



OPEN ACCESS

Clinical science

Deep learning model for extensive smartphone-based diagnosis and triage of cataracts and multiple corneal diseases

Yuta Ueno,¹ Masahiro Oda ,^{2,3} Takefumi Yamaguchi ,⁴ Hideki Fukuoka,⁵ Ryohei Nejima,⁶ Yoshiyuki Kitaguchi,⁷ Masahiro Miyake ,⁸ Masato Akiyama,⁹ Kazunori Miyata,⁶ Kenji Kashiwagi,¹⁰ Naoyuki Maeda ,⁷ Jun Shimazaki ,⁴ Hisashi Noma,¹¹ Kensaku Mori ,^{2,3,12} Tetsuro Oshika ¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bjo-2023-324488>).

For numbered affiliations see end of article.

Correspondence to

Dr Takefumi Yamaguchi, Ophthalmology, Tokyo Dental College Ichikawa General Hospital, Ichikawa, 272-8513, Japan; yamaguchit@tdc.ac.jp

YU and MO contributed equally.

Received 25 August 2023

Accepted 17 December 2023



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Ueno Y, Oda M, Yamaguchi T, et al. *Br J Ophthalmol* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bjo-2023-324488

ABSTRACT

Aim To develop an artificial intelligence (AI) algorithm that diagnoses cataracts/corneal diseases from multiple conditions using smartphone images.

Methods This study included 6442 images that were captured using a slit-lamp microscope (6106 images) and smartphone (336 images). An AI algorithm was developed based on slit-lamp images to differentiate 36 major diseases (cataracts and corneal diseases) into 9 categories. To validate the AI model, smartphone images were used for the testing dataset. We evaluated AI performance that included sensitivity, specificity and receiver operating characteristic (ROC) curve for the diagnosis and triage of the diseases.

Results The AI algorithm achieved an area under the ROC curve of 0.998 (95% CI, 0.992 to 0.999) for normal eyes, 0.986 (95% CI, 0.978 to 0.997) for infectious keratitis, 0.960 (95% CI, 0.925 to 0.994) for immunological keratitis, 0.987 (95% CI, 0.978 to 0.996) for cornea scars, 0.997 (95% CI, 0.992 to 1.000) for ocular surface tumours, 0.993 (95% CI, 0.984 to 1.000) for corneal deposits, 1.000 (95% CI, 1.000 to 1.000) for acute angle-closure glaucoma, 0.992 (95% CI, 0.985 to 0.999) for cataracts and 0.993 (95% CI, 0.985 to 1.000) for bullous keratopathy. The triage of referral suggestion using the smartphone images exhibited high performance, in which the sensitivity and specificity were 1.00 (95% CI, 0.478 to 1.00) and 1.00 (95% CI, 0.976 to 1.000) for 'urgent', 0.867 (95% CI, 0.683 to 0.962) and 1.00 (95% CI, 0.971 to 1.000) for 'semi-urgent', 0.853 (95% CI, 0.689 to 0.950) and 0.983 (95% CI, 0.942 to 0.998) for 'routine' and 1.00 (95% CI, 0.958 to 1.00) and 0.896 (95% CI, 0.797 to 0.957) for 'observation', respectively.

Conclusions The AI system achieved promising performance in the diagnosis of cataracts and corneal diseases.

INTRODUCTION

The cornea and crystalline lens are transparent media in the eye, which focus light on the retina and contribute to maintaining vision. The pathological conditions of these ocular media (ie, corneal opacity, infectious keratitis and cataracts) are the leading causes of vision impairment, affecting 75 million

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Artificial intelligence (AI) applications for single eye diseases, such as cataracts, diabetic retinopathy, retinopathy of prematurity and infectious keratitis, have been developed. Although these AI techniques have achieved good performance, the social application of these AI-based technologies has been limited by the difficulty of using AI to differentiate and diagnose from various pathologies in the real world.

WHAT THIS STUDY ADDS

⇒ This study developed an AI to diagnose multiple corneal diseases/cataracts. Furthermore, the AI installed in iPhone13 achieved a high performance to triage the diseases, based on the anterior segment photographs taken using iPhone13.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The AI-driven triage can help in early diagnosis/treatment by providing accurate medical information in the early stages of eye diseases, directly connecting patients and clinics anywhere in the world via smartphones and potentially prevent blindness due to cataract or corneal diseases.

people worldwide (15 million with blindness and 60 million with moderate-to-severe vision impairment).^{1,2} Corneal diseases and cataracts are considered as avoidable vision loss if they are diagnosed and treated properly.¹⁻³ However, the diagnosis is dependent on the availability of ophthalmologists. Thus, despite recent medical progress, the amount of avoidable blindness has continued to increase as the global population grows and ages owing to the limited number of experienced ophthalmologists.^{3,4}

Artificial intelligence (AI) provides a promising solution for disease diagnosis and triage based on medical imaging.^{5,6} In ophthalmology, AI applications for single diseases, such as cataracts, diabetic retinopathy, retinopathy of prematurity

and infectious keratitis, have been developed based on deep learning (DL).^{6–14} Although these AI techniques have achieved good performance, their ability to differentiate multiple corneal diseases remains limited.¹⁵ In this study, we sought to develop an AI-driven comprehensive diagnosis/triage system using anterior segment photographs. First, we developed a DL model for the extensive diagnosis of multiple corneal diseases and cataracts and compared its performance with that of board-certified corneal specialists and residents. Second, we evaluated the AI performance on photographs that were captured using smartphone cameras and successfully triaged multiple corneal diseases/cataracts based on the conditions.

