AMERICAN ACADEMY OF OPHTHALMOLOGY®

Comparison of Machine Learning Models to a Novel Score in the Identification of Patients at Low Risk for Diabetic Retinopathy

Amanda Luong, BS,[1,*] *Jesse Cheung, BS,*[1,*] *Shyla McMurtry, MD,*[2] *Christina Nelson, BS,*[1] *Tyler Najac, MD,*[3] *Philippe Ortiz, MD,*[4] *Stephen Aronoff, MD, MBA,*[5,6] *Jeffrey Henderer, MD,*[4] *Yi Zhang, MD, PhD*[4]

Purpose: To develop an easily applicable predictor of patients at low risk for diabetic retinopathy (DR).

Design: An experimental study on the development and validation of machine learning models (MLMs) and a novel retinopathy risk score (RRS) to detect patients at low risk for DR.

Subjects: All individuals aged ≥18 years of age who participated in the telemedicine retinal screening initiative through Temple University Health Systems from October 1, 2016 through December 31, 2020. The subjects must have documented evidence of their diabetes mellitus (DM) diagnosis as well as a documented glycosylated hemoglobin (HbA1c) recorded in their chart within 6 months of the retinal screening photograph.

Methods: The charts of 1930 subjects (1590 evaluable) undergoing telemedicine screening for DR were reviewed, and 30 demographic and clinical parameters were collected. Diabetic retinopathy is a dichotomous variable where low risk is defined as no or mild retinopathy using the International Clinical Diabetic Retinopathy severity score. Five MLMs were trained to predict patients at low risk for DR using 1050 subjects and further underwent 10-fold cross validation to maximize its performance indicated by the area under the receiver operator characteristic curve (AUC). Additionally, a novel RRS is defined as the product of HbA1c closest to screening and years with DM. Retinopathy risk score was also applied to generate a predictive model.

Main Outcome Measures: The performance of the trained MLMs and the RRS model was compared using DeLong's test. The models were further validated using a separate unseen test set of 540 subjects. The performance of the validation models were compared using DeLong's test and chi-square tests.

Results: Using the test set, the AUC for the RRS was not statistically different from 4 out of 5 MLM. The error rate for predicting low-risk patients using the RRS was significantly lower than the naive rate (0.097 vs. 0.19; $P < 0.0001$), and it was comparable to the error rates of the MLMs.

Conclusions: This novel RRS is a potentially useful and easily deployable predictor of patients at low risk for DR.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2025;5:100592 © 2024 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Diabetes mellitus (DM) is a systemic disease that is prevalent worldwide. As of 2019, 351.7 million individuals aged 20 to 64 years old worldwide were affected by DM; this number is projected to increase to 417.3 million by 2030 and 486.1 million by 2045.[1] Diabetic retinopathy (DR) has been identified in a third of individuals with DM, and it is a major vision-threatening complication that may lead to blindness.[2] Annual direct funduscopic examinations are currently recommended for adult patients with DM.[3] In 2020, only 58.3% of adults diagnosed with DM had an eye examination within the past 12 months, decreasing from 64.8% in 2019.[4] In a previous DR telemedicine project in the Temple University Health System, the estimated annual incidence of DR was 18% which suggests that the majority of examinations will yield negative results.[5] Given the expanding aging population and the projected increase in DM prevalence, it becomes increasingly challenging and less cost-effective to conduct annual diabetic eye evaluations for all affected individuals. Prior studies have recommended that annual screenings may not be necessary for every patient based on individual risk factors, and screenings may be safely delayed for up to 2 years.[6–8]

A targeted approach to screening aims to allocate limited clinical resources to individuals who are more likely to have eye disease. Several studies have employed machine learning techniques, utilizing moderate to large datasets and a substantial number of variables to develop predictive models of DR.[9–12] The reported models exhibit varying areas under the receiver operator characteristic curves (AUCs). Deployment of these models would require either a freestanding or web-based application for the input of patient variables.

https://doi.org/10.1016/j.xops.2024.100592
ISSN 2666-9145/24

A report from an international committee of experts identified elevated serum levels of glycosylated hemoglobin (HbA1c) as a risk factor for DR.[13] Additionally, the duration of disease has also been associated with the development of DR.[14,15] The purpose of this study is to determine if a novel retinopathy risk score (RRS) derived from the serum HbA1c concentration closest to the screening and the duration of disease can serve to be as effective of a predictor for patients at low risk for DR as compared with optimized machine learning algorithms.

Methods

The Temple University Institutional Review Board/Ethics Committee approval was obtained. The Institutional Review Board waived the requirements for informed consent given the nature of this study. The study methods adhered to the Declaration of Helsinki. All data manipulations, model building, and analyses were performed in the R environment (R Foundation for Statistical Computing).

