tvst

Article

Open-Source Automatic Biomarker Measurement on Slit-Lamp Photography to Estimate Visual Acuity in Microbial Keratitis

Jessica Loo[1], Maria A. Woodward[2,3,*], Venkatesh Prajna[4], Matthias F. Kriegel[2,5], Mercy Pawar[2], Mariam Khan[2], Leslie M. Niziol[2], and Sina Farsiu[1,6,*]

[1] Department of Biomedical Engineering, Duke University, Durham, NC, USA
[2] Department of Ophthalmology and Visual Sciences, Kellogg Eye Center, University of Michigan, Ann Arbor, MI, USA
[3] Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, USA
[4] Department of Cornea and Refractive Services, Aravind Eye Care System, Madurai, Tamil Nadu, India
[5] Department of Ophthalmology, Augenzentrum am St. Franziskus Hospital Münster, Münster, Germany
[6] Department of Ophthalmology, Duke University Medical Center, Durham, NC, USA

Correspondence: Sina Farsiu, 101 Science Drive, Campus Box 90281, Durham, NC 27708, USA. e-mail: sina.farsiu@duke.edu

Received: April 26, 2021
Accepted: September 7, 2021
Published: October 4, 2021

Keywords: cornea; microbial keratitis; slit-lamp photography; deep learning; machine learning

Citation: Loo J, Woodward MA, Prajna V, Kriegel MF, Pawar M, Khan M, Niziol LM, Farsiu S. Open-source automatic biomarker measurement on slit-lamp photography to estimate visual acuity in microbial keratitis. Transl Vis Sci Technol. 2021;10(12):2, https://doi.org/10.1167/tvst.10.12.2

Purpose: To assess clinical applicability of automatic image analysis in microbial keratitis (MK) by evaluating the relationship between biomarker measurements on slit-lamp photography (SLP) and best-corrected visual acuity (BCVA).

Methods: Seventy-six patients with MK with SLP images and same-day logarithm of the minimum angle of resolution (logMAR) BCVA were evaluated. MK biomarkers (stromal infiltrate, white blood cell infiltration, corneal edema, hypopyon, epithelial defect) were segmented manually by ophthalmologists and automatically by a novel, open-source, deep learning–based segmentation algorithm. Five measurements (presence, maximum width, total area, proportion of the corneal limbus area affected, centrality) were calculated. Correlations between the measurements and BCVA were calculated. An automatic regression model estimated BCVA from the measurements. Differences in performance between using manual and automatic measurements were evaluated using William's test (for correlation) and the paired-sample *t*-test (for absolute error).

Results: Measurements had high correlations of 0.86 (manual) and 0.84 (automatic) with true BCVA. Estimated BCVA had average (mean $\pm$ SD) absolute errors of 0.39 $\pm$ 0.27 logMAR (manual, median: 0.30) and 0.35 $\pm$ 0.28 logMAR (automatic, median: 0.30) and high correlations of 0.76 (manual) and 0.80 (automatic) with true BCVA. Differences between using manual and automatic measurements were not statistically significant for correlations of measurements with true BCVA ($P = .66$), absolute errors of estimated BCVA ($P = .15$), or correlations of estimated BCVA with true BCVA ($P = .60$).

Conclusions: The proposed algorithm measured MK biomarkers as accurately as ophthalmologists. Measurements were highly correlated with and estimative of visual acuity.

Translational Relevance: This study demonstrates the potential of developing fully automatic objective and standardized strategies to aid ophthalmologists in the clinical assessment of MK.

Introduction

Microbial keratitis (MK) is an infectious corneal disease manifesting in a diverse range of presentations. It is one of the leading causes of blindness worldwide, with many risk factors such as ocular trauma, contact lens use, hygiene, climate, and geography.[1–8] Ophthalmologists assess disease severity and recommend treatment regimens by evaluating the morphology of MK biomarkers. Key biomarkers are the stromal infiltrate (SI), epithelial defect (ED), and presence of a

hypopyon.[9–11] Other biomarkers such as white blood cell (WBC) infiltration and swelling due to surrounding corneal edema, if present, may also be indicative of disease severity.

The current methods to measure and record the morphology of MK biomarkers are subjective and include manual slit-lamp biomicroscope measurements using built-in calipers or manual rulers, free-text descriptions, and schematic drawings in the electronic health records (EHRs).[12] Measurements are not always consistently recorded in the EHR.[13,14] Additionally, patients can be seen by multiple ophthalmologists and have multiple follow-ups. Studies have shown that measurements differ among ophthalmologists, even in controlled settings.[15] Strategies have been developed that use measurements of the SI and hypopyon to aid ophthalmologists in the clinical assessment of MK.[16–18] However, these strategies do not use measurements of other biomarkers such as the ED, WBC infiltration, and corneal edema. Furthermore, to date, these strategies have not been widely adopted in clinical practice given the variability and subjectivity of manual measurements. Therefore, ophthalmologists do not use any image-based, let alone automatic, methods to measure and use MK biomarkers systematically for clinical assessment.

Image-based and computer-aided methods can reduce clinician-dependent variability and improve reliability of measurements.[19,20] Slit-lamp photography (SLP) is a low-cost technology universally available in eye clinics that provides high-resolution digital images of the eye.[21] Semiautomatic and fully automatic segmentation algorithms have been developed for SLP images of MK.[19,22–26] These are promising steps toward developing objective and standardized strategies to aid ophthalmologists in the clinical assessment of MK. However, some key questions that remain to be addressed are (1) how well these image-based measurements correlate with other important clinical measures such as visual function and (2) if these automatic segmentation algorithms can successfully substitute manual measurement methods to aid ophthalmologists in decision making for clinical care.[27] Often, it is not clear whether the performance of an automatic segmentation algorithm is good enough for clinical application. While performance metrics such as sensitivity, specificity, and the Dice similarity coefficient (DSC) are commonly used to measure the agreement between manual and automatic segmentations, there are no established thresholds that define clinically acceptable performance. Furthermore, such performance metrics are interpreted based on the assumption that manual segmentations are the gold standard and error free, which may not necessarily be the case, especially in complex or ambiguous cases. Therefore, an alternative approach to evaluate the performance of an automatic segmentation algorithm is to use both manual and automatic segmentations in a downstream task that uses an independent measure as the gold standard instead. The difference in performances in this downstream task can then be used to determine if the automatic segmentations are as good as manual segmentations.