METHODS

Image acquisition

As corneal diseases/cataracts are located anteriorly in the eye and can be imaged clearly using a smartphone camera without specific attachments, we aimed to develop an AI-assisted triage system using smartphone images. A total of 16 471 images (15 498 slit-lamp microscopy images and 973 iPhone 13 Pro images) were registered in the Japan Ocular Imaging Registry from 23 tertiary eye centres,¹⁶ to develop a comprehensive AI for diagnosing various anterior segment eye diseases. Slit-lamp microscopy images were retrospectively collected from tertiary centres. The slit-lamp images were captured with a diffuser light source at a magnification of ×10 or ×16 setting at 23 tertiary eye centres in Japan. All anterior segment images were obtained by corneal specialists using a camera-mounted slit-lamp microscope (Haag-Streit, Zeiss, Takagi) under diffuser illumination (without enhancing the slit beam) with a natural pupil (without mydriatic instillation) and saved in joint photographic expert group format. Smartphone images were prospectively captured in cornea clinics in each tertiary centre consecutively from 2019 to 2020 after obtaining informed consent. The smartphone images were captured using the super macro mode of the iPhone 13 Pro under the following conditions: (1) open built-in capture software (developed by YK) under standard room illumination (not in a dark room), (2) a distance of approximately 3–5 cm between the cornea and iPhone 13 Pro cameras; (3) a camera flash and (4) a clear focus on the cornea. The smartphone images

were deidentified and saved in portable network graphics format (size: 12–18 MB) with the information of the date, right or left eyes, diagnosis and hospital name.

Definition of clinical taxonomy

All images were carefully verified by two of four corneal specialists (YU, TY, HF and RN; ≥15 years of qualification in corneal specialty) after a 5-hour intensive tutorial and active discussion on the categorisation with a corneal specialist (NM with 30 years of experience as a corneal specialist). We confirmed the diagnosis made by tertiary centres and classified 36 corneal diseases/cataracts into 9 categories that cover the major diseases of the anterior segment of the eye (figure 1A, online supplemental table S1): ‘normal’, ‘cataract’, ‘infectious keratitis’, ‘immunological keratitis’, ‘corneal scar’, ‘corneal deposits’, ‘bullous keratopathy’, ‘ocular surface tumour’ and ‘primary angle-closure glaucoma’. A specific category for each slit-lamp image was provided by corneal specialists at tertiary university hospitals based on the medical records, clinical course, presentation, laboratory examinations (including culture, blood and PCR) and response to treatments. In images with multiple diseases, such as ‘acute angle-closure glaucoma’ and ‘cataract’, it was annotated based on the categories that needed to be treated with priority in clinical settings (in this case, ‘acute angle-closure glaucoma’).

Datasets

To avoid biased influences, we excluded 9728 duplicate and poor-quality images, which were defined as blurred, dark illumination, controversial diagnosis and slit-beam enhanced images. Most of the excluded images were duplicate images of specific patients with multiple similar images captured on the same day. Reasons for image exclusion were duplicate images from single patients taken on the same day (approximately 80%), followed by slit-beam enhanced images (15%) and images with fluorescein staining (3%), blurred or dim illumination images (2%) in slit-lamp images. In smartphone images, reasons for the image exclusion were duplicate images of single patients taken on the same day (approximately 80%), followed by decentred images (10%) and blurred or dim illumination images (10%). Finally, a total of

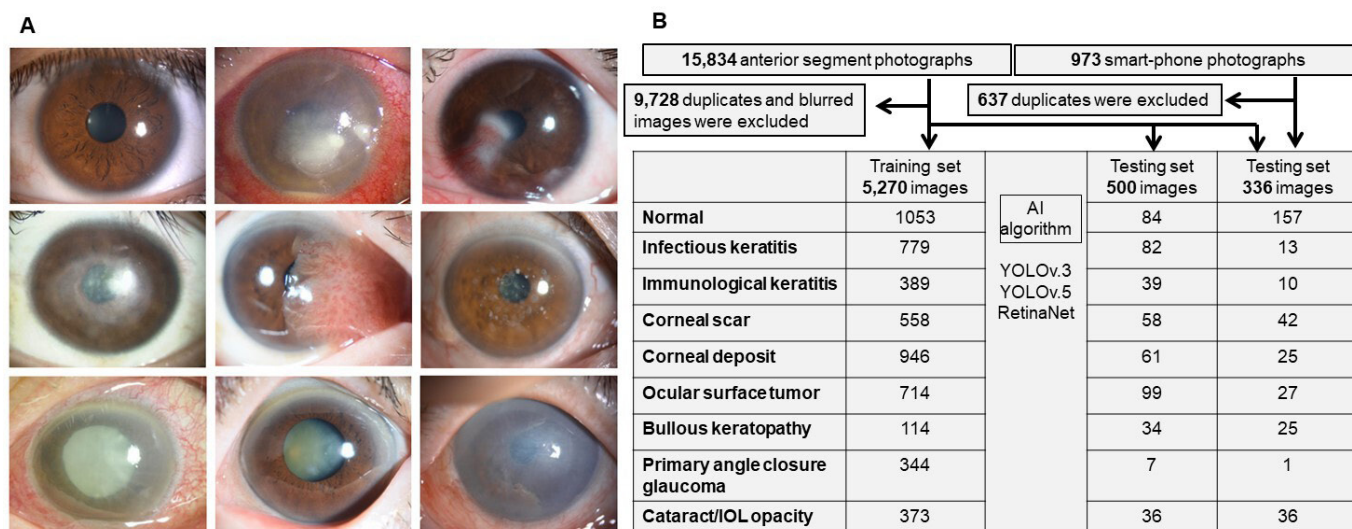


Figure 1 Selection process and sample size for training and testing sets. (A) Representative slit-lamp photographs of nine categories. (B) A total of 15 498 anterior segment photographs of 9 categories were captured using slit-lamp microscopy with diffuser light and 973 images were captured using an iPhone 13 Pro. IOL, intraocular lens.

6106 images were captured in a slit-lamp camera with diffuser light were used to develop the AI model (figure 1B and online supplemental table S2; 5270 images were randomly selected for the training dataset and 836 images for the testing dataset). The baseline information is shown in online supplemental table S3. However, the AI model was developed without any data on age, sex and race. Moreover, 336 images that were captured using smartphone cameras were used for the testing dataset to validate the DL model.

Training and testing protocol of AI models

A categorical annotation label and bounding box enclosing the corneal region were provided by experienced ophthalmologists for all images to perform the object detection process (online supplemental figure S1).