Subjects

The electronic medical record (EMR) database at Temple University Health System, serving North Philadelphia, was searched for patients who met the following criteria: (1) age $\geq$18 years; (2) carried a diagnosis of either type I or type II DM (International Classification of Diseases-10: E08.xxxx, E10.xxxx, E11.xxxx, E13.xxxx) at the time of their primary care visit; (3) did not have a dilated fundus exam for DR recorded in the EMR in the past year; (4) participated in the telemedicine retinal photographic screening initiative conducted by the outpatient internal medicine department between October 1, 2016 through December 31, 2020. The screening initiative consisted of trained medical assistants capturing retinal images in a primary care setting. Quality control was completed by reviewing the photos and providing feedback to the medical assistants; and (5) had a recorded measurement of serum HbA1c within 6 months of the retinal photograph. All the patients who were photographed received primary care services from the Temple University Health System and had supporting documentation in the EMR. Of note, specific provisions were not taken to ensure that patients did not have duplicate visits in the study period. If a patient previously had a positive screening result, they were referred to the ophthalmology department for further management rather than continuing to be screened in the primary care setting. Positive patients thus should have been removed from the screening pool. It is highly likely that negative patients were screened more than once. By doing so, it would help generate a more accurate cutoff score from tracking patients over time.

Data Extraction

Features. Predictive features for model development were extracted from each subject's medical record. In cases where multiple values were recorded over time, the measurement taken nearest to the time of the retinal photograph, plus or minus 6 months, was selected. The features extracted or engineered were: date of screening; age; gender; HbA1c; HbA1c group; disease duration (collected manually through chart review. Patients with more recent diagnoses were more accurately dated through chart review and associated labs, while the duration of others were based on documented patient recollection as the EMR at the institution was implemented in 2011); type of DM; body mass index; mean arterial pressure, systolic and diastolic blood pressure; history of: gastric bypass surgery, hypertension, hyperlipidemia, coronary artery disease, transient ischemic attack, cerebrovascular disease, peripheral artery disease, and/or angioplasty; serum high-density lipoprotein; total serum cholesterol; serum triglycerides; serum creatinine; serum blood urea nitrogen; chronic kidney disease stage; glomerular filtration rate; type of diabetes medication/management (diet alone, oral hypoglycemics, insulin, or combination of oral hypoglycemic and insulin); microalbuminuria and macroalbuminuria; and cardiovascular risk assessment score.

Response Variable. Retinal images for each subject were interpreted by optometrists associated with the Temple Department of Ophthalmology using the ETDRS Disease Severity Scale and the International Clinical Diabetic Retinopathy Severity Score. Grader reliability was not formally completed, but a retina specialist (Y.Z.) reviewed some of the optometrists' interpretations. Accuracy of the retina specialist was not formally reviewed. The final response variable was dichotomous; any image with an identifiable retinal abnormality consistent with more than mild DR was classified as positive (International Clinical Diabetic Retinopathy >1) and those without were classified as negative.

Feature Selection

Four approaches were taken to identify superfluous or poorly predictive features. First, any feature with values missing from >50% of the subjects was removed. Next, univariate comparisons of rate distributions for categorical variables and Gaussian distributions for continuous variables were performed; variables with P values > 0.1 based on these tests were removed. Finally, categorical variables with near zero variance, defined as >90% membership in a single class, and highly correlated continuous variables, defined as a correlation coefficient (r) >0.85, were also removed. Excluded features include the following: Atherosclerotic Cardiovascular Disease score, history of bariatric surgery, transient ischemic attack, body mass index, systolic blood pressure, diastolic blood pressure, triglycerides, high-density lipoprotein, low density lipoprotein, total cholesterol, blood urea nitrogen, microalbumin creatinine ratio, and type of DM (Figure 1).

Missing Values

Missing data in the aggregate dataset were imputed using multivariate imputation by chain equations (mice package).[16] For continuous variables, missing values were estimated by predictive mean matching; for categorical variables, class assignments were imputed by logistic regression (LR). Imputation was performed prior to data