In this study, we evaluate the clinical applicability of automatic image-based measurements on SLP images by comparing a novel, open-source, fully automatic, deep learning–based segmentation algorithm to manual segmentation by ophthalmologists. We evaluate (1) the correlation between image-based measurements of MK biomarkers and best-corrected visual acuity (BCVA) and (2) the ability to estimate BCVA from these image-based measurements using machine learning.

Methods

Data Set

The data set for this study was collected from patients at two academic eye hospitals—the University of Michigan Kellogg Eye Center (Ann Arbor, MI, USA) and Aravind Eye Care System (Madurai, India). Approval was obtained from the University of Michigan Institutional Review Board and Aravind Madurai Ethics Committee. The study adhered to the tenets of the Declaration of Helsinki.

Inclusion criteria were patients 15 years or older with MK as diagnosed by the treating clinician. Patients were excluded from participation if from a vulnerable population (i.e., hospitalized, imprisoned, institutionalized) or pregnant or not able to provide consent. Patients were also excluded if they had a corneal perforation, a prior corneal incisional surgery (i.e., corneal transplantation), or no light perception vision. Patients were not excluded for other preexisting health or ocular conditions.

Participants were photographed under diffuse white light illumination and/or diffuse blue light illumination with fluorescein staining following an SLP protocol.[19] Topical artificial tears were applied prior to imaging. For fluorescein staining, one drop of tetracaine was applied to a fluorescein strip, and the strip was then applied to the participant's lower fornix. The photographer allowed the participant to remove any excess tears and then immediately proceeded to imaging. Participants were photographed using a Canon EOS 7D camera (Tokyo, Japan) mounted on a Haag-Streit

International BX 900 model slit-lamp biomicroscope (Köniz, Switzerland). Ambient room lighting and a diffuse beam at maximum width (30 mm) of the light source with maximum light intensity as tolerated by the participant were used. Same-day BCVA was measured by the clinical standard of Snellen acuity and converted to logarithm of the minimum angle of resolution (logMAR) for analysis.[28]

Regions and Measurements of Interest

Pathologic regions of interest (ROIs) were evaluated as MK biomarkers. The SI, WBC infiltration, corneal edema, and hypopyon were segmented on diffuse white light images. The ED was segmented on diffuse blue light images. Nonpathologic ROIs were imaging-associated and anatomic structures. Ocular surface light reflexes, the pupil, and the corneal limbus were segmented on both diffuse white light and diffuse blue light images. The corneal limbus provided a reference for the true size of the ROIs, where the published average adult corneal limbus diameter of 11.7 mm was used to estimate the pixel resolution in millimeters to enable conversion from pixels to millimeters to obtain clinically meaningful measurements.[12,29]

Five measurements of interest were calculated for each pathologic ROI—the presence (absent or present), maximum width (mm), total area (mm^2), and proportion of the corneal limbus area affected (0–1) and centrality (mm^{-1}).[9,16–19]

The proportion of the corneal limbus area affected was calculated as

$$\text{Proportion of the corneal limbus area aflected} = \frac{\text{Total area of the ROI}}{\text{Total area of the corneal limbus}}$$

and centrality was calculated as

$$\text{Centrality} = \log\left(1 + \frac{1}{D}\right),$$

where D was the distance (mm) of the centroid of the ROI from the centroid of the corneal limbus. A logarithmic scale was used to avoid extreme values of centrality for ROIs very close to the center.

Manual Segmentation

Manual segmentation of the ROIs on the SLP images was performed by ophthalmologists (MFK, KHK) using ImageJ software (National Institutes of Health, Bethesda, MD, USA), as described in prior work.[12,26] Each image was segmented by one ophthalmologist. Segmentations were reviewed by a corneal specialist (MAW), who was also available for consultation and discussion involving difficult or ambiguous cases.

Automatic Segmentation

Automatic segmentation of the ROIs on the SLP images was performed by a novel modified version of our recently published, open-source, fully automatic, deep learning–based algorithm, SLIT-Net.[26] SLIT-Net is a convolutional neural network (CNN) designed and trained to identify and segment ROIs on SLP images. Separate networks were trained for diffuse white light and diffuse blue light images. The modifications made since the publication of SLIT-Net are summarized below. More details are available in Online Supplementary 1.

1. **Extended training:** The training procedure was extended from 100 epochs to 300 epochs.
2. **Threshold validation:** The validation procedure of retaining the weights of the best-performing epoch as the final weights of the network, by monitoring the performance on the hold-out validation set for the last 20 epochs at fixed classification and segmentation thresholds of 0.5, was replaced with a threshold validation procedure. For threshold validation, the weights of the final epoch were retained as the final weights of the network, and grid search was used to determine the best-performing classification and segmentation thresholds for each individual ROI class by monitoring the performance on the hold-out validation set.
3. **Limbus-Net:** The corneal limbus was segmented by Limbus-Net, a CNN designed and trained to specifically segment the corneal limbus on SLP images to obtain more accurate segmentation of this ROI. The images were cropped to the region automatically segmented by Limbus-Net before being provided as input to SLIT-Net for the subsequent segmentation of all other ROIs. This was done to focus the field of view as all other ROIs should occur only within the corneal limbus.

To evaluate the performance of the automatic segmentation algorithm, the DSC[30] was calculated to measure the proportion of overlap between the automatic segmentations and the gold-standard manual segmentations. The DSC ranges from 0 to 1, whereby a higher value indicates better performance. The absolute errors of the corneal limbus diameter and the resulting pixel resolution

obtained from the automatic segmentations were also calculated, using the corresponding values obtained from the gold-standard manual segmentations, whereby a lower value indicates better performance.

Correlation With BCVA

Pearson's correlation was calculated to analyze the relationship between the measurements of pathologic ROIs obtained from either manual or automatic segmentations and BCVA. For individual measurements, the correlation coefficient was calculated between each measurement of interest and BCVA. For multiple measurements, a multivariable linear regression model was fitted using the least squares approach with all five measurements of interest as the input variables and BCVA as the outcome variable. The multiple correlation coefficient was calculated as Pearson's correlation between the fitted values from the model and BCVA.[31,32] The multiple correlation analysis was performed for each individual pathologic ROI and the combination of all pathologic ROIs.

Estimation of BCVA

A radial basis function kernel epsilon–support vector machine (SVM)[33,34] for regression was trained to estimate BCVA using the measurements of pathologic ROIs obtained from either manual or automatic segmentations. The radial basis function kernel epsilon-SVM was selected based on early experiments comparing a range of linear and nonlinear models. These models included linear regression, ridge regression, decision tree, random forest, multilayer perceptron, and SVM models, and results indicated that the best performance could be achieved with the radial basis function kernel epsilon-SVM.