Training of nine-category classification AI models

We used You Only Look Once V.3 (YOLO V.3), YOLO V.5 and RetinaNet as the AI algorithms to perform the nine-category

classification. We used 5270 images with 36 eye diseases into 9 categories of annotations to train and test the models. Dataset separation and pretraining were performed in the training of YOLO V.3. The model parameters in YOLO V.5 were pretrained using the common object in context (COCO) dataset and subsequently fine-tuned using the training dataset. YOLO V.5 was trained for 200 epochs with a mini-batch size of 16. We used ResNet-101 as the backbone in the training of RetinaNet. RetinaNet was trained for 50 epochs with a mini-batch size of 8. The number of epochs and mini-batch sizes were selected experimentally so that each AI model achieved the highest performance.

Testing of nine-category classification AI models for comparison with ophthalmologists

In the testing process of the AI models, we obtained multiple estimated bounding boxes with estimated categories and the detection and classification systems for each test image from YOLO V.3, YOLO V.5 or RetinaNet. We implemented a programme that automatically selected the estimated category with the highest

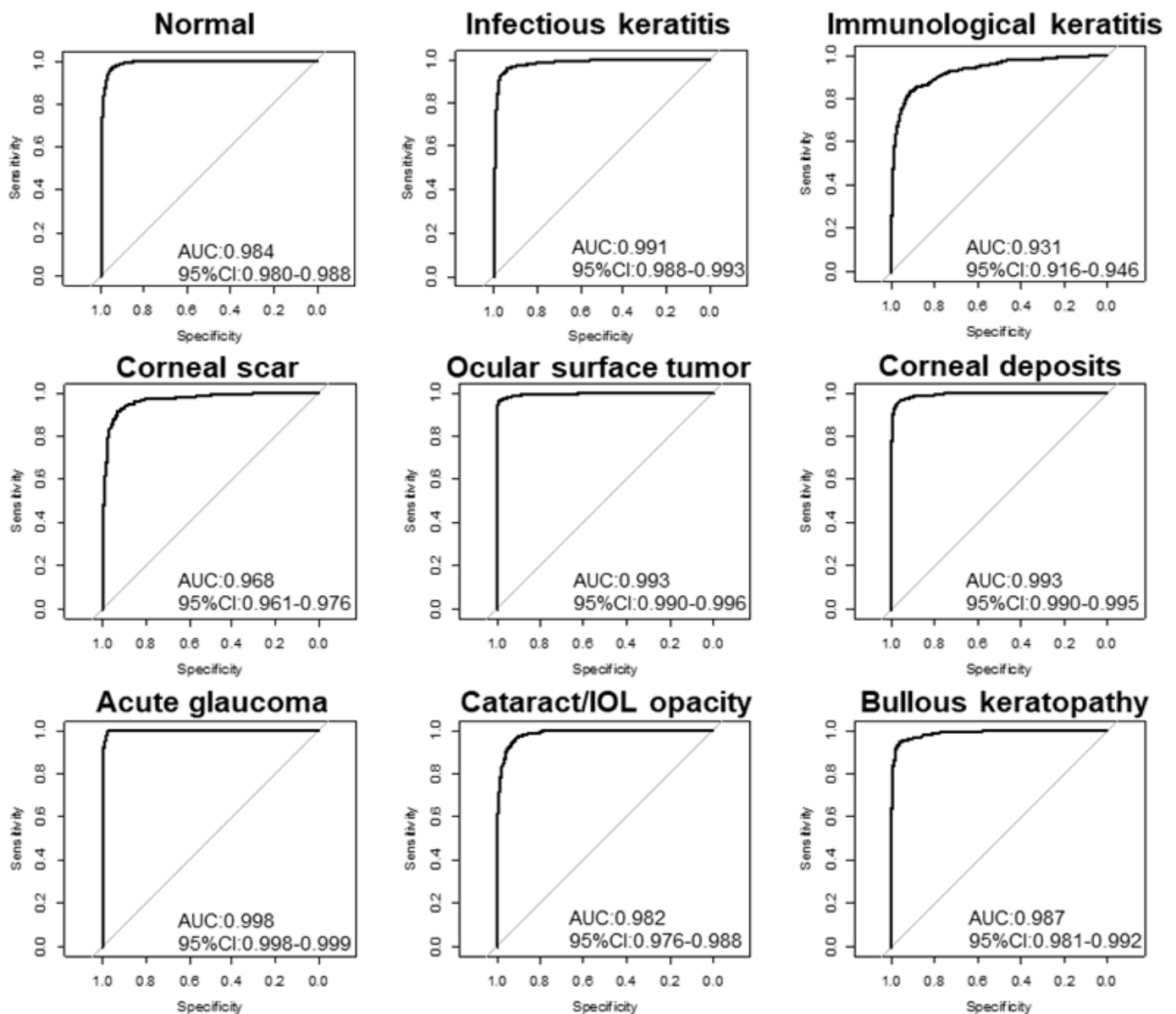


Figure 2 Performance of deep learning algorithm to classify cataract/cornea diseases into nine categories. Receiver operating characteristic curves indicating performance of YOLO V.5 for each category. The area under the curve (AUC) ranged from 0.968 to 0.998.

