



Published in final edited form as:

*Br J Ophthalmol.* 2023 April ; 107(4): 511–517. doi:10.1136/bjophthalmol-2021-319470.

## Generalizability and Performance of an OCT-Based Deep Learning Classifier for Community- and Hospital-Based Detection of Gonioscopic Angle Closure

Jasmeen Randhawa<sup>1</sup>, Michael Chiang<sup>2</sup>, Natalia Porporato<sup>3</sup>, Anmol A Pardeshi<sup>2</sup>, Justin Dredge<sup>2</sup>, Galo Apolo Aroca<sup>2</sup>, Tin A Tun<sup>3</sup>, Joanne HuiMin Quah<sup>4</sup>, Marcus Tan<sup>5</sup>, Risa Higashita<sup>6</sup>, Tin Aung<sup>3,5</sup>, Rohit Varma<sup>7</sup>, Benjamin Y Xu<sup>8</sup>

<sup>1</sup>Keck School of Medicine of the University of Southern California, Los Angeles, California, USA.

<sup>2</sup>Roski Eye Institute, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA.

<sup>3</sup>Singapore National Eye Centre, Singapore Eye Research Institute, Singapore.

<sup>4</sup>Outram Polyclinic, SingHealth Polyclinics, Singapore.

<sup>5</sup>Ophthalmology, National University of Singapore, Singapore.

<sup>6</sup>Tomey Corporation, Nagoya, Japan.

<sup>7</sup>Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles, California, USA.

<sup>8</sup>Keck School of Medicine of the University of Southern California, Los Angeles, California, USA.

### Abstract

**Purpose:** To assess the generalizability and performance of a deep learning classifier for automated detection of gonioscopic angle closure in anterior segment OCT (AS-OCT) images.

**Methods:** A convolutional neural network (CNN) model developed using data from the Chinese American Eye Study (CHES) was used to detect gonioscopic angle closure in AS-OCT images with reference gonioscopy grades provided by trained ophthalmologists. Independent test data was derived from the population-based CHES, a community-based clinic in Singapore, and a hospital-based clinic at the University of Southern California (USC). Classifier performance was evaluated with receiver operating characteristic curve (ROC) and area under the receiver operating characteristic curve (AUC) metrics. Inter-examiner agreement between the classifier and two human examiners at USC were calculated using Cohen's kappa coefficients.

---

**Corresponding Author:** Benjamin Xu, Department of Ophthalmology, Keck School of Medicine at the University of Southern California, 1450 San Pablo Street, 4th Floor, Suite 4700, Los Angeles, CA 90033 Phone number: 323-442-6780; Fax number: 323-442-6412, benjamin.xu@med.usc.edu.

Contributorship Statement

BYX and TA conceived of the presented study. AAP, JD, JHQ, MT, TA, RV, and BYX collected the data. JR, MC, NP, GAA, RH, and BYX analyzed the data. All authors discussed the results and contributed to the final manuscript.

Disclosures and Competing Interests Statement

The authors have no relevant financial disclosures or competing interests.

**Results:** The classifier was tested using 640 images (311 open, 329 closed) from 127 Chinese Americans, 10,165 images (9595 open, 570 closed) from 1,318 predominantly Chinese Singaporeans, and 300 images (234 open, 66 closed) from 40 multi-ethnic USC patients. The classifier achieved similar performance in the CHES (AUC=0.917), Singapore (AUC=0.894), and USC (AUC=0.922) cohorts. Standardizing the distribution of gonioscopy grades across cohorts produced similar AUC metrics (range 0.890–0.932). The agreement between the CNN classifier and two human examiners ( $\kappa=0.700$  and  $0.704$ ) approximated inter-examiner agreement ( $\kappa=0.693$ ) in the USC cohort.

**Conclusion:** An OCT-based deep learning classifier demonstrated consistent performance detecting gonioscopic angle closure across 3 independent patient populations. This automated method could aid ophthalmologists in the assessment of angle status in diverse patient populations.

## Introduction

Primary angle closure glaucoma (PACG), the most severe form of primary angle closure disease (PACD), is a leading cause of irreversible vision loss and blindness worldwide.<sup>1</sup> The number of people affected by PACG is projected to increase to 32 million by 2040 due to aging of the world's population.<sup>1</sup> PACG often develops asymptotically and causes significant vision loss in at least one eye before it is detected and treated by an eyecare provider.<sup>2</sup> Therefore, early detection of angle closure, the primary risk factor for PACG, is crucial so that treatments, such as laser peripheral iridotomy (LPI) and lens extraction, can be administered to prevent progression of PACD and development of glaucoma-associated vision loss.<sup>3, 4</sup>

Gonioscopy is currently the clinical standard for evaluating the anterior chamber angle and detecting angle closure. However, gonioscopy has several limitations that reduce its effectiveness for angle closure detection. Gonioscopy is subjective, expertise-dependent, and must be performed in-person by a trained examiner. It can also be time-intensive and uncomfortable for some patients, which discourages some eyecare providers.<sup>5,6</sup> Anterior segment optical coherence tomography (AS-OCT) is a non-contact in vivo imaging method that provides an alternative to gonioscopy for assessing the angle. AS-OCT imaging is fast, easily tolerated by most patients, and can be performed by a technician with limited training and experience. In addition, AS-OCT angle assessments have high degrees of intra- and inter-user reproducibility, and AS-OCT images can be saved for off-line interpretation.<sup>7-9</sup>

Manual interpretation of AS-OCT images to detect angle closure can be time- and expertise dependent, which limits the utility of AS-OCT imaging in clinical practice. Our research group recently developed and reported a deep learning classifier that detects gonioscopic angle closure and PACD based on automated analysis of AS-OCT images.<sup>10</sup> Deep learning algorithms hold the potential to increase accessibility to care and decrease the burden of disease detection in all fields of medicine.<sup>11-16</sup> However, these algorithms may only function in a narrow range of patients that closely resemble the training cohort.<sup>17</sup> While our deep learning classifier achieved favorable performance compared to previous manual and semi-automated methods, it was tested using data from the same population of Chinese Americans over the age of 50 who comprised the training cohort.<sup>18, 19</sup> In addition, reference

gonioscopy grades were provided by a single trained ophthalmologist. In this study, we use AS-OCT images and reference gonioscopy grades provided by multiple trained ophthalmologists at independent clinical sites to assess the generalizability and performance of our deep learning classifier for community- and hospital-based detection of gonioscopic angle closure.

