

UC San Diego

UC San Diego Previously Published Works

Title

Automated expert-level scleral spur detection and quantitative biometric analysis on the ANTERION anterior segment OCT system.

Permalink

<https://escholarship.org/uc/item/2nb96458>

Journal

British Journal of Ophthalmology, 108(5)

Authors

Bolo, Kyle

Apolo Aroca, Galo

Pardeshi, Anmol

et al.

Publication Date

2024-05-21

DOI

10.1136/bjo-2022-322328

Peer reviewed



OPEN ACCESS

Clinical science

Automated expert-level scleral spur detection and quantitative biometric analysis on the ANTERION anterior segment OCT system

Kyle Bolo ,¹ Galo Apolo Aroca ,¹ Anmol A Pardeshi,^{1,2} Michael Chiang,¹ Bruce Burkemper,^{1,2} Xiaobin Xie,³ Alex S Huang,⁴ Martin Simonovsky,⁵ Benjamin Y Xu ¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bjo-2022-322328>).

¹Roski Eye Institute, Department of Ophthalmology, University of Southern California Keck School of Medicine, Los Angeles, California, USA

²Department of Population and Public Health Sciences, University of Southern California Keck School of Medicine, Los Angeles, California, USA

³Eye Hospital of China Academy of Chinese Medical Sciences, Beijing, China

⁴Hamilton Glaucoma Center and Shiley Eye Institute, Department of Ophthalmology, University of California, San Diego, California, USA

⁵Heidelberg Engineering, Heidelberg, Germany

Correspondence to

Dr Benjamin Y Xu, Ophthalmology, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA; Benjamin.Xu@med.usc.edu

Received 7 August 2022

Accepted 14 June 2023

Published Online First

5 October 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: Bolo K, Apolo Aroca G, Pardeshi AA, et al. *Br J Ophthalmol* 2024;**108**:702–709.

ABSTRACT

Aim To perform an independent validation of deep learning (DL) algorithms for automated scleral spur detection and measurement of scleral spur-based biometric parameters in anterior segment optical coherence tomography (AS-OCT) images.

Methods Patients receiving routine eye care underwent AS-OCT imaging using the ANTERION OCT system (Heidelberg Engineering, Heidelberg, Germany). Scleral spur locations were marked by three human graders (reference, expert and novice) and predicted using DL algorithms developed by Heidelberg Engineering that prioritise a false positive rate <4% (FPR4) or true positive rate >95% (TPR95). Performance of human graders and DL algorithms were evaluated based on agreement of scleral spur locations and biometric measurements with the reference grader.

Results 1308 AS-OCT images were obtained from 117 participants. Median differences in scleral spur locations from reference locations were significantly smaller ($p < 0.001$) for the FPR4 ($52.6 \pm 48.6 \mu\text{m}$) and TPR95 ($55.5 \pm 50.6 \mu\text{m}$) algorithms compared with the expert ($61.1 \pm 65.7 \mu\text{m}$) and novice ($79.4 \pm 74.9 \mu\text{m}$) graders. Intergrader reproducibility of biometric measurements was excellent overall for all four (intraclass correlation coefficient range 0.918–0.997). Intergrader reproducibility of the expert grader (0.567–0.965) and DL algorithms (0.746–0.979) exceeded that of the novice grader (0.146–0.929) for images with narrow angles defined by OCT measurement of angle opening distance $500 \mu\text{m}$ anterior to the scleral spur (AOD500) $< 150 \mu\text{m}$.

Conclusions DL algorithms on the ANTERION approximate expert-level measurement of scleral spur-based biometric parameters in an independent patient population. These algorithms could enhance clinical utility of AS-OCT imaging, especially for evaluating patients with angle closure and performing intraocular lens calculations.

INTRODUCTION

The biometric properties of the anterior segment and its anatomical structures play an important role in the clinical care of patients with a range of ocular conditions. Specifically, anterior segment biometrics play an important role in the pathogenesis of primary angle closure disease (PACD), in which

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Deep learning (DL) algorithms can detect scleral spur locations in anterior segment optical coherence tomography (AS-OCT) images with expert-level performance; however, there is sparse information about the accuracy of AS-OCT measurements associated with these predicted scleral spur locations.

WHAT THIS STUDY ADDS

⇒ DL algorithms on the ANTERION OCT system (Heidelberg Engineering, Heidelberg, Germany) approximate expert-level detection of the scleral spur and measurement of anterior segment biometric parameters in a real-world clinical cohort. Performance of the algorithms generally exceeds that of a novice grader.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The automation of scleral spur detection and quantitative biometric analysis overcomes the time-dependent and expertise-dependent nature of AS-OCT imaging in the clinical setting. This technology provides clinicians with convenient access to data that could enhance care of patients with angle closure disease or patients receiving intraocular lens implantation.

aqueous humour outflow is impaired by apposition of the iris and trabecular meshwork, and closure of the anterior chamber angle (ACA).^{1–3} This process leads to primary angle closure glaucoma, a major cause of visual morbidity worldwide that currently affects more than 20 million people.^{4,5} In addition, the surgical treatment of eyes with cataract and high refractive error benefits from accurate biometric measurements when calculating the power and size of intraocular lenses (IOLs). Incorrect lens power leads to poor visual outcomes, and incorrect lens sizing can lead to harmful complications such as hyphema, uveitis, glaucomatous optic neuropathy or corneal decompensation.^{6,7}

There is a growing evidence that supports the clinical utility of anterior segment optical coherent tomography (AS-OCT) for measuring anterior

segment biometrics, many of which are based on scleral spur location. For example, angle opening distance (AOD) and trabecular iris space area (TISA) may find expanded roles in predicting progression of PACD and response to treatment with laser peripheral iridotomy (LPI).^{3 8 9} Quantitative OCT-based methods could complement gonioscopy, which remains the current standard for assessing the ACA despite being subjective, qualitative, variably reproducible and weakly correlated with AS-OCT measurements of angle width.^{10–16} In IOL selection, biometric parameters, including corneal curvature, anterior chamber depth (ACD) and lens thickness, are measured using optical or ultrasound methods and factored into modern IOL calculators.¹⁷ Anterior chamber width (ACW), also referred to as white-to-white distance, is important for sizing anterior chamber and phakic IOLs.^{18–21} Biometric parameters based on scleral spur location, such as lens vault (LV) and ACW, are potentially useful in IOL selection, but are difficult to measure and therefore rarely used in routine clinical practice.^{22 23}