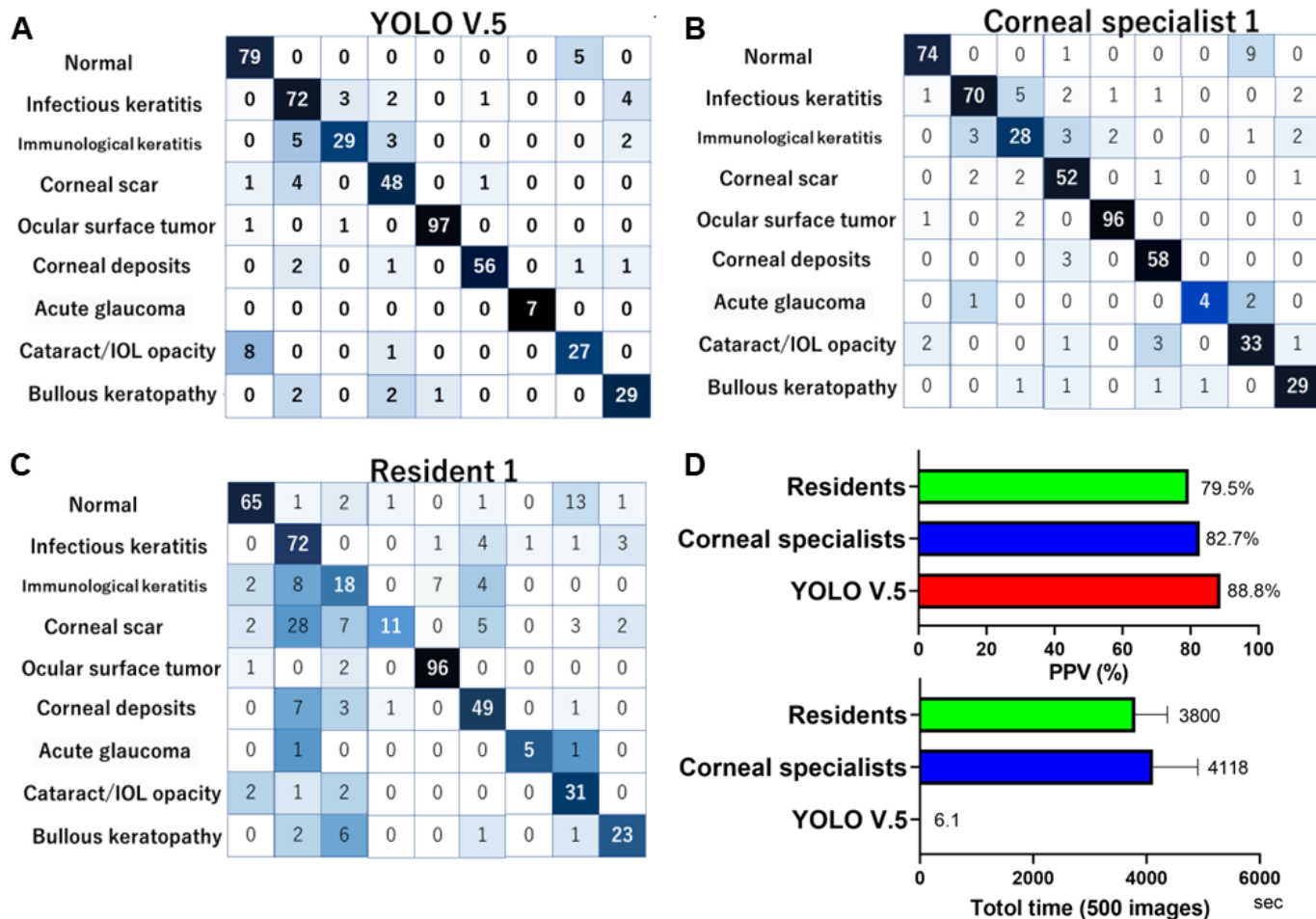


Figure 3 Comparison of diagnostic performance between YOLO V.5 and ophthalmologists. (A) Confusion matrices of image numbers in YOLO V.5 to classify 36 anterior segment eye diseases into 9 categories using testing dataset of anterior segment photographs without clinical data. Confusion matrices of board-certified corneal specialist (B) and ophthalmology resident (C) without clinical information. (D) YOLO V.5 required 6.1 s to complete 500 image classifications, whereas 4118 s were required for corneal specialists and 3800 s were required for residents. IOL, intraocular lens.

predictive score. By using the AI algorithms, we obtained an estimated category and its predictive score for each test image as the classification result of the AI algorithm. The model parameters in YOLO V.5 were pretrained using the COCO dataset and finetuned using the training dataset. YOLO V.5 was trained for 200 epochs with a mini-batch size of 16.

Testing of nine-category classification AI models using smartphone images

We performed testing using 336 smartphone images. The measurements of the processing times were performed on a computer equipped with an NVIDIA RTX A6000 graphics processing unit (NVIDIA, Santa Clara, California, USA) and Xeon Gold 5120 3.2 GHz CPU (Intel, Santa Clara, California, USA).

Predictive score calculation

The categories and corresponding predictive score of the estimated bounding boxes from a test image were estimated during the testing of the AI models. The estimated category with the highest predictive score was selected as the final estimation result. The predictive score is calculated using the sigmoid function in AI models. In the final layer (output layer) of DL-based AI models, the sigmoid function is applied to a feature value that is provided to the final layer, which is represented by:

$$s_{b,c}(x_{b,c}) = \frac{1}{1+e^{-x_{b,c}}}$$

where $s_{b,c}$ is the predictive score, $x_{b,c}$ is the feature value provided to the final layer, b is the index of the estimated bounding boxes and $c = 1, \dots, 9$ is the index of the categories.

DL versus corneal specialists and residents

We collaborated with 11 board-certified corneal specialists and 11 residents with 1–5 years of clinical experience to evaluate our DL model for classifying corneal diseases/cataracts. A testing dataset (500 slit-lamp images with diffuser light) was used to compare the performance of these participants to that of the DL model. The specialists and residents independently classified each image into one of the nine categories. Specialists and residents independently classified each image into one of the nine categories. In images with multiple diseases, such as active ‘infectious keratitis’ and ‘cataract’, they classified them based on clinically primary diseases which they needed to treat with priority.

