

An Artificial-Intelligence–Based Automated Grading and Lesions Segmentation System for Myopic Maculopathy Based on Color Fundus Photographs

Jia Tang^{1,2,*}, Mingzhen Yuan^{1,2,*}, Kaibin Tian^{3,*}, Yuelin Wang^{1,2,*}, Dongyue Wang^{1,2,*}, Jingyuan Yang^{1,2}, Zhikun Yang¹, Xixi He⁴, Yan Luo¹, Ying Li¹, Jie Xu⁵, Xirong Li^{3,6}, Dayong Ding⁴, Yanhan Ren⁷, Youxin Chen^{1,2}, Srinivas R. Sadda^{8,9}, and Weihong Yu^{1,2}

¹ Department of Ophthalmology, Peking Union Medical College Hospital, Beijing, China

² Key Laboratory of Ocular Fundus Diseases, Chinese Academy of Medical Sciences, Beijing, China

³ AI and Media Computing Lab, School of Information, Renmin University of China, Beijing, China

⁴ Vistel AI Lab, Visionary Intelligence, Beijing, China

⁵ Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Laboratory, Beijing, China

⁶ Key Laboratory of Data Engineering and Knowledge Engineering, Renmin University of China, Beijing, China

⁷ Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA

⁸ Doheny Eye Institute, Los Angeles, CA, USA

⁹ Department of Ophthalmology, University of California, Los Angeles, Los Angeles, CA, USA

Correspondence: Youxin Chen, Department of Ophthalmology, Peking Union Medical College Hospital, No. 1 Shuaifuyuan Wangfujing, Dongcheng District, Beijing 100730, China.

e-mail: chenyx@pumch.cn

Srinivas R. Sadda, Doheny Eye Institute, 1355 San Pablo Street, DVRC 211, Los Angeles, California 90033, USA.

e-mail: ssadda@doheny.org

Weihong Yu, Department of Ophthalmology, Peking Union Medical College Hospital, No. 1 Shuaifuyuan Wangfujing, Dongcheng District, Beijing 100730, China.

e-mail: yuwh@pumch.cn

Received: April 15, 2021

Accepted: May 18, 2022

Published: June 15, 2022

Keywords: artificial intelligence; deep learning; color fundus photograph; myopic maculopathy; pathologic myopia

Purpose: To develop deep learning models based on color fundus photographs that can automatically grade myopic maculopathy, diagnose pathologic myopia, and identify and segment myopia-related lesions.

Methods: Photographs were graded and annotated by four ophthalmologists and were then divided into a high-consistency subgroup or a low-consistency subgroup according to the consistency between the results of the graders. ResNet-50 network was used to develop the classification model, and DeepLabv3+ network was used to develop the segmentation model for lesion identification. The two models were then combined to develop the classification-and-segmentation–based co-decision model.

Results: This study included 1395 color fundus photographs from 895 patients. The grading accuracy of the co-decision model was 0.9370, and the quadratic-weighted κ coefficient was 0.9651; the co-decision model achieved an area under the receiver operating characteristic curve of 0.9980 in diagnosing pathologic myopia. The photograph-level F_1 values of the segmentation model identifying optic disc, peripapillary atrophy, diffuse atrophy, patchy atrophy, and macular atrophy were all >0.95 ; the pixel-level F_1 values for segmenting optic disc and peripapillary atrophy were both >0.9 ; the pixel-level F_1 values for segmenting diffuse atrophy, patchy atrophy, and macular atrophy were all >0.8 ; and the photograph-level recall/sensitivity for detecting lacquer cracks was 0.9230.

Conclusions: The models could accurately and automatically grade myopic maculopathy, diagnose pathologic myopia, and identify and monitor progression of the lesions.

Translational Relevance: The models can potentially help with the diagnosis, screening, and follow-up for pathologic myopia in clinical practice.

Citation: Tang J, Yuan M, Tian K, Wang Y, Wang D, Yang J, Yang Z, He X, Luo Y, Li Y, Xu J, Li X, Ding D, Ren Y, Chen Y, Sadda SR, Yu W. An artificial-intelligence-based automated grading and lesions segmentation system for myopic maculopathy based on color fundus photographs. *Transl Vis Sci Technol.* 2022;11(6):16, <https://doi.org/10.1167/tvst.11.6.16>

Introduction

Myopia is a serious public health burden, especially in East and Southeast Asia, where the prevalence of myopia among young individuals has reached 80% to 90% in some areas.¹ Due to an earlier onset age and faster rate of myopic progression, the prevalence of high myopia has reached 10% to 20% in young adults,^{1,2} further increasing the risk and prevalence of pathologic myopia (PM) or myopic maculopathy.^{3,4} As early as the beginning of the 21st century, PM had become one of the main causes of blindness and low vision in East Asia.⁵⁻⁷ Thus, PM is now a significant public health concern; consequently, screening, regular follow-up, and monitoring for progression of PM are in critical needs.

With the recent rapid development of artificial intelligence (AI), particularly in deep learning, AI is increasingly being applied in various healthcare disciplines, especially in fields such as ophthalmology where medical image assessment plays a key role.^{8,9} For example, in 2018, the U.S. Food and Drug Administration approved the world's first AI-based automated screening system for detection of referral-warranted diabetic retinopathy (IDx-DR), which demonstrated diagnostic sensitivity and specificity values of 87.2% and 90.7%, respectively.¹⁰ In terms of the automatic diagnosis of PM, there has also been some progress. Liu et al.¹¹ developed a system (PAMELA) to detect PM in color fundus photographs through the detection of peripapillary atrophy. Zhang et al.¹² used multiple kernel learning methods and proposed an AI diagnosis framework (PM-BMII) for the detection of PM. The framework uses a combination of heterogeneous biomedical information including fundus images and demographic, clinical, and genotyping information as its input data.¹² In 2019, the organizer of the International Symposium on Biomedical Imaging issued a challenge to train an AI model on a specified color fundus images dataset in order to (1) classify images as showing PM or non-PM, (2) locate the fovea, and (3) detect and segment the optic disc, patchy atrophy, and retinal detachment. Liu¹³ developed a diagnostic assist

system for the detection of myopic maculopathy that consisted of two levels of networks. The first network level was used to distinguish normal fundus images from myopic maculopathy images, and the second level of network further classified myopic maculopathy images into those with a tessellated fundus and those with atrophic lesions. However, to the best of our knowledge, there is currently no AI model that actually classifies or grades myopic maculopathy and segments the relevant lesions based on the latest and widely recognized grading system proposed by the META-PM (meta-analyses of pathologic myopia) study group in 2015.¹⁴

Therefore, in order to more precisely and automatically assess myopic maculopathy, including the determination of the stage of disease and quantification of the number and area of relevant lesions, our research aimed to develop an AI-based automated classification and lesion segmentation system for myopic maculopathy using color fundus photographs in accordance with META-PM study group diagnostic criteria.

Methods

Participants

This research was approved by the Institutional Review Board of the Peking Union Medical College Hospital (PUMCH), Beijing, China. The research was conducted in accordance with the tenets of the Declaration of Helsinki. As this research involved retrospective medical record reviews with no more than minimal risk to participants, it met all requirements for a waiver of informed consent per institutional policy. This research includes images collected from patients who presented to the outpatient clinic of PUMCH from July 2016 to December 2019 and the outpatient clinic of Beijing Tongren Hospital from January 2011 to December 2018. The inclusion criteria were clinical diagnosis of myopia (spherical equivalent $\leq -0.5D$) and the availability of macula-centered 45° color fundus photographs. Exclusion criteria were previous ocular surgery history, systemic diseases with ocular involvement, and

any concurrent retinal diseases aside from myopic maculopathy, including diabetic retinopathy, age-related macular degeneration, retinal vein occlusion, retinal artery occlusion, central serous chorioretinopathy, and glaucoma. All color fundus photographs were obtained by trained clinical faculty using a Kowa Nonmyd WX-3D retinal camera (Kowa Company Ltd., Nagoya, Japan) with a resolution of 2144×1424 pixels, Topcon TRC-50DX retinal fundus camera (Topcon, Tokyo, Japan) with a resolution of 1383×1182 to 2830×2410 pixels, or a Canon CR-DGi non-mydratic fundus camera (Canon Inc., Tokyo, Japan) with a resolution of 1924×1556 pixels. All identifying information for the patients was discarded.

Definition of Myopic Maculopathy

Myopic maculopathy was defined according to the META-PM study group.¹⁴ This classification system defines five categories of myopic maculopathy: no myopic retinal degenerative lesion (category 0), tessellated fundus (category 1), diffuse chorioretinal atrophy (category 2), extramacular patchy chorioretinal atrophy (or simply “patchy atrophy,” category 3), and macular patchy chorioretinal atrophy (or simply “macular atrophy,” category 4), as well as three kinds of “plus” lesions: lacquer cracks, myopic choroidal neovascularization (CNV), and Fuchs spots. Eyes of category 2 or greater, or with “plus” lesions, are classified as PM.¹⁴

Development of the Models

In this research, four ophthalmologists, who were all ophthalmological residents of our hospital interested in research into retinal diseases, were recruited to grade the images. Each ophthalmologist was randomly assigned 50 images after being trained on a standardized grading protocol, which included typical color fundus photographs of each grade of myopic maculopathy and “plus” lesions. Their grading results were compared with those of an experienced attending retinal ophthalmologist (Z.Y.). Only those who achieved a quadratic-weighted $\kappa \geq 0.75$ for classifying the five grades of myopic maculopathy were allowed to serve as graders for this research. All four ophthalmologists met this criterion after one or two attempts.

Color fundus photographs were randomly assigned to each ophthalmologist after anonymization, and an online grading platform was used to grade the images. Each image was graded by at least two ophthalmologists independently. For any image, if the grading results given by all ophthalmologists were consistent, then the image was included in the high-consistency subgroup. If there were differences in the grading results for an image, then the ophthalmologists partic-

ipating in the grading would discuss the case together in an attempt to adjudicate to a final consensus grading result for the image. If consensus could not be achieved, then the senior ophthalmologist (Z.Y.) would determine the final result. In such a situation, the image was still included in the high-consistency subgroup. However, if the senior ophthalmologist was uncertain and could not determine a final single grade, the results from both primary graders were retained, and the image was included in the low-consistency subgroup. The images in the high-consistency subgroup were stratified randomly and divided into a training set, validation set, or test set according to the grades. The distribution of the images in the training set, validation set, and test set was approximately 60%, 20%, and 20%. Also, considering that the quantity of data might be relatively small, a fivefold cross-validation was conducted to evaluate the generalization ability of the model. In the fivefold cross-validation, the images in the high-consistency subgroup were stratified randomly divided into five folds. Four folds were selected as the training set, and the other one was selected as the test set each time.

In addition to the primary maculopathy grade, the ophthalmologists were also required to locate or delineate the following structures and lesions: (1) fovea, the location of which was defined as the geometric center of it (Supplementary Fig. S1); (2) optic disc (Supplementary Fig. S1); (3) peripapillary atrophy, the outline of which was defined as the outline of its β zone + γ zone (Supplementary Fig. S1); (4) diffuse atrophy (Supplementary Fig. S2), patchy atrophy (Supplementary Fig. S3), and macular atrophy (Supplementary Fig. S4); (5) lacquer cracks (Supplementary Fig. S5); and (6) myopic CNV and Fuchs spots. We did not distinguish between the two lesions in this study, because Fuchs spots are essentially formed by the proliferation of retinal pigment epithelium around CNV, and they are similar in appearance (for an example, see Supplementary Fig. S6). In some cases, peripapillary atrophy might fuse with patchy atrophy or macular atrophy, in which case the boundaries between them were presumed by the ophthalmologists according to the trend of the outlines near the fusion area (Supplementary Fig. S7). To provide an indirect measure of the amount of tilt of the disk, the ovality index (the ratio of the minor axis to the major axis of the optic disc), as previously described,^{15–17} was calculated automatically by the grading platform after segmentation by the grader.

Two basic models were developed: the five-category classification model and the segmentation model. Three convolutional neural networks (CNNs), including ResNet-50, ResNext-50, and WideResNet-50, were compared in the development of the classification

algorithm, which could automatically grade the images. Other CNNs, such as AlexNet¹⁸ and VGG,¹⁹ were proposed earlier, in 2012 and 2014, respectively. Eventually, ResNet-50²⁰ was chosen to be used to develop the classification algorithm according to the comparison results in the pretest (see Table 4). Another CNN, DeepLabv3+,²¹ was used to develop the segmentation algorithm, which could detect and segment the fovea, optic disc, peripapillary atrophy, diffuse atrophy, patchy atrophy, macular atrophy, and lacquer cracks. The two models were both developed based on the training set and validation set and were evaluated on the test set, as well as on the low-consistency subgroup. Data augmentation algorithms included random rotation, random cropping, random inversion, random contrast variation, random brightness variation, and random saturation variation. Images were then downsized to a resolution of 224×224 pixels for the training, validation, and testing of the classification model or to a resolution of 512×512 pixels for the training, validation, and testing of the segmentation model.

All experiments were run with PyTorch. For ResNet-50 pretrained on ImageNet, stochastic gradient descent (SGD) was used for training, with a momentum of 0.9 and a weight decay of 0.0005. The initial learning rate was empirically set to 0.001. The training strategy was that, if the validation set was not improved in three times, then the learning rate would be reduced to 0.1 times the original. The lowest learning rate was set to 0.00001. The loss function used the cross-entropy loss function. In the validation set, if the accuracy was not improved for 10 consecutive times, then the best model would be saved. In each epoch, the batch size was set to 32. For the segmentation model DeepLabV3+, the optimizer used SGD. The learning rate was set to 0.01, the momentum of the optimizer was set to 0.95, and the weight decay was 0.0001. The training strategy was the same as ResNet-50. The loss function used the dice loss function. The batch size was set to 8.

In order to further improve the accuracy of the grading results of the system, we then developed a classification-and-segmentation-based co-decision model. In the co-decision model, a four-category classification model was first developed. Because patchy atrophy and macular atrophy are morphologically similar and only differ in their location (the former does not involve the fovea and the latter involves the fovea), we combined them into one category to reduce the five categories into a four-category classification model. We then amended the results of the four-category classification model based on the results of the segmentation model. The amendment strategies were as follows: (1) If the segmentation model did not detect diffuse atrophy but the result of the four-category classifica-

tion model was diffuse atrophy, then the final result was downgraded to tessellated fundus. (2) If the segmentation model did not detect patchy atrophy or macular atrophy but the result of the four-category classification model was patchy atrophy and macular atrophy, then the final result was downgraded to diffuse atrophy. (3) When the four-category classification model and the segmentation model both detected patchy atrophy and macular atrophy and if the fovea was located within the atrophic lesion, then the final result was classified as macular atrophy. If the fovea was not located within the atrophic lesion, then the final result was graded as patchy atrophy. The accuracy of the co-decision model was also tested on the test set and the low-consistency subgroup.

The co-decision model was transformed into a binary classification model to diagnose pathologic myopia. In the raw results of the four-category classification model, the probabilities of category 0 and category 1 were added together, and the probabilities of category 2 and categories 3 and 4 were added together. (If the result was amended according to amendment strategy 1 or 2, then the probability was set to 1 artificially.) If the sum probability of category 2 and categories 3 and 4 was larger than 0.5, then the image was classified as pathologic myopia.

Statistical Analyses

A confusion matrix was used to visualize and compare the grading results of the ophthalmologists and the classification model. The accuracy and quadratic-weighted κ were used to evaluate the grading accuracy of the algorithms. The precision, recall, F_1 value (the harmonic mean of precision and recall), and intersection over union (IOU) were calculated to evaluate the segmentation accuracy. Moreover, the results of the segmentation model were evaluated at the image level and pixel level. Image level refers to the evaluation of the results according to whether a certain type of lesion/region (for example, diffuse atrophy) of an image was identified by the model. Pixel level refers to the evaluation of the results according to the consistency of the pixels in the delineated lesion/region of the image between the segmentation model and the ophthalmologists. The sensitivity, specificity, Youden index, and area under the receiver operating characteristic curve (AUC) were used to evaluate the diagnostic accuracy for PM classification by the system. For analyses related to the segmentation model, the means and standard deviations (SDs) of three test results were calculated. For other analyses, 95% confidence intervals (CIs) were calculated. SPSS Statistics 26.0 (IBM, Armonk, NY) was used for the statistical analyses.

Results

This study included 1395 color fundus photographs and 895 patients; 1094 images of 604 patients were obtained from PUMCH, and 301 images of 291 patients were obtained from Beijing Tongren Hospital. The high-consistency subgroup included 1203 images, and 192 images were included in the low-consistency subgroup. The detailed characteristics of the high-consistency subgroup and the low-consistency subgroup are shown in Table 1 and Table 2, respectively.

The confusion matrix according to the test results for the basic five-category classification model on the test set is shown in Figure 1. The grading accuracy was 0.9076 (95% CI, 0.8616–0.9399). The quadratic-weighted κ was 0.9324 (95% CI, 0.8444–1.0000). The sensitivity, specificity, and their 95% CIs for each

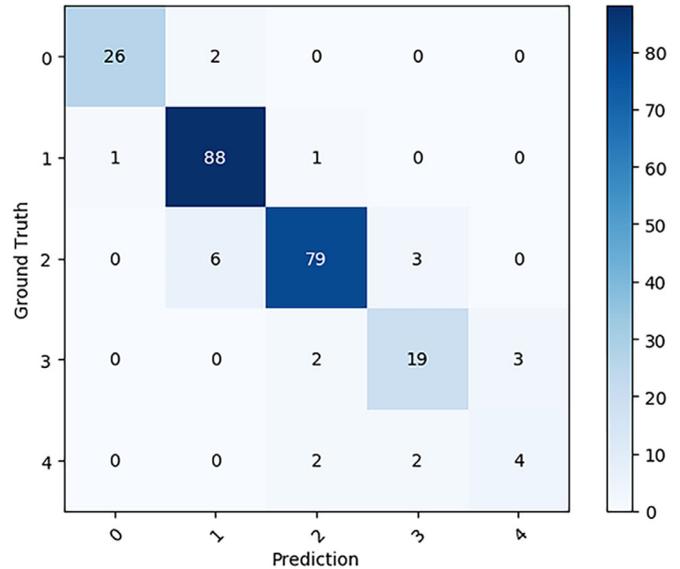


Figure 1. Confusion matrix according to the results of the basic five-category classification model on the test set. Ground truth represents the results of the ophthalmologists, and prediction represents the results of the model.

Table 1. Number of Images for Each Pathologic Myopia Category in the High-Consistency Subgroup

Category	Training Set	Validation Set	Test Set	Total
0	86	28	28	142
1	273	91	90	454
2	267	87	88	442
3	74	24	24	122
4	27	8	8	43
Total	727	238	238	1203

Table 2. Classification of Pathologic Myopia Category in the Low-Consistency Subgroup

Dilemmatic Grades	Number
Categories 0 and 1	80
Categories 1 and 2	80
Categories 2 and 3	21
Categories 3 and 4	11
Total	192

category are reported in Table 3. Beforehand, the three different CNNs were compared in a pretest, and the comparison results are shown in Table 4. The data distribution for the fivefold cross-validation is shown in Supplementary Table S1, and the grading accuracy of the fivefold cross-validation is shown in Supplementary Table S2. The mean accuracy was 0.9119 ± 0.0093 .

The results for the location of the fovea for each grade are shown in Table 5. Notice that in the low-grade images the location of the fovea as determined by the model was very close to the human grader-determined location. As the PM category increased, there were a few images where the model could not locate the fovea (for an example, see Supplementary Fig. S8), and overall the average deviation distance between the model and the human ground truth increased. The results of the segmentation model delineating

Table 3. Sensitivity, Specificity, and 95% CIs for Each Category of the Five-Category Classification Model for the Test Set

	Category				
	0	1	2	3	4
Sensitivity (95% CI)	0.9286 (0.7504–0.9875)	0.9778 (0.9144–0.9961)	0.8977 (0.8101–0.9493)	0.7917 (0.5729–0.9206)	0.5000 (0.1745–0.8255)
Specificity (95% CI)	0.9952 (0.9697–0.9998)	0.9459 (0.8927–0.9746)	0.9667 (0.9199–0.9877)	0.9766 (0.9433–0.9914)	0.9870 (0.9592–0.9966)

Table 4. Comparison Results of ResNet-50, ResNext-50, and WideResNet-50 in the Pretest

Network	ResNet-50	ResNext-50	WideResNet-50
Accuracy (95% CI)	0.9055 (0.8763–0.9341)	0.8992 (0.8664–0.9320)	0.8634 (0.8209–0.9059)
Quadratic-weighted κ (95% CI)	0.9307 (0.8988–0.9626)	0.9235 (0.8893–0.9577)	0.9029 (0.8696–0.9362)

Table 5. Location of Fovea for Each Pathologic Myopia Category

	Category				
	0	1	2	3	4
Number of images with missing location for the fovea/total number of images ± SD	0/28 ± 0.0	0/90 ± 0.0	2.3/88 ± 0.9	3.7/24 ± 0.5	3.0/8 ± 1.4
Average Euclidean distance (in pixels) ^a between automated and manually determined location ± SD	3.02 ± 0.44	3.49 ± 0.07	8.72 ± 0.82	23.28 ± 2.88	26.98 ± 5.03

^aImage size: 512 × 512 pixels.

Table 6. Results of the Segmentation Model

	Pixel Level ± SD				Image Level ± SD		
	Precision	Recall	F ₁	IOU	Precision	Recall	F ₁
Optic disc	0.9272 ± 0.0042	0.9661 ± 0.0033	0.9462 ± 0.0007	0.8979 ± 0.0012	1.0000 ± 0.0	1.0000 ± 0.0	1.0000 ± 0.0
Peripapillary atrophy	0.9029 ± 0.0099	0.8973 ± 0.0038	0.9001 ± 0.0040	0.8184 ± 0.0066	0.9576 ± 0.0047	1.0000 ± 0.0	0.9783 ± 0.0024
Lacquer cracks	0.2912 ± 0.0572	0.2006 ± 0.0031	0.2375 ± 0.0224	0.1370 ± 0.0141	0.4662 ± 0.1106	0.9230 ± 0.1332	0.6156 ± 0.0719
Diffuse atrophy	0.8876 ± 0.0148	0.8738 ± 0.0134	0.8808 ± 0.0141	0.7870 ± 0.0220	0.9600 ± 0.0088	1.0000 ± 0.0	0.9795 ± 0.0046
Patchy atrophy and macular atrophy	0.7598 ± 0.0236	0.8530 ± 0.0137	0.8036 ± 0.0167	0.6717 ± 0.0229	0.9150 ± 0.0301	1.0000 ± 0.0	0.9555 ± 0.0165

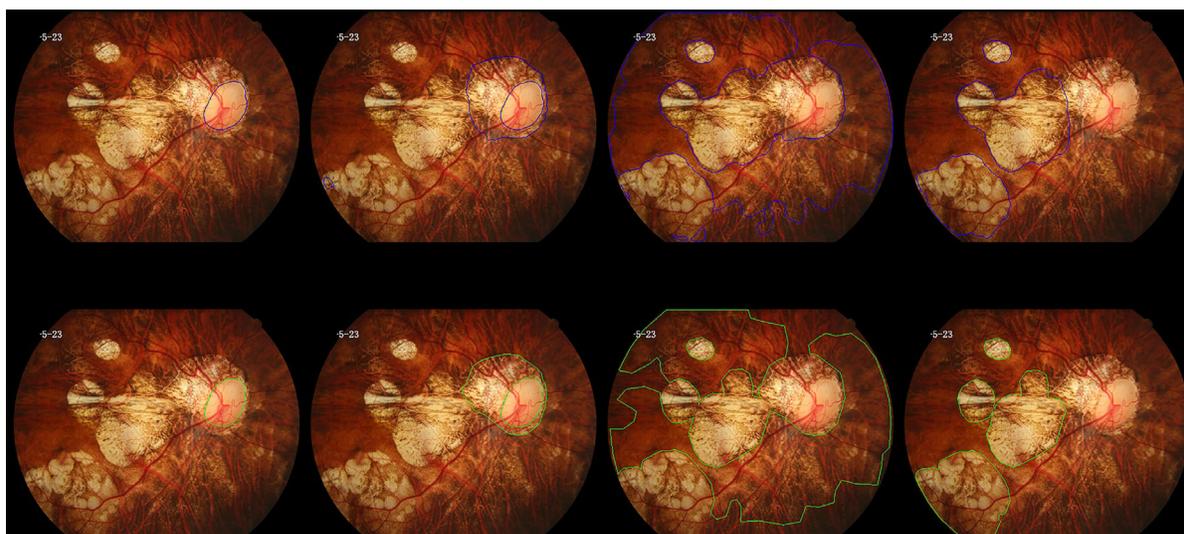


Figure 2. Segmentation results of a sample image. The results of the model are shown in the *upper row*, indicated by the area delineated with the *blue line*, and the manual annotation results are shown in the *bottom row*, with the *green line* delineating the segmented area. *From left to right*: optic disc, peripapillary atrophy, diffuse atrophy, patchy atrophy, and macular atrophy. Note that the model could distinguish the fused macular atrophy and peripapillary atrophy in the image.

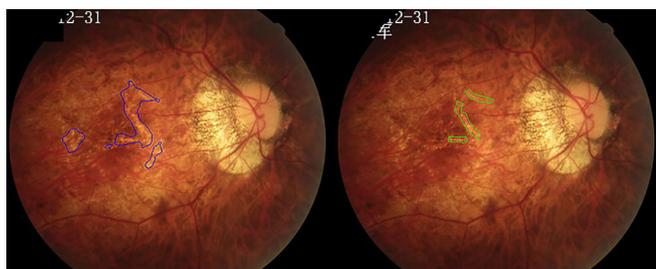


Figure 3. A sample image showing detection and segmentation of lacquer cracks. (*Left*) segmentation model results with the *blue line* delineating the segmentation. (*Right*) Manual annotation results with the *green line* delineating the segmentation.

the optic disc, peripapillary atrophy, lacquer cracks, diffuse atrophy, patchy atrophy, and macular atrophy are shown in Table 6, Figure 2, and Figure 3. Patchy atrophy and macular atrophy are combined and considered as one type of lesion in the analysis based on their similar morphology. Subsequently, these two categories of atrophy were distinguished according to the presence of foveal involvement as previously described. As for CNV and Fuchs spots, the segmentation model could not identify them accurately, with the precision and the recall both being zero, at the pixel level or image level. The average ovality indices of each grade calculated according to the segmentation model

Table 7. Average Ovality Index for Each Pathologic Myopia Category

	Ovality Index by Category				
	0	1	2	3	4
Segmentation model, average ± SD	0.8089 ± 0.0035	0.8043 ± 0.0019	0.7356 ± 0.0018	0.7191 ± 0.0028	0.7797 ± 0.0082
Ophthalmologists, average	0.8124	0.7957	0.7160	0.6979	0.7523

Table 8. Average Area (in Pixels)^a of Peripapillary Atrophy for Each Pathologic Myopia Category

	Area by Category				
	0	1	2	3	4
Segmentation model, average ± SD	1162 ± 79	1923 ± 21	11,235 ± 165	30,736 ± 199	23,505 ± 2137
Ophthalmologists, average	1235	2009	11,227	33,666	22,980

^aImage size: 512 × 512 pixels.

and based on the graders are shown in Table 7. The average areas of peripapillary atrophy for each PM category are shown in Table 8.

The confusion matrix according to the test results of the classification-and-segmentation-based co-decision model from the test set is shown in Figure 4. The grading accuracy was 0.9370 (95% CI, 0.8961–0.9631). The quadratic-weighted κ was 0.9651 (95% CI, 0.9479–0.9824). The receiver operating characteristic (ROC) curve for the diagnosis of PM for the co-decision model is shown in Figure 5. The AUC was 0.9980 (95% CI, 0.9954–1.0000), the sensitivity was 0.9667 (95% CI, 0.9117–0.9893), and the specificity was 0.9915 (95% CI, 0.9468–0.9996). The Youden index was 0.9582.

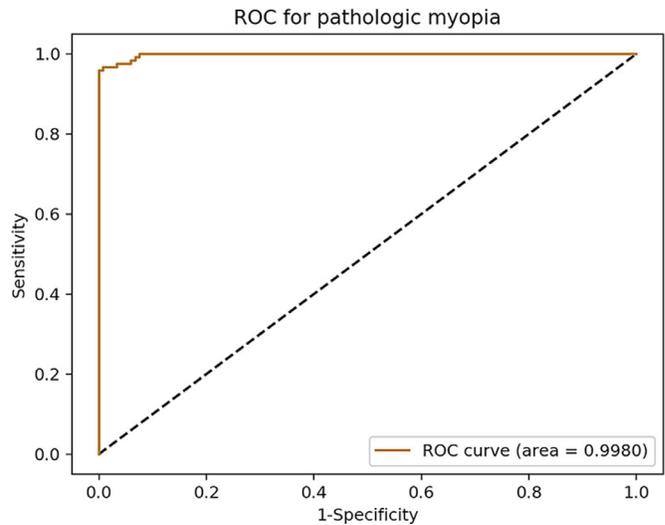


Figure 5. ROC curve for the diagnosis of PM of the co-decision model.

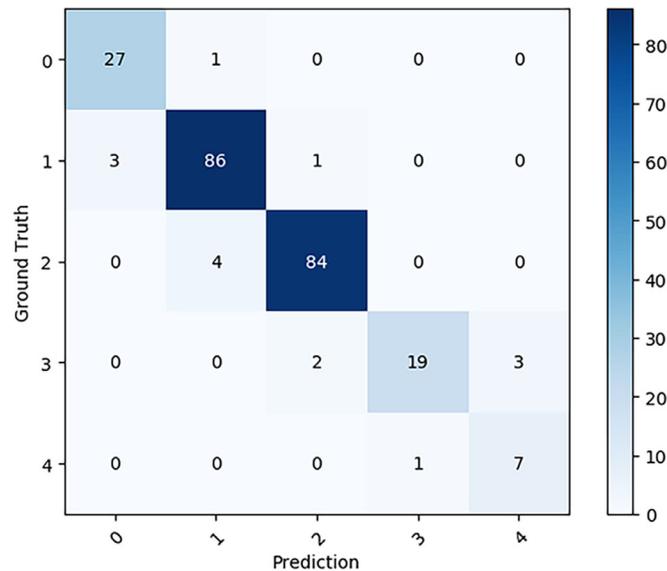


Figure 4. Confusion matrix according to the results of the classification-and-segmentation-based co-decision model on the test set. Ground truth represents the results of the ophthalmologists, and prediction represents the results of the model.

Table 9. Test Results of the Co-Decision Model for the Low-Consistency Subgroup

Dilemmatic Grades	Number	Test Results	Number
Categories 0 and 1	80	Category 0	46
		Category 1	34
Categories 1 and 2	80	Category 1	36
		Category 2	44
Categories 2 and 3	21	Category 2	17
		Category 3	4
Categories 3 and 4	11	Category 3	3
		Category 4	8

The two columns on the left are the grading results given by the ophthalmologists, and the two columns on the right are the grading results of the model.

The test result of the co-decision model for the low-consistency subgroup is shown in Table 9. Notice that, for any image, the grading result generated by the model falls within the range of the discrepant grades given by the ophthalmologists, which implies that the grading accuracy of the model was 1.0000 (95% CI, 0.9756–1.0000) for the low-consistency subgroup.

Discussion

In summary, our study demonstrated good accuracy for the categorization of pathologic myopia and the segmentation of pathologic myopic features using a deep learning model. Models such as the one described in this paper may prove valuable for clinical screening of PM and monitoring the progression of the disease. A significant advance of our project compared to previous studies is that our system not only classifies the myopic maculopathy according to the latest internationally accepted grading system but also segments the myopic lesions, which may further enhance the precision of the monitoring of disease progression. The two networks, ResNet-50 and DeepLabV3+, that we used to develop models have been widely used in other ophthalmological AI research.^{22–24}

The 1395 color fundus photographs used in the present study were divided into a high-consistency subgroup and a low-consistency subgroup according to the consistency among the graders. The images of the low-consistency subgroup accounted for 13.8% of total images. We utilized two subgroups in order to provide a group of images with a high level of confidence to provide a reliable ground truth to train the model while at the same time providing a group of indeterminate cases that reflected the challenging cases and diagnoses that may be presented in real-world clinical practice. Such challenging cases were called questionable according to the META-PM study group. For example, a questionable category 0 or category 1 existed when choroidal vessels could be observed around the arcade vessels but not in the macula.¹⁴ Specifically, when evaluating the reproducibility of the grading system, the META-PM study group used a dataset consisting of 100 myopic maculopathy images, of which 10% were classified as questionable images. Considering that the META-PM dataset did not contain images of category 0, the 13.8% proportion of low-consistency cases in our study would seem to be acceptable.

The use of three different fundus cameras in obtaining fundus photographs could help reduce overfitting. Also, the original resolutions of the three cameras were

different and large, and they required more computing resources. Therefore, the resolutions were downsized to 224×224 pixels for the classification model and 512×512 pixels for the segmentation model. The two sizes were appropriate for completing the network training while occupying fewer computing resources and preserving the features and details of the images to some extent.^{25,26} Data augmentation was used in the training stages to simulate the uncertainty in the image acquisition process by means of giving the augmented image the same label as the original image.

In reviewing the test results, the co-decision model demonstrated a better classification ability than the basic classification model, especially for high-grade/category images. Due to the morphologic similarities between patchy atrophy and macular atrophy, we anticipated that a basic classification model would have difficulty distinguishing between them (Fig. 1). After adding information regarding the location of the fovea with respect to the atrophy based on the segmentation model (amendment strategy 3), the classification ability was significantly improved in the co-decision model. In addition, because of the high recall/sensitivity (1.0000) of the segmentation model in detecting atrophic lesions (Table 6), the co-decision model also amended possible false-positive results that appeared in the four-category classification model (amendment strategy 1 and 2). In terms of diagnosing PM, the co-decision model achieved a good AUC (>0.99), sensitivity (>0.96), and specificity (>0.99). If the model is applied to a screening scenario where the sensitivity must be high, then we could achieve this by adjusting the probability threshold of diagnosis or the position on the ROC curve: for example, for a sensitivity of 1.0000, the specificity was 0.9237. For the low-consistency subgroup, all images were correctly classified by the co-decision model, which means that the classification result of the model for each image fell into one of the two adjacent grades labeled by the ophthalmologists. This indicates that when classifying questionable images that cannot be distinguished with certainty between two dilemmatic grades, the decision-making capability of the co-decision model is similar to that of the ophthalmologists.

The performance of the segmentation model was satisfactory for the detection and segmentation of the optic disc, peripapillary atrophy, and atrophic lesions, for which the image-level F_1 values were all >0.95 . The pixel-level F_1 values for segmentation of the optic disc and peripapillary atrophy were both >0.9 and were all >0.8 for the segmentation of the atrophic lesions. In terms of evaluating the tilting of the optic disc, the average optic disc ovality index for each PM category for the segmentation model was close to that of the

ophthalmologists, as shown in Table 7. In addition, the average area of the peripapillary atrophy for each PM category for the segmentation model was also close to that of the ophthalmologists, as shown in Table 8. Increasing optic disc tilt and peripapillary atrophy area may be a reflection of increasing severity of progression of myopia. Metrics such as these may allow the progression of PM to be monitored in a quantitative fashion.

Among the three plus lesions, Xu et al.²⁷ reported a detection rate of lacquer cracks on color fundus photographs of 98%. Although the lacquer crack itself has little effect on the vision, it often progresses to patchy atrophy or myopic CNV in the future.^{28,29} Therefore, it is important to detect lacquer cracks on color fundus photographs for risk prognostication. In this study, the image-level recall/sensitivity for the segmentation model to detect lacquer cracks was 0.9230, indicating that the model had a good ability to detect lacquer cracks. On the other hand, because lacquer cracks are linear lesions with an inhomogeneous distribution, it was relatively difficult for human graders to delineate their outlines accurately in the images, thus influencing the effectiveness of the training of the model, which could explain the low pixel-level results of the model (Fig. 3).

Myopic CNV is one of the most serious complications of PM and often leads to a sudden onset of decline in central vision.³⁰ Our segmentation model, however, could not accurately detect CNV or Fuchs spots, which might be related to the small number of CNV and Fuchs spots included in the dataset (23 images). On the other hand, although myopic CNV can appear as a gray-green membrane in color fundus images, imaging methods such as fluorescein fundus angiography and optical coherence tomography (OCT) are still necessary to confirm the diagnosis in clinical practice.^{31–33} New imaging methods such as OCT angiography are also very useful in the diagnosis of myopic CNV.^{34,35} Diagnosis through color fundus images alone often leads to misdiagnoses or missed diagnoses in clinical practice.³¹ Therefore, it may be more feasible to use other imaging information such as OCT images as training sources for such classifications.³⁶

This research has several limitations. First, although the co-decision model has greatly improved the ability to classify the high-category images, the quantity of the data, especially the high-category images and plus lesions in this study, was relatively small, which might lead to overfitting and reduce the generalization ability of the model. Therefore, we conducted a fivefold cross-validation to evaluate the generalization ability of the model. As shown

in Supplementary Table S2, the five test results of the fivefold cross-validation fluctuated within a small range, which proved that the model did not overfit in the current dataset to some extent. In a recent study that included 226,686 color fundus images, an AUC of 0.969 for the diagnosis of pathologic myopia was obtained.³⁷ Second, there was not an external test dataset in this research. However, as previously mentioned, one of the advantages of our research was that we developed a segmentation model that could help improve the classification accuracy of the system. For this reason, although it would be better to include an external test dataset, it was difficult to find an external dataset for which the images were delineated in the same detail as ours to test the segmentation model or the co-decision model. Third, we did not collect additional data such as the axial length of each patient or other imaging data such as OCT. If OCT had been included to quantitatively evaluate optic disc tilt or to measure the axial length, then the accuracy of the model would likely have been further improved. Fourth, for the peripapillary atrophy, we delineated only the outline of the β zone + γ zone and did not annotate each zone distinctly.³⁸ For the stage of tessellated fundus, the model lacks effective indicators to describe its progress. Yan et al.³⁹ proposed a method for further subclassifying a tessellated fundus which could be incorporated into future models.

In conclusion, deep learning models based on color fundus photographs could accurately and automatically diagnose and classify myopic maculopathy and quantify peripapillary atrophy, atrophic lesions, lacquer cracks, and optic disk tilting. Automated tools such as these may prove to be useful in screening programs for pathologic myopic and in monitoring disease progression over time.

Acknowledgments

Supported by the Chinese Academy of Medical Sciences Initiative for Innovative Medicine (CAMS-I2M, 2018-I2M-AI 001); the pharmaceutical collaborative innovation project of Beijing Science and Technology Commission (Z191100007719002); National Key Research and Development Project (SQ2018YFC200148); National Natural Science Foundation of China (81670879); and by the Beijing Natural Science Foundation (4202033). The sponsors or funding organizations had no role in the design or conduct of this research.

Discosure: **J. Tang**, None; **M. Yuan**, None; **K. Tian**, None; **Y. Wang**, None; **D. Wang**, None; **J. Yang**, None; **Z. Yang**, None; **X. He**, None; **Y. Luo**, None; **Y. Li**, None; **J. Xu**, None; **X. Li**, None; **D. Ding**, None; **Y. Ren**, None; **Y. Chen**, None; **S.R. Sadda**, None; **W. Yu**, None

* JT, MY, KT, YW and DW contributed equally to this work.

References

- Morgan IG, French AN, Ashby RS, et al. The epidemics of myopia: aetiology and prevention. *Prog Retin Eye Res.* 2018;62:134–149.
- Sun J, Zhou J, Zhao P, et al. High prevalence of myopia and high myopia in 5060 Chinese university students in Shanghai. *Invest Ophthalmol Vis Sci.* 2012;53(12):7504–7509.
- Verkicharla PK, Ohno-Matsui K, Saw SM. Current and predicted demographics of high myopia and an update of its associated pathological changes. *Ophthalmic Physiol Opt.* 2015;35(5):465–475.
- Wong Y-L, Saw S-M. Epidemiology of pathologic myopia in Asia and worldwide. *Asia Pac J Ophthalmol.* 2016;5(6):394–402.
- Hsu W-M, Cheng C-Y, Liu J-H, Tsai S-Y, Chou P. Prevalence and causes of visual impairment in an elderly Chinese population in Taiwan: the Shihpai Eye Study. *Ophthalmology.* 2004;111(1):62–69.
- Iwase A, Araie M, Tomidokoro A, Yamamoto T, Shimizu H, Kitazawa Y. Prevalence and causes of low vision and blindness in a Japanese adult population: the Tajimi Study. *Ophthalmology.* 2006;113(8):1354–1362.e1.
- Xu L, Wang Y, Li Y, et al. Causes of blindness and visual impairment in urban and rural areas in Beijing: the Beijing Eye Study. *Ophthalmology.* 2006;113(7):1134.e1–1134.e11.
- Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunović H. Artificial intelligence in retina. *Prog Retin Eye Res.* 2018;67:1–29.
- Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol.* 2019;103(2):167–175.
- Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med.* 2018;1:39.
- Liu J, Wong DWK, Lim JH, et al. Detection of pathological myopia by PAMELA with texture-based features through an SVM approach. *J Healthc Eng.* 2010;1(1):1–12.
- Zhang Z, Xu Y, Liu J, et al. Automatic diagnosis of pathological myopia from heterogeneous biomedical data. *PLoS One.* 2013;8(6):e65736.
- Liu L. Study on the *Automated Detection of Retinal Diseases Based on Deep Neural Networks*. Hefei, China: Information and Communication Engineering, University of Science and Technology of China; 2019. Thesis.
- Ohno-Matsui K, Kawasaki R, Jonas JB, et al. International photographic classification and grading system for myopic maculopathy. *Am J Ophthalmol.* 2015;159(5):877–883.e7.
- Tay E, Seah SK, Chan S-P, et al. Optic disk ovality as an index of tilt and its relationship to myopia and perimetry. *Am J Ophthalmol.* 2005;139(2):247–252.
- How ACS, Tan GSW, Chan Y-H, et al. Population prevalence of tilted and torted optic discs among an adult Chinese population in Singapore: the Tanjong Pagar Study. *Arch Ophthalmol.* 2009;127(7):894–899.
- Shin H-Y, Park H-YL, Park CK. The effect of myopic optic disc tilt on measurement of spectral-domain optical coherence tomography parameters. *Br J Ophthalmol.* 2015;99(1):69–74.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* 2012;1:1097–1105
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 2014, <https://doi.org/10.48550/arXiv.1409.1556>.
- Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. arXiv. 2017, <https://doi.org/10.48550/arXiv.1611.05431>.
- Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv. 2018, <https://doi.org/10.48550/arXiv.1802.02611>.
- Terry L, Trikha S, Bhatia KK, Graham MS, Wood A. Evaluation of automated multiclass fluid segmentation in optical coherence tomography images using the Pegasus fluid segmentation algorithms. *Transl Vis Sci Technol.* 2021;10(1):27.
- Kim KM, Heo TY, Kim A, et al. Development of a fundus image-based deep learning diagnos-

- tic tool for various retinal diseases. *J Pers Med*. 2021;11(5):321.
24. Hwang DD, Choi S, Ko J, et al. Distinguishing retinal angiomatous proliferation from polypoidal choroidal vasculopathy with a deep neural network based on optical coherence tomography. *Sci Rep*. 2021;11(1):9275.
 25. Pham Q, Ahn S, Song SJ, et al. Automatic drusen segmentation for age-related macular degeneration in fundus images using deep learning. *Electronics*. 2020;9(10):1617.
 26. Sreng S, Maneerat N, Hamamoto K, et al. Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images. *Appl Sci*. 2020;10(14):4916.
 27. Xu X, Fang Y, Uramoto K, et al. Clinical features of lacquer cracks in eyes with pathologic myopia. *Retina*. 2019;39(7):1265–1277.
 28. Hayashi K, Ohno-Matsui K, Shimada N, et al. Long-term pattern of progression of myopic maculopathy: a natural history study. *Ophthalmology*. 2010;117(8):1595–1611.e4.
 29. Ohno-Matsui K, Tokoro T. The progression of lacquer cracks in pathologic myopia. *Retina*. 1996;16(1):29–37.
 30. Cheung CMG, Arnold JJ, Holz FG, et al. Myopic choroidal neovascularization: review, guidance, and consensus statement on management. *Ophthalmology*. 2017;124(11):1690–1711.
 31. Ohno-Matsui K, Ikuno Y, Lai TYY, Gemmy Cheung CM. Diagnosis and treatment guideline for myopic choroidal neovascularization due to pathologic myopia. *Prog Retin Eye Res*. 2018;63:92–106.
 32. Ng DSC, Cheung CYL, Luk FO, et al. Advances of optical coherence tomography in myopia and pathologic myopia. *Eye (Lond)*. 2016;30(7):901–916.
 33. Battaglia Parodi M, Iacono P, Romano F, Bاندello F. Fluorescein leakage and optical coherence tomography features of choroidal neovascularization secondary to pathologic myopia. *Invest Ophthalmol Vis Sci*. 2018;59(7):3175–3180.
 34. Querques L, Giuffrè C, Corvi F, et al. Optical coherence tomography angiography of myopic choroidal neovascularisation. *Br J Ophthalmol*. 2017;101(5):609–615.
 35. Bruyère E, Miere A, Cohen SY, et al. Neovascularization secondary to high myopia imaged by optical coherence tomography angiography. *Retina*. 2017;37(11):2095–2101.
 36. Sogawa T, Tabuchi H, Nagasato D, et al. Accuracy of a deep convolutional neural network in the detection of myopic macular diseases using swept-source optical coherence tomography. *PLoS One*. 2020;15(4):e0227240.
 37. Tan TE, Anees A, Chen C, et al. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. *Lancet Digit Health*. 2021;3(5):e317–e329.
 38. Jonas JB, Wang YX, Zhang Q, et al. Parapapillary gamma zone and axial elongation-associated optic disc rotation: the Beijing Eye Study. *Invest Ophthalmol Vis Sci*. 2016;57(2):396–402.
 39. Yan YN, Wang YX, Xu L, Xu J, Wei WB, Jonas Jost B. Fundus tessellation: prevalence and associated factors: the Beijing Eye Study 2011. *Ophthalmology*. 2015;122(9):1873–1880.