## Materials and Methods

Approval for this study was granted by the University of Southern California Medical Center Institutional Review Board (IRB). All study procedures adhered to the recommendations of the Declaration of Helsinki. All study participants provided informed consent. Data for the Singapore cohort was obtained with permission from the Singapore Eye Research Institute (SERI), following approval by the ethics review board of Singhealth.<sup>20</sup>

The original dataset used for training and testing of the deep learning classifier was obtained retrospectively from the Chinese American Eye Study (CHES), a population-based epidemiological study of 4,572 Chinese participants 50 years and older living in Monterey Park, California, USA.<sup>21</sup> The second dataset used for classifier testing was obtained retrospectively as part of a community-based study in Singapore. A total of 2,052 participants 50 years and older were recruited from a community polyclinic providing primary care services in Singapore.<sup>20</sup> The final dataset used for classifier testing was obtained prospectively from a hospital-based glaucoma tertiary care referral center at the University of Southern California (USC) in Los Angeles, CA. A total of 40 participants 21 years and older presenting for routine glaucoma evaluations were recruited from USC eyecare facilities.

Inclusion criteria included receipt of gonioscopy and AS-OCT imaging. Exclusion criteria included prior eye surgery, penetrating eye injury, or media opacities that precluded visualization of anterior chamber angle structures. Subjects with a history of prior laser peripheral iridotomy (LPI) were not excluded.

## Clinical Assessment

The clinical assessment was similar across the three study cohorts. Each participant received a complete eye examination by a trained ophthalmologist including gonioscopy and AS-OCT imaging. Gonioscopy was performed in the seated position under dark ambient lighting (approximately 0.1 cd/m<sup>2</sup>) with a 1-mm light beam and a 4-mirror goniolens by at least one trained glaucoma specialist or ophthalmologist (D.W., an ophthalmologist in CHES; M.T., a glaucoma specialist in Singapore; B.Y.X., a glaucoma specialist at USC; J.D., a clinical glaucoma fellow at USC) masked to other examination findings. Gonioscopy by the glaucoma specialist and fellow at the USC site were performed in a semi-random order depending on their availability in clinic. Care was taken to avoid light falling on the pupil and inadvertent indentation of the globe. The gonioscopy lens could be tilted up to 10 degrees. The angle was graded in each quadrant according to the modified Shaffer classification system: grade 0, no structures visible; grade 1, non-pigmented TM visible; grade 2; pigmented TM visible; grade 3, scleral spur visible; grade 4, ciliary body visible. Gonioscopy grades were grouped into two categories: gonioscopic angle closure (grade 0 or

1), in which pigmented TM could not be visualized, and gonioscopic open angle (grades 2 to 4).

AS-OCT imaging was performed in the seated position under dark ambient lighting (approximately  $0.1 \text{ cd/m}^2$ ) after gonioscopy and prior to pupillary dilation by a trained ophthalmologist or technician with the Tomey CASIA SS-1000 swept-source Fourier-domain device (Tomey Corporation, Nagoya, Japan). 128 two-dimensional cross-sectional AS-OCT images were acquired per eye. During imaging, the eyelids were gently retracted taking care to avoid inadvertent pressure on the globe.

Raw AS-OCT image data was imported into the SS-OCT Viewer software (version 3.0, Tomey Corporation, Nagoya, Japan). Two images were exported in JPEG format per eye: one oriented along the horizontal (temporal-nasal) meridian and the other along the vertical (superior-inferior) meridian. Images were divided in two along the vertical midline, and right-sided images were flipped about the vertical axis to standardize images with the anterior chamber angle to the left and corneal apex to the right. No adjustments were made to image brightness or contrast. Eyes with corrupt images or images with significant lid artifacts precluding visualization of the anterior chamber angle were excluded from the analysis. Image manipulations were performed in MATLAB (Mathworks, Natick, MA).

### Deep Learning Classifier Development and Testing

A previously reported multi-class convolutional neural network (CNN) model based on the ResNet-18 architecture was used to classify the anterior chamber angle in individual AS-OCT images as either open or closed (Figure 1).<sup>10</sup> One AS-OCT scan was completed per quadrant, yielding one image per quadrant for assessment by the classifier. In brief, the classifier was trained using a dataset of 3,396 AS-OCT images with nearly equal distribution of open ( $N = 1,632$ ) and closed ( $N = 1,764$ ) angles. However, the distribution of gonioscopy grades was not balanced. Given an input image, the classifier produced a normalized probability distribution over Shaffer grades  $p = [p_0, p_1, p_2, p_3, p_4]$ . Binary probabilities for closed angle (grades 0 and 1) and open angle (grades 2 to 4) eyes were generated by summing probabilities over the corresponding grades, i.e.,  $p_{closed} = p_0 + p_1$  and  $p_{open} = p_2 + p_3 + p_4$ . Probabilities were thresholded in order to generate binary classifications where a positive detection event was defined as classification to grade 0 or 1. The threshold value is a tunable parameter that controls the tradeoff between false positive rate (FPR) and true positive rate (TPR).

Classifier performance was evaluated on unstandardized test data (i.e., test data where the distribution of grades has not been resampled) from each of the three cohorts (CHES, Singapore, and USC) using mean area under the receiver operating characteristic curve (AUC) metrics and receiver operating characteristic (ROC) curves. The classification of open angle (Schaffer grades 2, 3, and 4) and closed angle (Schaffer grades 0 and 1) was compared by plotting the ROC curve of the CNN classifier with the false positive rates (FPRs) and true positive rates (TPRs). ROC curves corresponding to the lower and upper bounds of the 95% confidence interval in AUC values were generated by bootstrapping to evaluate variability in CNN performance. Reference gonioscopy grades and labels of open or closed angle were provided by the human examiner(s) who performed manual gonioscopy

at each study site, which differed for each test dataset. Gonioscopy grades provided by the glaucoma specialist were used as the reference standard in the USC test dataset.

