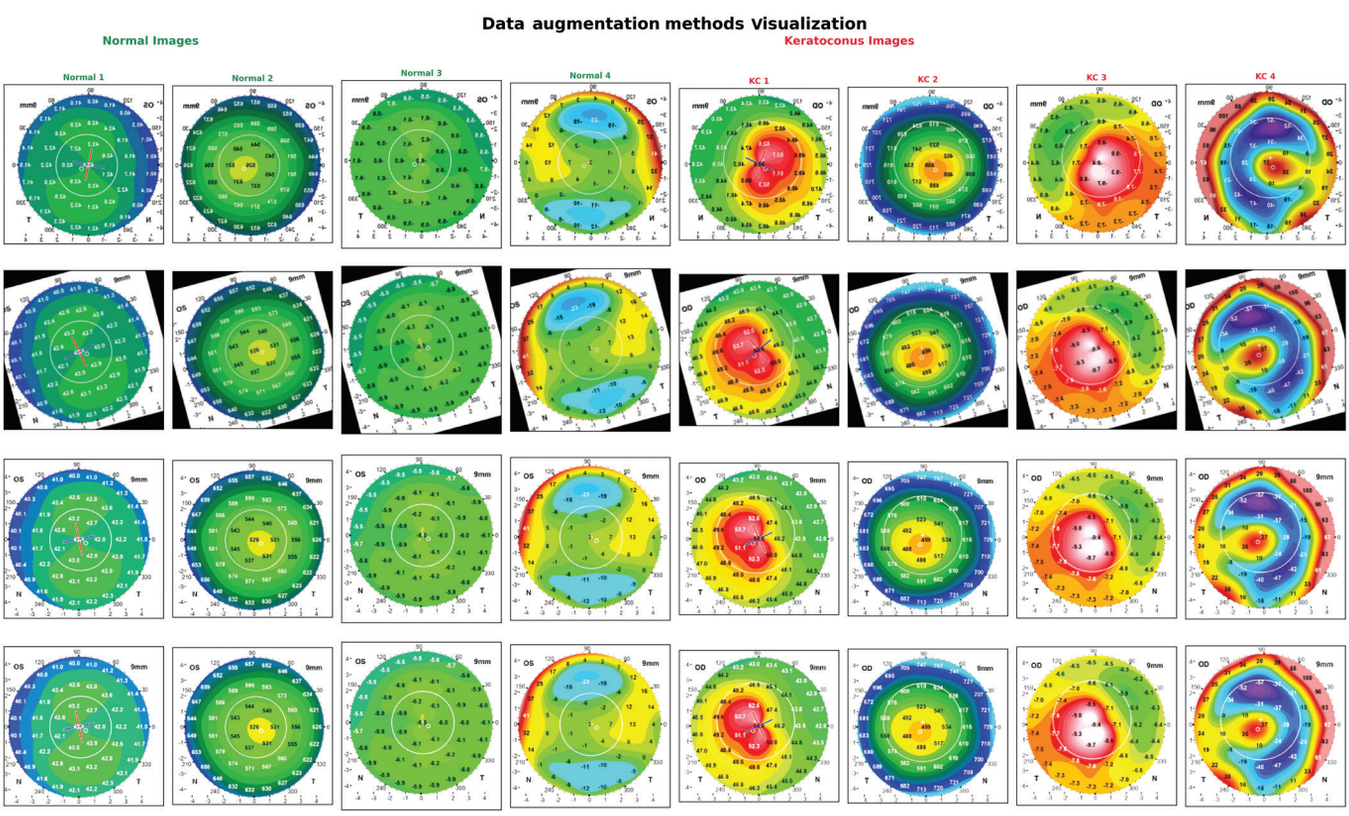

Figure 3 Visualization of augmentation techniques applied to KC and normal maps Horizonatl flip, rotation, brightness adjustment, contrast adjustment. KC: Keratoconus.

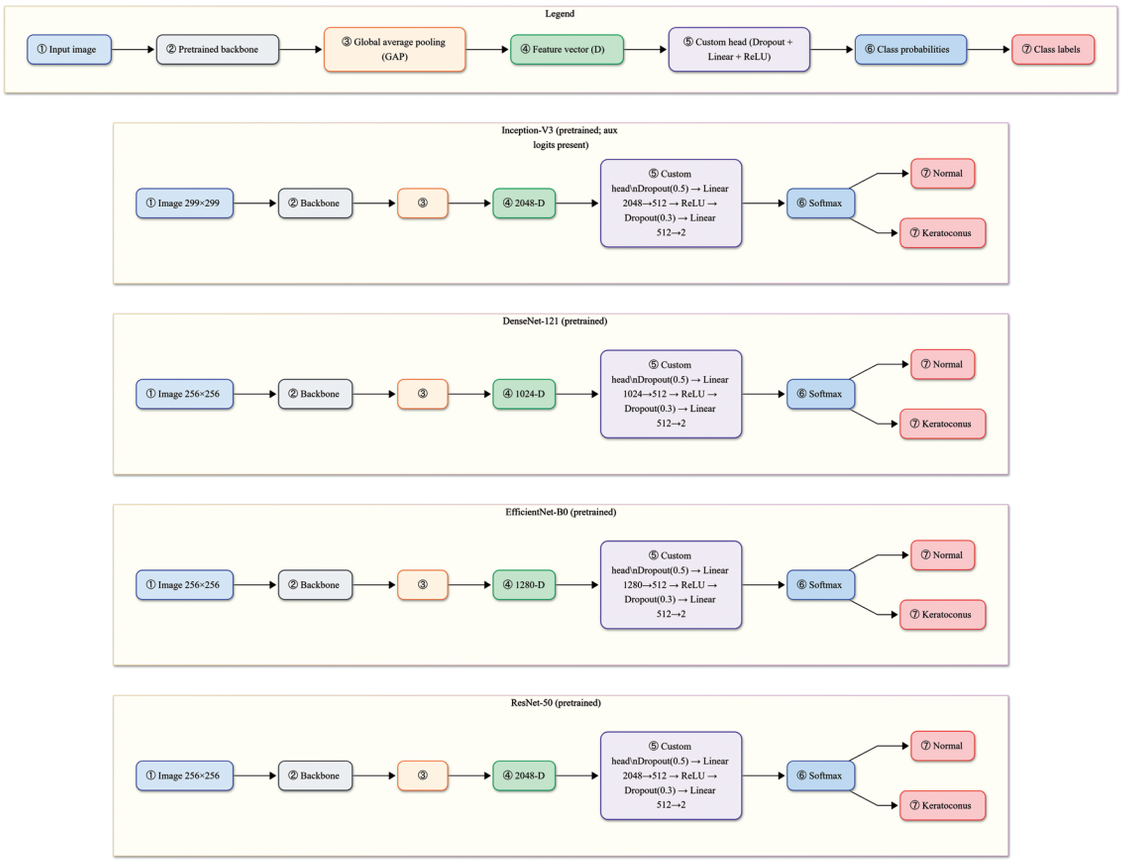

Figure 4 Architecture of CNN models for keratoconus prediction CNN: Convolutional neural network.