Statistical analyses

We determined the receiver operating characteristic curves for the predictive scores of the nine categories. We also provided the areas under the curves (AUCs) and their 95% CIs using the

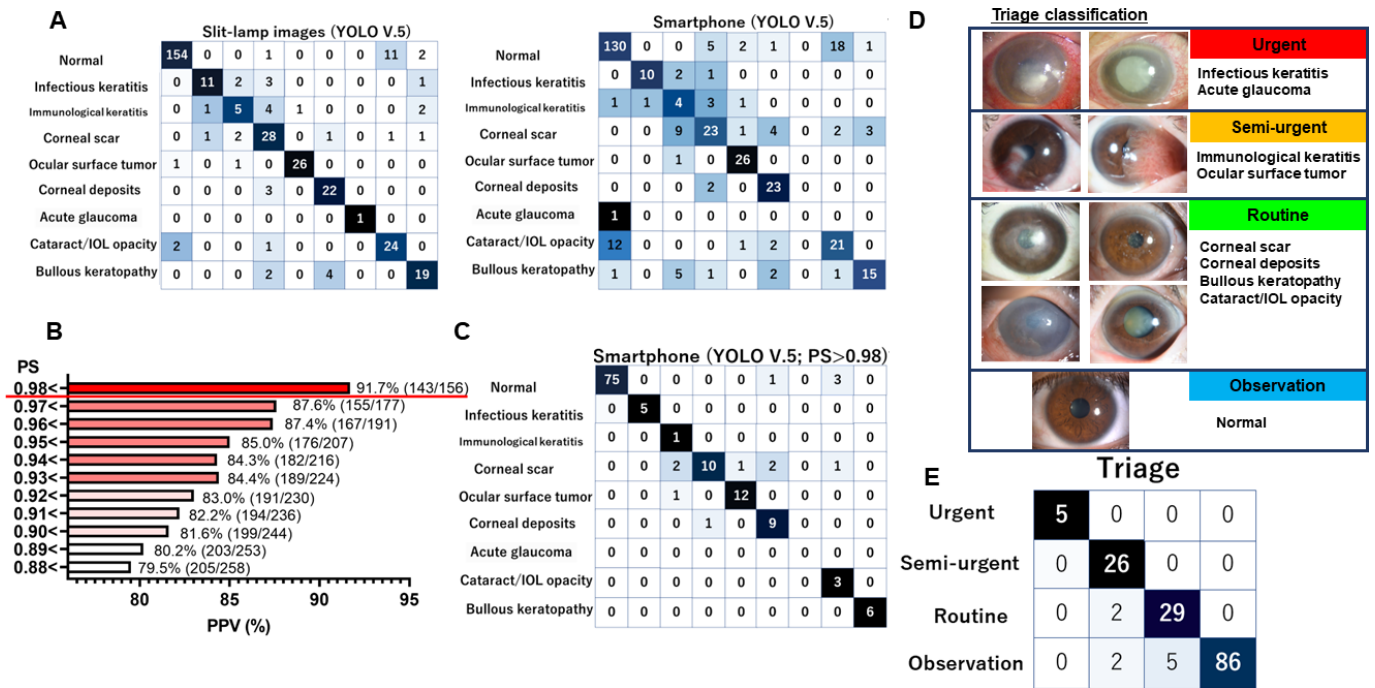


Figure 4 Artificial intelligence (AI) performance and triage using smartphone images. (A) Comparison of AI performance (YOLO V.5) between slit-lamp (diffuser light) and smartphone images (336 patients). Owing to the various conditions of the smartphone (online supplemental figure S4A), the positive predictive values (PPV) were lower for the smartphone images (75.0%) than for the slit-lamp images (88.8%). (B) As the AI performance was better in images with a higher predictive score in the smartphone images, the AI performance in the images with a predictive score of 0.98 or greater was expected to exceed that of the slit-lamp images. (C) Confusion matrices of image numbers in YOLO V.5 for classifying 36 anterior segment eye diseases into 9 categories using smartphone images with a predictive score of 0.98 or greater. The PPV was 91.0%, which exceeded the 82.7% for board-certified corneal specialists. (D) Definition of triage classification. (E) Confusion matrices of image numbers in triage using smartphone images. High-performance triage was obtained after stratifying the images based on a predictive score greater than 0.98. IOL, intraocular lens.

Delong method.¹⁷ Moreover, the sensitivities, specificities and diagnostic accuracies were calculated for the diagnostic algorithms. The 95% CIs were computed using the ‘exact’ Clopper-Pearson method. The statistical analyses were performed using R V.4.2.0 (R Foundation for Statistical Computing, Vienna, Austria)¹⁸ and Prism V. 6.04 for Windows software (Graphpad Software, San Diego, California, USA).

RESULTS

Performance of three AI models for classifying nine categories

AI models were developed using YOLO V.3, YOLO V.5 and RetinaNet based on 5270 images that were obtained using a slit-lamp microscope. YOLO V.5 achieved an AUC of 0.931–0.998 (figure 2, online supplemental figure S3), a sensitivity of 0.628–0.958 and a specificity of 0.969–0.998, thereby exhibiting the best performance, followed by YOLO V.3 and RetinaNet (online supplemental table S4).

Performance of YOLO V.5 on test datasets compared with corneal specialists and residents

In the test, dataset consisting of 500 images from a slit-lamp camera with diffuser light, YOLO V.5 exhibited sufficient performance in classifying the 9 categories (figure 3, source codes are available in <https://github.com/modafone/corneaai>). YOLO V.5 achieved an AUC of 0.996 (95% CI, 0.992 to 0.999) for normal eyes, 0.988 (95% CI, 0.978 to 0.997) for infectious keratitis, 0.960 (95% CI, 0.925 to 0.994) for immunological keratitis, 0.987 (95% CI, 0.978 to 0.996) for corneal scar, 0.997 (95% CI, 0.992 to 1.000) for ocular surface tumour, 0.993 (95% CI,

0.984 to 1.000) for corneal deposits, 1.000 (95% CI, 1.000 to 1.000) for acute angle-closure glaucoma, 0.992 (95% CI, 0.985 to 0.999) for cataracts and 0.993 (95% CI, 0.985 to 1.000) for bullous keratopathy (online supplemental table S5). In the test datasets, the positive predictive values (PPVs) were 88.8% for YOLO V.5, 82.2%±4.5% for the board-certified corneal specialists and 73.4%±8.4% for the residents, respectively (figure 3A–C). YOLO V.5 required 6.1 s to complete the diagnosis of 500 images, whereas 4118±2612s (ranging from 1200 to 10 800) was required for the corneal specialists and 3800±1971s (ranging from 1800 to 7200) was required for the residents to complete the same task (figure 3D).