Classifier performance was also evaluated on test data standardized so that its grade distribution matched that of the USC test dataset. Standardization was done by randomly selecting a given number of grade 0,1,2,3, and 4 images without replacement from the test data so its final distribution matched that of the USC clinical dataset. This process eliminates inter-cohort differences in angle grade distributions that might bias classifier performance. Test datasets were standardized to the USC test dataset since it contained the fewest images.

## Statistical Analyses

Continuous data were expressed as mean, standard deviation, and range values. Categorical data were expressed in percentages. Two-way tables and Cohen's kappa coefficients were calculated to assess the pairwise agreement between the CNN classifier and two human examiners at the USC site in the binary classification of open or closed angle. All analyses were performed using the R programming interface (version 4.0.3). Statistical analyses were conducted using a significance level of 0.05.

## Results

The CHES test dataset contained 640 AS-OCT images with corresponding grades from 127 subjects (Table 1, Figure 2). All images with eyelid artifacts had previously been removed from the dataset.<sup>21</sup> There were 311 images with open angles (48.6%) and 329 with closed angles (51.4%) (Table 2): grade 0 (N = 149), grade 1 (N = 180), grade 2 (N = 65), grade 3 (N = 175), grade 4 (N = 71). The mean age of participants was 62 years old, with 34% (N = 43) males and 66% (N = 84) females. All participants (N = 127) self-identified as Chinese.

The initial Singapore dataset contained 10,366 AS-OCT images with corresponding gonioscopy grades (Table 1, Figure 2). 201 images (1.9%) were excluded from analysis due to imaging and eyelid artifacts. The Singapore test dataset contained 10,165 AS-OCT images with corresponding grades from 1,318 subjects. There were 9,595 images with open angles (94.4%) and 570 with closed angles (5.6%) (Table 2): grade 0 (N = 196), grade 1 (N = 374), grade 2 (N = 1,174), grade 3 (N = 3,718), grade 4 (N = 4,703). The mean age of participants was 61 years old, with 37% (N = 484) males and 63% (N = 834) females. 1,115 participants identified as Chinese (87.6%), 93 as Indian (7.1%), 35 as Malay (2.7%), and 35 as other (2.7%).

The USC test dataset contained 300 AS-OCT images with corresponding grades from 40 subjects (Table 1, Figure 2). There were no eyelid artifacts among the images. There were 234 images with open angles (78.0%) and 66 with closed angles (22.0%) (Table 2): grade 0 (N = 30), grade 1 (N = 36), grade 2 (N = 32), grade 3 (N = 80), grade 4 (N = 122). The mean age of participants was 56 years old, with 45% (N = 18) males and 55% (N = 22) females. 13 patients identified as Hispanic (32.5%), 13 as Asian (32.5%), 9 as non-Hispanic White (22.5%), and 5 as Black (12.5%).

## Classifier Performance

The CNN classifier achieved an AUC of 0.917 (95% CI = 0.891–0.937) in detecting gonioscopic angle closure by quadrant in the CHES test dataset (Figure 3). The classifier achieved an AUC of 0.894 (95% CI = 0.882–0.905) and 0.922 (95% CI = 0.883–0.952) in the unstandardized Singapore and USC test datasets, respectively.

The CNN classifier achieved an AUC of 0.932 (95% CI = 0.900–0.960) in detecting gonioscopic angle closure by quadrant in the CHES test dataset and 0.890 (95% CI = 0.848–0.930) in the Singapore test dataset after distributions of gonioscopy grades were standardized to the distribution of the USC test dataset (Figure 4).

## Inter-Examiner and CNN-Human Agreement

Inter-examiner agreement in detecting angle closure in the USC cohort was assessed by pairwise comparisons between the two human examiners and the CNN classifier (Figure 5). The CNN classifier demonstrated similar levels of agreement with the glaucoma specialist ( $\kappa = 0.700$ , 95% CI = 0.593–0.792) and glaucoma fellow ( $\kappa = 0.704$ , 95% CI = 0.595–0.799) as between the two human examiners ( $\kappa = 0.693$ , 95% CI = 0.580–0.791) in detecting angle closure by quadrant. The CNN classifier produced a greater number of false positives (detecting angle closure when the reference examiner detected open angle) and a lower number of false negatives (detecting open angle when the reference examiner detected angle closure) when the glaucoma specialist was the reference examiner compared to when the glaucoma fellow was the reference examiner. The glaucoma fellow produced a more balanced number of false positives and false negatives when the glaucoma specialist was the reference examiner, with a sensitivity and specificity of 0.807 and 0.906, respectively, which fell within the 95% CI of the ROC curve by the CNN classifier (Figures 3 and 4).

## Discussion

In this study, we assessed the generalizability and performance of a deep learning classifier for detecting gonioscopic angle closure in AS-OCT images. We compared classifier performance in a Chinese American population-based cohort to performance in a predominantly Chinese Singaporean community-based cohort and a multi-ethnic hospital-based cohort. The classifier demonstrated excellent performance in all three test datasets despite differences in patient demographics, angle closure prevalence, ophthalmologists who provided reference gonioscopy grades, and practice settings. Classifier performance also approximated human inter-examiner agreement based on manual gonioscopy performed by a glaucoma specialist and fellow in the hospital-based cohort. Our findings demonstrate that an OCT-based deep learning classifier performing “automated gonioscopy” can be used to detect gonioscopic angle closure in a range of patient populations and care settings. This has important implications for increasing accessibility to care and reducing the burden on eyecare providers to perform gonioscopy in the general population.

Deep learning is an effective method for automating the analysis of medical images to detect a wide range of diseases.<sup>13, 16, 17</sup> However, the performance of deep learning algorithms can become degraded when the algorithms are applied to image datasets from populations that

differ from the training cohort.<sup>22</sup> For example, a CNN developed to detect glaucomatous optic neuropathy in fundus photographs initially demonstrated excellent performance (AUC = 0.996, 95% CI = 0.995–0.998) in local validation datasets that significantly decreased when the model was applied to a multi-ethnic dataset (AUC = 0.923, 95% CI = 0.916–0.930) and a dataset created with images of variable quality (AUC = 0.823, 95% CI = 0.787–0.855).<sup>23</sup> This and similar studies suggest that generalizability of deep learning algorithms cannot be assumed, and caution should be exercised when applying algorithms in different patient populations and care settings.<sup>17, 24</sup> Therefore, it is noteworthy our classifier demonstrates excellent performance that is consistent across three patient populations (AUC range from 0.894 to 0.922).