was adopted without an explicit freezing strategy. All models used standard PyTorch model implementations. To adapt DL models for classification of KC and normal eyes, a transfer learning approach was employed using pretrained ImageNet weights, which provided general feature extraction. A uniform learning rate strategy was adopted. For binary classification, the custom classifier head was trained from scratch with the addition of dropout layers to reduce overfitting. To ensure training stability in corneal image classification, the models were trained using carefully tuned hyperparameters, including a learning rate of 0.0001 and a weight decay of 0.001, providing balanced convergence and regularization. Training and validation accuracy and loss monitoring strategies were reported, as shown in Figure 4. For adaptive optimization and training stability, the Adam optimizer was employed.

Training was performed using a batch size of 16 for efficient GPU memory utilization. Each model was trained for a maximum of 50 epochs with early stopping (patience=30). The early stopping strategy was carefully adopted based on the complexity of the classification task and experimental validation. An epoch limit of 50 with a patience of 30 reflected the inherent complexity of the task, as binary classification has simpler decision boundaries, allowing faster convergence (within 30–40 epochs) and preventing overfitting.

Statistical Analysis For demographic analysis, a paired *t*-test was used to compare continuous parameters between the KC and normal groups. The Chi-square test was used to compare gender differences between the two groups, and all statistical analyses were carried out using Jeffrey's Amazing Statistics Program (JASP) software. A *P*-value of less than 0.05 was considered statistically significant. The performance of DL models was evaluated using the receiver operating characteristic (ROC) curve, confusion matrix, sensitivity, and specificity. Positive predictive value and negative predictive value were also calculated.

Accuracy: Proportion of correctly classified samples; Sensitivity: TP/TP+FN, measuring the ability to detect positive cases; Specificity: TN/TN+FP, evaluating correct classification of normal cases; AUC: quantifying discrimination across classifications thresholds.

RESULTS

The total number of participants in the KC group was 472 (790 KC eyes), including 287 males and 185 females, with a mean age of 22.00±8.02y, while the normal group consisted of 355 participants (621 normal eyes), including 174 males and 181 females, with a mean age of 23.18±8.57y. There was no statistically significant difference in age or gender between the KC and normal groups, indicating that the two groups were age- and gender-matched, as shown in Table 1. The comparison of topographic parameters between KC and

normal participants showed significant differences across all parameters between the two groups (Table 2).

Data Splitting A three-way data split strategy was implemented (training/validation/testing) with a 60/20/20 distribution, ensuring proper model evaluation. This study included a total of 827 participants (472 KC and 355 normal). A total of 1411 corneal images were collected (790 KC and 621 normal). Each image comprised four scans. Initially, a total of 5644 scans were extracted and poor image quality were excluded. Ultimately, 5602 high-quality scans were retained, including KC (3137) and normal (2465) scans. Augmentation was applied only to the training set (60%). A total of 10 083 images were utilized for model training (KC: 5646; normal: 4437). No augmentation was applied to the validation or testing datasets. Twenty percent of the data were allocated for validation (1120 images: 627 KC and 493 normal) and 20% for testing (1121 images: 628 KC and 493 normal), as depicted in Figure 5. This adopted approach contributed to robust model evaluation, reliability, and accuracy.

Four CNN models were utilized for the classification of KC and the normal group. Figure 6 shows training (blue) and validation (red) loss and accuracy for all models. All models showed strong performance, with high training and validation accuracy and minimal overfitting. The loss curves showed stable convergence with a minimal gap between training and validation. The AUC, as depicted in Figure 6, was 1.000 for DenseNet-121, 1.000 for ResNet-50, 0.998 for EfficientNet-B0, and 0.999 for Inception-V3.

A comparison of accuracy, sensitivity, specificity, positive predictive value, and negative predictive value is shown in Table 3. DenseNet-121 showed the highest accuracy 0.992 (with a sensitivity of 0.992 and specificity of 0.996), followed by ResNet-50 with an accuracy of 0.990 (sensitivity: 0.982; specificity: 0.997), Inception-V3 with an accuracy of 0.986 (sensitivity: 0.979; specificity: 0.990), and EfficientNet-B0 with an accuracy of 0.981 (sensitivity: 0.979; specificity: 0.982). All models exhibited acceptable performance, with the highest accuracy demonstrated by the DenseNet-121 model.

External validation was also performed using an independent dataset available from the Kaggle open data source for KC detection. The total number of participants was 280 (104 normal and 176 KC), and 542 corneal images were collected (338 KC and 204 normal). From each map, seven scans were extracted, resulting in a total of 3794 scans. These scans were used for model training. For validation and testing, another dataset was utilized using a three-split strategy, including a total of 85 participants (56 KC and 29 normal). A total of 150 corneal images were included (100 KC and 50 normal). The total number of scans extracted were 1050. A three-split technique was applied using a 60/20/20 training/validation/

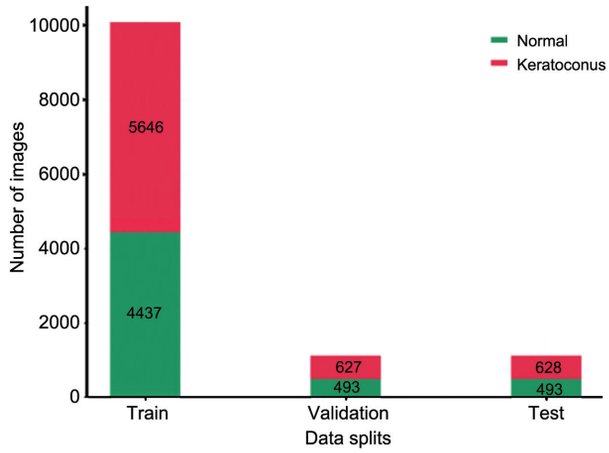

Figure 5 Details of splitting data for keratoconus and normal class distribution.

