



Published in final edited form as:

Ophthalmology. 2020 November ; 127(11): 1498–1506. doi:10.1016/j.ophtha.2020.04.033.

Using Deep Learning to Automate Goldmann Applanation Tonometry Readings

Ted Spaide, PhD¹, Yue Wu, PhD¹, Ryan T. Yanagihara, MD¹, Shu Feng, MD¹, Omar Ghabra, BA¹, Jonathan S. Yi, BA¹, Philip P. Chen, MD¹, Francy Moses¹, Aaron Y. Lee, MD, MSCI^{1,*}, Joanne C. Wen, MD^{2,*}

¹Department of Ophthalmology, University of Washington; 325 9th Ave, Box 359608; Seattle, WA 98104

²Duke Eye Center, Duke University; 2351 Erwin Road, Durham, NC 27707

Abstract

Purpose: To develop an objective and automated method for measuring intraocular pressure using deep learning and fixed-force Goldmann applanation tonometry (GAT) techniques.

Design: Prospective cross-sectional study

Subjects: Patients from an academic glaucoma practice

Methods: Intraocular pressure (IOP) was estimated by analyzing videos recorded using a standard slit lamp microscope and fixed-force GAT. Video frames were labeled to identify the outline of the reference tonometer and the applanation mires. A deep learning model was trained to localize and segment the tonometer and mires. IOP values were calculated from the deep learning predicted tonometer and mire diameters using the Imbert-Fick formula. A separate test set was collected prospectively where standard and automated GAT were collected in random order by two independent masked observers to assess the deep learning model as well as inter-observer variability.

Main Outcome Measures: IOP measurements between standard and automated methods were compared.

Results: 263 eyes from 135 subjects were included in the training and validation videos. For the test set, 50 eyes from 25 subjects were included. Each eye was measured by two observers, resulting in 100 videos. Within the test set, the mean difference between automated and standard GAT was -0.9 mmHg with 95% limits of agreement (LoA) of -5.4 to 3.6 mmHg. Mean difference between the two observers using standard GAT was 0.09 mmHg with LoA of -3.8 to 4.0 . Mean difference between the two observers using automated GAT videos was -0.3 mmHg with LoA of

Co-corresponding author and address for reprints: Joanne C. Wen, MD, Duke Eye Center, 2351 Erwin Road, Durham, NC 27705, joanne.wen@duke.edu, Aaron Y. Lee, MD, MSCI, University of Washington, 325 Ninth Ave, Box 359608, Seattle, WA 98104-2499, leeay@uw.edu.

*These authors contributed equally and should be considered co-senior authors

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

–4.1 to 3.5 mmHg. The coefficients of repeatability for automated and standard GAT were 3.8 and 3.9 mmHg, respectively. The bias for even numbered measurements was reduced when using automated GAT.

Conclusion: Preliminary measurements using deep learning to automate GAT demonstrate comparable results to standard GAT. Automated GAT has the potential to significantly improve upon our current GAT measurement standards by reducing bias and improving repeatability. In addition, ocular pulse amplitudes could be observed using this technique.

Abstract

An objective and automated method for measuring intraocular pressure using deep learning on videos of Goldmann applanation tonometry (GAT) was developed and demonstrated comparable intraocular pressure measurements to standard GAT.

Keywords

Tonometer; tonometry; smartphone; intraocular pressure; deep learning; Goldmann applanation tonometry

INTRODUCTION

Tonometry, the measurement of intraocular pressure (IOP), is an essential component of any ophthalmic exam. Tonometry is of particular importance in the diagnosis and management of glaucoma as elevated IOP is the greatest risk factor for the development of glaucoma, and glaucoma management relies on lowering IOP.

Fixed force tonometry was first introduced by Maklakov in 1885 and in the mid-20th century, commercially available fixed force tonometers were developed by Posner and Halberg.¹ With these tonometers, a known force was applied to the cornea and the resulting applanation surface diameter was measured to determine the IOP. These tonometers fell out of favor with the introduction of Goldmann Applanation Tonometry (GAT) in 1954.² GAT is currently considered to be the clinical gold standard and much of glaucoma research, including numerous large randomized multi-centered trials, have used GAT for their IOP measurements.^{3–5}

However, GAT has a number of limitations. It requires the subjective alignment of applanation mires, which can introduce systematic bias into the measurement depending on the user. For example, studies have reported a higher frequency of even numbered GAT IOP measurements compared to odd numbers, which corresponds to the even numbered tick marks on the GAT dial.^{6,7} Furthermore, a higher than expected incidence of symmetric GAT IOP measurements among paired eyes of the same patient has been observed when measured by GAT as compared to other methods.⁷ Additionally, a number of studies have demonstrated that the intra-observer and inter-observer coefficients of repeatability for GAT is as high as 4.3 and 5.7 mmHg respectively.⁸ Taken together, there is a need to make GAT more objective with improved reproducibility.

To overcome these limitations, we developed an objective and fully automated method for measuring IOP using deep learning with GAT techniques.

METHODS

Patients presenting to the University of Washington glaucoma clinics were recruited. This study was approved by the Human Subjects Division of the University of Washington (UW), it followed the tenets of the Declaration of Helsinki and was conducted in compliance with the Health Insurance Portability and Accountability Act. Informed consent was obtained from all subjects. Patients with a history of corneal scarring, corneal surgery or active infection of the eye were excluded. Demographic information including patient age, gender, and ocular history and exam findings were collected.

Data collection for algorithm generation

Subjects were randomized to receive either standard or automated GAT first. The GAT (Haag-Streit AG, Bern, Switzerland) was calibrated at the beginning of this study. One drop of topical Fluress® was applied to the eye. The applanation prism was rotated with its axis to 0 degrees and the tip was illuminated by the wide open slit beam using the blue filter at a 45 degree angle. For standard GAT, the prism was gently applanated against the corneal surface and the applanation mires were manually aligned, with the inner edges of the mires touching.

For automated GAT, an iPod Touch (Apple, Cupertino, CA, USA) was attached to the slit lamp ocular using a HookUpz 2.0 Smartphone Optics Adaptor (Carson, Ronkonkoma, NY, USA) (Figure 1). The slit lamp magnification was set to 10X. The GAT dial was set to 1.8 and the iPod Touch was set to record using the standard movie record function. The prism was gently applanated against the corneal surface until the appearance of applanation mires, and a 5–10 second video recording was made. The GAT dial was set at 1.8 (equivalent to an 18 mmHg IOP measurement) because at this force, a 6.12 mm tonometer tip should theoretically generate mires with diameters of reasonable resolution corresponding to IOP measurements ranging from 4.5 mmHg to approximately 42 mmHg, based on the Imbert-Fick formula. Each eye was measured once with standard GAT and once with automated GAT and measurements were all collected within a 5 minute time period.

Image Analysis and Deep Learning Methods

Videos were then split into training and validation videos in an 80:20 ratio. The videos taken were typically 13–26 seconds in length, which were captured at 24 to 30 frames/second. A typical frame during applanation is shown in Figure 1B. For each video, 50 frames were randomly sampled. For each frame the tonometer tip, left inner and right inner mires were manually fit if present (custom interface in MatLab R2019b); one circle delineating the circumference of the applanator tip and two for each of the semicircular mires (Figure 1C). When not all mires were present, these frames were also included in training as negative examples.

A modified PSPNet⁹ architecture with a ResNet50 backbone was used to detect the circles. The output layer was modified so that rather than a segmentation mask, the network output

nine numbers, predicting the x and y coordinates of the center of each circle and the radius of each circle. Furthermore, the class output was trained on the presence or absence of the inner mires. The loss function used was the sum of the Dice loss for the outer circle, the Dice loss for the inner circles if present, and the binary cross-entropy for the presence or absence of the inner mires. The Dice score between the ground truth and predicted circles was calculated analytically by calculating the areas of the regions given the locations and radii of the circles, rather than by counting pixels.

During training, data was augmented by randomly rotating images up to 15 degrees, rescaling the image by a random factor from 0.25 to 0.5, and cropping a 256×256 window containing most of the outer circle. This was done to broaden the set of training examples the network would see and ensure that the output would be robust to variations in angle, size, and location. In addition, synthetic tonometry images were generated and added to the training set (Supplemental Figure 1). While these are obviously distinguishable from real images, they gave the network a relatively easy but wide set of cases to learn and had a uniform standard for the locations of the circles. While we did not measure the exact effect of these augmentations, they have been generally observed to improve performance of deep learning models,¹⁰ and seem unlikely to have hindered performance in this case. This was only applied to the training of the network and not to the validation or the test set and performance metrics were done outside of these augmentations.

The diameter in mm of the semicircular mires was calculated based on the pixel ratio between the mires and the applanation tip and the known diameter (6.12 mm) of the applanation tip. The pixel ratios within the image are preserved, regardless of variability in magnification or minification. The diameter of the mire closest to the center of the applanation tip was taken as the representative diameter. This diameter was used to calculate the circular area of applanation and given the known force applied, an IOP was calculated using the Imbert-Fick equation ($IOP = \text{Force} / \text{Area}$).

When performing inference on videos, a 768×768 window of the video was cropped out and resized to a 256×256 image, which was input to the network. To apply the network properly, the cropped-out window needed to contain the applanator and ideally be centered on it. Since the position of the applanator was not known *a priori*, the following automated method for centering the the applanator in the window was created. In the first frame, an approximate x (resp. y) coordinate was chosen by maximizing the total brightness within the 768-pixel vertical (resp. horizontal) band centered on that coordinate. The window centered on those coordinates was segmented by the network, and the coordinates were readjusted based on the network output so the outer mire would be centered in the window. In subsequent frames, the window was moved so that the outer mire, as segmented by the network, would stay centered.

Within each video, each frame with a mire detected gave a pressure reading. Ideally, the reading from each frame would provide a precise and unbiased estimation of the patient's IOP. However, the readings of many frames were unreliable because the measurement procedure deviated from the ideal. Heuristics were developed using only data from the training set to filter out unreliable information and integrate temporal information of the IOP.

The primary cause of bad readings was the fact that the applanation was assumed to be at a fixed force, but the actual force would sometimes vary. Three heuristics were devised to remove frames with incorrect applanation forces. First, we observed that if the applanation tip only contacted the patient's eye briefly, the reading was generally inaccurate because the tonometer was not likely fully applanating the eye. Thus, the video was split into contiguous segments during which a mire was continuously detected, and any segment less than 20 frames was discarded to eliminate any prematurely short readings. Second, there is necessarily a brief transitional period between when the tonometer tip first makes contact with the eye and when the tonometer is applying the full applanation force (and likewise another transitional period as the applanation tip is being withdrawn). To ignore these periods, the first two and last two frames of each contiguous segment were discarded. Finally, we assumed *a priori* that a patient's IOP would not vary significantly over brief periods of time. Thus for each remaining frame, if the pressure fluctuated too much compared to the ten frames before and after that frame, it was discarded. This was measured by whether the maximum pressure in this window was greater than 1.2 times the minimum. Bad readings were also caused by the network incorrectly segmenting frames. We found that the preceding heuristics often removed incorrectly segmented frames, even though they were not specifically designed to do so. These heuristics were frozen and then applied to the validation and test sets. The IOP for the video was taken to be the median of the remaining pressure values. The resulting automated GAT IOP measurements from the training set were fit to the measured standard GAT IOP measurements by a least-squares linear fit, and these linear coefficients were used to calculate the IOP measurements of the test set (Supplemental Figure 2).

Test Set and Reproducibility Study

A separate set of patients were recruited prospectively for a reproducibility study and to act as the test set for the deep learning model. Subjects were randomized to technician or ophthalmologist first, and then randomized to the order of the measurement taken (standard versus automated GAT). Standard and automated GAT measurements were obtained using methods described above, however standard GAT tonometry measurements were made in a masked two-observer fashion (between the standard GAT measurements, the tonometer dial was reset to a random number and mire alignment was performed by one person and the number was read and recorded by a second person). Measurements taken by technician and ophthalmologist were performed within 30 minutes of one another. To compare inter-human segmentation, one frame containing mires from each of the test set videos was labeled by four different graders and compared. The four human segmentations were also compared against the deep learning segmentation.