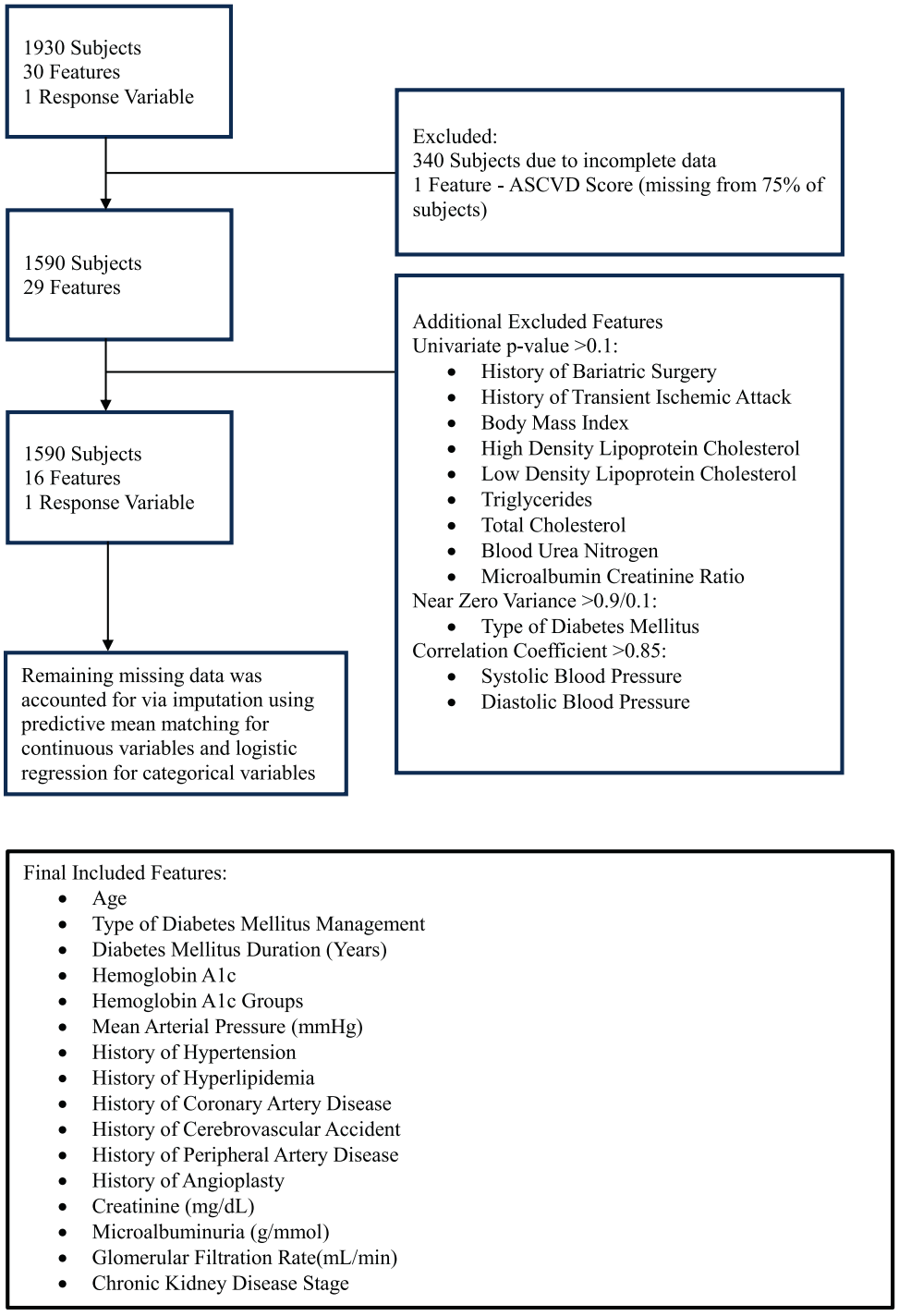

Figure 1. Dataset construction workflow. ASCVD = Atherosclerotic Cardiovascular Disease.

splitting to balance both sets. Subjects with missing response variables were removed.

Data Segregation

To provide a subset of subjects for model validation, the final dataset was split. Two-thirds of the subjects were randomly assigned to the training set which was used for tuning the predictive algorithms. The remaining one-third of patients were sequestered for validation of the optimized models using the createDataPartition function in the caret package.[17] This function creates training and test data sets by maintaining the distribution of the outcome variable. This is essential for predictive models since the utility of

Table 1. Subject Demographics and Features Utilized in Predictive Model Development