Racial homogeneity was a primary limitation of the original CHES data used to train and test the deep learning classifier. Biometric properties of the anterior segment vary by race, with shallower anterior chambers and thicker irises being more common in Asian compared to non-Asian eyes.<sup>25–27</sup> These differences likely contribute to differences in PACG prevalence observed between different racial populations.<sup>28</sup> The prevalence of PACG is highest in Asian populations<sup>28</sup>, approximating the prevalence of POAG.<sup>29</sup> Conversely, POAG is far more prevalent than PACG among Blacks, Hispanics, and non-Hispanic Whites.<sup>30–32</sup> However, these differences do not appear to affect classifier performance, as AUC metrics were similar in the racially homogenous CHES cohort and racially diverse USC clinical cohort; the original CHES cohort consisted entirely of Chinese Americans, whereas the majority of the USC cohort were non-Asian. Biometric properties of the anterior segment are more consistent within ethnically Chinese populations.<sup>27</sup> Therefore, race is less likely to contribute to a difference classifier performance between the CHES and Singapore cohorts, which were composed entirely of Chinese Americans and primarily of Chinese Singaporeans, respectively.

The prevalence of PACD varies widely by geographic location, and classifier performance could be affected by the distribution of gonioscopy grades and open/closed angles within a population.<sup>33</sup> While our deep learning classifier was originally developed using population-based data, the CHES training and test datasets were balanced based on the number of open and closed angles to avoid introducing biases in the detection of angle closure. However, individual angle grades were not balanced within these classes. While AUC metrics are not affected by different distributions of dichotomous states (open or closed angle), they are sensitive to different distributions of sub-classes that make up those states (angle grades). Therefore, we standardized to the gonioscopy grade distribution of the CHES and Singapore cohorts to that of the USC cohort, which was the smallest in size, to compare classifier performance directly across populations. The effect of standardization on AUC metrics was minimal. Performance on CHES data improved slightly, with an increase in AUC from 0.917 to 0.932, whereas performance on Singapore data was largely unaffected, with an AUC of 0.894 when unstandardized and 0.890 when standardized. These findings demonstrate that classifier performance is robust to not only differences in the distribution of open and closed angles within a population, but individual gonioscopy grades as well.

Gonioscopy is a subjective and expertise-dependent angle assessment method; therefore, the generalizability of the classifier could be limited by the quality of the reference gonioscopy

grades used to train the classifier. The deep learning classifier was trained using gonioscopy grades provided primarily by a single ophthalmologist, which could limit its performance when gonioscopy grades by other ophthalmologists are used as the reference standard. There are few studies on the inter-examiner reproducibility of angle closure detection based on manual gonioscopy, in part due to the time-consuming nature of having multiple eyecare providers perform gonioscopy on the same patient. We used the USC cohort to assess the inter-examiner agreement between a glaucoma specialist and fellow in detecting angle closure on manual gonioscopy. The agreement between the classifier and glaucoma specialist ( $\kappa = 0.700$ ) and glaucoma specialist ( $\kappa = 0.704$ ) closely approximated the agreement between the two human examiners ( $\kappa = 0.693$ ). The level of agreement is also similar to metrics of inter-examiner agreement in detecting gonioscopic angle closure by two trained glaucoma specialists during two separate clinic visits ( $\kappa = 0.66$  and  $0.69$ ).<sup>7</sup> These findings support the high quality of the reference gonioscopy grades used to train the classifier and the competency of the trained ophthalmologist who performed manual gonioscopy in CHES.

Data collection protocols and severity of angle closure are additional factors that did not appear to affect classifier performance across the three cohorts. The Singapore cohort was screened and examined at a busy, high-volume primary care community clinic. The CHES cohort was recruited through door-to-door solicitation and ocular examinations, including gonioscopy and AS-OCT imaging, were performed during a follow-up visit at a dedicated examination site. Finally, the USC cohort was recruited during routine glaucoma evaluations at a tertiary care referral center, and AS-OCT imaging performed as part of the standard clinical workflow. The three study cohorts also comprise a wider range of anatomical configurations and ocular biometrics than any single cohort, with previous studies finding milder PACD in the community and more severe PACD in the hospital.<sup>34–36</sup>

Our study has several limitations. First, the USC dataset is much smaller than both the CHES and Singapore datasets, which became a limiting factor when standardizing the datasets based on gonioscopy grades. All participants in the USC cohort received gonioscopy performed by the same two examiners. Consequently, recruitment ended early due to the onset of the COVID-19 pandemic and graduation of the second examiner (glaucoma fellow). However, the sample size of this cohort appears to be sufficient given the narrow confidence intervals associated with the ROC curve and AUC metric. Second, the study is limited to images taken on the Tomey CASIA SS-1000 AS-OCT device. While biometric measurements from the CASIA SS-1000 and newer swept-source AS-OCT devices are comparable, there are differences in image quality that would likely affect classifier performance across devices.<sup>37</sup> Therefore, additional generalizability studies and/or retraining of the classifier is required before the described method can be used on images acquired using other AS-OCT devices. Third, classifier predictions were based on a single AS-OCT scan per quadrant. This approach may miss localized regions of angle closure. Therefore, future studies utilizing the classifier may improve its sensitivity to gonioscopic angle closure by analyzing multiple images per quadrant. Fourth, a standalone AS-OCT device is expensive compared to a standard goniolens. However, the cost of OCT devices continues to decrease as technology evolves, and biometry functions in newer AS-OCT devices could increase the scope of their clinical utility and adoption. Fifth, the second



examiner in the USC cohort was a glaucoma fellow who was trained in and regularly performed gonioscopy. Therefore, the inter-examiner agreement, which was comparable to the human-CNN agreement, may not generalize to ophthalmologists with less experience performing gonioscopy. Finally, the classifier was trained using a dataset with a specific distribution of angle grades and open/closed angles. Therefore, our findings cannot be generalized to classifiers developing using other distributions of training data.