Statistical analysis was performed using python and scipy code (python3.5.5 and scipy1.1.0) and the statistics package Prism7 (Graphpad Software, La Jolla, CA). Inter-human consistency of segmentations was examined by taking each grader in turn as the ground truth and using this as the standard for every other grader and the algorithm's output. Bland-Altman plots were used to compare standard versus automated GAT measurements. Coefficients of repeatability were calculated for inter-observer and inter-video measurements. Statistical significance was defined at the $p < 0.05$ level.

RESULTS

The baseline characteristics of patients recruited for the training/validation and testing are shown in the Table. A total of 263 eyes of 135 subjects were recruited for the training and validation, and a separate group of 50 eyes from 25 subjects were recruited prospectively for the test and reproducibility set. While the mean age for the test set was older than the mean age of the training and validation set (68.4 ± 8.7 versus 61.3 ± 13.9 and 58.8 ± 16.9 years, respectively, $p = 0.02$), the remaining baseline characteristics, such as patient gender, mean standard GAT IOP measurement, central corneal thickness or distribution of underlying ocular disease, did not differ significantly.

A total of 4219 and 1040 frames from 211 and 52 videos were included in the training and validation sets, respectively, and used to train a deep learning model. Learning curves are supplied in Supplemental Figure 3, and additional training details are supplied in the Supplemental Table. The model was trained to take individual frames and provide the classification of presence/absence of the mires as well as the center coordinates and radii of the tonometer surface and both mires. For the test set, the deep learning predictions were compared against four masked, manual segmentations. The medians for the intersection over union for the tonometer tip, inner right mire, and inner left mire were 0.94, 0.93, and 0.90 for deep learning and 0.94, 0.95, and 0.94 for inter-human variance, respectively (Supplemental Figure 4).

In the prospectively collected test set, 100% were successfully processed with the deep learning algorithm. Mean standard GAT IOP was 15.4 ± 4.5 mmHg and the mean automated GAT IOP was 14.5 ± 3.7 mmHg. The mean difference between automated and standard GAT was -0.9 mmHg with 95% LoA of -5.4 to 3.6 mmHg (Figure 3A). Ninety percent of the automated GAT IOP measurements were within ± 4 mmHg of standard GAT measurements and 63% of automated GAT measurements were within ± 2 mmHg. Mean difference between the two observers using standard GAT was 0.09 mmHg with LoA of -3.8 to 4.0 (Figure 3B). Mean difference between the two observers using automated GAT videos was -0.3 mmHg with LoA of -4.1 to 3.5 (Figure 3c). The coefficients of repeatability for automated and standard GAT were 3.8 and 3.9 mmHg, respectively. The mean absolute inter-observer differences did not differ significantly between standard and automated GAT (1.5 ± 1.3 versus 1.4 ± 1.4 mmHg, respectively, $P=0.6$). When IOP measurements for entire videos were plotted, ocular pulsations could be observed (Figure 5, Supplemental Video).

Although both eyes of study subjects in the test set were included in the above analysis, similar results were obtained when only the right eye of each study subject was analyzed. The mean difference between automated and standard GAT when analyzing right eyes only was -0.9 mmHg with 95% LoA of -4.3 to 2.5 mmHg and the coefficients of repeatability for automated and standard GAT were 2.8 and 3.3 mmHg, respectively.

DISCUSSION

In this study, we developed a novel method of automating GAT measurements using fixed force applanation and standard ophthalmic exam equipment in conjunction with a camera system. Video recordings were obtained of the GAT mires generated using a fixed force, and using deep learning, we developed an algorithm to automatically segment the GAT applanation mires for each video frame. An IOP corresponding to the diameter of the mire was then calculated. The IOP measurements obtained using the automated GAT were comparable to standard GAT measurements with a mean bias of -0.9 mmHg and 95% LoA of -5.4 to 3.6 mmHg. Our mean bias and 95% LoA compare favorably to other tonometers in clinical use. In a large meta-analysis comparing various tonometers to GAT, rebound tonometers had a mean bias of $+0.9$ mmHg (95% LoA -4.3 to 6.1 mmHg) and Tono-Pens had a mean bias of -0.2 mmHg and (95% LoA -6.2 to 5.8 mmHg).¹¹ When compared to GAT, 52% of rebound tonometer measurements and 48% of Tono-Pen measurements were within 2.0 mmHg of the GAT measurement.¹¹ In our study, 63% of the automated IOP measurements were within 2.0 mmHg of the standard GAT measurement

We chose to use a deep learning approach for the automation of image segmentation because we felt it would be the most robust method for successfully segmenting videos obtained from different slit lamp machines under a range of imaging conditions. In a previous project, our group developed a smartphone-based tonometer with fixed force applanation and we used a machine learning approach with more traditional image processing methods to detect circular applanation mires.¹² However, when this method was applied to all videos, only 56.8% of videos were successfully segmented. From this study we concluded that a deep learning approach would be necessary to successfully process a greater number of videos given the variability present between videos. Of note, the deep learning approach in this study successfully segmented 100% of the videos in the test set.

While GAT is considered by many to be the clinical reference standard for IOP measurements, it is known to be susceptible to measurement bias. Recently, studies by Rozwat and Roberts⁶ and Buchanan et al. have described a bias to even numbered IOP measurements compared to odd, which they attribute to the tendency of the observer to round to the even numbered tick marks on the GAT dial. In an evaluation of 5 different tonometers, Kutzscher et al. likewise identified a preference for even numbers using GAT, which was not observed with the other tonometers that produced a digital output.¹³ Indeed, even within our training and validation data set, the frequency distribution for standard GAT showed a preference for IOP values 12, 14, 16 and 18 while the corresponding automated GAT measurements did not show this bias (Figure 4).

Another source of bias with standard GAT is the concern that a single observer aligning the applanation mires may be influenced by the IOP measurement on the dial. To avoid this bias, many large, multi-centered clinical trials such as the Ocular Hypertension Study³ and the Advanced Glaucoma Intervention Study⁴ use a two person method for their GAT measurements, whereby one observer views the applanation mires and adjusts the GAT dial and a separate observer reads and records the number off the dial. By removing the

subjective bias of standard GAT using our automated method, research resources may be better allocated such that a single observer can perform the tonometry measurement.

Another limitation of standard GAT is the relatively high degree of inter-observer variability. Though considered the reference standard because of its relatively lower degree of variability compared to other tonometers, the coefficient of repeatability for standard GAT has been reported to range from 2.3 mmHg to 5.7 mmHg.⁸ As part of this study, we sought to assess the inter-observer repeatability of standard and automated GAT and found that the coefficients of repeatability for automated and standard GAT were 3.8 and 3.9 mmHg, respectively. These values were comparable to the coefficients of repeatability reported in other studies.^{14–16} Furthermore, under ideal circumstances, it has been recommended that inter-operator GAT measurements should be within ± 4 mmHg of each other in 95% of eyes.¹⁷ Within the test set, 92% of the standard GAT measurements were within ± 4 mmHg compared to 96% of the automated GAT measurements. With only 50 eyes included in our test set, we were able to demonstrate comparable and even improved repeatability using the automated GAT method. With continued development and refinement of the algorithm, we hypothesize that the objective nature of the automated GAT method should eventually produce measurements with even better repeatability.

Of note, the Bland-Altman plot demonstrates significant proportional bias across varying IOP values (Figure 3A). This is not surprising as standard GAT and automated GAT rely on different applanation principles based on the Imbert-Fick equation. With standard GAT, force is adjusted to fit the mires to a set area while the automated GAT uses a set force and measures the resulting mire size. The Imbert-Fick relationship assumes a sphere of infinitely thin walls, devoid of many biomechanical properties that are known to affect applanation, such as corneal thickness and ocular rigidity. A comparison of the IOP difference between automated and standard GAT versus central corneal thickness (CCT) did not demonstrate a significant correlation ($r = -0.27$, $p = 0.06$). However, ocular rigidity is known to vary depending on the IOP.¹⁸ The changes in ocular rigidity at the upper and lower IOP extremes likely affects fixed force and fixed area applanation differently, leading to the observed proportional bias. We incorporated linear coefficients derived from the training data (Supplemental Figure 2) to adjust for some of the proportional bias, which should account for some of the overall biomechanical effects, including CCT and ocular rigidity.

However, the residual proportional bias in the test set suggests that further studies are needed to better account for changing ocular rigidity and other physical properties of the cornea when using fixed-force applanation to improve the accuracy of the automated GAT measurements compared to standard GAT.

When IOP measurements from sequential frames of the automated GAT videos were plotted, ocular pulsations could be seen (Figure 5). The relationship between ocular pulse amplitude (OPA) and glaucoma is less clearly understood though some studies report that OPA is significantly lower in glaucomatous eyes¹⁹ and may be correlated with glaucoma disease severity.²⁰ As few clinical devices exist to measure OPA, namely the pneumatonometer and the dynamic contour tonometer, our current understanding may simply reflect the limited means with which to study this parameter. The relatively low cost addition of a camera

system to standard ophthalmic equipment may allow us to more easily study the OPA and its relation to various ophthalmologic conditions.

The observed mean bias of -0.9 mmHg between the automated and standard GAT may have been affected by the presence of ocular pulsations. With the automated GAT, the median IOP was chosen as representative of the overall IOP measurement. However, when measuring with standard GAT, the observer must try to align the mires such that the midpoint of the mire excursion is approximated. This can be challenging and difficult to approximate, especially when the ocular pulsations are large. OPA values have been reported to be as high as 3.83 ± 1.27 mmHg.²¹ If the observer were to consistently align the mires at the upper or lower bounds of the excursion, then their measurements would be consistently higher or lower than the median IOP, which could account for the observed bias. Indeed, when the data from the 2 observers were analyzed separately, observer 1 had a lower bias of -0.7 mmHg (LoA -4.6 to 3.2 mmHg) compared to observer 2 who had a bias of -1.1 mmHg (LoA -6.2 to 4.0 mmHg). Because automated GAT is the median value of many IOP measurements throughout the pulse amplitude, it may more reliably identify the median than observers visually aligning the median of mire excursions and therefore reduce bias.

One limitation of this study is that all the training, validation and test videos were collected from two slit lamps by two observers. While the resulting deep learning algorithm was successfully applied to a test set of videos to demonstrate proof of concept, future refinement of the algorithm will necessitate obtaining more videos by more observers with different slit lamps and different mobile devices to improve the generalizability. There were few study subjects with higher IOP (>21 mmHg) recruited in the test set. As a proof-of-principle pilot study, we demonstrated comparable IOP measurements to standard GAT using automated GAT, especially in the more normal range of IOP values (< 21 mmHg). However, given the proportional bias observed at higher IOP values, the automated GAT may underestimate the IOP in this population. Future recruitment of higher IOP eyes will be needed to refine the algorithm to ensure accurate measurements at higher pressures. Another limitation is that study recruitment occurred at an academic glaucoma clinic and so most patients were either glaucoma suspects or had glaucoma. Further studies are needed to assess whether biases and limits of agreement differ when using automated GAT to measure IOP in patients with other ocular conditions.

In conclusion, IOP measurements taken with automated GAT using standard ophthalmic equipment and a deep learning based algorithm were comparable to measurements taken using standard GAT with similar inter-observer repeatability. Automated GAT reduces some of the biases introduced by standard GAT and may be a way to decrease research costs by allowing single observer GAT measurements. Lastly, the creation of near ubiquitous methods for measuring ocular pulse amplitude as outlined in this study may open new avenues of future glaucoma research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: Research to Prevent Blindness. University of Washington CoMotion Innovation Fund. AYL was supported by the NEI/NIH K23EY029246. The funding organizations had no role in the design or conduct of this research.

AYL has received grant support from Novartis and Carl Zeiss Meditec. AYL has received honoraria from Topcon, Genentech and Verana Health. AYL works with the US FDA.

References:

- Halberg GP. Hand applanation tonometer. *Trans Am Acad Ophthalmol Otolaryngol* 1968;72:112–114. [PubMed: 5637509]
- Goldmann H. [A new applanation tonometer]. *Bull Mem Soc Fr Ophtalmol* 1954;67:474–7; discussion, 477–8. [PubMed: 13284610]
- Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol* 1999;117:573–583. [PubMed: 10326953]
- Ederer F, Gaasterland DE, Sullivan EK, AGIS Investigators. The Advanced Glaucoma Intervention Study (AGIS): 1. Study design and methods and baseline characteristics of study patients. *Control Clin Trials* 1994;15:299–325. [PubMed: 7956270]
- Musch DC, Lichter PR, Guire KE, Standardi CL. The Collaborative Initial Glaucoma Treatment Study: study design, methods, and baseline characteristics of enrolled patients. *Ophthalmology* 1999;106:653–662. [PubMed: 10201583]
- Rozwat A, Roberts DK. Even-number Measurement Bias with Goldmann Applanation Tonometry. *J Glaucoma* 2019 Available at: 10.1097/IJG.0000000000001412.
- Buchan JC, Macleod D, Hickman W, Bastawrous A. Systematic bias in real-world tonometry readings based on laterality? *Eye* 2019 Available at: 10.1038/s41433-019-0558-0.
- Pearce JG, Maddess T. The Clinical Interpretation of Changes in Intraocular Pressure Measurements Using Goldmann Applanation Tonometry: A Review. *J Glaucoma* 2019;28:302–306. [PubMed: 30946710]
- Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017:2881–2890.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016: 236–238.
- Cook JA, Botello AP, Elders A, et al. Systematic review of the agreement of tonometers with Goldmann applanation tonometry. *Ophthalmology* 2012;119:1552–1557. [PubMed: 22578443]
- Wu Y, Luttrell I, Feng S, et al. Development and validation of a machine learning, smartphone-based tonometer. *Br J Ophthalmol* 2019 Available at: 10.1136/bjophthalmol-2019-315446.
- Kutzscher AE, Kumar RS, Ramgopal B, et al. Reproducibility of 5 Methods of Ocular Tonometry. *Ophthalmology Glaucoma* 2019;2:429–434. Available at: 10.1016/j.ogla.2019.07.006. [PubMed: 32672576]
- AlMubrad TM, Ogbuehi KC. The effect of repeated applanation on subsequent IOP measurements. *Clin Exp Optom* 2008;91:524–529. [PubMed: 18651843]
- Cheng J, Salam T, Russell PJ, et al. Dynamic contour tonometer and Goldmann applanation tonometer performance in a developing world setting: intraocular pressure measurement acquisition and precision. *J Glaucoma* 2013;22:736–739. [PubMed: 23708422]
- Kotecha A, White ET, Shewry JM, Garway-Heath DF. The relative effects of corneal thickness and age on Goldmann applanation tonometry and dynamic contour tonometry. *Br J Ophthalmol* 2005;89:1572–1575. [PubMed: 16299132]
- Weinreb RN, Brandt JD, Garway-Heath D, Medeiros FA. Intraocular Pressure: Measurement of Intraocular Pressure. In: *World Glaucoma Association 4th Consensus Meeting*.
- Ytteborg J. The effect of intraocular pressure on rigidity coefficient in the human eye. *Acta Ophthalmol* 1960;38:548–561. [PubMed: 13787537]
- Abegão Pinto L, Vandewalle E, Willekens K, et al. Ocular pulse amplitude and Doppler waveform analysis in glaucoma patients. *Acta Ophthalmol* 2014;92:e280–5. [PubMed: 24456194]

20. Vulsteke C, Stalmans I, Fieuws S, Zeyen T. Correlation between ocular pulse amplitude measured by dynamic contour tonometer and visual field defects. *Graefes Arch Clin Exp Ophthalmol* 2008;246:559–565. [PubMed: 18193267]
21. Kac MJ, Solari HP, Velarde GC, et al. Ocular pulse amplitude in patients with asymmetric primary open-angle glaucoma. *Curr Eye Res* 2011;36:727–732. [PubMed: 21780922]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

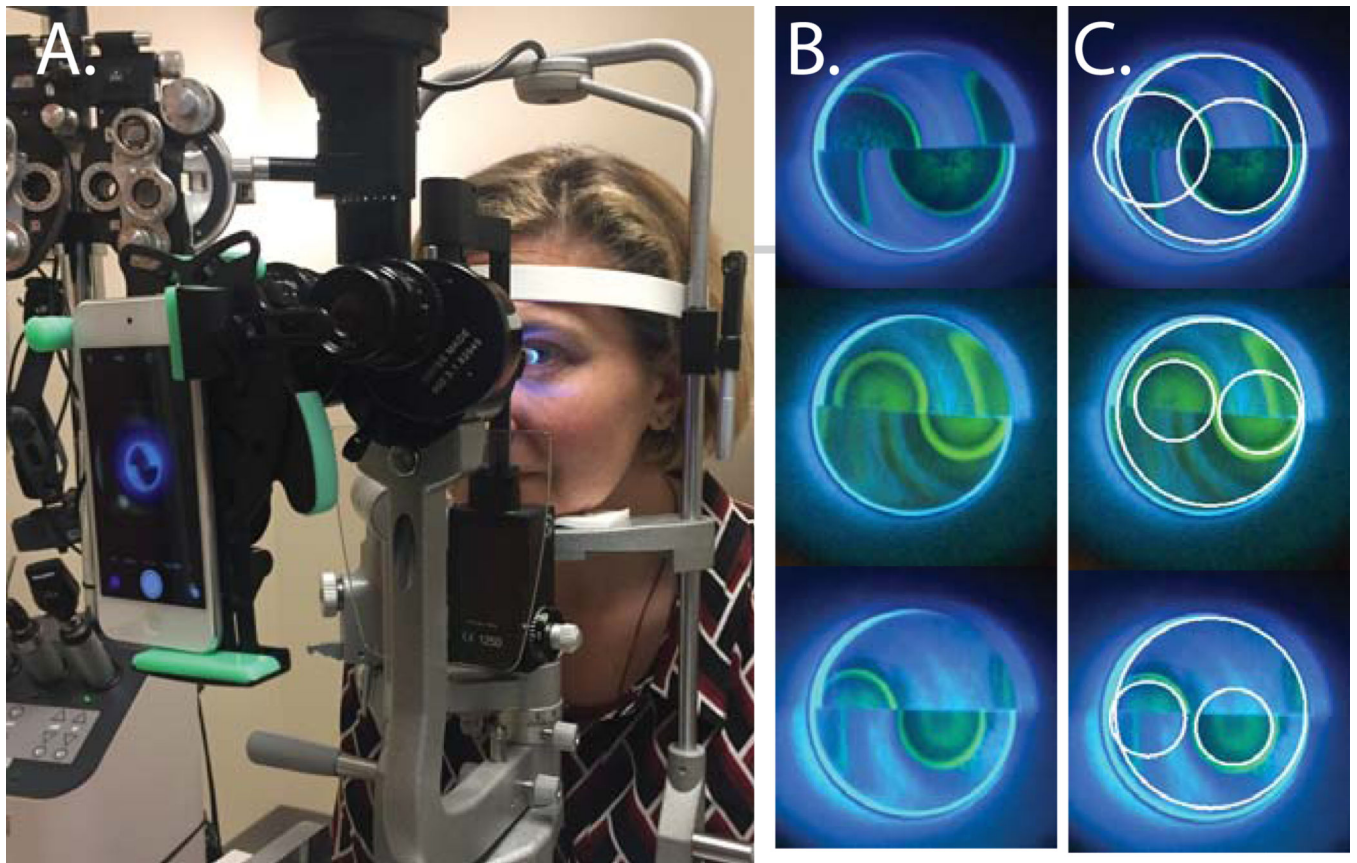


Figure 1. Representative images of the automated Goldman Applanation Tonometry (GAT) technique.

(A) The camera system set up on a slit lamp microscope. (B) Unlabeled frames from automated GAT video recordings. (C) Labeled frames with the tonometer tip (outer white circle) and 2 mires (smaller white circles) outlined. Because the size of the mire at a set force is what is being measured, no adjustment of mires for overlap is performed.

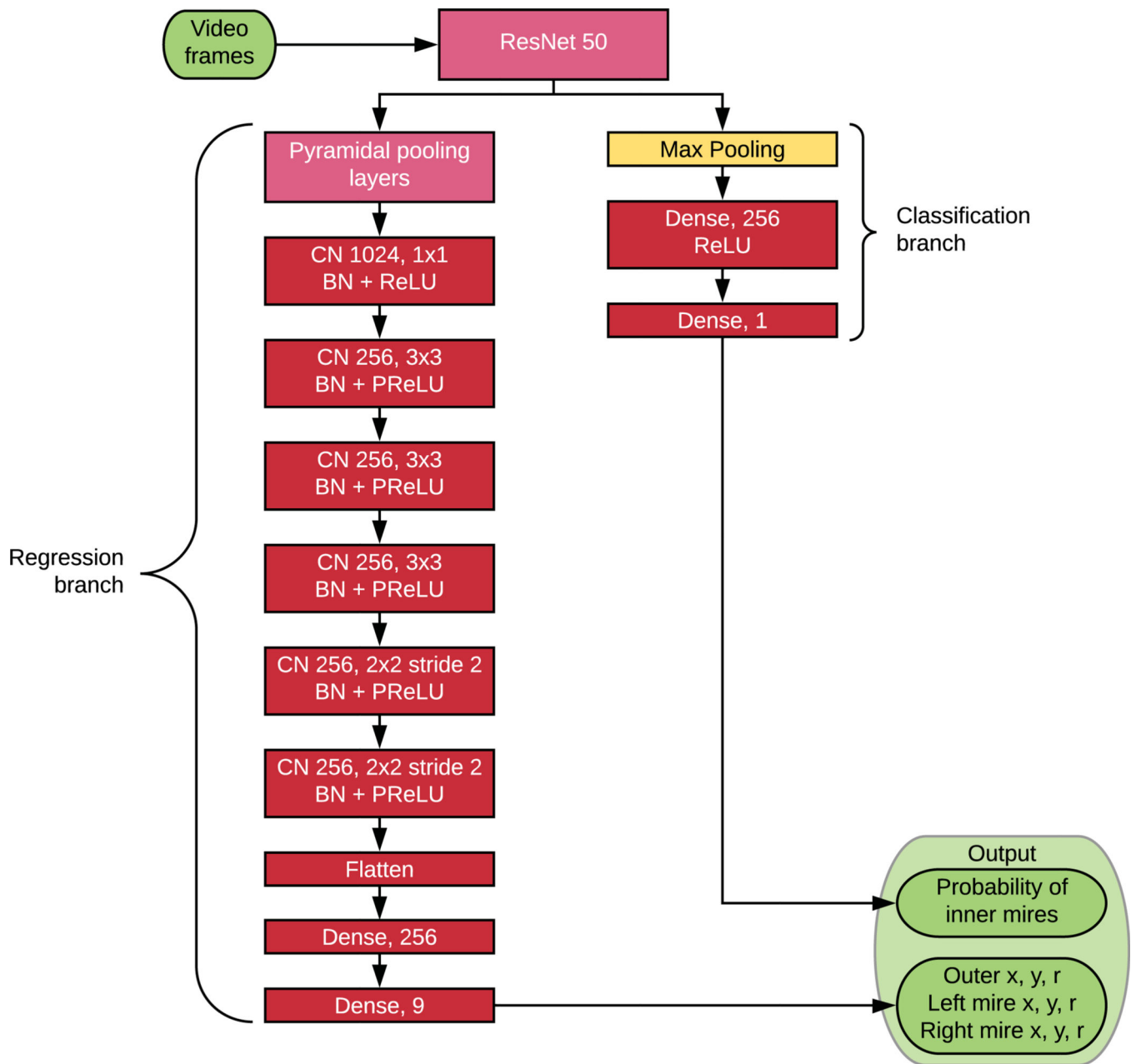


Figure 2. Neural Network Architecture.

Each video frame is separately processed by a deep learning model to directly output a classification loss to predict the presence or absence of the inner mires as well as the coordinates and radii of the tonometer tip and the two inner mires.

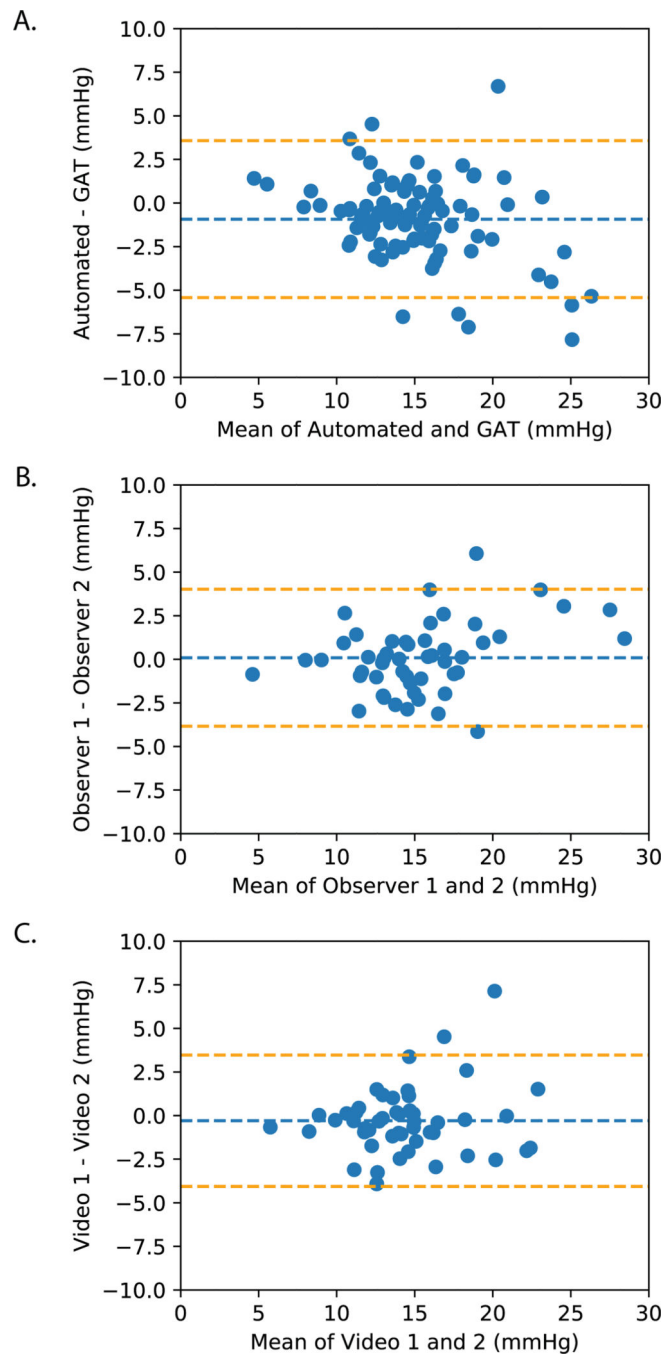


Figure 3. Bland-Altman plots comparing intraocular pressure (IOP) values from: (A) Standard versus automated Goldmann Applanation Tonometry (GAT) (B) Standard GAT by two observers (C) Automated GAT videos by two observers.

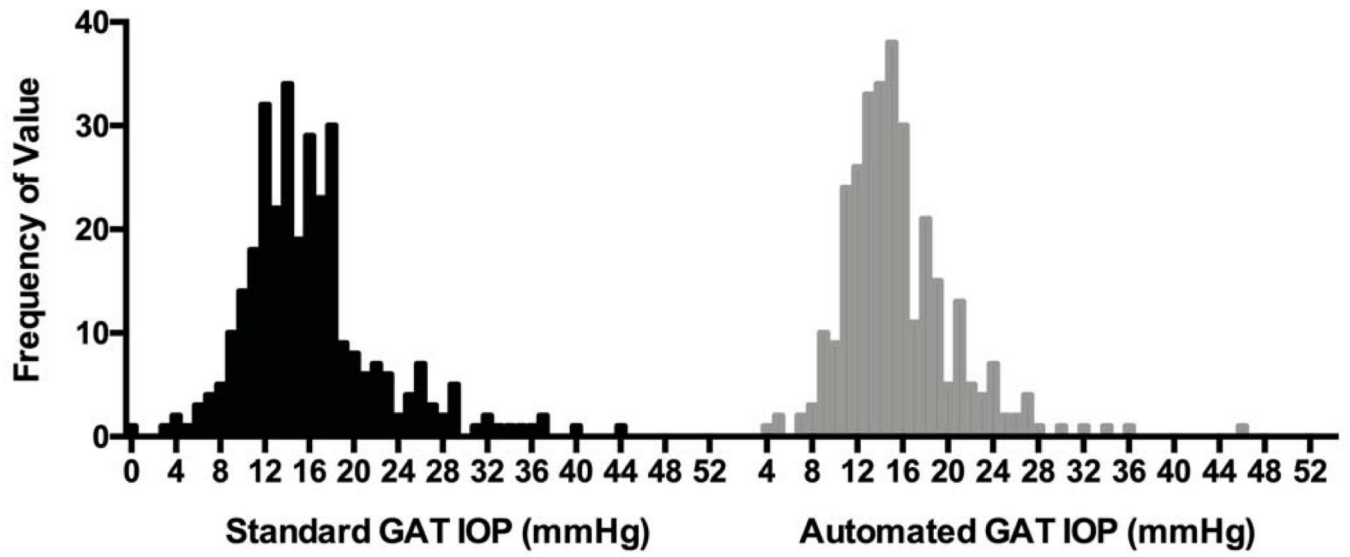


Figure 4. Intraocular Pressure (IOP) Frequency Distributions for standard and automated Goldmann Applanation Tonometry (GAT).

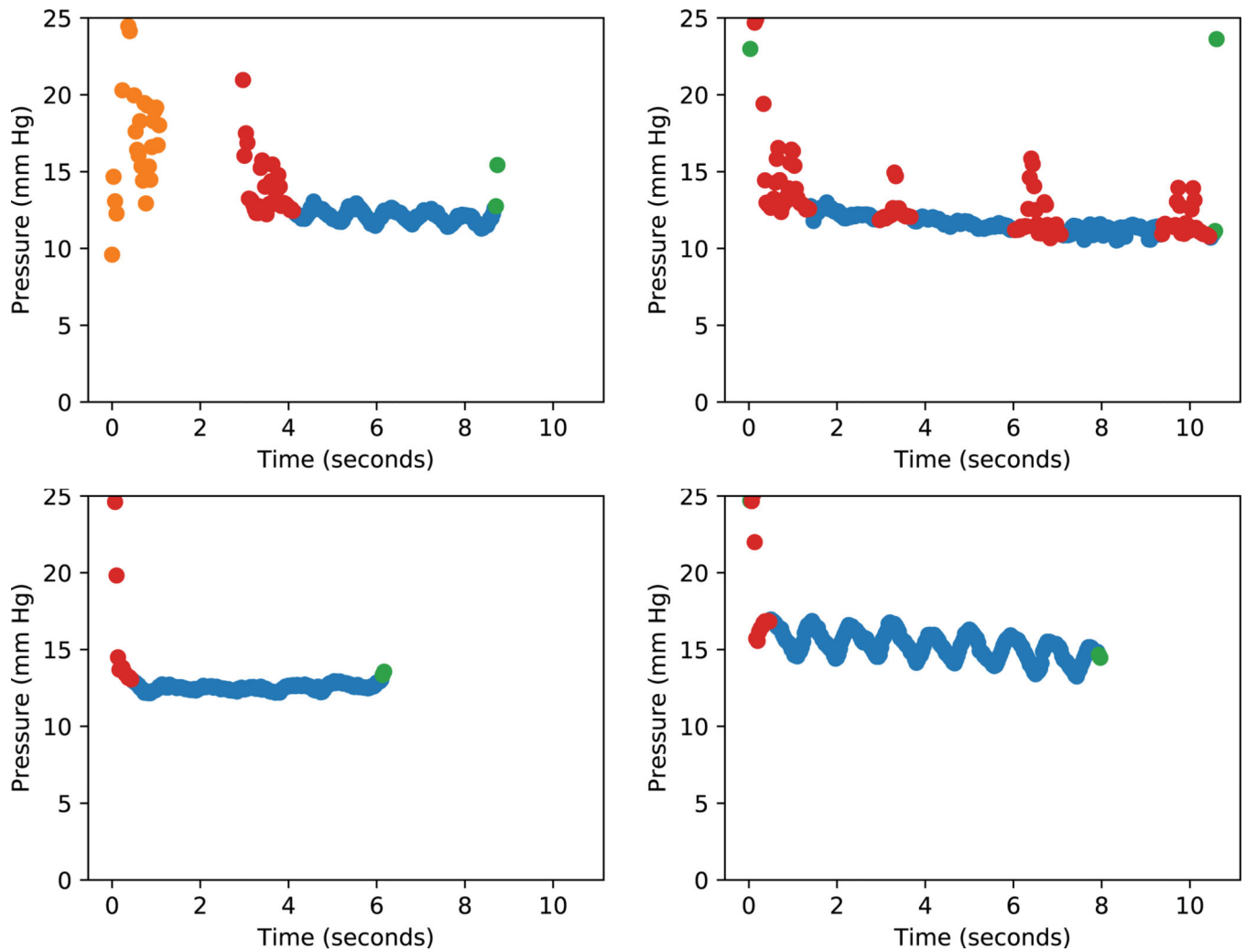


Figure 5. Examples of ocular pulse amplitudes detected by automated Goldmann applanation tonometry measurements.

Heuristics are applied for integrating temporal information. Dots are color coded as follows: Blue (measurement included), yellow (mire not present for at least 20 consecutive frames), green (a measurement immediately before or after an absent mire), red (the pressure measurement was more than 1.2 times the minimum measurement when compared to the ten frames before and after).

Table.

Demographics and baseline characteristics of the training, validation, and test sets.

Parameter	Training Set	Validation Set	Test Set	P-value
Number of Eyes	1211	52	50	-
Age, years \pm SD	61.3 \pm 13.9	58.8 \pm 16.9	68.4 \pm 8.7	0.02*
Sex, Number Female (%)	99 (47%)	24 (46%)	20 (40%)	0.79**
Mean GAT IOP \pm SD, mmHg	16.1 \pm 6.5	16.8 \pm 7.4	Obs1 15.3 \pm 4.2 Obs2 15.4 \pm 4.9	0.49*
Mean Central Corneal Thickness, μ m	547.1 \pm 52.7	541.4 \pm 46.8	548.1 \pm 44.6	0.54*
Diagnosis, % (n)				0.24***
Primary Open Angle Glaucoma	38% (80)	40% (21)	52% (26)	
Glaucoma Suspect	22% (47)	15% (8)	22% (11)	
Normal	19% (40)	25% (13)	8% (4)	
Primary Angle Closure (Glaucoma)	8% (17)	2% (1)	4% (2)	
Ocular Hypertension	10% (22)	12% (6)	14% (7)	
Neovascular Glaucoma	1% (2)	4% (2)	0% (0)	
Uveitic Glaucoma	1% (3)	2% (1)	0% (0)	

* One-way ANOVA

** Fisher Exact Test

*** Chi-Square Test