The SVM was trained for 3000 iterations with an epsilon of 0.05 and L_2 weight regularization applied with a factor of 2. All five measurements of interest were used for all pathologic ROIs. Continuous measurements were normalized by subtracting the mean and dividing by the standard deviation to ensure that all measurements were on a similar scale. All estimated values were rounded to one decimal place and capped to the minimum (0.0 logMAR) and maximum (2.3 logMAR) values in the data set.

To evaluate the performance of the BCVA estimation algorithm, Pearson's correlation between the true BCVA and estimated BCVA and the absolute error of the estimated BCVA were calculated. Performance metrics were calculated separately for the models using measurements obtained from manual and automatic segmentations.

Statistical Significance Tests

To determine the statistical significance of the difference between using manual and automatic segmentations, William's test was used for the differences in correlation, and the paired-sample *t*-test was used for the differences in absolute error.

Cross-Validation

Sixfold cross-validation was used to evaluate the performance of the automatic segmentation and BCVA estimation algorithms on all available data to avoid selection bias and to ensure independence between the training and testing sets. For each algorithm, the participants were randomly divided into six groups of approximately equal size. Five groups were designated as the training set, and the remaining group was designated as the testing set. The groups were rotated such that each group was used once for testing. For validation procedures, one group from the training set was designated as the hold-out validation set.

Implementation

The data set was prepared using ImageJ and MATLAB (MathWorks, Natick, MA, USA). The automatic segmentation and BCVA estimation algorithms were implemented in Python using the TensorFlow[35] (version 1.5.1), Keras[36] (version 2.0.8), and Scikit-learn[37] (version 0.22.2) libraries. Statistical analysis was performed using MATLAB[38] (version 9.5.0 R2018b). The data set and new and improved algorithms are available on GitHub at https://github.com/jessicaloohw/SLIT-Net_v2.

Results

Data Set

The data set consisted of 76 participants with SLP images and same-day BCVA. Twenty-seven participants (35.5%) were from the United States, and 49 participants (64.5%) were from India. Each participant had one diffuse white light image and one diffuse blue light image available from one eye. BCVA ranged from 0.0 to 2.3 logMAR, and the mean ± SD BCVA was 1.3 ± 0.74 (median, 1.4) logMAR. Five participants had missing demographic information from the clinical chart. The remaining 71 participants ranged from 19 to 88 years of age and were on average (mean ± SD) 49.6 ± 15.2 (median, 49) years. Twenty-five

participants (35%) were female. The cause of MK was fungi in 32 cases (45.1%), bacteria in 16 cases (22.5%), a mixture of fungi and bacteria in 2 cases (2.8%), acanthamoeba in 2 cases (2.8%), and indeterminant in 19 cases (26.8%) due to either no culture being performed or no growth in the culture.

One hundred and thirteen additional participants with SLP images were included for the development of the automatic segmentation algorithm. Fifty-nine participants had both diffuse white light and diffuse blue light images available, 44 participants had only diffuse white light images available, and 10 participants had only diffuse blue light images available. Ten participants had more than one image available for a specific illumination. In total, 195 diffuse white light images and 148 diffuse blue light images from 189 participants were available for the development of the automatic segmentation algorithm.

Automatic Segmentation

The 189 participants with SLP images were used for training and testing the automatic segmentation algorithm. Table 1 shows the average DSC of the automatic segmentation algorithm. Table 2 shows the average absolute error of the corneal limbus diameter and resulting pixel resolution obtained from the automatic segmentation algorithm. Overall, the performance of the automatic segmentation algorithm was good, with moderate to high DSCs and low absolute errors.[39] All DSCs had a mean above 0.6 and a median above 0.8. The DSCs for the corneal limbus were at least 0.9, indicating very high performance, and the absolute errors of the resulting pixel resolution were less than half a micron, indicating accurate conversion from pixels to millimeters to obtain clinically meaningful measurements. The performance of this novel modified version of the automatic segmentation algorithm was significantly better than the original version. More details are available in Online Supplementary 2.

Correlation With BCVA

The 76 participants with SLP images and same-day BCVA were used for the correlation analysis. Table 3 shows the correlation coefficients between individual measurements of interest and BCVA. Overall, the correlation coefficients ranged from −0.18 to 0.70, and most correlations were statistically significant, except for some measurements of corneal edema and some measurements of presence. The highest correlations were most frequently observed for measurements of maximum width, whereas the lowest correlations were most frequently observed for measurements of presence. There were no statistically significant differences between using manual and automatic segmentations for any measurements of the SI, WBC infiltration, hypopyon, and ED. There was a statistically significant increase in correlation using automatic segmentations for all measurements of corneal edema, except presence. More details are available in Online Supplementary 3.

Table 4 shows the multiple correlation coefficients between all five measurements of interest and BCVA. Overall, the correlation coefficients ranged from 0.26 to 0.86, and all correlations were statistically significant. While the correlations were low to moderate for individual pathologic ROIs, the correlation was high for the combination of all pathologic ROIs.[39] There were no statistically significant differences between

Table 1. Average (Mean ± SD, Median) DSC of the Automatic Segmentation Algorithm on Diffuse White Light and Diffuse Blue Light Images

Illumination	Type of ROI	ROI	DSC
Diffuse white light	Pathologic	SI	0.68 ± 0.30, 0.81
		WBC infiltration	0.62 ± 0.40, 0.83
		Corneal edema	0.72 ± 0.45, 1.00
		Hypopyon	0.91 ± 0.25, 1.00
	Nonpathologic	Light reflexes	0.74 ± 0.28, 0.86
		Pupil	0.76 ± 0.35, 0.94
		Corneal limbus	0.93 ± 0.17, 0.97
Diffuse blue light	Pathologic	ED	0.78 ± 0.33, 0.93
	Nonpathologic	Light reflexes	0.80 ± 0.20, 0.87
		Pupil	0.67 ± 0.40, 0.87
		Corneal limbus	0.90 ± 0.14, 0.94

Table 2. Average (Mean ± SD, Median) Absolute Error of the Corneal Limbus Diameter and Resulting Pixel Resolution Obtained From the Automatic Segmentation Algorithm on Diffuse White Light and Diffuse Blue Light Images

Illumination	Absolute Error of Corneal Limbus Diameter (Pixels)	Absolute Error of Pixel Resolution (10^{-3} mm)
Diffuse white light	57 ± 64, 40	0.23 ± 0.47, 0.08
Diffuse blue light	164 ± 182, 105	0.38 ± 0.57, 0.19

Six diffuse white light and three diffuse blue light images were excluded from analysis due to the corneal limbus not being segmented manually.

Table 3. Correlation Coefficients between the Individual Measurements of Interest and BCVA

	Manual Segmentations					Automatic Segmentations				
ROI	P	W	A	%	C	P	W	A	%	C
SI	**0.27**	**0.65**	**0.49**	**0.47**	**0.37**	**0.36**	**0.70**	**0.58**	**0.58**	**0.52**
WBC infiltration	0.11	**0.46**	**0.51**	**0.51**	**0.40**	**0.28**	**0.55**	**0.61**	**0.60**	**0.47**
Corneal edema	−0.18	−0.14	−0.07	−0.08	−0.13	0.14	0.21	**0.26**	**0.25**	0.20
Hypopyon	**0.55**	**0.57**	**0.45**	**0.46**	**0.57**	**0.58**	**0.58**	**0.52**	**0.53**	**0.59**
ED	0.21	**0.54**	**0.44**	**0.42**	**0.34**	**0.31**	**0.57**	**0.46**	**0.44**	**0.37**

Statistically significant ($P \leq 0.05$) correlation coefficients are shown in bold.

William's test was used to determine the statistical significance between manual and automatic segmentations. Statistically significant ($P \leq 0.05$) differences are underlined. % indicates the proportion of corneal limbus area affected. A, total area; C, centrality; P, presence; W, maximum width.

Table 4. Multiple Correlation Coefficients Between All Five Measurements of Interest (Presence, Maximum Width, Total Area, Proportion of Corneal Limbus Area Affected, and Centrality) and BCVA for Each Individual Pathologic ROI and the Combination of all Pathologic ROIs From the Multivariable Linear Regression Model

ROI	Manual Segmentations	Automatic Segmentations
SI	0.71	0.74
WBC infiltration	0.62	0.66
Corneal edema	0.26	0.38
Hypopyon	0.59	0.59
ED	0.64	0.61
Combination	0.86	0.84

All correlation coefficients were statistically significant ($P \leq 0.05$).

William's test was used to determine the statistical significance between manual and automatic segmentations. There were no statistically significant differences between using manual and automatic segmentations.

using manual and automatic segmentations. More details are available in Online Supplementary 3.

Estimation of BCVA

The 76 participants with SLP images and same-day BCVA were used for training and testing the BCVA estimation algorithm. Overall, the performance of the BCVA estimation algorithm was good with high correlations and moderate absolute errors.[39,40] Using the measurements of pathologic ROIs obtained from manual segmentation, the algorithm achieved a correlation of 0.76 ($P \leq 0.001$) and an average absolute error of 0.39 ± 0.27 (median, 0.30) logMAR. Using the measurements of pathologic ROIs obtained from automatic segmentation, the algorithm achieved a correlation of 0.80 ($P \leq 0.001$) and an average absolute error of 0.35 ± 0.28 (median:, 0.30) logMAR. While the performance was better using automatic segmentations, this difference was not statistically significant ($P = .60$ for the correlation and $P = .15$ for the absolute error). Figure 1 shows the relationship between the true BCVA and estimated BCVA of the algorithm using the measurements obtained from either manual

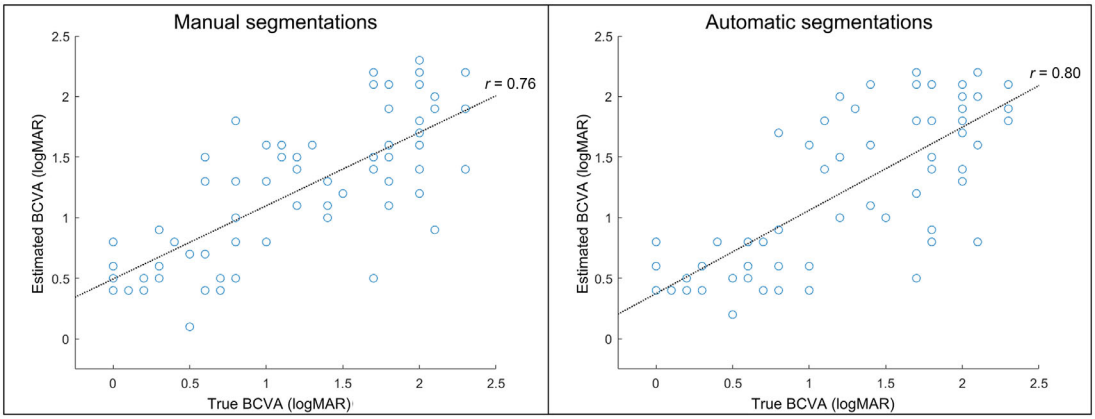

Figure 1. Scatterplot showing the relationship between the true BCVA and estimated BCVA of the BCVA estimation algorithm. The difference in performance between using manual and automatic segmentations was not statistically significant.