Demographic/Clinical Feature	Training Set		Test Set	
	No	*Yes*	*No*	*Yes*
Presence of diabetic retinopathy (N, %)	851 (81.0)	199 (19.0)	438 (81.1)	102 (19.0)
Age				
Mean (SD)	58.7 (12.9)	59.2 (11.7)	59.3 (11.8)	56.9 (12.0)
Median [min, max]	59.0 [1.00, 118]	59.0 [25.0, 89.0]	61.0 [28.0, 91.0]	57.0 [29.0, 83.0]
Type of diabetes mellitus management				
Insulin	431 (50.6%)	65 (32.7%)	218 (49.8%)	33 (32.4%)
Oral hypoglycemics	97 (11.4%)	50 (25.1%)	47 (10.7%)	26 (25.5%)
Insulin and oral medications	170 (20.0%)	70 (35.2%)	97 (22.1%)	36 (35.3%)
Diet alone/no medications	153 (18.0%)	14 (7.0%)	76 (17.4%)	7 (6.9%)
Diabetes mellitus duration (yrs)				
Mean (SD)	7.48 (6.83)	12.4 (10.1)	7.78 (7.17)	13.2 (9.05)
Median [min, max]	6.00 [0, 58.0]	9.00 [0, 58.0]	6.00 [0, 41.0]	10.0 [0, 49.0]
Hemoglobin A1c				
Mean (SD)	7.67 (2.13)	8.63 (2.08)	7.57 (1.93)	9.03 (2.42)
Median [min, max]	6.90 [4.50, 15.5]	8.20 [4.70, 15.0]	6.90 [4.70, 14.8]	8.80 [5.00, 14.9]
Hemoglobin A1c groups				
<5.7	77 (9.0%)	6 (3.0%)	32 (7.3%)	7 (6.9%)
5.7-6.4	222 (26.1%)	17 (8.5%)	107 (24.4%)	12 (11.8%)
6.5-10	427 (50.2%)	130 (65.3%)	243 (55.5%)	48 (47.1%)
>10	125 (14.7%)	46 (23.1%)	56 (12.8%)	35 (34.3%)
Mean arterial pressure (mmHg)				
Mean (SD)	87.5 (31.9)	97.6 (21.9)	87.4 (30.9)	95.6 (19.0)
Median [min, max]	95.3 [0, 144]	97.3 [0, 166]	95.0 [0, 130]	97.0 [0, 132]
History of hypertension				
No	205 (24.1%)	21 (10.6%)	101 (23.1%)	18 (17.6%)
Yes	646 (75.9%)	178 (89.4%)	337 (76.9%)	84 (82.4%)
History of hyperlipidemia				
No	347 (40.8%)	51 (25.6%)	169 (38.6%)	31 (30.4%)
Yes	504 (59.2%)	148 (74.4%)	269 (61.4%)	71 (69.6%)
History of coronary artery disease				
No	737 (86.6%)	162 (81.4%)	373 (85.2%)	74 (72.5%)
Yes	114 (13.4%)	37 (18.6%)	65 (14.8%)	28 (27.5%)
History of cerebrovascular accident				
No	794 (93.3%)	181 (91.0%)	404 (92.2%)	86 (84.3%)
Yes	57 (6.7%)	18 (9.0%)	34 (7.8%)	16 (15.7%)
History of periphery artery disease				
No	814 (95.7%)	188 (94.5%)	417 (95.2%)	91 (89.2%)
Yes	37 (4.3%)	11 (5.5%)	21 (4.8%)	11 (10.8%)
History of angioplasty				
No	780 (91.7%)	175 (87.9%)	402 (91.8%)	84 (82.4%)
Yes	71 (8.3%)	24 (12.1%)	36 (8.2%)	18 (17.6%)
Creatinine (mg/dL)				
Mean (SD)	1.10 (0.787)	1.57 (1.66)	1.10 (0.796)	1.62 (1.92)
Median [min, max]	0.960 [0.290, 11.9]	1.07 [0.370, 12.2]	0.940 [0.0800, 9.60]	1.04 [0.590, 11.0]
Glomerular filtration rate (mL/min)				
Mean (SD)	75.6 (47.9)	71.3 (46.2)	79.0 (45.9)	75.9 (45.5)
Median [min, max]	92.0 [1.00, 138]	79.0 [1.00, 138]	96.0 [1.00, 138]	80.5 [1.00, 137]
Chronic kidney disease stage				
1	285 (33.5%)	51 (25.6%)	137 (31.3%)	31 (30.4%)
2	387 (45.5%)	80 (40.2%)	212 (48.4%)	33 (32.4%)
3a	103 (12.1%)	24 (12.1%)	49 (11.2%)	15 (14.7%)
3b	46 (5.4%)	19 (9.5%)	27 (6.2%)	10 (9.8%)
4	21 (2.5%)	14 (7.0%)	7 (1.6%)	7 (6.9%)
5	9 (1.1%)	11 (5.5%)	6 (1.4%)	6 (5.9%)
Microalbuminuria (g/mmol)				
Mean (SD)	0.250 (0.544)	0.482 (0.744)	0.251 (0.546)	0.588 (0.788)
Median [min, max]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]	0 [0, 2.00]
HbA1c x diabetes duration (yrs)				
Mean (SD)	57.6 (55.3)	107 (92.2)	59.6 (56.7)	121 (97.0)
Median [min, max]	44.8 [0, 435]	79.2 [0, 583]	45.3 [0, 299]	91.5 [0, 573]

AUC = area under the receiver operator characteristic curve; DUR = duration of disease in years; HbA1c = glycosylated hemoglobin.
SD = standard deviation.

Table 2. Comparison of the Receiver Operator Characteristics of 6 Combinations of Serum HbA1c and Disease Duration

Model	AUC	*P* Value*
HbA1c × DUR	0.703	-
HbA1c + 0.25 (DUR)	0.711	0.57
HbA1c + 0.5 (DUR)	0.712	0.31
HbA1c + DUR	0.707	0.39
HbA1c + 2.5 (DUR)	0.700	0.07
HbA1c + 5 (DUR)	0.690	0.005

AUC = area under the receiver operator characteristic curve; DUR = duration of disease in years; HbA1c = glycosylated hemoglobin. SD = standard deviation.
*DeLong's test using the product of serum HbA1c and duration of disease as the reference.

the models will vary based on the prior outcome rates. Alternative methods of data splitting, such as random splitting, run the risk of changing the outcome rates.