In conclusion, our deep learning classifier for detecting gonioscopic angle closure in AS-OCT images demonstrated consistent and robust performance across a range of patients and patient care settings, approximating that of a trained ophthalmologist performing manual gonioscopy. This has important implications for the clinical utility of AS-OCT imaging as a method for performing “automated gonioscopy” to detect patients who would benefit from more thorough clinical assessments by trained eyecare providers. Gonioscopy appears to be underperformed by eyecare providers, despite recommendations by the American Academy of Ophthalmology (AAO),<sup>38</sup> perhaps because it is time-consuming and should be performed prior to pupillary dilation. This deficiency is especially problematic given that the treatment of open angle and angle closure forms of glaucoma differ substantially.<sup>39</sup> While the recent landmark Zhongshan Angle Closure Prevention (ZAP) Trial suggested that the majority of patients with early angle closure may not benefit from treatment with laser peripheral iridotomy (LPI), angle closure can nevertheless lead to significant ocular morbidity if left undetected.<sup>40</sup> Therefore, automated methods to detect angle closure may yet play a vital role in improving efficiency of eyecare and decreasing the burden of angle closure detection on healthcare providers and systems worldwide.

## Acknowledgements

This work was supported by grants U10 EY017337 and K23 EY029763 from the National Eye Institute, National Institute of Health, Bethesda, Maryland; a Young Clinician Scientist Research Award from the American Glaucoma Society, San Francisco, CA; a Grant-in-Aid Research Award from Fight for Sight, New York, NY; a SC-CTSI Clinical and Community Research Award from the Southern California Clinical and Translational Science Institute, Los Angeles, CA; and an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, New York, NY.

## References

1. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. Nov 2014;121(11):2081–90. [PubMed: 24974815]
2. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. May 2014;311(18):1901–11. [PubMed: 24825645]
3. Azuara-Blanco A, Burr J, Ramsay C, et al. Effectiveness of early lens extraction for the treatment of primary angle-closure glaucoma (EAGLE): a randomised controlled trial. *Lancet*. Oct 2016;388(10052):1389–1397. [PubMed: 27707497]
4. Radhakrishnan S, Chen PP, Junk AK, Nouri-Mahdavi K, Chen TC. Laser Peripheral Iridotomy in Primary Angle Closure: A Report by the American Academy of Ophthalmology. *Ophthalmology*. 07 2018;125(7):1110–1120. [PubMed: 29482864]
5. Riva I, Micheletti E, Oddone F, et al. Anterior Chamber Angle Assessment Techniques: A Review. *J Clin Med*. Nov 2020;9(12)
6. Hertzog LH, Albrecht KG, LaBree L, Lee PP. Glaucoma care and conformance with preferred practice patterns. Examination of the private, community-based ophthalmologist. *Ophthalmology*. Jul 1996;103(7):1009–13. [PubMed: 8684788]

7. Rigi M, Bell NP, Lee DA, et al. Agreement between Gonioscopic Examination and Swept Source Fourier Domain Anterior Segment Optical Coherence Tomography Imaging. *J Ophthalmol.* 2016;2016:1727039. [PubMed: 27990300]
8. Maram J, Pan X, Sadda S, Francis B, Marion K, Chopra V. Reproducibility of angle metrics using the time-domain anterior segment optical coherence tomography: intra-observer and inter-observer variability. *Curr Eye Res.* May 2015;40(5):496–500. [PubMed: 24955626]
9. Cumba RJ, Radhakrishnan S, Bell NP, et al. Reproducibility of scleral spur identification and angle measurements using fourier domain anterior segment optical coherence tomography. *J Ophthalmol.* 2012;2012:487309. [PubMed: 23209880]
10. Xu BY, Chiang M, Chaudhary S, Kulkarni S, Pardeshi AA, Varma R. Deep Learning Classifiers for Automated Detection of Gonioscopic Angle Closure Based on Anterior Segment OCT Images. *Am J Ophthalmol.* 12 2019;208:273–280. [PubMed: 31445003]
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* May 2015;521(7553):436–44. [PubMed: 26017442]
12. Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA.* 12 2017;318(22):2211–2223. [PubMed: 29234807]
13. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 12 2016;316(22):2402–2410. [PubMed: 27898976]
14. Wang XN, Dai L, Li ST, Kong HY, Sheng B, Wu Q. Automatic Grading System for Diabetic Retinopathy Diagnosis Using Deep Learning Artificial Intelligence Software. *Curr Eye Res.* Dec 2020;45(12):1550–1555. [PubMed: 32410471]
15. Brinker TJ, Hekler A, Enk AH, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer.* 09 2019;119:11–17. [PubMed: 31401469]
16. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 02 2017;542(7639):115–118. [PubMed: 28117445]
17. AlBadawy EA, Saha A, Mazurowski MA. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Med Phys.* Mar 2018;45(3):1150–1158. [PubMed: 29356028]
18. Nongpiur ME, Haaland BA, Friedman DS, et al. Classification algorithms based on anterior segment optical coherence tomography measurements for detection of angle closure. *Ophthalmology.* Jan 2013;120(1):48–54. [PubMed: 23009888]
19. Nongpiur ME, Haaland BA, Perera SA, et al. Development of a score and probability estimate for detecting angle closure based on anterior segment optical coherence tomography. *Am J Ophthalmol.* Jan 2014;157(1):32–38.e1. [PubMed: 24210768]
20. Porporato N, Baskaran M, Tun TA, et al. Assessment of Circumferential Angle Closure with Swept-Source Optical Coherence Tomography: a Community Based Study. *Am J Ophthalmol.* 03 2019;199:133–139. [PubMed: 30502338]
21. Varma R, Torres M, McKean-Cowdin R, et al. Prevalence and Risk Factors for Refractive Error in Adult Chinese Americans: The Chinese American Eye Study. *Am J Ophthalmol.* Mar 2017;175:201–212. [PubMed: 27769895]
22. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health.* 09 2020;2(9):e489–e492. [PubMed: 32864600]
23. Liu H, Li L, Wormstone IM, et al. Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs. *JAMA Ophthalmol.* Sep 2019;
24. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* 11 2018;15(11):e1002683. [PubMed: 30399157]
25. Spaeth G Angle-closure glaucoma in East Asian and European people. Different diseases? *Eye (Lond).* Jan 2007;21(1):99–100; author reply 100. [PubMed: 16680097]