Full biometric analysis of AS-OCT images on commercial devices currently requires specialised software and manual marking of scleral spurs, which is expertise-dependent and time-consuming, thereby presenting a barrier to widespread implementation.^{24 25} Prior studies have established the accuracy of scleral spur detection automated using deep learning (DL), a form of artificial intelligence.^{26 27} In this study, we investigate if biometric measurements associated with scleral spur locations predicted by DL algorithms on the Heidelberg ANTERION V.1.4 swept-source OCT system (Heidelberg Engineering, Heidelberg, Germany) approximate interexpert reproducibility in an independent patient population and clinical environment.

METHODS

Scleral spur detection algorithm

DL algorithms to automate scleral spur detection were developed and tested internally by Heidelberg Engineering (Heidelberg, Germany) prior to this study. While these algorithms are proprietary, some information was provided by Heidelberg Engineering about their development. In brief, a set of 4798 ANTERION AS-OCT images from 1 or both eyes of 360 patients were evaluated by an expert ophthalmologist to identify scleral spur locations. These images were divided into non-overlapping training (3810 images; 80%) and test (979 images; 20%) datasets. The training dataset was used to train a convolutional neural network (CNN) based on the M2U-Net architecture that predicts scleral spur location within a predefined region of interest (ROI).²⁸ The ROI is a 256×256 pixel area around the ACA determined heuristically based on the posterior boundary of the cornea and the anterior boundary of the iris as defined by the ANTERION's segmentation algorithms. Reference scleral spur locations were transformed into reference heatmaps containing a Gaussian function with SD of 10 pixels centred on the reference location. Data augmentation, including affine deformations, noising and blurring, was used to increase the robustness of the CNN. The subpixel-refined position and intensity of the strongest peak in the predicted heatmap were used to estimate the position and confidence (level of certainty ranging from 0 to 1) of the scleral spur. The test dataset was used to select two operating points along the receiver operating characteristic curve (online supplemental figure 1) for further analysis, one more conservative to limit the false positive rate (FPR; scleral spur marked by the algorithm but not the ophthalmologist) below 4% (FPR4 algorithm) and the other more aggressive to ensure a true positive rate (TPR;

scleral spur marked by the algorithm and ophthalmologist) above 95% (TPR95 algorithm).

Acquisition and analysis of validation dataset

Patients 18 years of age and older undergoing routine eye examinations were prospectively and consecutively recruited from attending glaucoma clinics at the Roski Eye Institute at the USC and attending comprehensive ophthalmology clinics at the Doheny Eye Institute at the University of California Los Angeles. Each of these clinics had approximately 20 patients per clinic, the majority of whom were follow-ups. Recruitment occurred from March 2021 to August 2021. Exclusion criteria included corneal opacities that precluded AS-OCT imaging and prior history of ocular trauma.

AS-OCT imaging was performed using the ANTERION and Metrics Application. All images were obtained by trained technicians following a standardised imaging protocol. Imaging of both eyes was performed in the seated position prior to pupillary dilation in a dark room under standardised lighting conditions (<0.01 lux) at the imaging plane. Participants were instructed to maintain fixation on the internal fixation target with their eyelids open without retraction by the technician.

The scleral spur was identified as the inward projection at the junction of the sclera and cornea.²⁹ Scleral spur locations in all six B-scans (separated by 30°, creating 12 angle sectors at 0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300° and 330°) were marked by three human graders: (1) an expert trained grader (AAP; reference grader) with experience marking over 40 000 scleral spurs after a 5-hour training period of marking 500 scleral spurs under the supervision of two glaucoma specialists; (2) an expert glaucoma specialist with experience marking over 10 000 scleral spurs (BYX; expert grader); (3) a novice trained grader (ASH; novice grader) with experience marking fewer than 100 scleral spurs. The reference grader previously demonstrated low intragrader variability in scleral spur locations and AS-OCT measurements of biometric parameters.^{27 30} Scleral spur locations were also predicted by the FPR4 and TPR95 algorithms.

The anterior and posterior boundaries of the cornea and lens, and the anterior boundary of the iris, were computed automatically by the ANTERION's segmentation algorithm. The reference grader made minor segmentation adjustments of the angle recess, including the posterior cornea and anterior iris, in fewer than 15 images (1.1% of total) prior to obtaining biometric measurements. After the scleral spurs were marked, eight scleral spur-based biometric parameters were measured in an automated fashion by the ANTERION software: AOD, TISA and scleral spur angle (SSA) at 500 and 750 µm from the scleral spur, ACW, and LV. AOD500/750 was defined as the perpendicular distance from the TM at 500 or 750 µm anterior to the scleral spur to the anterior iris surface. TISA500/750 was defined as the area bounded anteriorly by AOD500/750; posteriorly by a line drawn from the scleral spur perpendicular to the plane of the inner scleral wall to the opposing iris; superiorly by the inner corneal-scleral wall; and inferiorly by the iris surface. SSA500/750 was defined as the angles formed by lines originating at the scleral spur and terminating at the TM or anterior iris surface 500 or 750 µm anterior to the scleral spur. ACW was defined as the distance between scleral spurs. LV was defined as the perpendicular distance from the apex of the anterior lens surface to a line between scleral spurs.