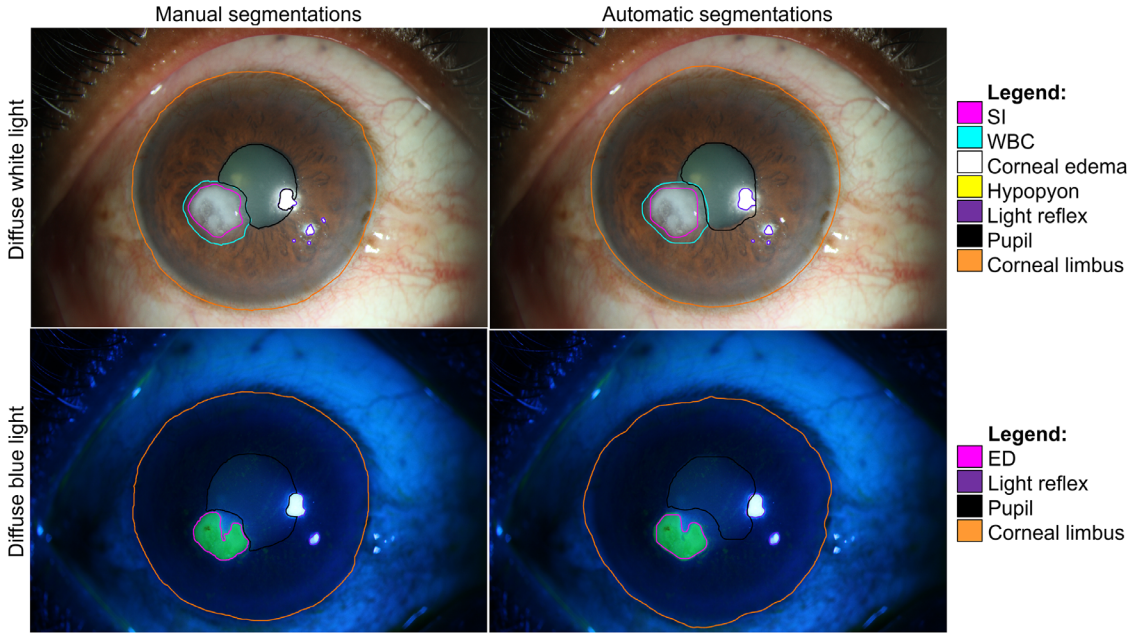

Figure 2. Example of SLP images with manual and automatic segmentations. The estimated BCVA was 0.7 logMAR and 0.5 logMAR using the measurements obtained from manual and automatic segmentations, respectively. The true BCVA was 0.5 logMAR.

or automatic segmentations. Figures 2 to 5 show examples of the segmented images and the corresponding estimated BCVA.

Discussion

MK is an infectious corneal disease and one of the leading causes of blindness worldwide. The purpose of this work is to aid ophthalmologists in the clinical assessment of MK by developing automatic decision aid tools and algorithms that analyze MK biomarkers linked to clinical outcomes. Five MK biomarkers (SI, WBC infiltration, corneal edema, hypopyon, and ED) were analyzed based on five measurements of interest (presence, maximum width, total area, proportion of the corneal limbus area affected, and centrality). Measurements of interest were calculated using two segmentation methods—manual segmentation by ophthalmologists and automatic segmentation by a novel modified version of our open-source, fully automatic, deep learning–based segmentation algorithm. The automatic segmentations overlapped with the manual segmentations with moderate to high DSCs ranging from 0.62 to 0.93. There were also low

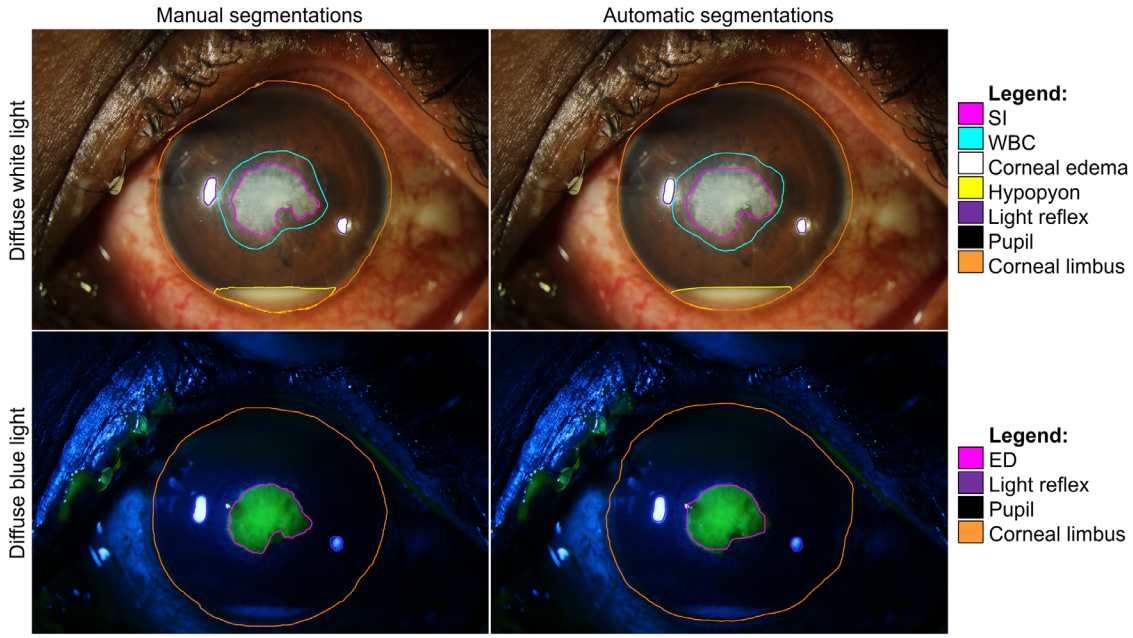

Figure 3. Example of SLP images with manual and automatic segmentations. The estimated BCVA was 1.7 logMAR using both measurements obtained from manual and automatic segmentations. The true BCVA was 2.0 logMAR.

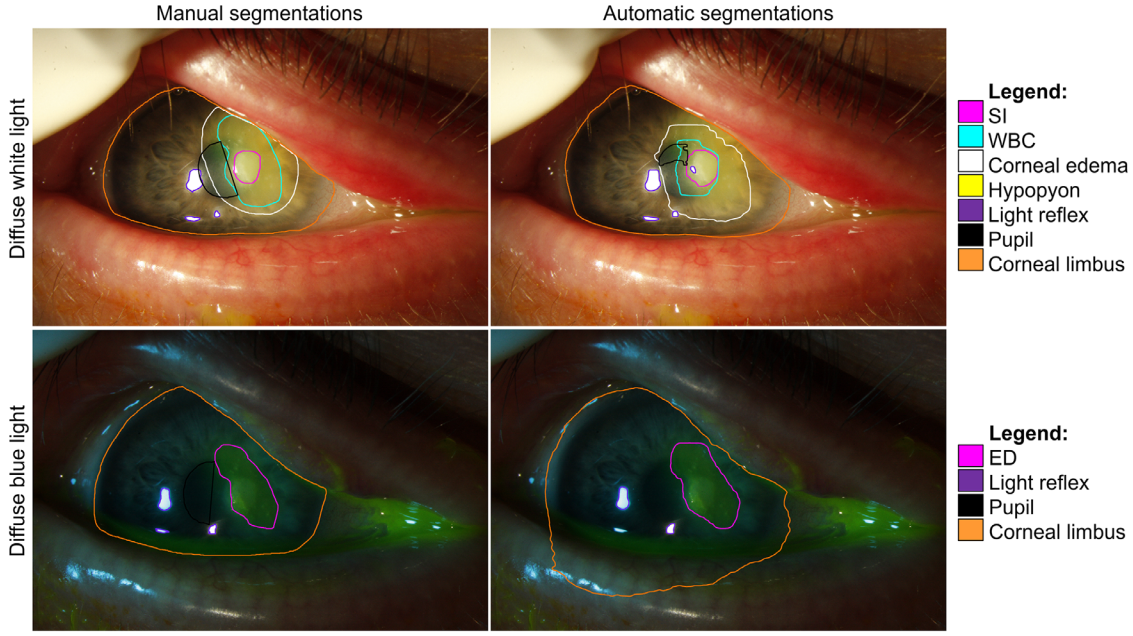

Figure 4. Example of SLP images with manual and automatic segmentations. The estimated BCVA was 0.8 logMAR and 0.9 logMAR using the measurements obtained from manual and automatic segmentations, respectively. The true BCVA was 0.4 logMAR.

absolute errors of less than half a micron in estimating the pixel resolution when converting from pixels to millimeters to obtain clinically meaningful measurements. The performance of this novel modified version of the algorithm was significantly better than the original version.[26] The measurements had good estimation of same-day visual acuity, indicating important structure–function correlation.

The analysis of the correlation between individual measurements and BCVA showed that the highest correlations were most often observed for the measurements of maximum width, and the lowest correlations

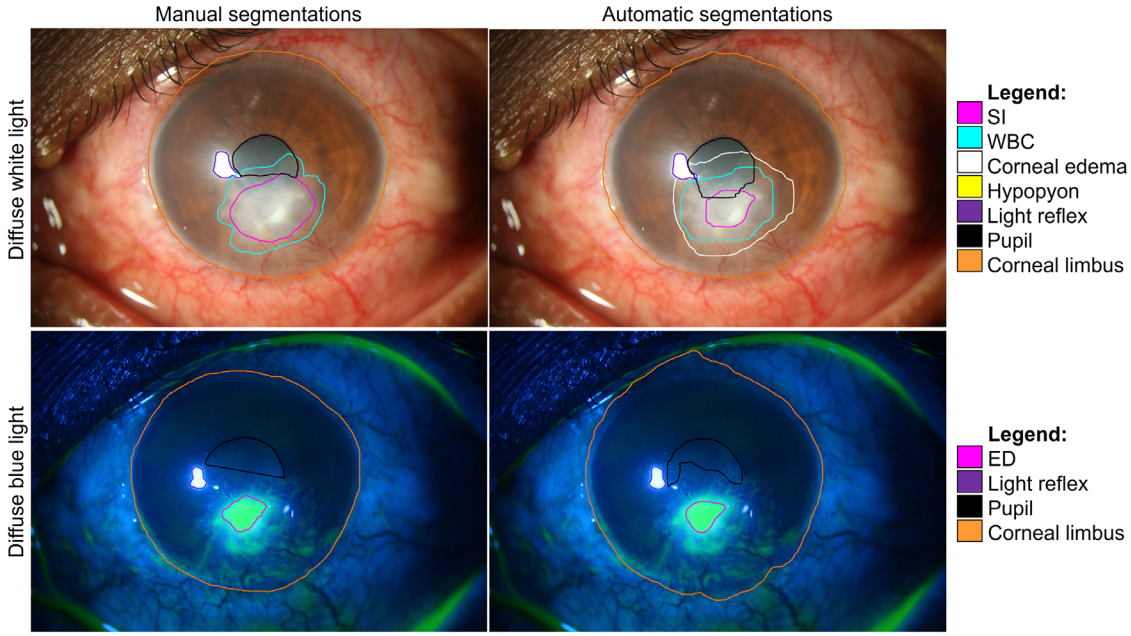

Figure 5. Example of SLP images with manual and automatic segmentations. The estimated BCVA was 1.1 logMAR using both measurements obtained from manual and automatic segmentations. The true BCVA was 0.8 logMAR.

were most often observed for measurements of presence of a biomarker. This indicates that measurements of the size, rather than the presence of a biomarker, are more correlated with visual acuity. The moderate correlations observed for measurements of centrality may in part be due to the logarithmic scale used in the calculations. We will further explore alternative approaches for calculating centrality to determine if improved correlation with BCVA can be achieved, such as using a different scaling function or calculating centrality based on the edge of the ROI instead of the centroid of the ROI. There were no statistically significant differences between using manual and automatic segmentations for any measurements of the SI, WBC infiltration, hypopyon, and ED. Interestingly, all measurements of corneal edema had negative correlations with BCVA using manual segmentations but positive correlations using automatic segmentations, and these differences were statistically significant for all measurements except presence. Prior work has shown that corneal edema is a difficult biomarker to identify and segment even by experienced ophthalmologists, and using automatic segmentations may improve segmentation consistency.[26] This indicates that assessing corneal edema may play a role in assessing visual function in corneal imaging of MK. While not necessarily the most prominent biomarker in MK, corneal edema certainly affects visual acuity across many anterior segment conditions, such as Fuchs dystrophy.

The multiple correlation analysis showed that for individual pathologic ROIs, measurements of the SI and ED had the highest correlations with BCVA, consistent with the current clinical knowledge and understanding of the disease.[9,10,41] However, adding measurements of other biomarkers, including WBC infiltration, corneal edema, and the hypopyon, increased the correlations up to 20%. This demonstrates the importance of including not only the SI and ED in assessment but other biomarkers as well. Overall, for the combination of all pathologic ROIs, the measurements had high correlations of 0.86 and 0.84 with BCVA, using manual and automatic segmentations, respectively. There were no statistically significant differences between using manual and automatic segmentations.

Using an SVM, BCVA could be automatically estimated from the measurements. The estimated BCVA had high correlations of 0.76 and 0.80 with the true BCVA and average absolute errors of 0.39 ± 0.27 (median, 0.30) logMAR and 0.35 ± 0.28 (median, 0.30) logMAR, using manual and automatic segmentations, respectively. While the performance using automatic segmentations was better than using manual segmentations, this difference was not statistically significant, most likely due to the small sample size. The performance was considered good as the median absolute errors were approximate to the minimum standard of 15 letters on the Early Treatment Diabetic Retinopathy Study (ETDRS) chart (corresponding

to 0.3 logMAR) often used in the field of ophthalmology.[42–46] The mean absolute error using automatic segmentations was also less than half the SD of BCVA (0.5 SD = 0.37 logMAR), whereby half the SD may be considered the minimally important difference for health-related quality of life.[40] However, while the levels of correlation reflect a positive trend, some of the individual absolute errors may still be considered large by clinical standards. Therefore, there is certainly room for improvement of the algorithm before it can be implemented for real-world clinical use. Optimizing the algorithm and improving the size and robustness of the data set will likely reduce the absolute error, ideally to within five letters on the ETDRS chart (corresponding to 0.1 logMAR). As BCVA depends on many other factors besides corneal features, the inclusion of additional clinically relevant features and demographics such as age, race, gender, and location will be important. In particular, BCVA would also be affected by the presence of other ocular diseases that do not manifest in the SLP images, such as retinal or optic nerve diseases. Thus, this algorithm is not expected to be applicable to all patient populations, such as a much older patient population, which may also be affected by age-related macular degeneration. The inclusion of ocular history, demographics, or biomarkers obtained from images of other parts of the eye, such as optical coherence tomography, will also be important to improve the performance of the algorithm and extend the target population of the algorithm beyond patients with MK.

Overall, the accuracy and utility of using image-based measurements for assessing functional severity in MK was demonstrated by the significant correlations and the ability to estimate BCVA from these measurements. There was no statistically significant difference between using manual and automatic segmentations in most cases, as well as statistically significant improvement in some cases, demonstrating the accuracy and clinical applicability of the automatic segmentation algorithm in lieu of time-consuming and labor-intensive manual segmentation by ophthalmologists. While this was a pilot study based on a limited data set of 76 participants, these results can have important clinical implications in the future as the data set expands and the algorithms are improved. For example, if remote health care personnel could take a photograph of a cornea and be able to automatically assess the patient's visual acuity with some certainty, the health care team would have better knowledge about triage and referral of the patient to tertiary eye care centers. However, the automatic algorithm would only be intended for use as a decision support tool to aid health care personnel and should not be used in a standalone manner. Any estimations provided by the algorithm would need to be interpreted based on the specific patient's context and health care personnel's expertise. The standard questions pertaining to the patient's ocular history and major population health ocular conditions such as diabetic eye disease, macular degeneration, cataracts, and glaucoma should be asked and considered accordingly. Further research is necessary before this technology can be implemented for real-world clinical use. As the research progresses, standardized strategies for quantified imaging are adopted, and our understanding of the structure–function relationship in MK evolves, we expect that it will ultimately lead to improved diagnosis, prognosis, and treatment, similar to what has been observed in other diseases such as macular degeneration, diabetic macular edema, and glaucoma over the past three decades.[47–50]

There are limitations to this study. Since the enrollment was a pilot study, it is based on a limited data set of 76 participants with an MK diagnosis. However, the MK population can be much more diverse than the study population. Therefore, care must be taken when extrapolating the results from this study to the larger MK population. We will further investigate the structure–function relationship in MK as we expand the data set in our ongoing prospective clinical study (Automated Quantitative Ulcer Analysis Study [AQUA], NCT04420962). We also note that the current study indicates only correlation, not causation, between the image-based measurements and BCVA. We will delve deeper into descriptive analysis and model interpretation in future work. For example, we will use univariate analysis followed by a feature selection procedure to select potential predictors and confounders to build a final model to provide a clearer picture regarding any significant associations of the individual predictors or lack thereof. We will also investigate the relationship between the morphology of MK biomarkers and other important factors such as pathogenesis, patient-reported symptoms and discomfort, and other clinical findings. We will continue to improve the algorithms as well, including exploring the use of images taken under different illuminations.

In conclusion, image-based and computer-aided measurements of the morphology of MK biomarkers show a high correlation with visual function and can be used to assess the disease, paving the way for the development of objective and standardized strategies to aid ophthalmologists in the clinical assessment of MK. The performance of the novel modified version of our open-source, fully automatic, deep learning–based segmentation algorithm was as accurate as manual

segmentation by ophthalmologists, demonstrating the potential of these strategies being fully automatic and requiring only minimal manual intervention from ophthalmologists for clinical use. To promote future research in this area, we have made our data set and new and improved algorithms open source and freely available online.

Acknowledgments

The authors thank Kyeong Hwan Kim, Megan Tuohy, Tomas Meijome, Joseph Pongrac, and Kassim Salami for their help with the preparation and manual segmentation of part of the data set.

Supported by the National Institutes of Health (R01 EY031033-01, P30 EY005722), Research to Prevent Blindness Career Advancement Award (MAW), Research to Prevent Blindness Unrestricted Grant to Duke University, and Michigan Institute for Clinical & Health Research Pilot Grant (UL1TR002240). The sponsor or funding organization had no role in the design or conduct of this research.

Disclosure: **J. Loo**, None; **M.A. Woodward**, None; **V. Prajna**, None; **M.F. Kriegel**, None; M. **Pawar**, None; **M. Khan**, None; **L.M. Niziol**, None; **S. Farsiu**, None

* MAW and SF are joint senior authors.

References

1. Whitcher JP, Srinivasan M, Upadhyay MP. Corneal blindness: a global perspective. *Bull WHO*. 2001;79:214–221.
2. Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob Health*. 2013;1(6):e339–e349.
3. Flaxman SR, Bourne RR, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health*. 2017;5(12):e1221–e1234.
4. Musch DC, Sugar A, Meyer RF. Demographic and predisposing factors in corneal ulceration. *Arch Ophthalmol*. 1983;101(10):1545–1548.
5. Dart J, Stapleton F, Minassian D. Contact lenses and other risk factors in microbial keratitis. *Lancet*. 1991;338(8768):650–653.
6. Lam D, Houang E, Fan D, et al. Incidence and risk factors for microbial keratitis in Hong Kong: comparison with Europe and North America. *Eye*. 2002;16(5):608.
7. Gopinathan U, Sharma S, Garg P, Rao GN. Review of epidemiological features, microbiological diagnosis and treatment outcome of microbial keratitis: experience of over a decade. *Indian J Ophthalmol*. 2009;57(4):273.
8. Gorski M, Genis A, Yushvayev S, et al. Seasonal variation in the presentation of infectious keratitis. *Eye Contact Lens*. 2016;42(5):295–297.
9. Krachmer J, Mannis M, Holland E. *Cornea*. Maryland Heights, MO: Elsevier; 2005.
10. Miedziak AI, Miller MR, Rapuano CJ, et al. Risk factors in microbial keratitis leading to penetrating keratoplasty. *Ophthalmology*. 1999;106(6):1166–1171.
11. Lalitha P, Prajna NV, Kabra A, et al. Risk factors for treatment outcome in fungal keratitis. *Ophthalmology*. 2006;113(4):526–530.
12. Kriegel MF, Loo J, Farsiu S, et al. Measurement reliability for keratitis morphology. *Cornea*. 2020;39(12):1503–1509.
13. Ballouz D, Maganti N, Tuohy M, et al. Medication burden for patients with bacterial keratitis. *Cornea*. 2019;38(8):933.
14. Maganti N, Tan H, Niziol LM, et al. Natural language processing to quantify microbial keratitis measurements. *Ophthalmology*. 2019;126(12):1722–1724.
15. Parikh PC, Valikodath NG, Estopinal CB, et al. Precision of epithelial defect measurements. *Cornea*. 2017;36(4):419.
16. McLeod SD, LaBree LD, Tayyanipour R, et al. The importance of initial management in the treatment of severe infectious corneal ulcers. *Ophthalmology*. 1995;102(12):1943–1948.
17. McLeod SD, Kolahdouz-Isfahani A, Rostamian K, et al. The role of smears, cultures, and antibiotic sensitivity testing in the management of suspected infectious keratitis. *Ophthalmology*. 1996;103(1):23–28.
18. Vital MC, Belloso M, Prager TC, Lanier JD. Classifying the severity of corneal ulcers by using the "1, 2, 3" rule. *Cornea*. 2007;26(1):16–20.
19. Patel TP, Prajna NV, Farsiu S, et al. Novel image-based analysis for reduction of clinician-dependent variability in measurement of the corneal ulcer size. *Cornea*. 2018;37(3):331–339.
20. Toutain-Kidd CM, Porco TC, Kidd EM, et al. Evaluation of fungal keratitis using a newly

developed computer program, Optscore, for grading digital corneal photographs. *Ophthalmic Epidemiol.* 2014;21(1):24–32.
21. Martonyi CL, Bahn CF, Meyer RF. *Slit Lamp: Examination and Photography*. Yavapai, AZ: Time One Ink, Ltd.; 2007.
22. Sun Q, Deng L, Liu J, et al. Patch-based deep convolutional neural network for corneal ulcer area segmentation. In: Cardoso J, Arbel T, Melbourne A, Bogunovic H, Moeskops P, Chen X, Schwartz E, Garvin MK, Robinson E, Trucco E, Ebner M, Xu Y, Makropoulos A, Desjardin A, Vercauteren T, eds. *Fetal, Infant and Ophthalmic Medical Image Analysis*. Basel, Switzerland: Springer; 2017:101–108
23. Deng L, Huang H, Yuan J, Tang X. Automatic segmentation of corneal ulcer area based on ocular staining images. In: Gimi B, Krol A, eds. *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Bellingham, WA: International Society for Optics and Photonics; 2018:10578.
24. Deng L, Huang H, Yuan J, Tang X. Superpixel based automatic segmentation of corneal ulcers from ocular staining images. *IEEE 23rd International Conference on Digital Signal Processing (DSP)*. New York City, NY: IEEE; 2018:1–5
25. Liu Z, Shi Y, Zhan P, et al. Automatic corneal ulcer segmentation combining Gaussian mixture modeling and Otsu method. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. New York City, NY: IEEE; 2019:6298–6301.
26. Loo J, Kriegel MF, Tuohy MM, et al. Open-source automatic segmentation of ocular structures and biomarkers of microbial keratitis on slit-lamp photography images using deep learning. *IEEE J Biomed Health Inform*. 2021;25(1):88–99.
27. Loo J, Clemons TE, Chew EY, et al. Beyond performance metrics: automatic deep learning retinal OCT analysis reproduces clinical trial outcome. *Ophthalmology*. 2020;127(6):793–801.
28. Holladay JT. Proper method for calculating average visual acuity. *J Refract Surg*. 1997;13(4):388–391.
29. Baumeister M, Terzi E, Ekici Y, Kohnen T. Comparison of manual and automated methods to determine horizontal corneal diameter. *J Cataract Refract Surg*. 2004;30(2):374–380.
30. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297–302.
31. Abdi H. Multiple correlation coefficient. In: Salkind NJ, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: SAGE Publications, Inc.; 2007;648–651.
32. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Statistical Models*. New York, NY: McGraw Hill; 1996.
33. Fan R-E, Chang K-W, Hsieh C-J, et al. LIBLINEAR: a library for large linear classification. *J Mach Learn Res*. 2008;9:1871–1874.
34. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2(3):1–27.
35. Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning. *OSDI*. 2016;5:16.
36. Chollet F, et al. Keras. 2015, https://keras.io/getting_started/faq/#how-should-i-cite-keras.
37. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
38. MATLAB. version 9.5.0 (R2018b). Natick, MA: MathWorks; 2018.
39. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24(3):69–71.
40. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582–592.
41. Bartimote C, Foster J, Watson S. The spectrum of microbial keratitis: an updated review. *Open Ophthalmol J*. 2019;13(1):100–130.
42. Rosenfeld PJ, Brown DM, Heier JS, et al. Ranibizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2006;355(14):1419–1431.
43. Brown DM, Kaiser PK, Michels M, et al. Ranibizumab versus verteporfin for neovascular age-related macular degeneration. *N Engl J Med*. 2006;355(14):1432–1444.
44. Heier JS, Brown DM, Chong V, et al. Intravitreal aflibercept (VEGF trap-eye) in wet age-related macular degeneration. *Ophthalmology*. 2012;119(12):2537–2548.
45. Gillespie BW, Musch DC, Niziol LM, Janz NK. Estimating minimally important differences for two vision-specific quality of life measures. *Invest Ophthamol Vis Sci*. 2014;55(7):4206–4212.
46. Kaiser PK. Prospective evaluation of visual acuity assessment: a comparison of Snellen versus ETDRS charts in clinical practice (an AOS thesis). *Trans Am Ophthalmol Soc*. 2009;107:311.

47. Malik R, Swanson WH, DF Garway-Heath. 'Structure–function relationship' in glaucoma: past thinking and current concepts. *Clin Exp Ophthalmol*. 2012;40(4):369–380.
48. Yohannan J, Boland MV. The evolving role of the relationship between optic nerve structure and function in glaucoma. *Ophthalmology*. 2017;124(12):S66–S70.
49. Group CR. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2011;364(20):1897–1908.
50. Soheilian M, Ramezani A, Yaseri M, et al. Initial macular thickness and response to treatment in diabetic macular edema. *Retina*. 2011;31(8):1564–1573.