26. Congdon NG, Foster PJ, Wamsley S, et al. Biometric gonioscopy and the effects of age, race, and sex on the anterior chamber angle. *Br J Ophthalmol*. Jan 2002;86(1):18–22. [PubMed: 11801496]
27. Wang D, Qi M, He M, Wu L, Lin S. Ethnic difference of the anterior chamber area and volume and its association with angle width. *Invest Ophthalmol Vis Sci*. May 2012;53(6):3139–44. [PubMed: 22531702]
28. Cook C, Foster P. Epidemiology of glaucoma: what's new? *Can J Ophthalmol*. Jun 2012;47(3):223–6. [PubMed: 22687296]
29. Song P, Wang J, Bucan K, Theodoratou E, Rudan I, Chan KY. National and subnational prevalence and burden of glaucoma in China: A systematic analysis. *J Glob Health*. Dec 2017;7(2):020705. [PubMed: 29302324]
30. Varma R, Ying-Lai M, Francis BA, et al. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: the Los Angeles Latino Eye Study. *Ophthalmology*. Aug 2004;111(8):1439–48. [PubMed: 15288969]
31. Tielsch JM, Sommer A, Katz J, Royall RM, Quigley HA, Javitt J. Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey. *JAMA*. Jul 1991;266(3):369–74. [PubMed: 2056646]
32. Friedman DS, Jampel HD, Muñoz B, West SK. The prevalence of open-angle glaucoma among blacks and whites 73 years and older: the Salisbury Eye Evaluation Glaucoma Study. *Arch Ophthalmol*. Nov 2006;124(11):1625–30. [PubMed: 17102012]
33. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw*. Oct 2018;106:249–259. [PubMed: 30092410]
34. 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 Nov-Dec 2018;1(3):158–166. [PubMed: 31025032]
35. Guzman CP, Gong T, Nongpiur ME, et al. Anterior segment optical coherence tomography parameters in subtypes of primary angle closure. *Invest Ophthalmol Vis Sci*. Aug 2013;54(8):5281–6. [PubMed: 23788370]
36. Moghimi S, Vahedian Z, Fakhraie G, et al. Ocular biometry in the subtypes of angle closure: an anterior segment optical coherence tomography study. *Am J Ophthalmol*. Apr 2013;155(4):664–673, 673.e1. [PubMed: 23246271]
37. Pardeshi AA, Song AE, Lazkani N, Xie X, Huang A, Xu BY. Intradvice Repeatability and Interdevice Agreement of Ocular Biometric Measurements: A Comparison of Two Swept-Source Anterior Segment OCT Devices. *Transl Vis Sci Technol*. 08 2020;9(9):14.
38. Coleman AL, Yu F, Evans SJ. Use of gonioscopy in medicare beneficiaries before glaucoma surgery. *J Glaucoma*. Dec 2006;15(6):486–93. [PubMed: 17106360]
39. Lee DA, Higginbotham EJ. Glaucoma and its treatment: a review. *Am J Health Syst Pharm*. Apr 2005;62(7):691–9. [PubMed: 15790795]
40. He M, Friedman DS, Ge J, et al. Laser peripheral iridotomy in primary angle-closure suspects: biometric and gonioscopic outcomes: the Liwan Eye Study. *Ophthalmology*. Mar 2007;114(3):494–500. [PubMed: 17123610]

### Synopsis

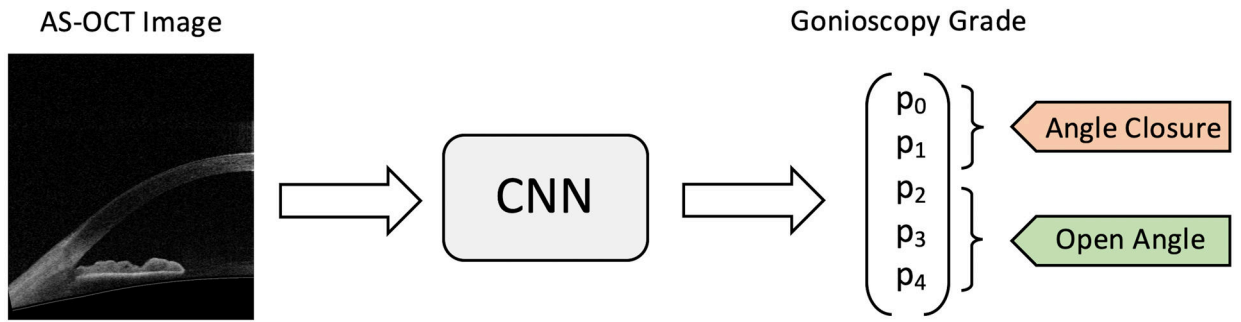
An OCT-based deep learning classifier demonstrates consistent and robust performance, approximating inter-examiner agreement between trained ophthalmologists, in detecting gonioscopic angle closure across a range of patient populations and care settings.