A subset of images was classified as having narrow angles, defined as an AOD500 measurement <150 µm by the reference

grader. This threshold was chosen to define narrow angles due to its high sensitivity and specificity of less strict AOD500 thresholds for detecting gonioscopic angle closure in a prior study.³¹ Narrow angles were not defined based on gonioscopy for several reasons: (1) the majority of patients were not glaucoma patients, and therefore, did not receive gonioscopy; (2) there is inter-grader variability in the detection of gonioscopic angle closure and (3) angle widths associated with gonioscopic angle closure vary significantly by quadrant.^{10 31 32}

Images with borderline or poor interpretability due to eyelid and other imaging artefacts were included in the analysis so that false negative rates (FNRs) and FPRs could be calculated for the expert and novice graders and both DL algorithms. In addition, human graders were not provided specific instruction about what constituted a gradable scleral spur; the decision to grade an image was left to the discretion of each grader. A reference false negative (FN_{ref}) was defined as a scleral spur identified by the reference grader but not by another grader or algorithm. A reference false positive (FP_{ref}) was defined as a scleral spur marked by another grader or algorithm but not by the reference grader. A consensus false negative was defined as a scleral spur marked by all three human graders but not by a DL algorithm. A consensus false positive was defined as a scleral spur marked by a DL algorithm but not any of the three human graders.

Statistical analysis

Scleral spur location differences were calculated as the Euclidean distance between scleral spur locations by the reference grader and second human grader or DL algorithm. Normality testing was performed on scleral spur location differences using the Kolmogorov-Smirnov test. Medians and IQRs were calculated based on non-normality of the data. Scleral spur location differences were grouped by grader or algorithm and compared using the Kruskal-Wallis test. Pairwise comparisons of scleral spur location differences between groups (six comparisons in total) were performed using the post hoc Dunn's test adjusted for multiple comparisons at a significance level of 0.05. Intraclass correlation coefficients (ICCs) were calculated for each biometric parameter measured in all AS-OCT images to assess the intergrader agreement between the reference grader and a second human grader (expert or novice) or DL algorithm (FPR4 and TPR95). ICCs were also calculated for each biometric parameter measured in a single sector (superior or temporal) of only one eye per participant to eliminate intraeye and intraparticipant correlations. Bland-Altman plots were generated for AOD500 to assess intergrader agreement across the entire range of angle widths. All analyses were performed by using the R statistical package (V4.0.3) at a significance level of 0.05.

RESULTS

In total, 1308 AS-OCT images were obtained from 117 participants, which included 2616 potential scleral spurs; however, not all of the scleral spurs were gradable due to eyelid or other imaging artefacts. Mean age was 52.1 ± 17.6 years with 59 males (50.4%) and 58 females (49.6%). Among all participants, 50 (42.7%) were Caucasian, 32 (27.4%) were Hispanic, 21 (17.9%) were Asian, and 7 (6.0%) were black, and 7 (6.0%) had unknown race/ethnicity.

In total, the reference grader marked 1504 spurs, the expert grader marked 1726 spurs, the novice grader marked 1622 spurs, the FPR4 algorithm marked 1459 spurs and the TPR95 algorithm marked 1722 spurs. Given that the reference grader detected fewer scleral spurs than other graders, the images

marked by the expert grader but not reference grader were reviewed. Among these 237 images, the large majority included eyelid artefacts ($N=222$, 93.4%) or shadowing by eyelashes or pterygia ($N=12$, 5.1%) that partially obscured the angle recess. Distributions of scleral spur location differences compared with the reference grader varied by grader or algorithm (figures 1 and 2). Median and IQR of scleral spur location differences were $61.1 \pm 65.7 \mu\text{m}$ for the expert grader, $79.4 \pm 74.9 \mu\text{m}$ for the novice grader, $52.6 \pm 48.6 \mu\text{m}$ for the FPR4 algorithm and $55.5 \pm 50.6 \mu\text{m}$ for the TPR95 algorithm. There were significant differences ($p < 0.001$) among the four groups of scleral spur location differences. Pairwise comparisons demonstrated a non-significant difference in scleral spur location differences between the DL algorithms ($p = 0.33$) and significant differences between all other pairs of graders and algorithms ($p < 0.001$).

There was a wide range of angle widths (mean 0.41 ± 0.25 mm) based on the distribution of AOD500 measurements by the reference grader (online supplemental figure 2). Measurement agreement between the reference grader and the expert grader or either algorithm was excellent (ICC range 0.955–0.997) and similar for all parameters (table 1). Measurement agreement for the novice grader was lower but still excellent for all parameters (ICC range 0.918–0.994). Bland-Altman plots for AOD500 reflected consistent agreement across the entire range of AOD500 measurements for all four (figure 3). ICCs of measurements from only superior or temporal sectors from one eye per participants showed similar trends (online supplemental tables 1 and 2) as the primary analysis.

Among the 1504 AS-OCT images graded by the reference grader, 198 (13.2%) had narrow angles ($\text{AOD500} < 150 \mu\text{m}$). Among the subset of participants who received gonioscopy as part of their clinical examination, 9 of 36 (25%) had gonioscopic angle closure (inability to visualise the pigmented trabecular meshwork) in at least two quadrants. Among measurements from these images with narrow angles, ICCs for ACW and LV were similar to those for the overall study population (ICC range 0.856–0.979) whereas ICCs for angle width measurements tended to be lower (ICC range 0.146–0.878) (table 2). ICC is defined as: $\text{intereye variance} / (\text{intereye variance} + \text{intraeye variance})$. Therefore, the lower ICC values likely reflect the lower inter-eye variance of angle width measurements associated with narrow angles. In contrast, the Bland-Altman plots demonstrate consistent limits of agreement for AOD500 measurements below and above the AOD500 threshold for narrow angles (figure 3). Bland-Altman plots for the TISA500, ACW and LV also showed good interexpert agreement across the range of measurements (online supplemental figures 3–5). The difference in ICC (agreement with the reference grader) between the expert and novice graders was more pronounced, favouring the expert grader, in the subset of narrow angle images compared with all images. The DL algorithms matched if not exceeded the agreement between the expert and reference graders in the subset of narrow angle images (table 2).

Rates of FN_{ref} and FP_{ref} differed by grader and algorithm (figure 4). The expert and novice graders and TPR95 algorithm all had $FN_{ref} < 3.0\%$ and $FPR_{ref} > 10.0\%$ whereas the FPR4 algorithm had $FN_{ref} = 12.6\%$ and $FPR_{ref} = 9.6\%$. Compared with the consensus, the FN_{con} of the FPR4 algorithm (12.3%) was higher than the TPR95 algorithm (2.7%) whereas the difference in the FPR_{con} was smaller (FPR4 1.1% vs TPR95 4.1%). On visual inspection of misclassified images by the TPR95 algorithm, many of the images had obvious lid, shadowing or motion artefacts that make scleral spur detection difficult (online supplemental figure 6).

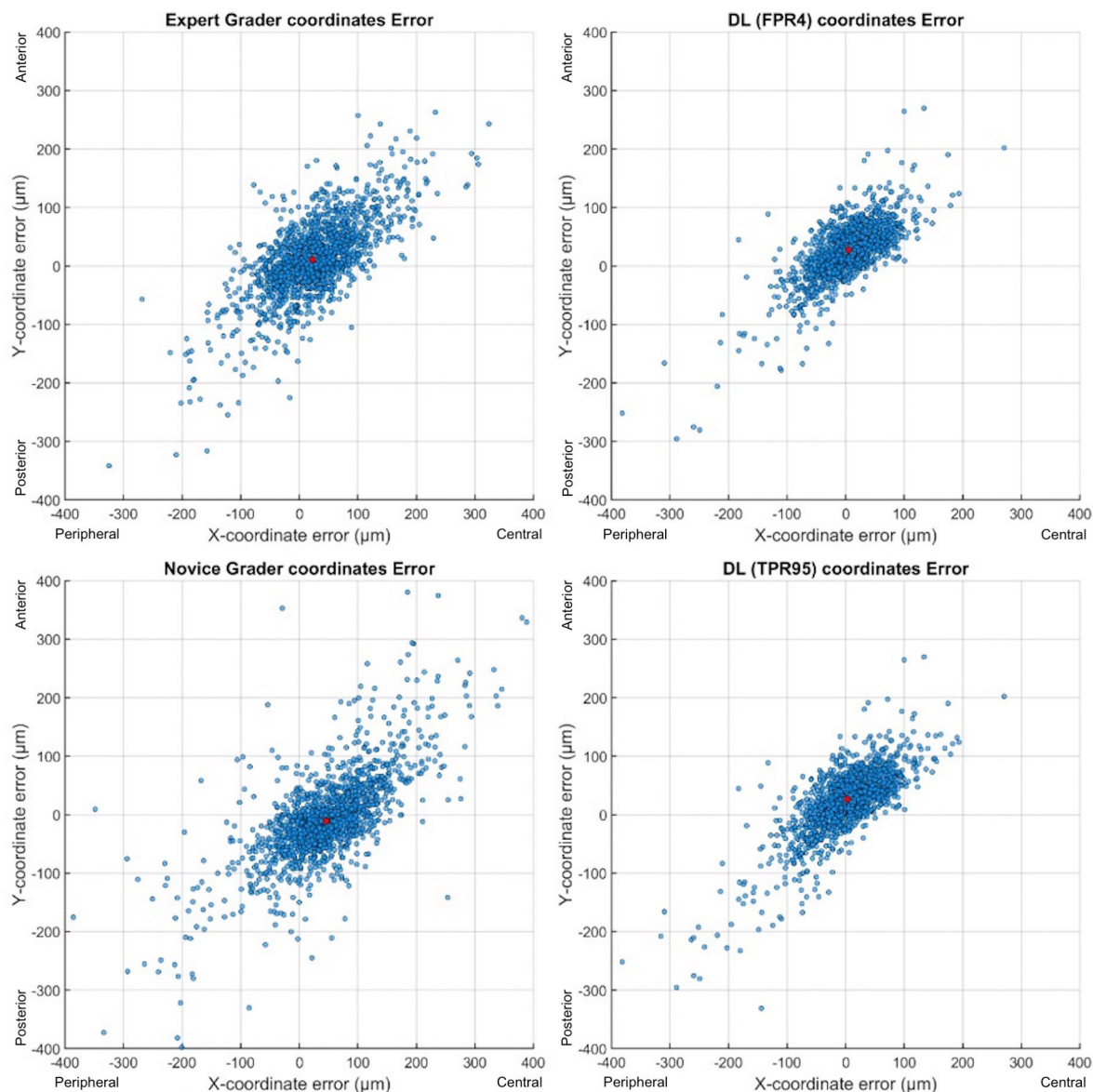


Figure 1 Human-human and human-machine differences in scleral spur locations. Scatter plots showing X-coordinate and Y-coordinate errors in comparison to the reference grader for the expert grader (top left), novice grader (bottom left), FPR4 algorithm (top right) and TPR95 algorithm (bottom right). Red dots indicate median X-coordinate and Y-coordinate differences. DL, deep learning; FPR4, false positive rate <4%; TPR95, true positive rate >95%.

DISCUSSION

In this study, DL algorithms for the ANTERION OCT system achieved expert-level performance predicting scleral spur locations and measurements of scleral spur-based biometric parameters in a large set of AS-OCT images from an independent patient population and clinical environment. Both the conservative (FPR4) and aggressive (TPR95) algorithms generally approximated the performance of the expert grader and exceeded that of the novice grader, especially among images with narrow angles. The TPR95 algorithm more closely approximated the FNR and FPR of the human graders, while the FPR4 algorithm made substantially fewer predictions. These findings support the implementation of the TPR95 algorithm for scleral spur detection and automated biometric analysis of ANTERION images, which in turn could greatly enhance the accessibility and utility of quantitative AS-OCT imaging.

Measurements of scleral spur-based biometric parameters are dependent on accurate identification of scleral spur location, which is variable even among experienced graders.^{24 25} Both the FPR4 and TPR95 algorithms produced similar accuracy in predicting scleral spur locations relative to the reference grader, with median differences that were smaller than those of the expert and novice graders (<60 μm for both algorithms). This performance is comparable to that of a DL algorithm developed by Xu *et al* for the Tomey CASIA SS-1000, in which the mean human-machine scleral spur location difference was $73.08 \pm 52.06 \mu\text{m}$.²⁷ Pham *et al* developed a different DL algorithm for the CASIA SS-1000 and plots of human-human and human-machine differences are on a similar scale to those from this study.²⁶ These findings suggest that the FPR4 and TPR95 algorithms achieve expert-level performance in scleral spur detection that approximates if not exceeds the agreement between two experienced graders.