Comorbidity; limitation of AI application in real world

When assigning a task to diagnosing images with typical features, YOLO V.5 performed successful diagnosis with a very high predictive score of 99.9% (online supplemental figure S3A). However, we observed that the AI tended to misdiagnose specific types of slit-lamp photographs. We analysed such images and found that the predictive scores were low, as the images presented multiple clinical findings, which is one of the significant challenges when applying AI to clinical practice in the real world.^{19–21} For example, in infectious keratitis, corneal infiltration is gradually converted into scarring with resolution. When the predictive score was assessed, YOLO V.5 appropriately listed both corneal scars and infectious keratitis (online supplemental figure S3B). In the testing datasets, 93% of the images had only 1 category, whereas 6% had 2 and 1% had 3 or more categories (online supplemental figure S3C). Therefore, we analysed the three categories with 3 largest predictive scores in each image,

Table 1 Performance of YOLO V.5 for nine categories and triage using smartphone images

	Sensitivity (95% CI)	Specificity (95% CI)
Nine categories		
Normal	1.000 (0.952 to 1.000)	0.932 (0.835 to 0.981)
Infectious keratitis	1.000 (0.478 to 1.000)	1.000 (0.972 to 1.000)
Immunological keratitis	0.250 (0.006 to 0.806)	1.000 (0.972 to 1.000)
Corneal scar	0.909 (0.587 to 0.998)	0.951 (0.897 to 0.982)
Ocular surface tumour	0.923 (0.640 to 0.998)	0.992 (0.955 to 1.000)
Corneal deposits	0.750 (0.428 to 0.945)	0.992 (0.428 to 0.945)
Primary angle-closure glaucoma	1.000 (0.025 to 1.000)	1.000 (0.973 to 1.000)
Cataract/IOL opacity	0.429 (0.099 to 0.816)	1.000 (0.971 to 1.000)
Bullous keratopathy	1.000 (0.541 to 1.000)	1.000 (0.972 to 1.000)
Triage		
Urgent	1.000 (0.478 to 1.000)	1.000 (0.976 to 1.000)
Semi-urgent	0.867 (0.693 to 0.962)	1.000 (0.971 to 1.000)
Routine	0.853 (0.689 to 0.950)	0.983 (0.942 to 0.998)
Observation	1.000 (0.958 to 1.000)	0.896 (0.797 to 0.957)
IOL, intraocular lens.		

which revealed that the PPV of the correct diagnosis was 88.8% within the largest predictive score, 95.6% within the 2 largest predictive scores and 98.0% within the 3 largest predictive scores (online supplemental figure S3D and table S6).

Triage using smartphone images and AI model

Smartphone corneal images of 36 anterior segment diseases and those obtained using slit-lamp microscopy with a diffuser light source (online supplemental figure S4A; 336 images of the same subjects) were categorised into 9 clinical categories using YOLO V.5. Although the YOLO V.5 performance in the slit-lamp images was comparable to that of our previous analyses (figure 4A; PPV: 88.8%), that of the smartphone images was 75% (online supplemental video S1). The predictive score was significantly lower for the smartphone than for the slit-lamp images (0.906 ± 0.145 vs 0.964 ± 0.082 , $p < 0.001$). This was attributable to the varying qualities of the smartphone camera images. Thus, we analysed the YOLO V.5 performance stratified based on the predictive score of each image, which demonstrated that the smartphone images with high predictive scores were diagnosed correctly (figure 4B). When we set the cut-off value of the predictive score to 0.98 or more, the smartphone-based YOLO V.5 performance achieved better outcomes (figure 4C, table 1: predictive score of 0.98 or more; online supplemental figure S4B,C: predictive score of 0.96 or more). To establish AI-driven triage using the smartphone images, we classified the nine categories as 'urgent', 'semi-urgent', 'routine' and 'observation' (figure 4D). The sensitivity and specificity were 1.00 (95% CI, 0.478 to 1.00) and 1.00 (95% CI, 0.976 to 1.00) for 'urgent', 0.867 (95% CI, 0.683 to 0.962) and 1.00 (95% CI, 0.971 to 1.00) for 'semi-urgent', 0.853 (95% CI, 0.689 to 0.95) and 0.983 (95% CI, 0.942 to 0.998) for 'routine' and 1.00 (95% CI, 0.958 to 1.00) and 0.896 (95% CI, 0.797 to 0.957) for 'observation', respectively (table 1, figure 4E).

DISCUSSION

The high performance of AI for automated diagnosis in ophthalmology has been reported in single diseases, such as cataracts, age-related macular degeneration, diabetic retinopathy, glaucoma and corneal diseases, using fundus photographs, optical coherence tomography and anterior segment images.^{8–15} These

methods can successfully differentiate one disease from a normal condition. However, AI algorithms need to diagnose various diseases in the real world.

Millions of people lose their vision due to cataracts and corneal diseases annually.^{1–4} Even if the infections or inflammatory lesions are minor, once they expand and cover the pupil area, corneal diseases result in severe visual impairment. In comparison with other eye diseases, such as glaucoma and retinal diseases, most corneal diseases can be avoided using low-cost medicine if they are treated in the early stages.¹ Furthermore, ocular surface tumours, such as conjunctival squamous cell carcinoma, may metastasise and cause death.²² The implementation of DL in medicine has advanced as it demonstrates significant prospects for improving clinical outcomes and public health.²³ It is noteworthy that our DL model exhibited high performance in triaging corneal diseases using smartphone cameras. Furthermore, this AI will be accessible anywhere in the world provided that smartphone cameras and the internet are available (online supplemental video S1).

Among the three AI algorithms (YOLO V.3, YOLO V.5 and RetinaNet), the YOLO algorithms performed better than RetinaNet. The performance differences were owing to differences in (1) the DL model structure and (2) the data augmentation method. As DL model structures for image feature extraction, YOLO V.3 and YOLO V.5 employ Darknet-53 and its improved version,²⁴ respectively, which were designed to identify the location of the target in the image and its category. Such models are effective for identifying the corneal location in an image and a category that corresponds to corneal diseases. RetinaNet employs ResNet-101 for image feature extraction.²⁵ However, as ResNet-101 was designed for image classification, it is not effective for identifying categories based on local image patterns in the corneal regions. Furthermore, RetinaNet uses a simple data augmentation method (horizontal flipping) during its training process, whereas YOLO V.3 and YOLO V.5 employ various data augmentation methods in their training to achieve high robustness against geometrical and colour changes. Therefore, YOLO V.3 and YOLO V.5 could achieve higher performance than RetinaNet in classifying the 36 anterior segment eye diseases.