Author Manuscript

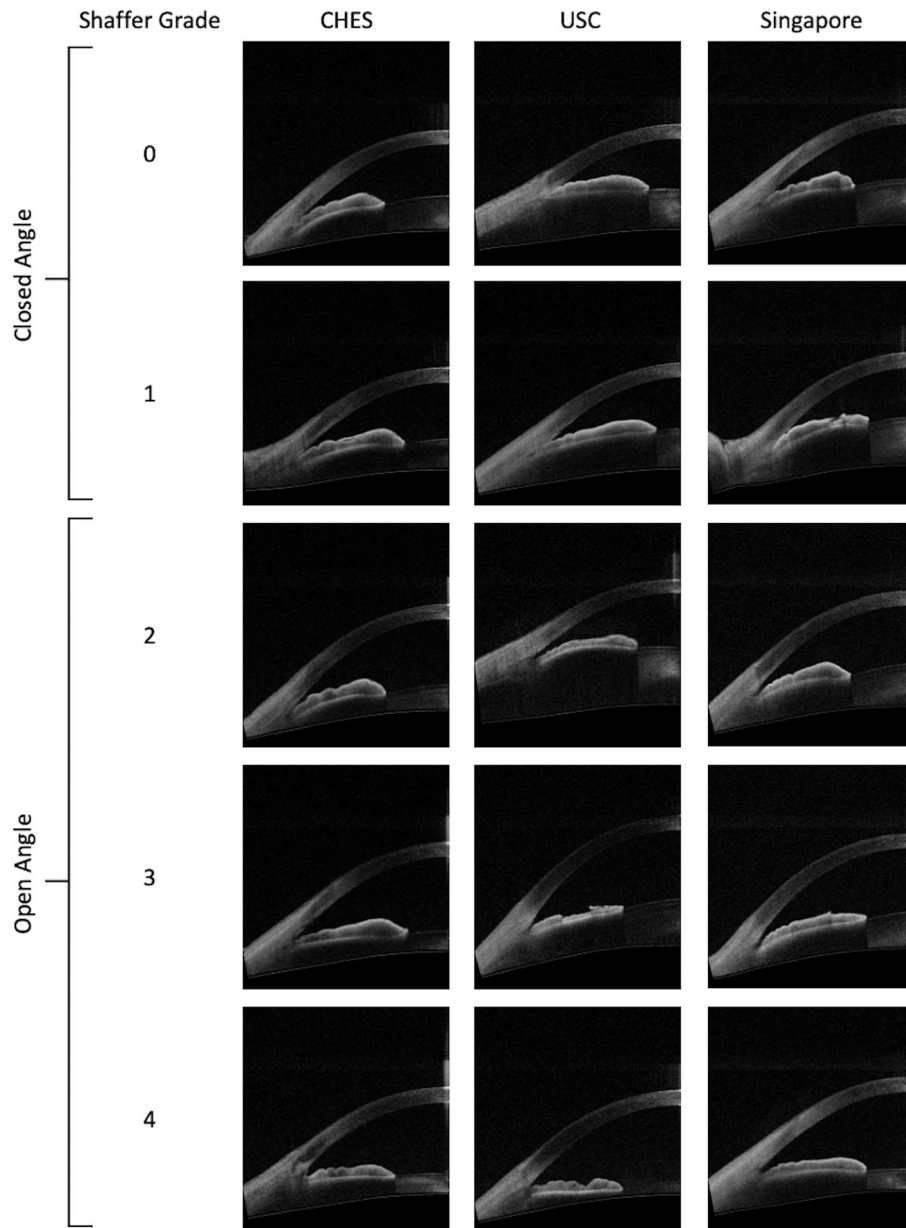
Author Manuscript

Author Manuscript

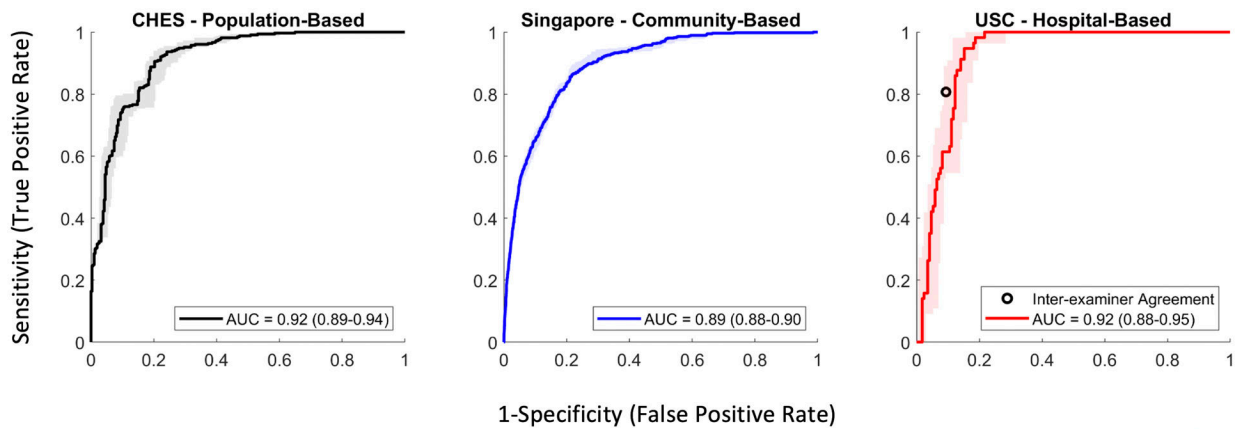
Author Manuscript



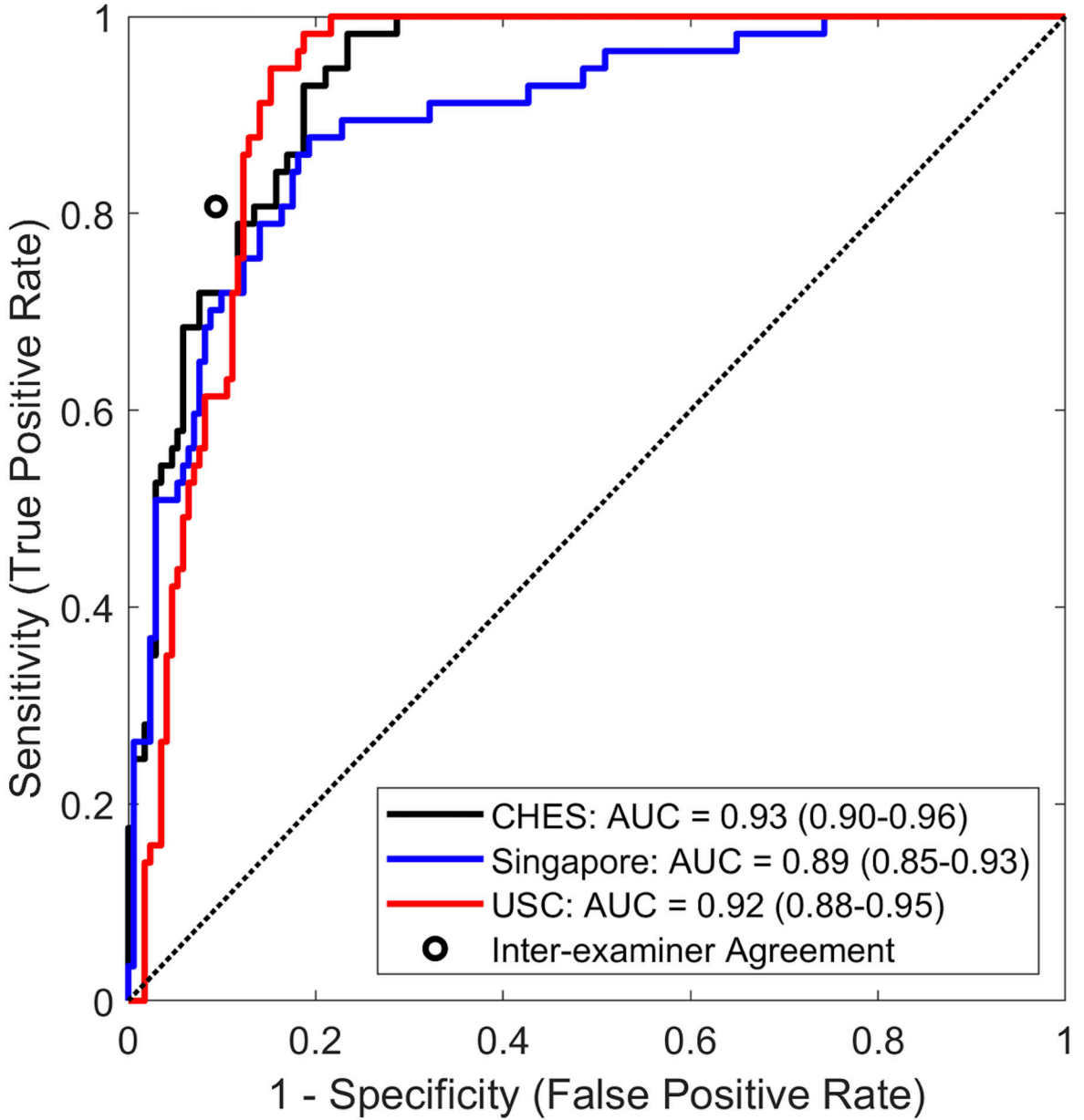
**Figure 1.** Schematic of binary classification process. Unmarked AS-OCT images were used as inputs to the CNN classifier. Gonioscopy grade probabilities (P0 to P4) were summed to make the binary prediction of angle status: angle closure = grades 0 and 1, open angle = grades 2, 3, and 4.



**Figure 2.** Representative AS-OCT images from CHES, USC, and Singapore cohorts corresponding to open and closed angles based on manual gonioscopy.



**Figure 3.** Receiver operating characteristic (ROC) curves with 95% confidence intervals (colored bars) and area under the curve (AUC) metrics for detecting angle closure in three independent test datasets with unstandardized distributions of gonioscopy grades. A second ophthalmologist performed manual gonioscopy in the USC cohort to compare with classifier performance (black point).



**Figure 4.** Receiver operating characteristic (ROC) curves and area under the curve (AUC) metrics for detecting angle closure in three independent test datasets with standardized distributions of gonioscopy grades. A second ophthalmologist performed manual gonioscopy in the USC cohort to compare with classifier performance (black point).



**CNN and Specialist**

		Reference	
		Open	Closed
Predicted	Open	144	5
	Closed	27	52

$K = 0.700$   
(0.593 – 0.792)

**CNN and Ophthalmologist**

		Reference	
		Open	Closed
Predicted	Open	143	23
	Closed	6	56

$K = 0.704$   
(0.595 – 0.799)