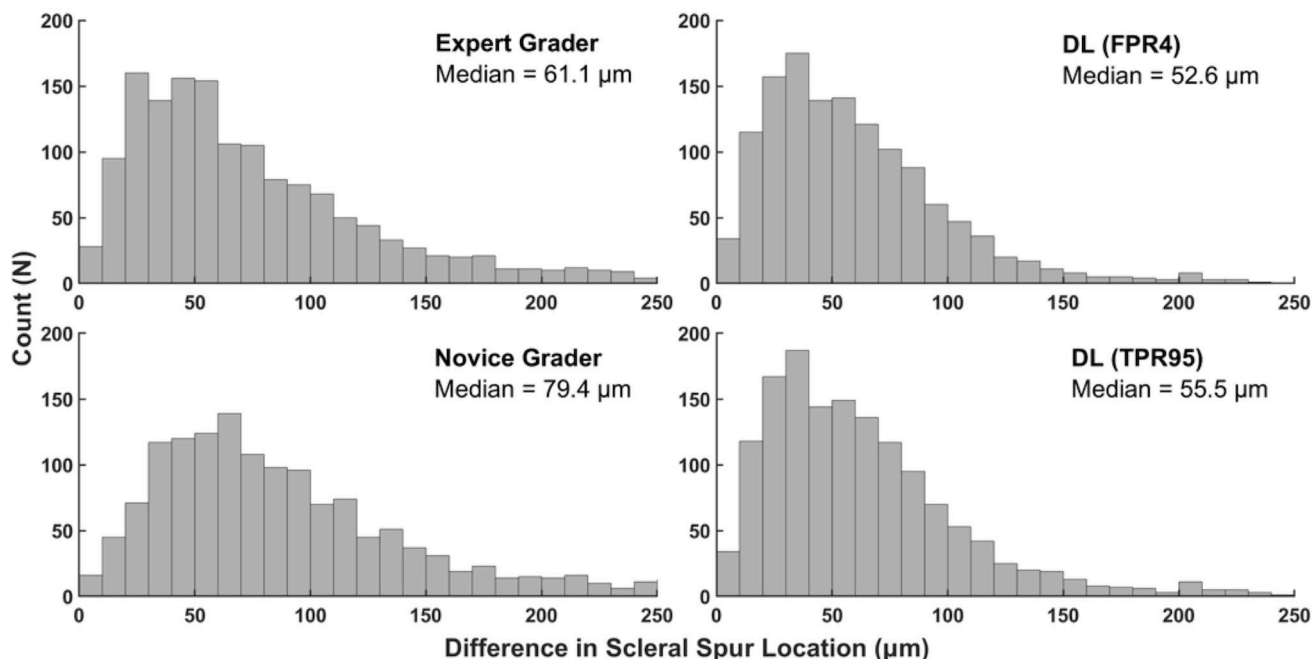


Figure 2 Human-human and human-machine differences in scleral spur locations. Histograms showing the Euclidean distance between scleral spur locations by the reference grader and expert grader (top left), novice grader (bottom left), FPR4 algorithm (top right) and TPR95 algorithm (bottom right). DL, deep learning; FPR4, false positive rate <4%; TPR95, true positive rate >95%.

Limited access to quantitative measurements of scleral spur-based biometric parameters has hindered the development and implementation of novel clinical methods for evaluating and treating a range of ocular conditions, including PACD, refractive error and cataract. Our findings suggest that biometric measurements associated with scleral spur predictions by both algorithms are highly correlated with measurements by the reference grader and approximate the agreement between two experienced human graders, including in eyes with narrow angles. An automated method that provides access to expert-level measurements of scleral spur-based biometric measurements could help modernise the clinical evaluation and management of patients with PACD. Measurements of AOD and TISA are associated with IOP and anatomical variations in PACD eyes and may predict a higher risk of PACD progression or poor angle widening after LPI.^{8 15 16} In addition, automated measurements of ACW and LV could be beneficial for IOL selection: ACW is helpful in sizing anterior chamber and phakic IOLs, and there is evidence that LV could play an important role in determining effective lens position and calculating IOL power.^{18–23}

Our results demonstrate that rates of scleral spur detection are highly variable under real-world conditions without eyelid retraction during imaging, even among experienced graders. This point, which has not been previously studied, suggests there is differing confidence among graders when deciding whether to mark a scleral spur. Based on number of scleral spurs marked, the reference grader appeared the most conservative and the expert grader the most aggressive among human graders. This trend reflects the graders' individual thresholds for identifying scleral spurs in the context of eyelid and other imaging artefacts that partially obscure the angle recess. The reference grader, having been trained to detect scleral spurs for scientific studies, marked fewer images with artefact, whereas the expert grader, a clinician, was less conservative in marking scleral spurs in the presence of imaging artefacts. The TPR95 algorithm approximated the FNR and FPR of the expert grader (1.0% and 15.8% vs 2.9% and 17.4%). While the more conservative FPR4 algorithm had a lower FPR compared with the TPR95 algorithm (9.6% vs 17.4%), this came at the expense of a higher FNR (12.6% vs 2.9%). Despite the greater number of scleral spurs

Table 1 Human-human and human-machine reproducibility of measurements of scleral spur-based biometric parameters

	Expert grader	Novice grader	FPR4	TPR95
ACW	0.966 (0.962 to 0.970)	0.943 (0.936 to 0.949)	0.981 (0.979 to 0.984)	0.976 (0.973 to 0.978)
LV	0.996 (0.995 to 0.996)	0.994 (0.993 to 0.994)	0.997 (0.997 to 0.998)	0.997 (0.996 to 0.997)
AOD500	0.961 (0.957 to 0.965)	0.925 (0.917 to 0.932)	0.976 (0.974 to 0.979)	0.973 (0.970 to 0.975)
AOD750	0.976 (0.973 to 0.978)	0.950 (0.945 to 0.955)	0.981 (0.979 to 0.983)	0.977 (0.975 to 0.979)
TISA500	0.965 (0.961 to 0.968)	0.918 (0.909 to 0.926)	0.980 (0.977 to 0.982)	0.976 (0.974 to 0.979)
TISA750	0.974 (0.971 to 0.977)	0.934 (0.926 to 0.94)	0.984 (0.982 to 0.986)	0.981 (0.979 to 0.983)
SSA500	0.955 (0.95 to 0.959)	0.928 (0.92 to 0.935)	0.970 (0.967 to 0.973)	0.968 (0.964 to 0.971)
SSA750	0.969 (0.966 to 0.972)	0.945 (0.939 to 0.95)	0.979 (0.976 to 0.981)	0.973 (0.971 to 0.976)

Intraclass correlation coefficients with 95% CIs comparing measurements from all sectors by the reference grader and a second human grader or DL algorithm.