Table 1 Age and gender distribution between KC and normal participants

Variables	Keratoconus (*n*=472)	Normal (*n*=355)	*P*
Age, mean±SD	22.00±8.02	23.18±8.57	0.83
Gender, *n* (%)			
Male	287 (60.85)	174 (49.01)	0.81
Female	185 (39.20)	181 (50.98)	

KC: Keratoconus; SD: Standard deviation.

Table 2 Topographic parameter between keratoconus and normal groups mean±SD

Variables	Keratoconus (*n*=472)	Normal (*n*=355)	*P*
Mean K			
OD	49.65±3.81	43.52±1.04	<0.001
OS	49.15±3.30	43.48±1.11	<0.001
SRI			
OD	1.50±0.39	0.84±0.44	<0.001
OS	1.48±0.41	0.81±0.23	<0.001
SAI			
OD	3.58±1.70	0.80±0.50	<0.001
OS	3.64±2.03	0.75±0.31	<0.001
KPI			
OD	74.75±29.05	3.88±5.33	<0.001
OS	74.12±28.99	4.03±5.94	<0.001
IS			
OD	5.74±3.69	0.82±1.08	<0.001
OS	8.82±4.56	0.84±0.72	<0.001
PPK.X			
OD	98.80±6.60	19.86±38.05	<0.001
OS	72.98±38.64	23.02±39.45	<0.001

K: Keratometry; SRI: Surface regularity index; SAI: Surface asymmetry index; KPI: Keratoconus predictive index; IS: Inferior superior index; PPK.X: Percent probability of keratoconus on X axis; OD: Right eye; OS: Left eye; SD: Standard deviation.

Table 3 Results of CNN models showing accuracy, sensitivity, and specificity

Models	Accuracy	Sensitivity	Specificity	PPV	NPV
DenseNet-121	0.992	0.992	0.996	0.994	0.991
ResNet-50	0.990	0.982	0.997	0.996	0.986
Inception-V3	0.986	0.979	0.990	0.987	0.983
EfficientNet-B0	0.981	0.979	0.982	0.977	0.984

CNN: Convolutional neural network; PPV: Postive predictive value; NPV: Negative predictive value.

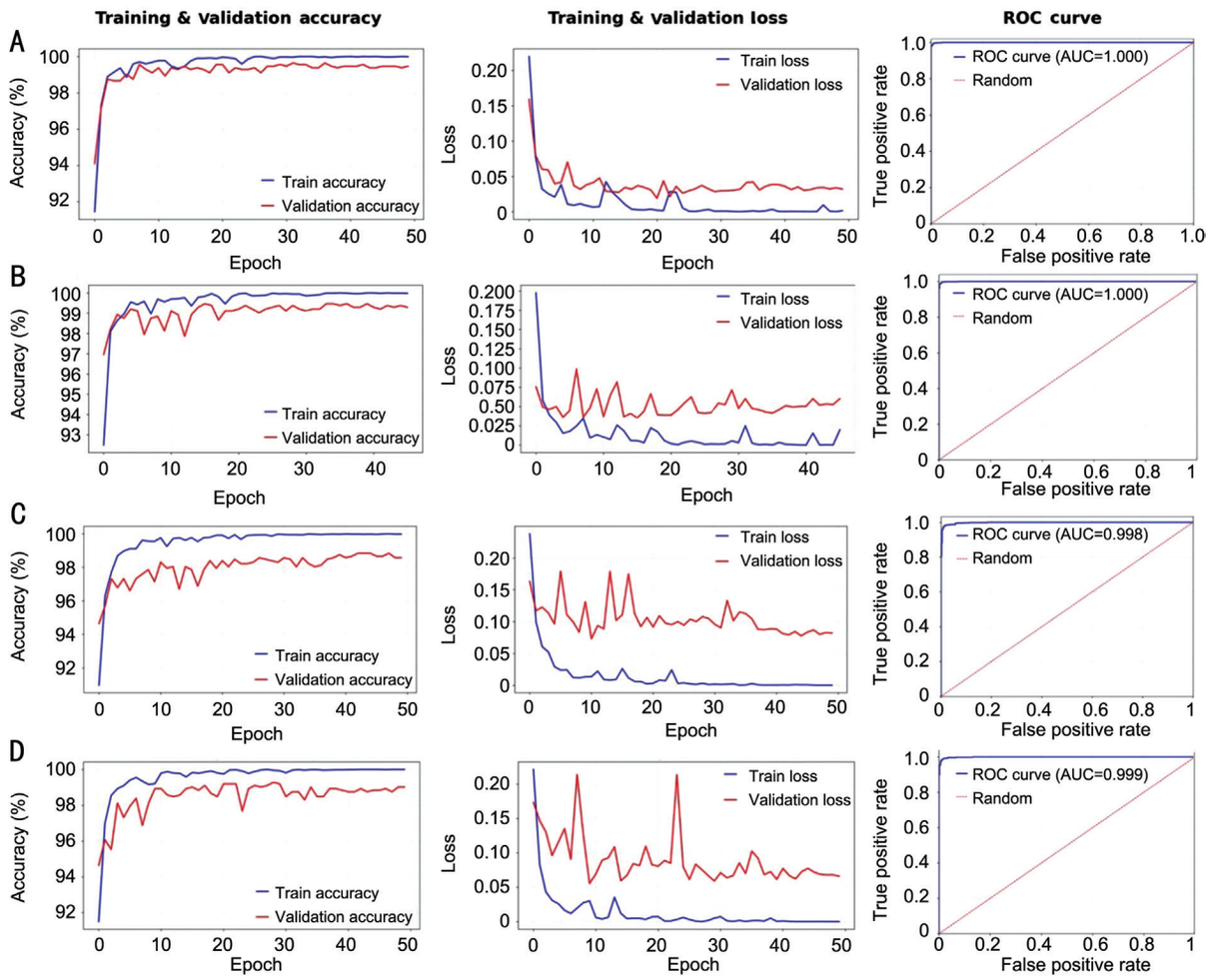

Figure 6 Training and validation accuracy, and loss curves, receiver operating characteristic curves of CNN models A: Training and validation accuracy curves, training and validation loss curves and AUC of DenseNet-121 model demonstrate stable learning behavior and high classification performance; B: Training and validation accuracy curves, training and validation loss curves and AUC of ResNet-50 model showed consistent convergence predicting discrimination between KC and normal eyes; C: Training and validation accuracy curves, training and validation loss curves and AUC of EfficientNet-B0 model demonstrate effective feature learning with minimal overfitting; D: Training and validation accuracy curves, training and validation loss curves and AUC of Inception-V3 model highlights stable training performance and efficient diagnostic performance. CNN: Convolutional neural network; AUC: Area under the curve; KC: Keratoconus; ROC: Receiver operating characteristic.