In recent years, smartphone applications ('apps') with AI have been used increasingly in healthcare applications. Regarding skin cancer, two apps, namely 'SkinScan' and 'SkinVision', have recently become available for download. 'SkinVision' achieves a considerably high sensitivity of 95% and specificity of 78% in identifying malignant and premalignant lesions.²⁶ However, careful interpretation is required when applying a smartphone in real-world settings. Even when smartphone apps with a high sensitivity of 95% are used for a disease with low incidence, such as infectious keratitis (with an incidence of 30 persons per 100 000 annually),^{27,28} an app with hypothetical sensitivity/specificity of 90% (those of the AI of the current study) would have a PPV of only 0.28%, with approximately 10 000 false positive results per 100 000 users. The potential overload on healthcare owing to false positive results will become considerably large. Therefore, we considered presenting the top three differential diagnoses with an indicator, namely the predictive score, in the AI system. Moreover, the current AI system offers several advantages (online supplemental video S1). As opposed to skin carcinoma, anterior segment eye disorders have various subjective symptoms, such as 'blurred vision', 'ocular pain', 'itching' and 'red eye'.²⁹ The performance was poor in ophthalmology-board-certified specialists, as it was assessed based on strictly image-based diagnosis. Indeed, the PPV for human specialists increased from 82.2% to 91.3% when clinical data (such as visual acuity)

were added to the images (data are not shown). For smartphone application by non-healthcare professional, we will need image quality control procedures to avoid misclassification and confusion by selecting images with sufficient quality. As shown in figure 4B,C, if we selected the images with high predictive scores, the AI performance improved. Therefore, the predictive score indicates the certainty of the classification and can be used to avoid poor images and misclassification, although we need to substantiate the results in a larger number of subjects.

This study exhibits several limitations. First, therapeutic measures are lacking. However, information on the pathophysiology, treatment, and prognosis of the disease can easily be provided by linking the AI to the appropriate database. Furthermore, through triaging disorders and facilitating persons with potential urgent/semi-urgent conditions, we believe that early diagnosis and treatment will aid in preventing avoidable blindness. Second, we did not evaluate the accuracy of the AI in determining the disease severity. However, surprisingly, it could detect the diseases from the early stages or very small lesions of the diseases with very high predictive scores of 0.997 to 0.999 (online supplemental figure S5). Third, we did not include other common corneal/external eye diseases, such as lid disorders (chalazion, entropion, etc), hyposphagma, conjunctivitis and dry eyes in the current study, as we aimed to focus on cataracts and corneal diseases, which are the leading causes of blindness. Fourth, the AI model was developed based solely on Japanese people with brown iris colour. We will need to evaluate its performance in Caucasian people with blue iris, and if the performance is poor, we will need to develop other AI models for different races. Fifth, in the current study, image capture and testing was performed using the iPhone 13Pro smartphone. However, to apply extensively in the real world, the AI performance using other types of smartphone cameras and image quality taken by non-healthcare professional people will need to be substantiated in the future.

In conclusion, we established a high-performance DL model to detect, categorise and triage cataracts and multiple corneal diseases. Its diagnostic accuracy is comparable to that of corneal specialists. We believe that the AI using smartphone images can be applied for efficient triaging of anterior segment diseases.

Author affiliations

¹Department of Ophthalmology, University of Tsukuba, Tsukuba, Japan

²Information Technology Center, Nagoya University, Nagoya, Japan

³Graduate School of Informatics, Nagoya University, Nagoya, Japan

⁴Department of Ophthalmology, Tokyo Dental College Ichikawa General Hospital, Ichikawa, Japan

⁵Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan

⁶Miyata Eye Hospital, Miyakonojo, Japan

⁷Department of Ophthalmology, Osaka University Graduate School of Medicine, Osaka, Japan

⁸Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan

⁹Department of Ocular Pathology and Imaging Science, Kyushu University, Fukuoka, Japan

¹⁰Department of Ophthalmology, University of Yamanashi, Kofu, Japan

¹¹Department of Data Science, Institute of Statistical Mathematics, Tokyo, Japan

¹²National Institute of Informatics, Tokyo, Japan

Twitter Masahiro Oda @moda0, Takefumi Yamaguchi @TakefuYamaguchi and Masahiro Miyake @eyemiyake

Acknowledgements We thank Drs D Miyazaki, T Chikama, T Usui, Y Okada, H Eguchi, F Hotta, K Kamiya, J Yoshida, A Kobayashi, H Yokogawa, M Yamada, C Shigeyasu, H Mitamura, Y Hara, Y Yoshinaga, Y Hori, K Kakisu, S Takahashi, T Inomata, K Harada and K Shinozaki for providing a large number of anterior segment images. We also thank Mr. T Mihashi for providing the test software and Editage for the English language editing.

Contributors Concept and design: TO, KMori, YU, MO and TY. Acquisition of patient photograph: YU, TY, HF, RN, KMiyata and JS. Development of the network architectures as well as training and testing models, and evaluation of their performance: MO and KMori. Software engineering: YK. Critical revision of the manuscript for important intellectual content: NM, MM, MA, JS, KK and KMiyata. Management of this project: MM, MA, KMori and TO. Statistical analysis: HN. Obtained funding: KMori and TO. Administrative, technical or material support: MM, MA, KK, KMori and TO. Supervision: KMori and TO. Guarantor: TY and TO.

Funding This study was supported by the Japan Agency for Medical Research and Development (TO 19lk1010024h0003) and (YU 22hma322004h0001).

Competing interests TY: Grants (Novartis Pharma); honoraria for lectures (Alcon Japan, HOYA, Novartis Pharma, AMO Japan, Santen Pharmaceuticals, Senju Pharmaceutical, Johnson & Johnson), MM: Grants (Novartis Pharma); honoraria for lectures (Bayer Yakuhin, Kowa Pharmaceutical, Alcon Japan, HOYA, Novartis Pharma, AMO Japan, Santen Pharmaceutical, Senju Pharmaceutical, Johnson & Johnson K.K., Japan Ophthalmic Instruments Association). MA: Grants (Novartis, Santen), honoraria for lectures (Novartis, Takeda, Senju, Chugai, Kowa), Support for attending meetings and travel (Wakamoto, Novartis), Endowed (NIDEK).

Patient consent for publication The review committee stated that patient consent was not required for the retrospective study of slit-lamp microscopy images, because all slit-lamp images used in the study were deidentified. Informed consent was obtained from all participants for the prospective study of anterior segment images captured using smartphones.

Ethics approval This study involves human participants and was approved by the institutional ethics review board of all tertiary university hospitals (institution review board of Japanese Ophthalmological Society; protocol number: 15000133-20001). All the procedures conformed to the tenets of the Declaration of Helsinki and the Japanese Guidelines for Life Science and Medical Research. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. YOLO V5 exhibited sufficient performance in classifying the nine categories (figure 3, source codes are available in <https://github.com/modafone/corneaaai>).

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

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

ORCID iDs

Masahiro Oda <http://orcid.org/0000-0001-7714-422X>

Takefumi Yamaguchi <http://orcid.org/0000-0002-2546-2813>

Masahiro Miyake <http://orcid.org/0000-0001-7410-3764>

Naoyuki Maeda <http://orcid.org/0000-0001-6509-9054>

Jun Shimazaki <http://orcid.org/0000-0002-0435-3095>

Kensaku Mori <http://orcid.org/0000-0002-0100-4797>

Tetsuro Oshika <http://orcid.org/0000-0002-9071-7018>

REFERENCES

- Flaxman SR, Bourne RRA, Resnikoff S, *et al*. Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. *Lancet Glob Health* 2017;5:e1221-34.
- Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. *Br J Ophthalmol* 2012;96:614-8.
- GBD 2019 Blindness and Vision Impairment Collaborators on behalf of the Vision Loss Expert Group of the Global Burden of Disease Study. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study. *Lancet Glob Health* 2021;9:e130-43.
- Stevens GA, White RA, Flaxman SR, *et al*. Global prevalence of vision impairment and blindness: magnitude and temporal trends, 1990-2010. *Ophthalmology* 2013;120:2377-84.
- Kudo S-E, Ichimasa K, Villard B, *et al*. Artificial intelligence system to determine risk of T1 colorectal cancer metastasis to lymph node. *Gastroenterology* 2021;160:1075-84.

- 6 De Fauw J, Ledsam JR, Romera-Paredes B, *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50.
- 7 Li Z, Jiang J, Chen K, *et al.* Preventing corneal blindness caused by keratitis using artificial intelligence. *Nat Commun* 2021;12:3738.
- 8 Wang J, Ji J, Zhang M, *et al.* Automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening. *JAMA Netw Open* 2021;4:e218758.
- 9 Koyama A, Miyazaki D, Nakagawa Y, *et al.* Determination of probability of causative pathogen in infectious keratitis using deep learning algorithm of slit-lamp images. *Sci Rep* 2021;11:22642.
- 10 Gomez Rossi J, Rojas-Perilla N, Krois J, *et al.* Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. *JAMA Netw Open* 2022;5:e220269.
- 11 Kuo M-T, Hsu BW-Y, Yin Y-K, *et al.* A deep learning approach in diagnosing fungal keratitis based on corneal photographs. *Sci Rep* 2020;10:14424.
- 12 Hu S, Luan X, Wu H, *et al.* ACCV: automatic classification algorithm of cataract video based on deep learning. *Biomed Eng Online* 2021;20:78.
- 13 Fang X, Deshmukh M, Chee ML, *et al.* Deep learning algorithms for automatic detection of Pterygium using anterior segment photographs from slit-lamp and hand-held cameras. *Br J Ophthalmol* 2022;106:1642–7.
- 14 Wu X, Huang Y, Liu Z, *et al.* Universal artificial intelligence platform for collaborative management of cataracts. *Br J Ophthalmol* 2019;103:1553–60.
- 15 Gu H, Guo Y, Gu L, *et al.* Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs. *Sci Rep* 2020;10:17851.
- 16 Miyake M, Akiyama M, Kashiwagi K, *et al.* Japan ocular imaging registry: a national ophthalmology real-world database. *Jpn J Ophthalmol* 2022;66:499–503.
- 17 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.
- 18 Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;26:404–13.
- 19 He M, Li Z, Liu C, *et al.* Deployment of artificial intelligence in real-world practice: opportunity and challenge. *Asia Pac J Ophthalmol (Phila)* 2020;9:299–307.
- 20 Kann BH, Hosny A, Aerts H. Artificial intelligence for clinical oncology. *Cancer Cell* 2021;39:916–27.
- 21 He J, Baxter SL, Xu J, *et al.* The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25:30–6.
- 22 Santoni A, Thariat J, Maschi C, *et al.* Management of invasive squamous cell carcinomas of the conjunctiva. *Am J Ophthalmol* 2019;200:1–9.
- 23 Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of medicine. *JAMA* 2020;323:509–10.
- 24 Redmon J, Farhadi A. YOLOv3: an incremental improvement. *arXiv* 2018.
- 25 He K, Zhang X, Ren S, *et al.* YOLOv3: an incremental improvement. *IEEE Conference on Computer Vision and Pattern Recognition*; 2018
- 26 Freeman K, Dinnes J, Chuchu N, *et al.* Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ* 2020;368:m127.
- 27 Whitcher JP, Srinivasan M, Upadhyay MP. Corneal blindness: a global perspective. *Bull World Health Organ* 2001;79:214–21.
- 28 Liesegang TJ, Melton LJ, Daly PJ. Epidemiology of ocular herpes simplex. *Arch Ophthalmol* 1989;107:1155.
- 29 Inomata T, Nakamura M, Iwagami M, *et al.* Symptom-based stratification for hay fever: a crowdsourced study using the smartphone application AllerSearch. *Allergy* 2021;76:3820–4.