ACW, anterior chamber width; AOD500/750, angle opening distance 500/750 μm; DL, deep learning; FPR4, false positive rate <4%; LV, lens vault; SSA500/750, scleral spur angle 500/750 μm; TISA500/750, trabecular Iris Space Area 500/750 μm; TPR95, true positive rate >95%.

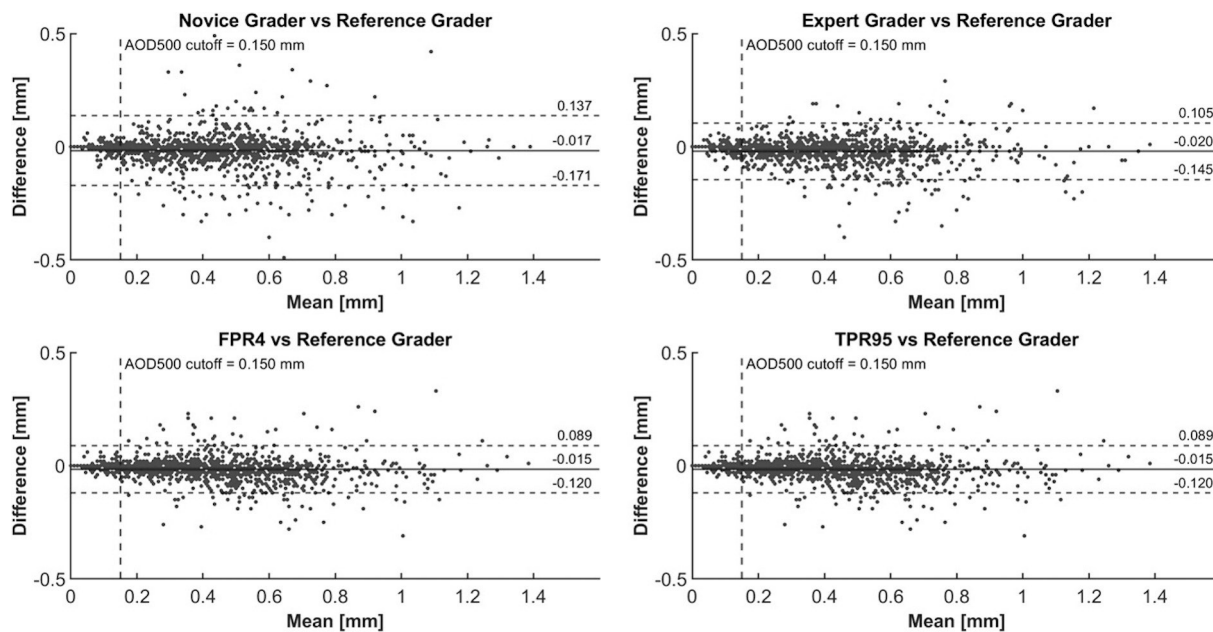


Figure 3 Bland-Altman plots of human-human and human-machine comparisons of AOD500 measurements. Vertical dotted line indicates cut-off (AOD500<150 μm) for narrow angles. AOD, angle opening distance; FPR4, false positive rate <4%; TPR95, true positive rate >95%.

identified by the TPR95 algorithm, measurement agreement between the reference grader and both algorithms were similar. In a busy clinical environment, the higher TPR of the TPR95 algorithm is likely of greater utility than the lower FPR of the FPR4 algorithm as it is more convenient to ignore a questionably marked scleral spur than to manually mark a more obvious one.

Our study has several strengths compared with prior studies on automated scleral spur detection.^{26 27} First, DL algorithms maintained expert-level performance in a real-world clinical environment, defined as a diverse cohort of patients of various ages and races who were recruited from comprehensive and glaucoma clinics during routine delivery of eye care. This validation cohort and setting was completely independent from the cohort and environment in which the algorithm was developed. These findings support the generalisability and widespread implementation of DL algorithms in diverse practice settings, while prior studies that used smaller and more homogenous cohorts do not.^{26 27} Second, images with eyelid or other imaging artefacts were not omitted from in the validation dataset. This approach allowed us to assess variability in human grader and algorithm confidence in scleral spur detection and evaluate its effect on

detection rates and measurement agreement. It also avoids introducing biases associated with analysing only a subset of images and applying arbitrary definitions of image quality that may be difficult in real-world practice environments. Third, all images were graded by a novice grader in addition to a second expert grader, which allowed us to determine that there is a benefit to using DL algorithms over a trained but inexperienced grader.

Our study also has several limitations. First, the reference grader was relatively conservative and marked fewer images than the other human graders and TPR95 algorithm. Post hoc analysis revealed eyelid or other shadowing artefacts in over 98% of these images. Second, while the overall number of images analysed was large, only 13.2% had narrow angles, which contributed to wider CIs in the ICC analysis of this subset of images. In the future, a larger cohort would be beneficial for more detailed study of narrow angles and individual sectors of the eye. Third, less than half of participants received gonioscopy; therefore, we were unable to assess algorithm performance based on gonioscopic angle status. Our OCT-based definition of narrow angles has relatively high (>80%) sensitivity and specificity for gonioscopic angle closure and our use of a quantitative OCT-based

Table 2 Human-human and human-machine reproducibility of measurements of scleral spur-based biometric parameters in narrow angles (AOD500 less than 150 μm)

	Expert grader	Novice grader	FPR4	TPR95
ACW	0.931 (0.907 to 0.949)	0.856 (0.807 to 0.894)	0.972 (0.958 to 0.981)	0.959 (0.944 to 0.970)
LV	0.965 (0.952 to 0.975)	0.929 (0.904 to 0.949)	0.979 (0.970 to 0.986)	0.973 (0.963 to 0.981)
AOD500	0.548 (0.435 to 0.644)	0.146 (-0.002 to 0.288)	0.777 (0.699 to 0.836)	0.746 (0.668 to 0.807)
AOD750	0.764 (0.698 to 0.817)	0.528 (0.416 to 0.623)	0.861 (0.812 to 0.898)	0.821 (0.767 to 0.863)
TISA500	0.762 (0.682 to 0.824)	0.267 (0.105 to 0.416)	0.852 (0.790 to 0.897)	0.796 (0.721 to 0.852)
TISA750	0.792 (0.721 to 0.847)	0.257 (0.094 to 0.406)	0.878 (0.826 to 0.916)	0.824 (0.759 to 0.873)
SSA500	0.567 (0.458 to 0.660)	0.219 (0.073 to 0.355)	0.782 (0.706 to 0.841)	0.750 (0.673 to 0.811)
SSA750	0.765 (0.699 to 0.818)	0.574 (0.469 to 0.662)	0.854 (0.803 to 0.893)	0.816 (0.761 to 0.859)

Intraclass correlation coefficients with 95% CIs comparing measurements from all sectors by the reference grader and a second human grader or DL algorithm.

ACW, anterior chamber width; AOD500/750, angle opening distance 500/750 μm; DL, deep learning; FPR95, false positive rate <4%; LV, lens vault; SSA500/750, scleral spur angle 500/750 μm; TISA500/750, trabecular iris space area 500/750 μm; TPR95, true positive rate >95%.

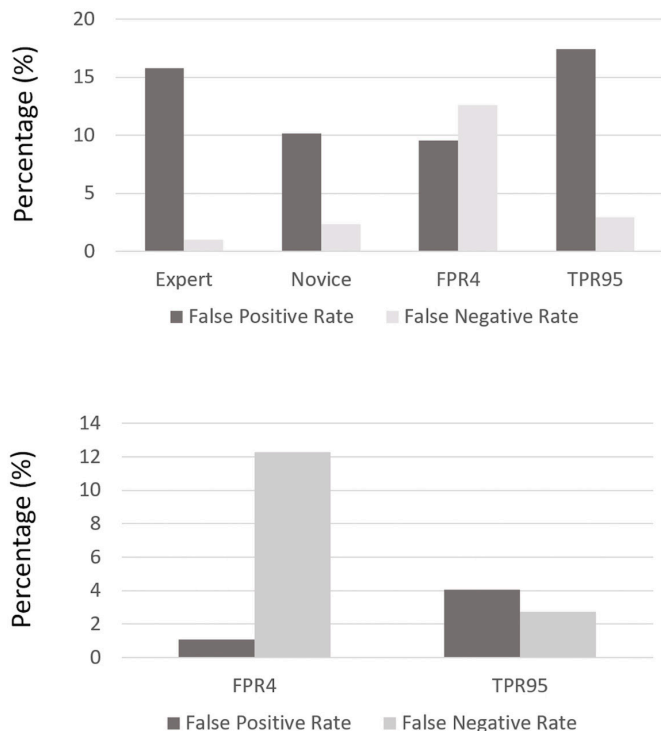


Figure 4 False positive rate and false negative rates relative to the reference grader (top) and consensus between all three human graders (bottom). FPR4, false positive rate <4%; TPR95, true positive rate >95%.

definition of narrow angles may be more appropriate for evaluating the performance of quantitative analysis algorithms; however, further study is needed to determine if human-human and human-machine limits of agreement are sufficient for detection and evaluation of narrow angles. Finally, the described algorithms are only available for images acquired on the ANTERION OCT system, and their expert-level performance would likely not generalise to images acquired on other AS-OCT devices.

In conclusion, DL algorithms provide expert-level scleral spur detection and biometric analysis in a large set of AS-OCT images from a diverse clinical cohort. There appears to be a benefit to using the TPR95 algorithm compared with grading by a novice in terms of the number of scleral spurs identified and the accuracy of biometric measurements. This study supports the implementation of the TPR95 algorithm in diverse patient populations and real-world practice settings. While this technology has the potential to expand the clinical utility of AS-OCT imaging and modernise the care of ocular conditions dependent on accurate anterior segment biometry, further studies are needed to help guide its use in routine clinical practice and decision-making processes.

X Benjamin Y Xu @BenXuLab

Contributors GAA, AAP, ASH and BYX evaluated participants and AS-OCT images. KB, GAA, MC, BB and BYX completed the statistical analysis. KB and BYX prepared the manuscript. MS contributed to DL algorithm design and its description in the Methods section. XX contributed background information and to study design. BYX is guarantor.

Funding This work was supported by grant K23 EY029763 from the National Eye Institute, National Institute of Health, Bethesda, Maryland and an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, New York, New York, USA.

Competing interests BYX and ASH receive research support from Heidelberg Engineering. MS is employed by Heidelberg Engineering.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the University of Southern California (USC) Institutional Review Board (IRB number HS-17-00684). All study procedures adhered to the recommendations of the Declaration of Helsinki. Written informed consent was obtained from all participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

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

Kyle Boló <http://orcid.org/0000-0001-7885-192X>
 Galo Apolo Aroca <http://orcid.org/0000-0001-5751-6102>
 Benjamin Y Xu <http://orcid.org/0000-0003-1573-988X>