Table 4 Results of CNN models using independent data set obtained from Pentacam instrument

Models	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
DenseNet-121	0.983	0.9714	0.9952	0.9951	0.9721	0.992
EfficientNet-B0	0.981	0.9762	0.9857	0.9856	0.9764	0.998
ResNet-50	0.971	0.9571	0.9857	0.9853	0.9583	0.997

CNN: Convolutional neural network; PPV: Postive predictive value; NPV: Negative predictive value; AUC: Area under the curve.

testing distribution. This included 630 scans for training, 210 (20%) for validation, and 210 (20%) for testing.

Three of these models were assessed for generalizability to other clinical settings using 1050 corneal images from 85 participants obtained from the Pentacam system. The highest accuracy was obtained by DenseNet-121 (0.983), followed by EfficientNet-B0 (0.981). The differences in results were attributed to dataset variations and image variability associated with the Pentacam instrument. However, all models showed satisfactory results when applied to different datasets (Table 4).

The following Figure 7 shows a visual comparison of test dataset original images and predicted classes using the DenseNet-121 model as an example. The DenseNet-121 model correctly classified different KC maps, such as thickness maps showing clear thinning, anterior axial curvature maps showing asymmetric bow-tie patterns, and posterior elevation maps showing increased elevation, as well as normal maps, where thickness maps showed no thinning and anterior axial curvature maps showed symmetric patterns, with a confidence score of 1.000.

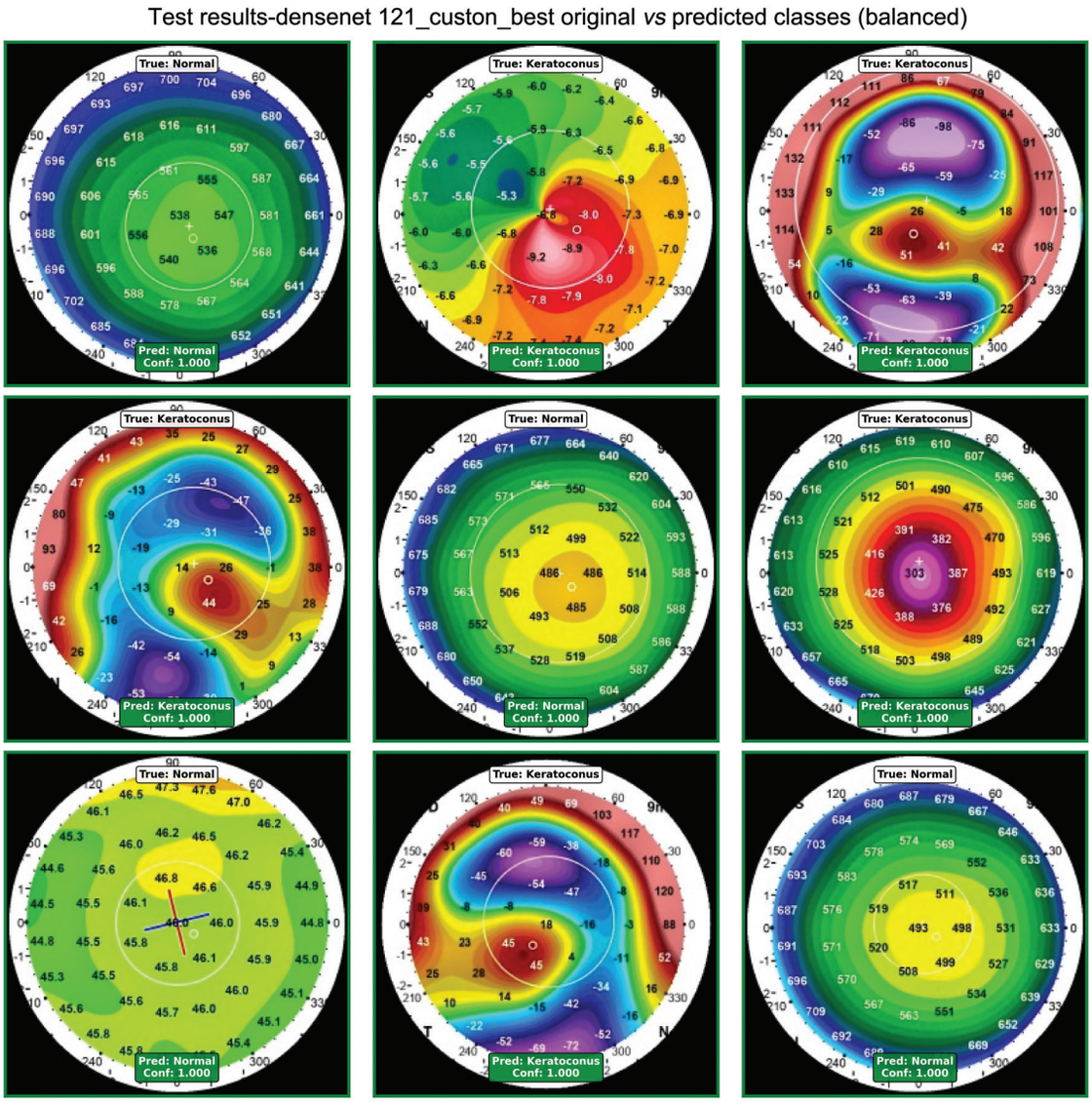

Figure 7 Comparison of original test data set (true classes) and predicted classes by CNN DenseNet-121 model CNN: Convolutional neural network.

Confusion matrices were also calculated for all models to assess model performance. Figure 8 shows that, out of 1121 test images, the DenseNet-121 model incorrectly classified nine images, with six misclassified as KC and three as normal. The ResNet-50 model incorrectly classified 11 images, of which 9 were misclassified as KC and 2 as normal. EfficientNet-B0 incorrectly classified 21 images, while Inception-V3 incorrectly classified 16 images.

The gradient-weighted class activation mapping (Grad-CAM) technique was adopted to provide detailed model interpretation and explanation. Heat maps generated by Grad-CAM highlighted regions of the input images, providing better visualization of feature maps. Grad-CAM works by computing gradients of the target class score with respect to the feature maps in the last convolutional layer. These gradients were used to weight feature maps, creating a coarse localization map that displayed essential regions in the image. Its implementation resulted in detailed visualizations that helped clinicians understand model decisions with respect to clinical markers (Figure 9).

DISCUSSION

In this study, the results showed that the use of DL with color-coded corneal maps obtained from Galilei (Ziemer, Port, Switzerland) Scheimpflug-based corneal tomography demonstrated high accuracy in distinguishing KC from normal eyes. Four DL models DenseNet-121, ResNet-50, Inception-V3, and EfficientNet-B0 were developed, and all models showed excellent accuracy in discriminating between KC and normal eyes. DenseNet-121 showed the highest accuracy (0.992) among all models, followed by ResNet-50 (0.990), Inception-V3 (0.986), and EfficientNet-B0 (0.981).

This study included a large customized dataset of 5602 images from 827 participants, which is appropriate for training DL models. Manual augmentation strategies were applied only

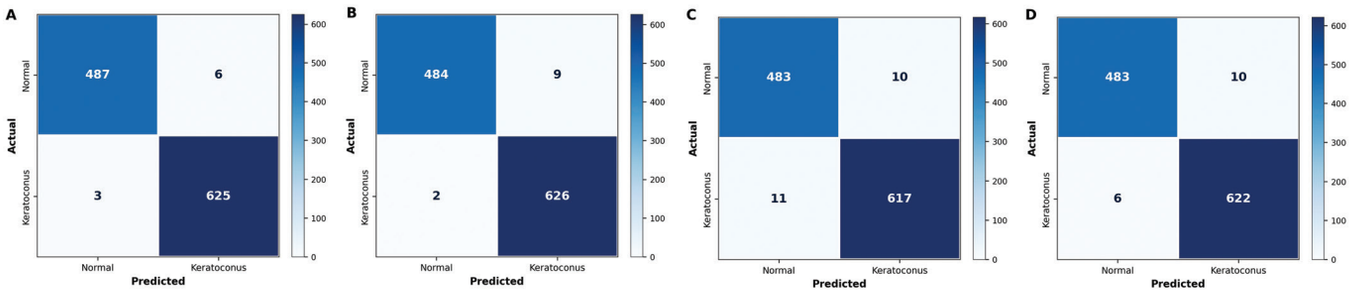

Figure 8 Confusion matrices of all models for classification of KC A: Confusion matrix of DenseNet-121 demonstrated high number of correctly classfied KC and normal groups achieving strong classification performance; B: Confusion matrix of ResNet-50 showed predicted results with minimal misclassifications between KC and normal groups; C: Confusion matrix of EfficientNet-B0 revealed discriminative ability with less proportion of fasle positive and false negative; D: Confusion matrix of Inception-V3 highlights effective classification outcomes with correct cases of KC and normal. KC: Keratoconus.

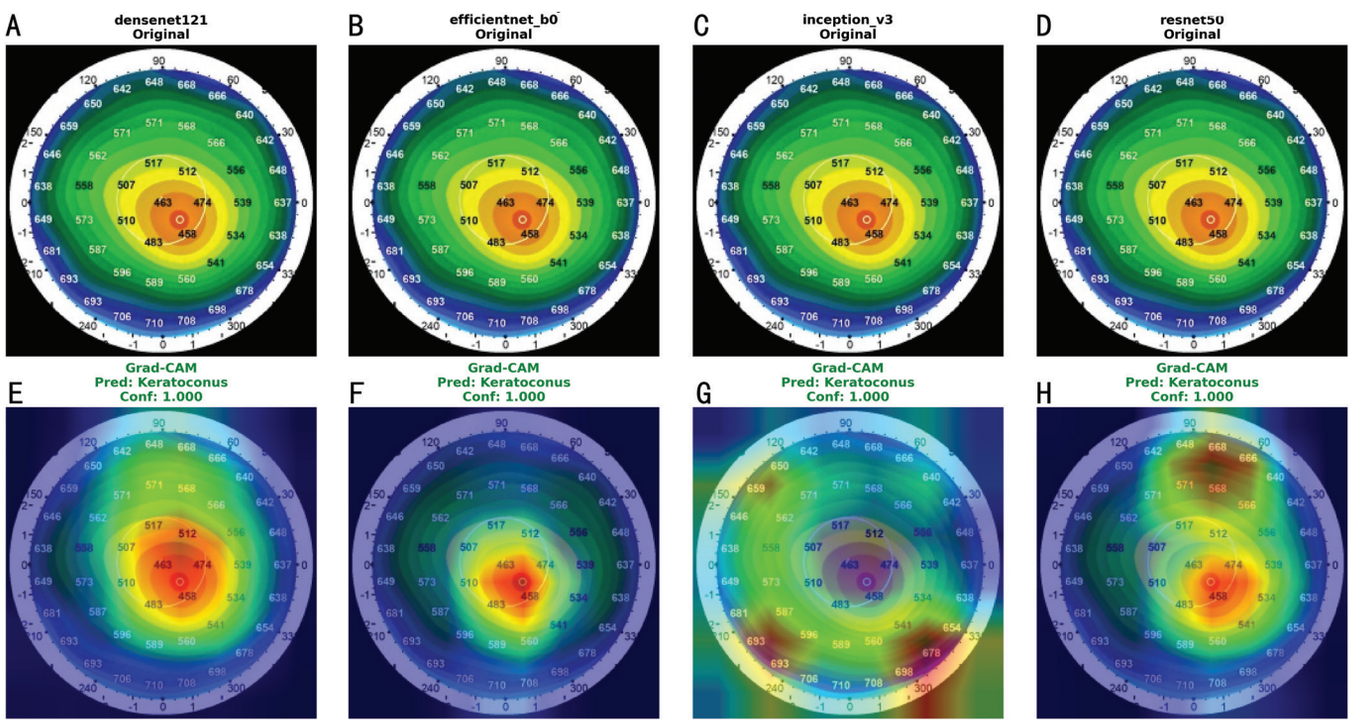

Figure 9 Multi-CNN model Grad-CAM visualization of topography images A-D: The original KC image predicted by all models; E-H: The Grad-CAM heat-map overlay highlighting major regions in the topography image. CNN: Convolutional neural network; Grad-CAM: Gradient-weighted class activation mapping; KC: Keratoconus.

to the training dataset using real-world data. As a result of the augmentation techniques, a total of 10 083 images were utilized for model training, ensuring a balance between data diversity and preservation of diagnostic features, as depicted in Figure 3, which improved model performance on unseen testing data. The models showed excellent results in terms of accuracy, sensitivity, and specificity, as displayed in Table 3.

Several studies have utilized different data augmentation techniques, such as variational autoencoders (VAEs), to synthesize and augment corneal images. These approaches have a greater ability for feature extraction and can generate outputs from continuous distributions while processing various inputs to produce the desired outcomes[29]. Generative adversarial networks (GANs) were also utilized in some studies to synthesize high-quality images[35-36]. Recent study showed that using GANs for synthetic data augmentation enhance diagnostic accuracy by 13% to traditional augmentation during corneal disease classification. The Xception classifier achieved the highest accuracy of 90.5% when trained using synthetic data[37]. Abdelmotaal *et al*[27] used a DL convolutional neural network (VGG-16) model to detect KC using corneal topography, achieving an accuracy of 99.78% by training on a combination of synthetic and original images[27-29].

In contrast, the current study used manual augmentation applied only to the training dataset without depending on synthetic image generation. The DenseNet-121 model achieved a high accuracy of 0.992 when applied to testing data without the use of augmentation techniques. The use of synthetic

images can improve model performance; however, their implementation in clinical settings using real-world images from different topography systems may affect performance and make generalization more challenging.

Different ML approaches have been employed using various indices for KC detection. Kovács *et al*[38] used automatic classifiers trained on tomographic KC indices and achieved higher accuracy and AUC values in discriminating KC from normal eyes (AUC: 0.96 and 0.88). They reported significantly higher keratometry, KC indices, and pachymetry values in KC eyes compared with controls, which parallels the findings of our study, as depicted in Table 1.

Ruiz Hidalgo *et al*[21] used support vector machine algorithms to detect KC and normal eyes based on 22 parameters obtained from the Pentacam system. The overall accuracy obtained for KC and normal eyes was 98.9%, with a sensitivity of 99.1% and specificity of 98.5%. In another study, Ruiz Hidalgo *et al*[39] utilized the KC Assistant, which uses ML algorithms and 25 parameters derived from the Pentacam instrument. They demonstrated excellent results, with the KC Assistant showing agreement with the clinical diagnosis in 92.6% of cases. Yousefi *et al*[40] developed unsupervised ML algorithms using multiple corneal parameters, and based on the ectasia static index, they reported a specificity of 94.1% and sensitivity of 97.7% for identifying KC from normal eyes.

dos Santos *et al*[41] designed a custom neural network, CorneaNet, to segment healthy and KC images using an optical coherence tomography (OCT) system, achieving a high accuracy of 99.56%. Issarti *et al*[22] developed a computer-aided diagnosis system combining a feed-forward neural network to detect clinical KC, suspect KC, and normal groups. The computer-aided diagnosis (CAD) system detected KC with an accuracy of 96.56% (sensitivity: 97.78%; specificity: 95.56%). This system was highly accurate and provided a stable screening platform to assist ophthalmologists in the detection of KC. Different ML models have been adapted to extract features from corneal topography, such as neural networks, logistic regression[42-43], decision trees[44], random forest[45].

Study design, sample size, disease categories, ML approaches and corneal parameters such as keratometry, I-S value, astigmatism, skewed radial axis (KISA) and IS values varied across studies[12,42]. These differences limit direct comparison between previous studies and the current study, and the utilization of such parameters remains challenging due to complexity and the potential for false-positive results, which can complicate clinical interpretation and integration into clinical work flow.

Several studies employed DL models that are pretrained and operate without manual feature extraction, learning features directly from the data[22,26,28,41]. These studies employed color-coded maps, which provide better visualization than numerical parameters due to their complexity. The findings of these studies are concurrent with present study in which whole color-coded images were used to train the DL models, which showed high areas under the ROC curve; DenseNet-121 and ResNet-50 achieved an AUC of 1.000, followed by Inception-V3 (0.999) and EfficientNet-B0 (0.98). Agharezaei *et al*[29] employed VGG 16 model demonstrated highest AUC of 0.99 (sensitivity 0.994), and out of 800 images only there were 13 misclassifications.

However, in contrast to present work, similarly high performance was observed, but difference was observed in sensitivity-related outcomes. The high sensitivity of the models indicates a low rate of false-negative predictions; in our study, DenseNet-121 showed high sensitivity, with only 9 misclassifications out of 1121 images, which may be attributed to similar pattern changes observed in astigmatism across axial curvature maps of KC and normal eyes. Some studies utilized single-map analysis, in which models such as VGG-16 were trained on individual maps, including SAG, front elevation, back elevation, and corneal thickness maps, yielding accuracies of 88.8%, 98.9%, 94.8%, and 94.5%, respectively. Model evaluation in these studies was performed using separate training and testing datasets without a validation set[46].

Kamiya *et al*[26] utilized six corneal maps to train a ResNet model, achieving an accuracy of 0.991 for the posterior curvature map, followed by the anterior elevation map (0.983), corneal pachymetry map (0.982), refractive power map (0.978), and anterior curvature map (0.976). Their study did not include subclinical KC or other corneal diseases, limiting evaluation of model performance in these conditions. While the current study did not involve single-map analysis, as color-coded maps were used as a single composite image including anterior axial curvature, pachymetry, posterior axial curvature, and posterior elevation maps; Similarly, subclinical KC and KC suspects were not included in our study due to limited sample availability and external validation was performed on an independent dataset for generalization of models

Zéboulon *et al*[47] used a relatively large sample size of 3000 corneal images to discriminate between KC, normal eyes, and eyes with a history of refractive surgery, achieving high accuracy but did not include suspect KC cases; similarly, the current study included large sample size demonstrated highest accuracy however, suspect class was not included.

The current study employed Grad-CAM techniques, which generate heat maps highlighting major regions of the input images, providing better visualization of map features, as shown in Figure 8, and revealing that the models focused on areas relevant to clinical markers. This technique has also been adopted in other studies to enhance visualization[29-47].

Table 5 Comparison of previous studies using DL models with current study

Authors	Labelling protocol	Diagnostic imaging modality	Sample data set	Model evaluation protocol	AI techniques employed	Model evaluation performance (accuracy)
Kamiya *et al*[26]	4 grades of KC and normal	Tomey CASIA	543 cases	Fivefold CV	ResNet-18	99.0%
Kuo *et al*[12]	KC and normal	Tomey TMS 4	354 cases	Training, testing and subclinical testing	VGG16	93.1%
					Inception V3	93.1%
					Resnet 152	95.8%
Lavric and Valentin[28]	KC and normal	Synthesized images	145	Training testing and validation	Keratodetect CNN algorithm	99.33%
Zéboulon *et al*[47]	KC, history of refractive surgery, and normal	Baush and LomB Orbscan	3000	Tenfold CV	CNN	99.3%
Al-Timemy *et al*[48]	KC and normal	Oculus Pentacam	542 cases	Training validation and testing	EDTL with Alex net and Product Fusion	98.3%
Al-Timemy *et al*[49]	KC, suspected KC, and normal	Oculus Pentacam	692	Training validation and independent testing	EfficientNet-B0 DL with SVM	Two class 98%, three class 81.6%
Agharezaei *et al*[29]	KC and normal	Tomey TS4 topography	1758 eyes	Training validation and testing	VGG 16	99.3%
					Resnet 152-V2	95.9%
					Efficient Net-Bo	95.2%
					Customized CNN	97.4%
Current study	KC and normal	Galilei (Zeimer S Port Switzerland)	1411 eyes	Training validation and testing independent testing	DenseNet 121	99.2%
					ResNet 50	98.4%
					EfficientNet-B0	98.1%
					Inception V3	98.6%

DL: Deep learning; KC: Keratoconus; CV: Cross validation; CNN: Convolutional neural network; VGG: Visual geometry group; SVM: Support vector machine; EDTL: Ensemble deep transfer learning.

Below is a summary of previous literature in comparison to the current study describing corneal topography images collected from different instruments, the types of models used, sample sizes, and evaluation methods (Table 5)[12,26,28-29,47-49].

We utilized a relatively large dataset compared with previous studies. The models were trained and evaluated using an independent dataset obtained from the Pentacam instrument. The results showed excellent accuracy when generalizing the models to datasets obtained from Pentacam, Sirius, and Orbscan corneal topography systems. Complex corneal indices and multi-model parameter in previous studies may limited seamless integration in routine clinical workflow and may increase burden interpretation. However, the current approach employs color coded maps which help clinician in decisions making and support easy clinical adoption. Additionally, generalization across different imaging devices addresses a key challenges related to device variability and supports regulatory and cost-effective implementation.

This study has certain limitations, as subclinical KC, forme fruste KC, and other corneal diseases were not included due to limited sample availability. The second limitation relates to differences in thickness maps between KC and normal participants with regular astigmatism; incorporating the thickness speed progression index[50] and conducting longitudinal studies over three and six months would allow more accurate labeling and model training. Furthermore, external validation was performed with limited data set using independent data source but large multicenter data are required to confirm generalizability across populations and imaging modalities.

Corneal tomography maps contain numerical indices but these were not used for detection of KC. Color coded maps were used to train DL models. Numerical values were analyzed separately to document topographic differences between KC and normal groups. Analysis was performed on eye-based scans with both eyes included. Despite the bilateral asymmetric nature of KC, participants level evaluation was not performed. Future studies could explore and enhance early detection.

In conclusion, four DL (CNN) models were developed to extract deep features from corneal topography maps and were trained on a large dataset of 5602 corneal topography images. All models exhibited excellent diagnostic accuracy in differentiating KC from normal eyes based on topography images. These findings provide strong potential for clinical implementation, helping ophthalmologists validate clinical decisions and manage KC participants more precisely at earlier stages. The diagnostic accuracy of these models may also help clinicians reduce disease burden, minimize the need for advanced treatment options such as corneal transplantation, reduce healthcare costs, and improve patient outcomes.

ACKNOWLEDGEMENTS

Authors' Contributions: Conceptualization and study design: Qayyum S, Arshad M; Data collection/investigation: Qayyum S; Statistical analysis/interpretation: Qayyum S; Manuscript

drafting: Qayyum S; Manuscript drafting and editing: Qayyum S, Saeed H; Critical reviewing: Arshad M, Saeed H.

Conflicts of Interest: Qayyum S, None; **Arshad M,** None; **Saeed H,** None.

REFERENCES

1 Tan ZP, Chen X, Li KS, *et al*. Artificial intelligence–based diagnostic model for detecting keratoconus using videos of corneal force deformation. *Trans Vis Sci Tech* 2022;11(9):32.

2 Sriranganathan A, Chan CC, Dhillon J, *et al*. Global incidence and prevalence of keratoconus: a systematic review and meta-analysis. *Cornea* 2025. Online ahead of print.

3 Gomes JAP, Rodrigues PF, Lamazales LL. Keratoconus epidemiology: a review. *Saudi J Ophthalmol* 2022;36(1):3-6.

4 Alhazmi A, Alsomali A, Algarni S, *et al*. Pediatric keratoconus in a tertiary eye hospital in Eastern province, Saudi Arabia: Patient characteristics and management outcomes. *Saudi J Ophthalmol* 2024;38(3):261-267.

5 Doroodgar F, Alizadeh F, Niazi S, *et al*. Inflammatory and genomic interactions within keratoconus susceptible patients: a nationwide registered case–control study. *Eye Vis (Lond)* 2024;11(1):40.

6 Yang KL, Li D, Xu LY, *et al*. Independent and interactive effects of eye rubbing and atopy on keratoconus. *Front Immunol* 2022;13:999435.

7 Mohammadpour M, Heidari Z. Updates on keratoconus diagnosis. In: Mohammadpour M, Khorrami-Nejad M, editors. *Keratoconus: Optical and Surgical Management*. 1st ed. Boca Raton, FL: CRC Press; 2024:21-41.

8 Goto S, Maeda N. Corneal topography for intraocular lens selection in refractive cataract surgery. *Ophthalmology* 2021;128(11):e142-e152.

9 Xing L, Dai WJ, Zhang YS. Scheimpflug camera-based technique for multi-point displacement monitoring of bridges. *Sensors (Basel)* 2022;22(11):4093.

10 Sun BQ, Hu XJ, Peng XL, *et al*. Comparative study of the corneal morphological characteristics of keratoconus evaluated using Pentacam, Sirius, and CASIA2. *Photodiagn Photodyn Ther* 2025;53:104612.

11 Heidari Z, Mohammadpour M, Amanzadeh K, *et al*. Evaluation of corneal topographic, tomographic and biomechanical indices for detecting clinical and subclinical keratoconus: a comprehensive three-device study. *Int J Ophthalmol* 2021;14(2):228-239.

12 Kuo BI, Chang WY, Liao TS, *et al*. Keratoconus screening based on deep learning approach of corneal topography. *Trans Vis Sci Tech* 2020;9(2):53.

13 Ipp E, Liljenquist D, Bode B, *et al*. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. *JAMA Netw Open* 2021;4(11):e2134254.

14 Gao YD, Xiong F, Xiong J, *et al*. Recent advances in the application of artificial intelligence in age-related macular degeneration. *BMJ Open Ophthalmol* 2024;9(1):e001903.

15 Durmaz Engin C, Ozturk T, Ozkan O, *et al*. Prediction of retinopathy of prematurity development and treatment need with machine learning models. *BMC Ophthalmol* 2025;25(1):194.

16 Ting DSJ, Foo VH, Yang LWY, *et al*. Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology. *Br J Ophthalmol* 2021;105(2):158-168.

17 Ting DSW, Pasquale LR, Peng L, *et al*. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103(2):167-175.

18 Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019;19(1):64.

19 Syta A, Podkowiński A, Chorągiewicz T, *et al*. Machine learning-assisted early detection of keratoconus: a comparative analysis of corneal topography and biomechanical data. *Sci Rep* 2025;15:24399.

20 Du KY, Peng RM, Chen YG, *et al*. Enhancing early keratoconus detection with multimodal machine learning: integrating tomography, biomechanics, and clinical risk factors. *Am J Ophthalmol* 2025;280:334-346.

21 Ruiz Hidalgo I, Rozema JJ, Saad A, *et al*. Validation of an objective keratoconus detection system implemented in a scheimpflug tomographer and comparison with other methods. *Cornea* 2017;36(6):689-695.

22 Issarti I, Consejo A, Jiménez-García M, *et al*. Computer aided diagnosis for suspect keratoconus detection. *Comput Biol Med* 2019;109:33-42.

23 de Almeida GC, Guido RC, Neto JS, *et al*. Corneal Tomography Multivariate Index (CTMVI) effectively distinguishes healthy corneas from those susceptible to ectasia. *Biomed Signal Process Control* 2021;70:102995.

24 Hwang ES, Perez-Straziota CE, Kim SW, *et al*. Distinguishing highly asymmetric keratoconus eyes using combined scheimpflug and spectral-domain OCT analysis. *Ophthalmology* 2018;125(12):1862-1871.

25 Lopes BT, Ramos IC, Salomão MQ, *et al*. Enhanced tomographic assessment to detect corneal ectasia based on artificial intelligence. *Am J Ophthalmol* 2018;195:223-232.

26 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.

27 Abdelmotaal H, Mostafa MM, Mostafa ANR, *et al*. Classification of color-coded scheimpflug camera corneal tomography images using deep learning. *Trans Vis Sci Tech* 2020;9(13):30.

28 Lavric A, Valentin P. KeratoDetect: keratoconus detection algorithm using convolutional neural networks. *Comput Intell Neurosci* 2019;2019:8162567.

29 Agharezaei Z, Firouzi R, Hassanzadeh S, *et al*. Computer-aided diagnosis of keratoconus through VAE-augmented images using deep learning. *Sci Rep* 2023;13:20586.

30 Rabinowitz YS, McDonnell PJ. Computer-assisted corneal topography in keratoconus. *J Refract Surg* 1989;5(6):400-408.

31 Shah Z, Purohit D, Danayak P. Keratoconus characteristics and associations: a cross-sectional keratoconus study in western India (CKSWI). *Indian J Ophthalmol* 2024;72(5):704-711.

32 Singh K, Pant H, George J, *et al*. Exploiting EfficientNetB3, InceptionV3, and DenseNet121 for enhanced eye disease classification. *2024 International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS).* December 27-28, 2024. Pattaya, Thailand. IEEE, 2024:368-374.

33 Kaur A, Kukreja V, Kumar M, *et al*. RetDense: a fine-tuned DenseNet121 model for retinal eye disease detection. *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI).* March 14-16, 2024. Gwalior, India. IEEE, 2024:1-5.

34 Ahmed N, Rahman MM, Ishrak MF, *et al.* Comparative performance analysis of transformer-based pre-trained models for detecting keratoconus disease. arXiv. 2024;2408:09005.

35 Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal* 2019;58:101552.

36 Abdelmotaal H, Abdou AA, Omar AF, *et al*. Pix2pix conditional generative adversarial networks for scheimpflug camera color-coded corneal tomography image generation. *Trans Vis Sci Tech* 2021;10(7):21.

37 Jameel SK, Aydin S, Ghaeb NH, *et al*. Exploiting the generative adversarial network approach to create a synthetic topography corneal image. *Biomolecules* 2022;12(12):1888.

38 Kovács I, Miháltz K, Kránitz K, *et al*. Accuracy of machine learning classifiers using bilateral data from a Scheimpflug camera for identifying eyes with preclinical signs of keratoconus. *J Cataract Refract Surg* 2016;42(2):275-283.

39 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.

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

41 dos Santos VA, Schmetterer L, Stegmann H, *et al*. CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning. *Biomed Opt Express* 2019;10(2):622-641.

42 Kojima T, Nishida T, Nakamura T, *et al*. Keratoconus screening using values derived from auto-keratometer measurements: a multicenter study. *Am J Ophthalmol* 2020;215:127-134.

43 Velázquez-Blázquez JS, Bolarín JM, Cavas-Martínez F, *et al*. EMKLAS: a new automatic scoring system for early and mild keratoconus detection. *Trans Vis Sci Tech* 2020;9(2):30.

44 Mosa ZM, Ghaeb NH, 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.

45 Chandapura R, Salomão MQ, Ambrósio R Jr, *et al*. Bowman's topography for improved detection of early ectasia. *J Biophotonics* 2019;12(10):e201900126.

46 Salih N, Akgün Ö, Hussein N. Building smart algorithm to extract features of topographic images of a human eye. Aksaray Üniversitesi Fen Bilimleri Enstitüsü 2019.

47 Zéboulon P, Debellemanière G, Bouvet M, *et al*. Corneal topography raw data classification using a convolutional neural network. *Am J Ophthalmol* 2020;219:33-39.

48 Al-Timemy AH, Ghaeb NH, Mosa ZM, *et al*. Deep transfer learning for improved detection of keratoconus using corneal topographic maps. *Cogn Comput* 2022;14(5):1627-1642.

49 Al-Timemy AH, Mosa ZM, Alyasseri Z, *et al*. A hybrid deep learning construct for detecting keratoconus from corneal maps. *Trans Vis Sci Tech* 2021;10(14):16.

50 Awwad ST, Hammoud B, Assaf JF, *et al*. Thickness speed progression index: machine learning approach for keratoconus detection. *Am J Ophthalmol* 2025;271:188-201.