Derivation of a Novel RRS

A previous study found a strong correlation between serum HbA1c concentration, duration of disease, and risk of DR.[10] With the intention of being easily applicable and convenient for users, 6 combinations of HbA1c and duration of disease in years (DUR) were trialed as candidates for a RRS: (1) HbA1c × DUR, (2) HbA1c + 0.25 × DUR, (3) HbA1c + 0.5 × DUR, (4) HbA1c + DUR, (5) HbA1c + 2.5 × DUR, and (6) HbA1c + 5 × DUR. The values were calculated from the training dataset and receiver operator characteristic curves (ROCs) were generated and compared using the value from the product score (HbA1c × DUR) as the reference. The final RRS was either the product score or the score that produced a statistically significantly higher AUC than the product score. Using the ROC of the final RRS generated from the training data set, the threshold value for separating low risk (no retinopathy predicted) and high risk (retinopathy predicted) patients was determined by identifying the point on the ROC that maximized Cohen's kappa statistic.

Derivation of Machine Learning Models

Five algorithms were chosen: 2 parametric algorithms (LR and support vector machine [SVM] using the radial kernel), 1 nonparametric algorithm (recursive partitioning [RPART]), and 2 ensemble algorithms (random forest and gradient boosted machine). Each algorithm was tuned to maximize the AUC using 10-fold cross-validation and the train function from the caret package. From the collection of tuned models produced for each algorithm, resampled data were analyzed and the parameter or hyperparameter values for the model with the greatest ROC were selected as the final model for the given algorithm.

Validation of the Machine Learning Models and Comparison with the RRS

Using the fine-tuned models of each algorithm, the test subset of data and the predict function from the caret package, a vector of probabilities of no retinopathy was determined using each subject and an ROC was constructed using the known outcomes of the test subject. The threshold value for separating low-risk and high-risk patients was determined in an identical manner as for the RRS except that the ROCs developed from the test data set were used. Finally, an ROC was constructed for the RRS using the values from subjects in the test dataset.

Confusion matrices were constructed for each model and the RRS using the previous determined threshold values. The rows corresponding to the low-risk predictions were assembled into a matrix and compared.

Statistical Analyses

For feature selections, differences in rate distributions were determined by the chi-square test while differences in Gaussian distributions of continuous variables were determined by *t* tests; features were eliminated if the *P* value exceeded 0.1.

Differences in ROCs were determined in a pairwise fashion using DeLong's test for correlated ROC. Comparisons of risk assignment were determined by chi-square tests of the pertinent portions of confusion matrices generated by

Table 3. Mean Values of Area Under the Receiver Operator Characteristic Curves, Sensitivity, and Specificity Using Resampled Data for 5 Optimized Predictive Algorithms

Statistical Value	GBM	LR	RF	SVM	RPART
Area under the receiver operator characteristic curve	0.772	0.759	0.763	0.693	0.677
P value	Reference	1	1	0.425	0.072
Sensitivity (predicting NORMAL retinas)	0.985	0.961	0.975	0.978	0.947
P value	Reference	0.412	1	1	0.628
Specificity (predicting ABNORMAL retinas)	0.121	0.247	0.167	0.106	0.212
P value	Reference	0.059	1	1	1

GBM = gradient boosted machine; LR = logistic regression; RF = random forest; RPART = recursive partitioning; SVM = support vector machine.
Mean values of each parameter determined by 10-fold cross validation.
P values by *t* test vs. GBM value.

each model and the RRS. In all cases, P values <0.05 were considered significant.

Results

Dataset Construction

The raw dataset had 1930 subjects, 30 predictive features, and 1 response variable. The 340 subjects with missing response values were eliminated from the dataset. In the culling process, 1 variable with 75% of values missing was dropped; 9 variables with P values >0.1 by univariate comparison, 1 categorical variable with near zero variance, and 2 variables with correlation coefficients >0.85 were dropped. The remaining 16 variables were: age (in years), type of diabetic medications, disease duration (years), HbA1c, HbA1c group, mean arterial blood pressure, diagnosis of hypertension and/or hyperlipidemia, history of (1) coronary artery disease; (2) cerebrovascular accident; (3) peripheral artery disease and/or; (4) angioplasty, serum creatinine, glomerular filtration rate, stage of chronic renal disease, and microalbuminuria. Since none of the remaining predictors had missing values >25%, all of them were retained for analysis. The rate of DR in the final data set of 1590 subjects and 16 predictive variables was 19%.

The final dataset was divided into training (n = 1050) and test (n = 540) subsets. The training and test data sets were comparable on the basis of demographics and features as shown in Table 1.

Derivation of a Novel RRS

The iterations of the RRS combining the serum HbA1c concentration closest to the screening with illness duration were used to develop an individual ROC. Using the product of HbA1c and disease duration as the reference, none of the other formulations were statistically superior by comparison of ROC (Table 2); the combination of serum HbA1c and 5× the duration of disease produced a statistically poorer model than the RRS ($P = 0.005$). Based on these observations, the product of HbA1c and disease duration was selected as the RRS. A cut off value of 53.9 was determined by maximizing the Cohen kappa statistic.

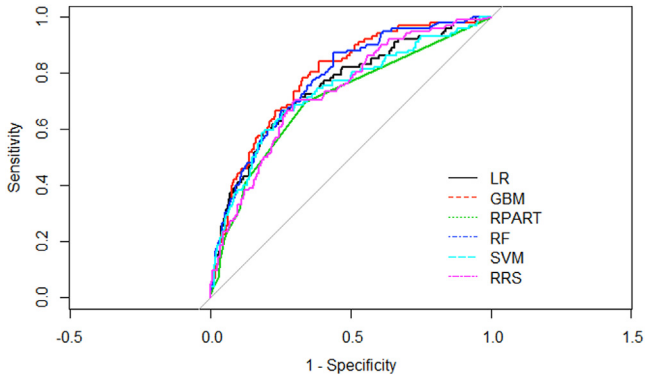

Figure 2. Receiver operator characteristic curves for 5 machine learning models and the retinopathy risk score using 540 test subjects. GBM = gradient boosted machine; LR = logistic regression; RF = random forest; RPART = recursive partitioning; RRS = retinopathy risk score; SVM = support vector machine.

Derivation of Machine Learning Predictive Models of DR

Using 5 algorithms (LR, SVM, RPART, random forest, and gradient boosted machine) and the training dataset, 5 predictive models were optimized using 10-fold cross validation. The performance of these models, using resampled data from the training set, is shown in Table 3. The AUCs ranged from 0.677 to 0.772 and were not statistically significantly different. The sensitivity, defined as the correctly predicted rate of patients with normal retinas, ranged from 0.947 to 0.985 and showed no statistical difference among the 5 models; however, specificity was poor, ranging from 0.106 to 0.247.

Validation and Comparison of Machine Learning Models and the RRS

Each subject in the test set of 540 patients was classified as low risk (normal retina examination) or high risk (abnormal retina examination) by the 5 machine learning models (MLMs). Using the predictions and known responses, ROC curves were generated; the AUC was calculated for each ROC. A similar curve was constructed for the RRS using the same dataset. Following the identification of the cutoff value on the ROC curve that maximized Cohen's kappa statistic, confusion matrices were created for the predictions from each model. A confusion matrix was constructed for the RRS using the cut off value determined with the training data.

The ROC curves for the 5 MLMs and the RRS are shown in Figure 2; AUCs are shown in Table 4. The AUC for the gradient boosted machine was significantly larger than the AUC for RRS (0.789 vs. 0.733; $P = 0.011$); the AUCs for the remaining models were not statistically different from the AUC for the RRS.

The performances of the 5 MLMs and the RRS to identify diabetic patients without retinopathy are shown in

Table 4. Comparison of the Areas Under the Receiver Operator Characteristic Curves for 5 Machine Learning Models and the Novel Retinopathy Risk Score Using 540 Test Subjects

Model	AUC	*P* Value*
RRS	0.73347	REF
RPART	0.705133	0.35
SVM	0.732832	0.993
LR	0.750179	0.526
RF	0.773693	0.052
GBM	0.785769	0.011

AUC = area under the receiver operator characteristic curve; GBM = gradient boosted machine; LR = logistic regression; RF = random forest; RPART = recursive partitioning; RRS = retinopathy risk score; SVM = support vector machine.
*By DeLong's test.

Table 5. With the exception of RPART, the remaining 4 MLMs and the RRS had significantly lower error rates than the baseline rate of DR (naive or no information rate) in the test set. When the no information rate and the RPART predictions were removed, the remaining 4 MLMs and the RRS were not significantly different ($P = 0.106$).

The performance of the RRS is shown in Figure 3. The plot demonstrates the increasing rate of abnormal retina examinations (error rate) with increasing values of the RRS. The functional portion of the curve appears to lie between 0 and 100. The red line demonstrates the cutoff value for the score determined by maximizing Cohen's kappa statistic. As the error rate falls, the proportion of patients identified as low risk decreases.

Discussion

Since DR is a leading cause of vision impairment and blindness,[18] an annual dilated fundus examination is recommended for patients with type II DM.[3] In 2020, the estimated prevalence of DM in Philadelphia County, Pennsylvania was estimated to be 13.1%. Compared with the county rate, 3 economically disadvantaged communities in North Philadelphia served by the Temple University Health System had higher estimated DM prevalences of 16.1%, 17.9%, and 18.2%.[19] Given the higher prevalence of DM within this population, following the current annual screening guidelines results in a large expenditure of scarce resources. Therefore, an effective screening method to identify patients at low risk for DR would optimize primary care physicians to identify patients who require less frequent screening. By reducing the frequency of screening low-risk patients, this will allow primary care physicians and ophthalmologists to redirect resources to those patients at higher risk of disease.

A number of studies have employed machine learning and large datasets to create predictive models. Li et al used the machine learning algorithms XGBoost, LR, SVM, and random forest along with 17 predictive features to predict the risk of DR among a cohort of diabetic patients in China.[9] Dividing the dataset into training (4875 subjects) and validation (27 577 subjects) subsets, this study found that XGBoost achieved the highest AUC of 0.90, while the other models had AUCs of 0.83, 0.79, and 0.87 respectively. Predictive features such as HbA1c, serum creatinine levels, presence of nephropathy, insulin treatment, and diabetic lower extremity arterial disease were associated with an increased risk of DR based on the Shapley Additive exPlanation in the models.[8] Unlike the present study, Li et al had an extremely low DR rate (6.3% vs. 19%), severely limiting the comparison of predictive results. In a study reported in 2015, Ogunyemi et al developed predictive DR MLM using clinical data from 6 federally qualified primary care clinics in south Los Angeles, California (Federally Qualified Health Centers) (513 subjects, DR rate = 25%) and public health data from the National Health and Nutrition Examination Survey conducted by the National Center for Health Statistics (1239 subjects, DR rate = 13%); 80% of each dataset was used to train the models and the remainder was used for validation.[10] With validation data, RUSBoost and AdaBoost.M1 models using the Federally Qualified Health Centers dataset had AUCs of 0.72 and 0.60, respectively; with the National Health and Nutrition Examination Survey dataset the AUCs were uninformative (0.48 and 0.51, respectively).[9] In another study by the same group reported in 2019, a dataset of 27 116 diabetic subjects from the Los Angeles County Department of Health Services' Teleretinal Diabetic Retinopathy Screening Program and Reading Center was used (18 077 in the training set and 9039 subjects in the validation set).[11] The rate of DR in the combined dataset was 34.1%.[11] The rate of DR was adjusted in the training set using synthetic minority oversampling. Eight predictive features were used to identify patients at high risk for DR. Three algorithms (penalized LR, SVM, and artificial neural networks) were trained; the validation dataset produced ROCs with AUCs of 0.752, 0.745, and 0.754, respectively.[10] Van der Heijden et al conducted a systematic review that encompassed 12 studies developing

Table 5. Predictions of Patients at Low Risk for Diabetic Retinopathy by 5 Machine Learning Models and an RRS Using a Test Set of 540 Subjects

Algorithm	Normal Retina	Abnormal Retina	Number Classified	Abnormal Retina Rate	P Value*	
Naive/no information†	438	102	540	0.189	REF	-
RPART	438	102	540	0.189	1.00	-
SVM	352	41	393	0.104	1.61E-12	REF
RRS	251	27	278	0.097	1.61E-12	0.106
LR	322	34	356	0.096	1.61E-12	0.106
GBM	293	22	315	0.070	1.61E-12	0.106
RF	243	13	256	0.051	1.61E-12	0.106

GBM = gradient boosted machine; LR = logistic regression; RF = random forest; RPART = recursive partitioning; RRS = retinopathy risk score; SVM = support vector machine.
*By chi-square test.
†Baseline rate of diabetic retinopathy in 540 subjects with diabetes.

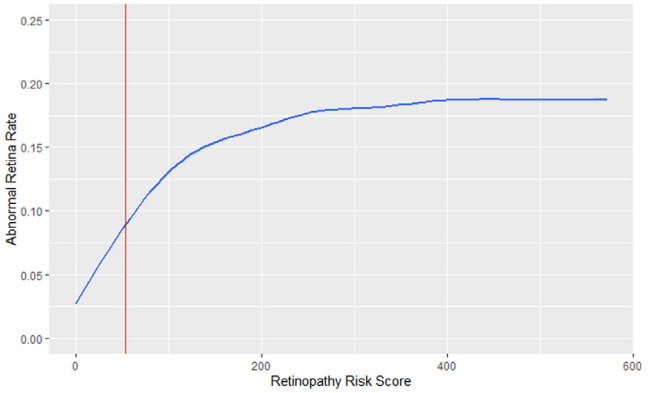

Figure 3. Retinopathy risk score performance as a low-risk predictor of diabetic retinopathy. The x-axis is the product of A1c and years with diabetes mellitus. Vertical line cutoff = 53.9.

prognostic prediction models for DR in individuals with type II DM.[20] Seven studies reported validation results with AUCs ranging from 0.55 to 0.84.[20]

The present study utilized a dataset of patients with DM who live in an underserved, economically disadvantaged community in North Philadelphia with a DR rate of 19%, comparable to most of the other reported studies. Our validated MLMs produced AUCs similar to those reported from previous studies with comparable rates of DR in the target population. In addition, the RRS, defined as the product of disease duration and HbA1c closest to the screening, produced an AUC that was statistically indistinguishable from the AUCs of most of the MLM. Finally, the error rate of the RRS was not statistically different from the error rates of the best MLM used to predict patients at low risk for DR.

There are several limitations of this study. First, the dataset primarily included patients from an economically disadvantaged, historically underserved, low-income population, which may limit the generalizability of these results. Second, the sample size was relatively small, consisting of 1930 subjects, which may have impacted the effectiveness of the machine learning algorithms. Finally, the duration of DM in years was missing for 33% of the subjects in the total dataset of 1590 subjects to account for missing data. Diabetic mellitus duration was imputed for these patients using predictive mean matching, a standard technique for estimating missing continuous data using multiple variables.[21] Another potential limitation in regard to DM duration is that these data were manually collected through chart review. Some patients had more recent diagnoses indicated by previous encounters and corresponding laboratory values, while the time of diagnosis for others was estimated based on anecdotal patient recollection documented in the EMR. These limitations should be considered when interpreting the results of this study.

The novel RRS we propose, easily calculated as the product of the HbA1c closest to the screening and the duration of disease in years, appears to be as effective as MLM that utilize a large number of predictive features in identifying patients with DM at low risk for DR. Previous studies have suggested the cost effectiveness and safety of decreasing the frequency of recommended diabetic eye examinations.[20] This RRS has the potential to act as a bridge between primary care, where patients with DM are routinely managed, and ophthalmology, ensuring that individuals at low risk for DR are not over-screened while optimizing resources for those at higher risk for developing vision-threatening DR. This study suggests that the RRS has potential as a unique metric for predicting the risk of DR. Before this model could be safely used clinically, a retrospective study using prospectively collected data on patients that have undergone screenings over several years would need to show a low risk of developing DR before considering a large prospective study to determine the rate of patients without DR and a low RRS of developing DR.

Footnotes and Disclosures

Originally received: January 3, 2024.
Final revision: July 28, 2024.
Accepted: July 30, 2024.
Available online: August 3, 2024. Manuscript no. XOPS-D-24-00001.

[1] Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania.

[2] Department of Internal Medicine at University of Kentucky HealthCare, Lexington, Kentucky.

[3] Department of Emergency Medicine at St. Barnabas Hospital System, Bronx, New York.

[4] Department of Ophthalmology, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania.

[5] Department of Pediatrics, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania.

[6] Department of Medical Education and Data Science, Lewis Katz School of Medicine at Temple University, Philadelphia, Pennsylvania.

*A.L. and J.C. contributed equally to this work.

Presented at the Association for Research in Vision and Ophthalmology 2023 Annual Meeting.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s):

Y.Z.: Others — Abbvie INC.

Supported by the Knights of Templar Eye Foundation Travel Grant for attending Association for Research in Vision and Ophthalmology 2023 Annual Meeting (A.L., J.C., and C.N.).

HUMAN SUBJECTS: Human subjects were included in this study. The Temple University Institutional Review Board (IRB)/Ethics Committee approval was obtained. The IRB waived the requirements for informed consent given the nature of this study. The study methods adhered to the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Luong, Cheung, McMurtry, Ortiz, Aronoff, Henderer, Zhang

Data collection: Luong, Cheung, Nelson, McMurtry, Najac, Ortiz, Aronoff, Henderer, Zhang

Analysis and interpretation: Luong, Cheung, McMurtry, Aronoff, Henderer, Zhang

Obtained funding: N/A

Overall responsibility: Luong, Cheung, Aronoff, Henderer, Zhang

Abbreviations and Acronyms:
AUC = area under the receiver operator characteristic curve; **DM** = diabetes mellitus; **DR** = diabetic retinopathy; **DUR** = duration of disease in years; **EMR** = electronic medical record; **HbA1c** = glycosylated hemoglobin; **LR** = logistic regression; **MLM** = machine learning model; **ROC** = receiver operator characteristic curve; **RPART** = recursive partitioning; **RRS** = retinopathy risk score; **SVM** = support vector machine.

Keywords:
Diabetic retinopathy, Machine learning, Predictive modeling.

Correspondence:
Yi Zhang MD, PhD, Lewis Katz School of Medicine at Temple University, 3401 N. Broad Street, Parkinson Pavilion, Suite 630, Philadelphia, PA 19140. E-mail: yi.zhang2@tuhs.temple.edu.

References

1. *IDF Diabetes Atlas*. 9th Edn. Brussels, Belgium: International Diabetes Federation, 2019; 2019.
2. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet*. 2010;376:P124−P136.
3. Flaxel CJ, Adelman RA, Bailey ST, et al. Diabetic retinopathy preferred practice pattern. *Ophthalmology*. 2020;127:P66−P145.
4. Increase the proportion of adults with diabetes who have a yearly eye exam - d-04. Healthy People. https://health.gov/healthypeople/objectives-and-data/browse-objectives/diabetes/increase-proportion-adults-diabetes-who-have-yearly-eye-exam-d-04; 2023. Accessed August 23, 2024.
5. Benjamin JE, Sun J, Cohen D, et al. A 15 month experience with a primary care-based telemedicine screening program for diabetic retinopathy. *BMC Ophthalmol*. 2021;21:70.
6. Scanlon PH. Screening intervals for diabetic retinopathy and implications for care. *Curr Diab Rep*. 2017;17:96.
7. Lund SH, Aspelund T, Kirby P, et al. Individualised risk assessment for diabetic retinopathy and optimisation of screening intervals: a scientific approach to reducing healthcare costs. *Br J Ophthalmol*. 2016;100:683−687.
8. Aspelund T, Pórisdóttir Ó, Ólafsdottir E, et al. Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy. *Diabetologia*. 2011;54:2525−2532.
9. Li W, