JAMA Ophthalmology | Original Investigation

Automated Machine Learning for Predicting Diabetic Retinopathy Progression From Ultra-Widefield Retinal Images

Paolo S. Silva, MD; Dean Zhang; Cris Martin P. Jacoba, MD; Ward Fickweiler, MD, BM; Drew Lewis, BS; Jeremy Leitmeyer, BS; Katie Curran, PhD; Recivall P. Salongcay, MD, MPM; Duy Doan, BS; Mohamed Ashraf, MD, PhD; Jerry D. Cavallerano, OD, PhD; Jennifer K. Sun, MD, MPH; Tunde Peto, MD, PhD; Lloyd Paul Aiello, MD, PhD

IMPORTANCE Machine learning (ML) algorithms have the potential to identify eyes with early diabetic retinopathy (DR) at increased risk for disease progression.

OBJECTIVE To create and validate automated ML models (autoML) for DR progression from ultra-widefield (UWF) retinal images.

DESIGN, SETTING AND PARTICIPANTS Deidentified UWF images with mild or moderate nonproliferative DR (NPDR) with 3 years of longitudinal follow-up retinal imaging or evidence of progression within 3 years were used to develop automated ML models for predicting DR progression in UWF images. All images were collected from a tertiary diabetes-specific medical center retinal image dataset. Data were collected from July to September 2022.

EXPOSURE Automated ML models were generated from baseline on-axis 200° UWF retinal images. Baseline retinal images were labeled for progression based on centralized reading center evaluation of baseline and follow-up images according to the clinical Early Treatment Diabetic Retinopathy Study severity scale. Images for model development were split 8-1-1 for training, optimization, and testing to detect 1 or more steps of DR progression. Validation was performed using a 328-image set from the same patient population not used in model development.

MAIN OUTCOMES AND MEASURES Area under the precision-recall curve (AUPRC), sensitivity, specificity, and accuracy.

RESULTS A total of 1179 deidentified UWF images with mild (380 [32.2%]) or moderate (799 [67.8%]) NPDR were included. DR progression was present in half of the training set (590 of 1179 [50.0%]). The model's AUPRC was 0.717 for baseline mild NPDR and 0.863 for moderate NPDR. On the validation set for eyes with mild NPDR, sensitivity was 0.72 (95% CI, 0.57-0.83), specificity was 0.63 (95% CI, 0.57-0.69), prevalence was 0.15 (95% CI, 0.12-0.20), and accuracy was 64.3%; for eyes with moderate NPDR, sensitivity was 0.80 (95% CI, 0.70-0.87), specificity was 0.72 (95% CI, 0.66-0.76), prevalence was 0.22 (95% CI, 0.19-0.27), and accuracy was 73.8%. In the validation set, 6 of 9 eyes (75%) with mild NPDR and 35 of 41 eyes (85%) with moderate NPDR progressed 2 steps or more were identified. All 4 eyes with mild NPDR that progressed within 6 months and 1 year were identified, and 8 of 9 (89%) and 17 of 20 (85%) with moderate NPDR that progressed within 6 months and 1 year, respectively, were identified.

CONCLUSIONS AND RELEVANCE This study demonstrates the accuracy and feasibility of automated ML models for identifying DR progression developed using UWF images, especially for prediction of 2-step or greater DR progression within 1 year. Potentially, the use of ML algorithms may refine the risk of disease progression and identify those at highest short-term risk, thus reducing costs and improving vision-related outcomes.

+ Invited Commentary

+ Multimedia

+ Supplemental content

Author Affiliations: Beetham Eye Institute, Joslin Diabetes Center, Boston, Massachusetts (Silva, Zhang, Jacoba, Fickweiler, Doan, Ashraf, Cavallerano, Sun, Aiello); Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts (Silva, Jacoba, Fickweiler, Ashraf, Cavallerano, Sun, Aiello); Estenda Solutions, Conshohocken, Pennsylvania (Lewis, Leitmeyer); Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom (Curran, Salongcay, Peto).

Corresponding Author: Paolo S. Silva, MD, Beetham Eye Institute, Joslin Diabetes Center, One Joslin Place, Boston, MA 02215 (paoloantonio.silva@joslin.harvard. edu).

E stimating the risk of diabetic retinopathy (DR) progression is one of the most important and challenging tasks clinicians face when caring for individuals with diabetic eye disease.[1] The current DR severity scales are based on the presence and extent of defined retinal features and inform clinicians regarding the risk of progression, appropriate follow-up intervals, and need for treatment.[2] However, especially with early DR severity levels, estimated progression risks are relatively broad, and each specific severity level may not represent a homogenous group with similar long-term risks. An estimation of progression risk is fundamental for clinical care decisions. The use of artificial intelligence (AI) algorithms may improve this process by providing additional information.[3] AI deep-learning systems have been shown to predict DR development or progression using color fundus photographs, and these systems were independent of, and more informative than, other available risk factors.[4,5]

Automated machine learning allows the development of code-free deep-learning models at minimal cost, decreasing the barriers to using AI for addressing relevant clinical needs.[6] The customization and accuracy of automated ML allows the development of AI models based on specific clinical requirements and has been useful in various medical fields, including ophthalmology.[6]

The vast majority of AI applications for DR have been based on standard 30° to 60° retinal images that focus on the posterior 30% of the retinal surface. However, the use of UWF imaging has demonstrated that retinal lesions can appear or develop within the peripheral retinal area not captured by standard retinal imaging and that these findings alter disease risk.[7,8] Although there has been increasing use of UWF imaging in both clinical and DR screening programs, to our knowledge, none of the current published predictive AI models have been used on UWF images.

The purpose of this study is to prospectively create and validate a code-free automated ML model for predicting DR progression from UWF retinal images. The development of this type of risk stratification tool would help to optimize examination intervals and thus potentially reduce costs and improve vision-related outcomes in diabetes eye care.

Methods

This was a prospective development and validation study of automated ML models for predicting DR progression based on existing UWF retinal images. All images were taken with the California retinal imager (Optos). Deidentified retinal images with diabetes were used in the development of the automated ML models. Additional unique and deidentified images with mild NPDR and moderate NPDR from the same population that were not used to train, optimize, or test the algorithm were included in a validation set. Only 1 image for each eye was used in the set. Eyes with progression (progressors) had matched longitudinal retinal images taken 0.2 to 4.9 years later (mean [SD] time, 2.2 [1.1] years). Eyes without progression (nonprogressors) had matched longitudinal retinal images taken from 2.5 to 5.3 years (mean [SD] time, 3.1 [0.5] years). For the entire study set (progressors and nonprogressors), the time for follow-up was 0.2 to 5.3 years (mean [SD]

Key Points

Question Can automated machine learning models using ultra-widefield retinal images predict diabetic retinopathy (DR) progression?

Findings In this diagnostic study including 1179 deidentified ultra-widefield retinal images, the performance of the algorithms to identify DR progression matched or exceeded previously published performance of bespoke artificial intelligence models. In the validation set, 77.5% of eyes with mild nonproliferative DR (NPDR) and 85.4% of eyes with moderate NPDR that progressed 2 steps or more were identified by the model, and all eyes with mild NPDR and 85.0% of eyes with moderate NPDR that progressed within 1 year were identified.

Meaning In the future, automated machine learning models may refine the risk of disease progression, inform personalized screening intervals, and improve vision-related outcomes.

time, 2.7 [1.0] years). All images were taken from January 2016 to June 2022 at the Beetham Eye Institute of the Joslin Diabetes Center, Boston, Massachusetts. The study design complied with the ethical standards stated in the 1964 Declaration of Helsinki, and the study protocol was approved by the institutional review board of Joslin Diabetes Center, which waived the need to obtain informed consent from the patients. The reporting of study data are in accordance with the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline.

Assessment of DR Severity and Progression

Previously acquired 200° UWF retinal images were identified using *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-10*) billing diagnosis codes from the electronic medical records. All eyes with mild or moderate NPDR with at least 2 visits and photographically documented progression within 4 years by *ICD-10* codes were retrieved. Eyes with a prior history of intravitreal injections were excluded from the study. Eyes that received panretinal photocoagulation were graded to have proliferative DR (Early Treatment Diabetic Retinopathy Study [ETDRS] level 60). Evidence of macular laser scars was not used in the assessment of DR severity. Nonprogressors with at least 2 visits and photographically documented 3-year nonprogression were matched based on age and diabetes duration and selected by DR severity using *ICD-10* codes. All images were deidentified and stereographically projected and then assessed at a centralized reading center using high-resolution, high-definition LCD computer displays. Standardized grading was performed independently by certified graders using the clinical ETDRS severity scale (C. M. P. J. and W. F.). Mean (SD) weighted κ for graders at the reading center was 0.78 (0.3), showing substantial agreement. The clinical ETDRS severity of DR (no DR, mild nonproliferative DR [NPDR], moderate NPDR, severe NPDR, or proliferative DR [PDR]) was assessed from the entire 200° UWF image.[9] Based on the reading center assessment of DR, only eyes with mild or moderate NPDR at baseline were included in the training and validation datasets. Mild NPDR was defined as ETDRS levels 20 to 35. Moderate NPDR was defined as ETDRS levels 43 to 47. An image was considered ungradable if there was inadequate photo-

Table 1. Demographic Characteristics and Nonproliferative Diabetic Retinopathy (NPDR) Distribution of the Training Set

Characteristic	Mild NPDR (n = 380)		Moderate NPDR (n = 799 eyes)	
	Progressor (n = 190 eyes)[a]	Nonprogressor (n = 190 eyes)	Progressor (n = 400 eyes)[a]	Nonprogressor (n = 399 eyes)
Age, mean (SD), y	47.4 (16.0)	51.4 (15.0)	57.5 (15.9)	59.9 (13.4)
Self-reported sex, No. (%)				
Female	126 (66.3)	130 (68.4)	202 (50.5)	190 (47.6)
Male	64 (33.7)	60 (31.6)	198 (49.5)	209 (52.4)
Diabetes duration, mean (SD), y	28.4 (13.8)	30.7 (13.4)	26.3 (10.2)	27.9 (10.9)

[a] Progressors were defined as eyes with progression of 1 or more steps on the clinical Early Treatment Diabetic Retinopathy Study severity scale over 3 years.

graphic quality or if media opacity made it impossible to determine whether DR lesions were present in at least 50% of the image or if the disc and macula were not visible. All ungradable images were excluded from the training data and validation data.

To assess DR progression, both baseline and follow-up retinal images were reviewed side by side after assessing DR severity. Progression for this study was defined as 1 or more steps of worsening on the clinical ETDRS severity scale. Images were classified as progressors when DR progression was present and a snonprogressors when DR progression was not observed. Uncertain or questionable DR severity grades were adjudicated by a senior retina specialist (P. S. S.), and the adjudicated grade was considered the final assessment.

Automated ML

All baseline images were labeled as either nonprogressors or progressors, as defined above. All baseline gradable labeled images were uploaded to the Google AutoML platform (Google) to build the model using the recommended node hours, which were then automatically split into training (80%), optimization (10%), and testing (10%) sets. The AutoML platform automatically preprocesses the data while identifying the best network architecture and hyperparameters to detect patterns in the training subset, while using the optimization set to fine-tune the algorithm further. The testing set was used to evaluate model performance. Validation was performed using a previously adjudicated and curated dataset of 328 images (including 50 [15.2%] progressors) with mild NPDR and 425 images (including 95 [22.4%] progressors) with moderate NPDR from the same population that were not used to train, optimize, or test the algorithm.

Statistical Analysis

The automated ML platform natively provides detailed statistics for model performance based on the testing set. This AI model is a binary classifier that predicts categories based on a generated probability it assigns to each image. Confidence thresholds, also called decision thresholds, are the minimum confidence levels the algorithm should have to classify an image under one category based on the assigned probability. The area under the precision-recall curve (AUPRC), a measurement of the mean accuracy of the model, is generated by evaluating the testing set at various confidence levels from 0.0 to 1.0. Every point on the curve shows a pair of precision-recall values at different confidence thresholds. Modifiable confidence thresholds are

available, allowing different levels of precision and recall at varying confidence levels, as shown in the precision-recall curve. Confusion matrices are also generated, which allow users to see true positives, true negatives, false positives, and false negatives. Sensitivity, specificity, positive predictive value, and negative predictive value for the internal and external validation sets were also computed with confidence intervals.

The required training sample size for ML models applied to medical imaging data are not completely defined.[10] Based on prior work, we estimated that the minimum number of images needed to train the automated ML model was at least 1000.[11] Furthermore, the ratio of progressors (589) to nonprogressors (590) was 1:1, which is the ideal ratio for database creation of ML models.[11] Statistical analyses not provided in automated ML were performed using SAS software version 9.4 (SAS Institute).
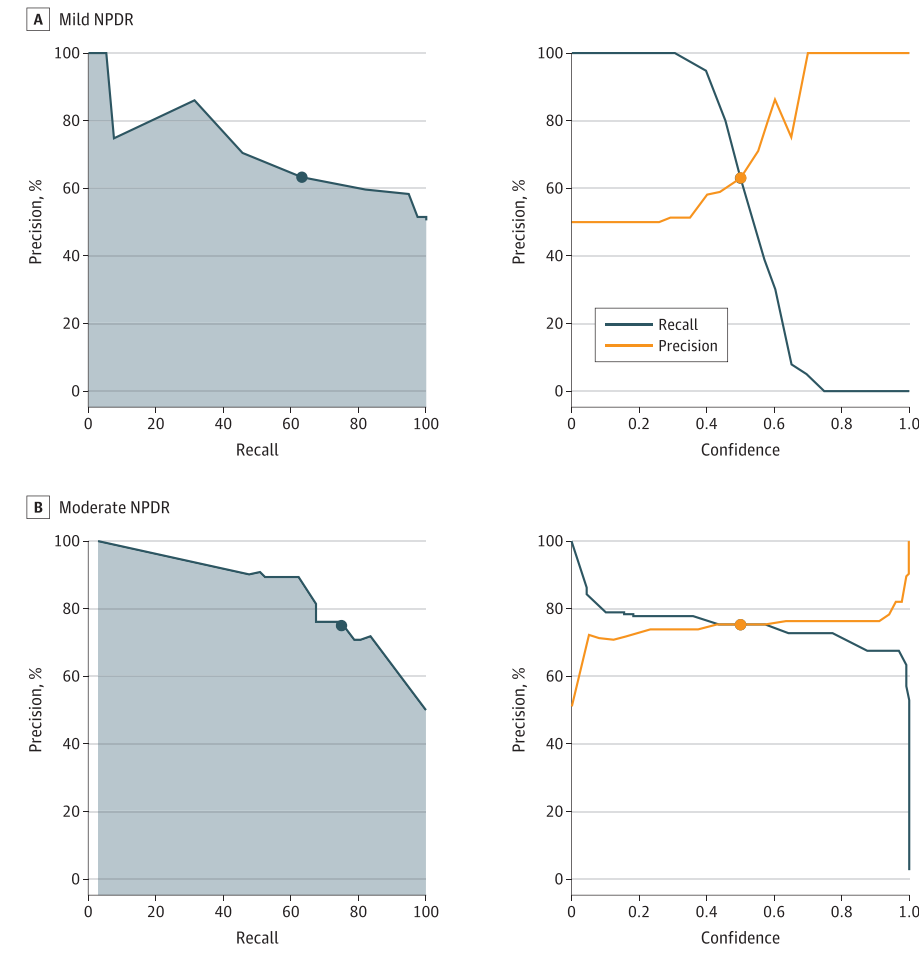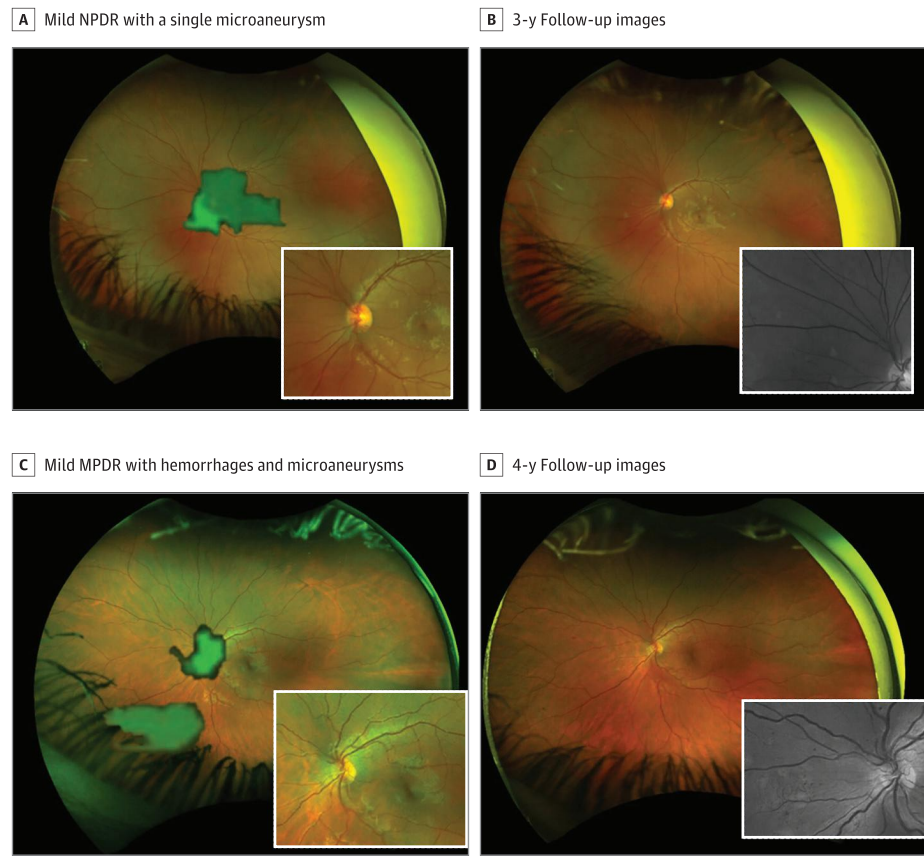
Creation of Saliency Maps

Google cloud supports 2 different saliency map techniques: XRAI and Integrated Gradients. The latter was chosen for detailed pixel-based attributions as opposed to the larger region-based alternative. Based on our best automated ML models, we generated an AI model using the tensorflow framework and Python to best visualize the attention maps. Only attributions with a score of more than 95% significance were used. The maps were created for images from the validation set and analyzed for significant findings.

Results

A total of 1179 deidentified UWF images with mild (380 [32.2%]) or moderate (799 [67.8%]) NPDR were included. In the training sets, progression was present in 589 eyes (50.0%), and the demographic distribution is presented in **Table 1**. The AUPRC of the model for mild NPDR was 0.717 with a precision and recall of 53.16% (**Figure 1**A), and the AUPRC for moderate NPDR was 0.863 with a precision and recall of 75% (Figure 1B). The confusion matrix generated from the testing set of 38 eyes with mild NPDR showed 12 true positives (63%), 7 false negatives (37%), 12 true negatives (63%), and 7 false positives (37%). For 80 eyes with moderate NPDR, there were 30 true positives (75%), 10 false negatives (25%), 10 true negatives (75%), and 30 false positives (25%). Performance on the testing set for eyes with mild NPDR was 0.63 (95% CI, 0.39-0.83) for sensitivity, 0.63 (95% CI, 0.39-0.83) for specificity, 0.63 (95% CI, 0.39-

Figure 1. Precision-Recall Curves for Eyes With Mild and Moderate Nonproliferative Diabetic Retinopathy (NPDR)

The data point indicates the 0.50 score threshold.

Table 2. Demographic Characteristics and Nonproliferative Diabetic Retinopathy (NPDR) Distribution of the Validation Set

Characteristic	Mild NPDR (n = 328 eyes)		Moderate NPDR (n = 425 eyes)	
	Progressor (n = 50 eyes)[a]	Nonprogressor (n = 278 eyes)	Progressor (n = 95 eyes)[a]	Nonprogressor (n = 330 eyes)
Age, mean (SD), y	49.4 (17.1)	49.7 (15.7)	58.4 (17.8)	60.8 (13.2)
Self-reported sex, No. (%)				
Female	30 (60)	175 (62.9)	41 (43)	150 (45.5)
Male	20 (40)	103 (37.1)	54 (57)	180 (54.5)
Diabetes duration, mean (SD), y	28.4 (13.2)	29.1 (13.7)	26.0 (9.5)	26.7 (11.1)

[a] Progressors were defined as eyes with progression of 1 or more steps on the clinical Early Treatment Diabetic Retinopathy Study severity scale over 3 years.

0.83) for positive predictive value, 0.63 (95% CI, 0.39-0.83) for negative predictive value, 0.50 (95% CI, 0.37-0.66) for prevalence, and 63.2% for accuracy. Performance for eyes with moderate NPDR was 0.75 (95% CI, 0.58-0.87) for sensitivity, 0.75 (95% CI, 0.58-0.87) for specificity, 0.75 (95% CI, 0.58-0.87) for positive predictive value, 0.75 (95% CI, 0.58-0.87) for negative predictive value, 0.50 (95% CI, 0.39-0.61) for prevalence, and 75% for accuracy.

For the validation sets, 328 images with mild NPDR (including 50 [15.2%] progressors) and 425 images with moder-

ate NPDR (including 95 [22.3%] progressors) were used. The demographic characteristics of the validation sets are presented in Table 2. The reported rates of progression in eyes with mild and moderate NPDR from clinical trials were used to determine the percentage of progressors for the mild and moderate validation sets.[2] The mean (SD) time to progression was 2.3 (0.9) years in eyes with mild NPDR and 1.8 (1.0) years in eyes with moderate NPDR. Mean (SD) follow-up for nonprogressors was 3.1 (0.5) years in eyes with mild NPDR and 3.2 (0.5) years in eyes with moderate NPDR.

Figure 2. Saliency Heat Maps of Eyes With Nonproliferative Diabetic Retinopathy (NPDR) Identified as Progressors

A Mild NPDR with a single microaneurysm

B 3-y Follow-up images

C Mild MPDR with hemorrhages and microaneurysms

D 4-y Follow-up images

Ultra-widefield color images and heat maps. A and B, Progression from mild NPDR to moderate NPDR. A, Mild NPDR with a single microaneurysm identified in field 2. The overlay shows heat maps generated by the automated machine learning algorithm focused on the optic disc. The inset shows the optic disc and vessels in the area identified by the heat map showing dilated retinal vessels. B, Three-year follow-up images showing the development of cotton-wool spots and intraretinal microvascular abnormalities less than standard 8A. C and D, Progression from mild NPDR to PDR. C, Mild NPDR with hemorrhages and microaneurysms less than standard 1A and a single microaneurysm in the nasal peripheral field 7. The overlay shows heat maps generated by the automated ML algorithm with focus on the optic disc. The inset shows the optic disc and vessels in the area identified by the heat map showing dilated retinal vessels. D, Four-year follow-up images showing development of new vessels on the disc and intraretinal microvascular abnormalities greater than 8A. The inset shows the magnified red-free images of optic disc and adjacent areas with new vessels and intraretinal microvascular abnormalities.

Performance for 1-step DR progression on the validation set for eyes with mild NPDR was 0.72 (95% CI, 0.57-0.83) for sensitivity, 0.63 (95% CI, 0.57-0.69) for specificity, 0.26 (95% CI, 0.19-0.34) for positive predictive value, 0.92 (95% CI, 0.88-0.96) for negative predictive value, 0.15 (95% CI, 0.12-0.20) for prevalence, and 64.3% for accuracy. Performance for eyes with moderate NPDR was 0.80 (95% CI, 0.70-0.87) for sensitivity, 0.72 (95% CI, 0.66-0.76) for specificity, 0.45 (95% CI, 0.37-0.53) for positive predictive value, 0.92 (95% CI, 0.88-0.95) for negative predictive value, 0.22 (95% CI, 0.19-0.27) for prevalence, and 73.8% for accuracy. In the validation set for eyes with mild NPDR, 6 of 9 eyes (75%) that progressed 2 or more steps were identified. In eyes with mild NPDR, all 4 eyes (100%) with mild NPDR that progressed within 6 months and 1 year were identified. For moderate NPDR, 35 of 41 eyes (85%) that progressed 2 or more steps were identified. In eyes with moderate NPDR at baseline, 8 of 9 (89%) and 17 of 20 (85%) that progressed within 6 months and 1 year, respectively, were identified. Automated ML saliency heat maps of eyes progressing from mild to more severe NPDR are shown in **Figure 2**.

Discussion

This study evaluated the accuracy of automated ML generated AI algorithms to identify DR progression based on UWF im-ages. The model was developed and tested on standardized UWF retinal images acquired from a tertiary diabetes-specific academic medical center. Although there are no established standards for diagnostic accuracy on predicting DR progression, the performance of the algorithm (AUPRC: mild NPDR, 0.717; moderate NPDR, 0.860) to identify DR progression match or exceed previously published performance of bespoke AI models based on fundus photography (area under the receiver operating curve, 0.710[4] and 0.79[5]). Furthermore, in the validation set, among eyes that progressed 2 or more steps, 75% with mild NPDR at baseline and 85% with moderate NPDR at baseline were identified. All eyes with mild NPDR that progressed within 6 months and 1 year were identified, as were 89% and 85% with moderate NPDR that progressed within 6 months and 1 year, respectively.

These deep-learning AI models based on UWF color retinal images were able to predict the progression of DR within 3 years in eyes with only early diabetic retinal changes. This AI model for risk stratification might allow a more personalized follow-up interval, potentially allowing patients with low risks of progression to be evaluated at longer intervals and those at higher risk to be evaluated sooner. Indeed, the model was particularly accurate at predicting progression risks within 6 months or 1 year of imaging (100% for mild NPDR and 88.9% and 85.0%, respectively, for moderate NPDR). Eyes with such mild baseline retinopathy severity are generally evaluated at more extended intervals due to presumed limited progres-

Table 3. Comparison With Published Nonautomated Machine Learning (ML) Models

Progression prediction[a]	Automated ML UWF	Rom et al[13]	Bora et al[14]
Retina image type			
Degree	200°	45°	45°
Fields	1	1 to 3	1 to 3
Images/eyes in dataset, No.	1179	156 363	575 431
AUC			
No DR	NA	0.67	0.79
Mild NPDR[b]	0.72	0.63	NA
Moderate NPDR[b]	0.86	0.75	NA
Prediction time frame, y	3	3	2

Abbreviations: AUC, area under the receiver operating characteristic curve; DR, diabetic retinopathy; NA, not applicable; NPDR, nonproliferative diabetic retinopathy; UWF, ultra-widefield.

[a] Each of models were evaluated using different image sets and grading methods. The difference in the performance measures may be due to

difference in the image sets, grading methods, algorithm threshold, or a combination of various factors. These are presented as a reference and generally do not represent comparative performance.

[b] Mild NPDR is defined as ETDRS levels 20 and 35. Moderate NPDR is defined as the ETDRS levels 43 and 47.

sion risk, so identifying the subset at risk of rapid progression is particularly pertinent.

This risk stratification model would be particularly beneficial in DR screening programs, where a large portion of patients have early DR. For eyes with early DR, a fixed yearly interval DR screening is capital and labor intensive for those not progressing and inadequate for those progressing. This deep-learning AI system might provide an individualized risk-based DR screening interval that would allow timely referral and treatment for high-risk individuals and extend up to 3-year intervals for low-risk individuals. Based on these findings, it is calculated such individualized evaluation could lead to a more than 20% reduction in cost and approximately 40% fewer appointments. It is important to emphasize that regulatory hurdles to the use of automated ML applications remain. A prospective clinical trial tailored to the requirements of the US Food and Drug Administration will be needed for fully autonomous use in the US prior to general clinical use by clinicians. Furthermore, studies will be required to determine whether implementation of point-of-care risk assessment in early NPDR will result in improved clinical outcomes and if this approach will translate into more cost-effective care.

The Diabetes Control and Complications Trial (DCCT) and Epidemiology of Diabetes Interventions and Complications (EDIC) study has provided a mean 23.5 years of follow-up data on DR progression starting from early DR severity.[12] In the DCCT/EDIC, among almost 24 000 standardized retinopathy examinations documented by photography, 14.5% showed worsening from the previous visit, 7.8% showed improvement, and 77.7% showed no change. These data highlight the relatively low rates of progression over time in eyes with early DR and highlight the need to identify approaches that can better predict the risk of worsening. It is essential that we further refine methods to identify eyes at increased risk that require closer follow-up while simultaneously identifying low-risk eyes where time to follow-up can be safely lengthened and patient and heath care burdens reduced.

Our exploratory evaluation of retinal saliency heat maps generated by automated ML point to regions of interest within the image that have not been routinely evaluated in these patient populations with early retinopathy. In these eyes with

early DR, the optic disc and the retinal vessels around the optic disc were highlighted as a region of interest by the AI algorithm (Figure 2). Although neovascularization of the disk is a clear risk factor for visual loss in diabetes, the morphology of disk vessels and surrounding retinal area are not generally recognized to provide substantial progression risk information in early NPDR. This finding emphasizes the potential for AI algorithms to identify retinal features that go beyond the currently defined DR risk parameters.

While automated ML significantly simplifies the AI development process, the platform does not remove the critical task of data preparation and prospective clinical trial validation. Clinicians and researchers must properly curate the datasets used as inputs to create AI models. Flawed and inconsistent labeling of training data will lead to poor-quality outputs, undermining the potential clinical benefits of automated decision support systems. The accurate labeling of medical images in a validated and standardized process is required to achieve optimal model performance, a process that should use representative images that capture the full range of image quality, photographic artifacts, patient anatomy, and disease pathology in the clinical settings where the AI system will be used.

Strengths and Limitations

The strengths of this study include the curation of the dataset used to train the automated ML algorithms and its unique application to longitudinal data in UWF retinal images. The process included the use of a standardized image acquisition and grading protocol for UWF field images with secondary adjudication to establish an accurate ground truth for images. Prior nonautomated ML models focused on standard 1-field or 2-field retinal images that focus only on the posterior retina. Our data suggest that despite a smaller number of images used to train the model, the performance of high-resolution UWF images was comparable with published nonautomated ML models (**Table 3**).[13,14]

Limitations of this study include the size of the dataset used, the lack of visual acuity outcomes, and the identification of any progression in only eyes with mild and moderate NPDR. These limitations should be considered if attempting to extrapolate these results to other populations.

Regulatory hurdles to the use of automated ML applications remain, and a prospective clinical trial tailored to the requirements of the Food and Drug Administration will be needed for fully autonomous use in the US. Furthermore, studies will be required to determine whether implementation of point-of-care risk assessment in early NPDR will result in improved clinical outcomes and if this approach will translate into more cost-effective care.[15]

Conclusions

This sudy demonstrates the accuracy and feasibility of an automated ML model for identifying DR progression in eyes with early DR using standard clinical UWF imaging. Although there are no defined standards for the prediction of DR progression, the performance of these initial automated ML models are comparable with published diagnostic accuracy metrics of bespoke AI models for non-UWF retinal imaging. Furthermore, the automated ML model is particularly effective in predicting short-term risks of DR progression in early mild to moderate NPDR. Although prospective validation and regulatory approval is required before these AI models are made available to physicians for clinical use, our results highlight the increasing accessibility of ML applications to address unmet clinical needs that may improve screening and vision outcomes for patients with diabetes.

REFERENCES

1. Jacoba CMP, Celi LA, Silva PS. Biomarkers for progression in diabetic retinopathy: expanding personalized medicine through integration of AI with electronic health records. *Semin Ophthalmol*. 2021;36(4):250-257. doi:10.1080/08820538.2021.1893351

2. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. *Ophthalmology*. 1991;98(5)(suppl):823-833. doi:10.1016/S0161-6420(13)38014-2

3. Lin WC, Chen JS, Chiang MF, Hribar MR. Applications of artificial intelligence to electronic health record data in ophthalmology. *Transl Vis Sci Technol*. 2020;9(2):13. doi:10.1167/tvst.9.2.13

4. Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit Health*. 2021;3(1):e10-e19. doi:10.1016/S2589-7500(20)30250-8

5. Arcadu F, Benmansour F, Maunz A, Willis J, Haskova Z, Prunotto M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med*. 2019;2:92. doi:10.1038/s41746-019-0172-3

6. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health*. 2019;1(5):e232-e242. doi:10.1016/S2589-7500(19)30108-6

7. Silva PS, Cavallerano JD, Haddad NM, et al. Peripheral lesions identified on ultrawide field imaging predict increased risk of diabetic retinopathy progression over 4 years. *Ophthalmology*. 2015;122(5):949-956. doi:10.1016/j.ophtha.2015.01.008

8. Silva PS, Cavallerano JD, Sun JK, Soliman AZ, Aiello LM, Aiello LP. Peripheral lesions identified by mydriatic ultrawide field imaging: distribution and potential impact on diabetic retinopathy severity. *Ophthalmology*. 2013;120(12):2587-2595. doi:10.1016/j.ophtha.2013.05.004

9. Silva PS, Salongcay RP, Aquino LA, et al. Intergrader agreement for diabetic retinopathy (DR) using hand-held retinal imaging. *Invest Ophthalmol Vis Sci*. 2021;62(8):1896.

10. Balki I, Amirabadi A, Levman J, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. *Can Assoc Radiol J*. 2019;70(4):344-353. doi:10.1016/j.carj.2019.06.002

11. Google. AutoML beginner's guide. Accessed January 9, 2024. https://cloud.google.com/vertex-ai/docs/beginner/beginners-guide

12. Nathan DM, Bebu I, Hainsworth D, et al; DCCT/EDIC Research Group. Frequency of evidence-based screening for retinopathy in type 1 diabetes. *N Engl J Med*. 2017;376(16):1507-1516. doi:10.1056/NEJMoa1612836

13. Rom Y, Aviv R, Ianchulev T, Dvey-Aharon Z. Predicting the future development of diabetic retinopathy using a deep learning algorithm for the analysis of non-invasive retinal imaging. *BMJ Open Ophth*. 2022;7(1):e001140. doi:10.1136/bmjophth-2022-001140

14. Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit Health*. 2021;3(1):e10-e19. doi:10.1016/S2589-7500(20)30250-8

15. Gomez Rossi J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. *JAMA Netw Open*. 2022;5(3):e220269. doi:10.1001/jamanetworkopen.2022.0269