REFERENCES

- Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA* 2014;311:1901–11.
- Foster PJ, Buhrmann R, Quigley HA, et al. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol* 2002;86:238–42.
- Shan J, DeBoer C, Xu BY. Anterior segment optical coherence tomography: applications for clinical care and scientific research. *Asia Pac J Ophthalmol (Phila)* 2019;8:146–57.
- Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006;90:262–7.
- Tham Y-C, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 2014;121:2081–90.
- Durr GM, Ahmed IK. Intraocular lens complications: decentration, uveitis-glaucoma-hyphema syndrome, opacification, and refractive surprises. *Ophthalmology* 2021;128:e186–94.
- Shen JF, Deng S, Hammersmith KM, et al. Intraocular lens implantation in the absence of zonular support: an outcomes and safety update: a report by the American Academy of Ophthalmology. *Ophthalmology* 2020;127:1234–58.
- Xu BY, Friedman DS, Foster PJ, et al. Ocular biometric risk factors for progression of primary angle closure disease: the Zhongshan angle closure prevention trial. *Ophthalmology* 2022;129:267–75.
- Xu BY, Friedman DS, Foster PJ, et al. Anatomic changes and predictors of angle widening after laser peripheral Iridotomy: the Zhongshan angle closure prevention trial. *Ophthalmology* 2021;128:1161–8.
- Phu J, Wang H, Khuu SK, et al. Anterior chamber angle evaluation using gonioscopy: consistency and agreement between optometrists and ophthalmologists. *Optom Vis Sci* 2019;96:751–60.
- Chiang M, Guth D, Pardeshi AA, et al. Glaucoma expert-level detection of angle closure in gonioscopic images with convolutional neural networks: the Chinese American Eye study. *Am J Ophthalmol* 2021;226:100–7.
- Pardeshi AA, Song AE, Lazkani N, et al. Intradevice repeatability and interdevice agreement of ocular biometric measurements: a comparison of two swept-source anterior segment OCT devices. *Transl Vis Sci Technol* 2020;9:14.
- Fukuda S, Kawana K, Yasuno Y, et al. Repeatability and reproducibility of anterior ocular biometric measurements with 2-dimensional and 3-dimensional optical coherence tomography. *J Cataract Refract Surg* 2010;36:1867–73.
- Baskaran M, Iyer JV, Narayanaswamy AK, et al. Anterior segment imaging predicts incident gonioscopic angle closure. *Ophthalmology* 2015;122:2380–4.
- Randhawa J, Chiang M, Porporato N, et al. Generalisability and performance of an OCT-based deep learning classifier for community-based and hospital-based detection of gonioscopic angle closure. *Br J Ophthalmol* 2023;107:511–7.
- Porporato N, Tun TA, Baskaran M, et al. "Towards "automated gonioscopy": a deep learning algorithm for 360° angle assessment by swept-source optical coherence tomography". *Br J Ophthalmol* 2022;106:1387–92.
- Kane JX, Chang DF. Intraocular lens power formulas, biometry, and intraoperative aberrometry: a review. *Ophthalmology* 2021;128:e94–114.

- 18 Bruner C, Skanchy DF, Wooten JP, *et al.* Anterior chamber lens SIZING: comparison of white-to-white and scleral spur-to-Scleral spur methods. *J Cataract Refract Surg* 2020;46:95–101.
- 19 Deshpande K, Shroff R, Biswas P, *et al.* Phakic Intraocular lens: getting the right size. *Indian J Ophthalmol* 2020;68:2880–7.
- 20 Montés-Micó R, Tañá-Rivero P, Aguilar-Córcoles S, *et al.* Angle-to-angle and spur-to-spur distance analysis with high-resolution optical coherence tomography. *Eye Vis (Lond)* 2020;7:42.
- 21 Nemeth G, Hassan Z, Szalai E, *et al.* Comparative analysis of white-to-white and angle-to-angle distance measurements with partial coherence interferometry and optical coherence tomography. *J Cataract Refract Surg* 2010;36:1862–6.
- 22 Takagi Y, Kojima T, Nishida T, *et al.* Prediction of anterior chamber volume after implantation of posterior chamber phakic Intraocular lens. *PLoS One* 2020;15:e0242434.
- 23 Yan C, Yao K. Effect of lens vault on the accuracy of intraocular lens calculation formulas in shallow anterior chamber eyes. *Am J Ophthalmol* 2022;233:57–67.
- 24 Maram J, Pan X, Sadda S, *et al.* Reproducibility of angle metrics using the time-domain anterior segment optical coherence tomography: intra-observer and inter-observer variability. *Curr Eye Res* 2015;40:496–500.
- 25 Kim DY, Sung KR, Kang SY, *et al.* Characteristics and reproducibility of anterior chamber angle assessment by anterior-segment optical coherence tomography. *Acta Ophthalmol* 2011;89:435–41.
- 26 Pham TH, Devalla SK, Ang A, *et al.* Deep learning algorithms to isolate and quantify the structures of the anterior segment in optical coherence tomography images. *Br J Ophthalmol* 2021;105:1231–7.
- 27 Xu BY, Chiang M, Pardeshi AA, *et al.* Deep neural network for scleral spur detection in anterior segment OCT images: the Chinese American Eye study. *Transl Vis Sci Technol* 2020;9:18.
- 28 Laibacher T, Weyde T, Jalali S. M2U-Net: effective and efficient retinal vessel segmentation for resource-constrained environments. 2018. Available: <https://arxiv.org/abs/1811.07738v3> [Accessed 9 Dec 2021].
- 29 Ho S-W, Baskaran M, Zheng C, *et al.* Swept source optical coherence tomography measurement of the Iris-trabecular contact (ITC) index: a new parameter for angle closure. *Graefes Arch Clin Exp Ophthalmol* 2013;251:1205–11.
- 30 Xu BY, Burkemper B, Lewinger JP, *et al.* Correlation between intraocular pressure and angle configuration measured by OCT: the Chinese American Eye study. *Ophthalmol Glaucoma* 2018;1:158–66.
- 31 Narayanaswamy A, Sakata LM, He M-G, *et al.* Diagnostic performance of anterior chamber angle measurements for detecting eyes with narrow angles: an anterior segment OCT study. *Arch Ophthalmol* 2010;128:1321–7.
- 32 Xu BY, Pardeshi AA, Burkemper B, *et al.* Differences in anterior chamber angle assessments between gonioscopy, eyecam, and anterior segment OCT: the Chinese American Eye study. *Transl Vis Sci Technol* 2019;8:5.