**Specialist and Ophthalmologist**

		Reference	
		Open	Closed
Predicted	Open	155	11
	Closed	16	46

$K = 0.693$   
(0.580 - 0.791)

**Figure 5.** Statistical measures of agreement between the CNN classifier and two human examiners in the USC cohort.

**Table 1.**

Patient demographics by cohort and dataset

	CHES Train (N=664) (%)	CHES Test (N=127) (%)	Singapore (N=1318) (%)	USC (N=40) (%)
<b>Gender</b>				
Male	209 (31.5)	43 (33.9)	484 (36.7)	18 (45.0)
Female	455 (68.5)	84 (66.1)	834 (63.3)	22 (55.0)
<b>Age</b>				
<50	0	0	0	13 (32.5)
50–60	374 (56.3)	61 (48.0)	615 (46.7)	8 (20.0)
61–70	201 (30.3)	44 (34.6)	556 (42.2)	8 (20.0)
70+	89 (13.4)	22 (17.3)	147 (11.2)	11 (27.5)
<b>Race</b>				
Chinese	664 (100)	127 (100)	1155 (87.6)	13 (32.5) *
Indian	0	0	93 (7.1)	0
Malay	0	0	35 (2.7)	0
Non-Hispanic White	0	0	0	9 (22.5)
Black	0	0	0	5 (12.5)
Hispanic	0	0	0	13 (32.5)
Other	0	0	35 (2.7)	0

\* USC patients self-identified as Asian.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2.**

Distributions of angle status and grades by cohort and dataset

	CHES Train (N=3396) (%)	CHES Test (N=640) (%)	Singapore (N=10165) (%)	USC (N=300) (%)
<b>Angle Status</b>				
<b>Closed</b>	1764 (51.9)	329 (51.4)	570 (5.6)	66 (22.0)
<b>Open</b>	1632 (48.1)	311 (48.6)	9595 (94.4)	234 (78.0)
<b>Gonioscopy Grade</b>				
<b>0</b>	808 (23.8)	149 (23.3)	196 (1.9)	30 (10.0)
<b>1</b>	956 (28.2)	180 (28.1)	374 (3.7)	36 (12.0)
<b>2</b>	360 (10.6)	65 (10.2)	1174 (11.5)	32 (10.7)
<b>3</b>	898 (26.4)	175 (27.3)	3718 (36.6)	80 (26.7)
<b>4</b>	374 (11.0)	71 (11.1)	4703 (46.3)	122 (40.7)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript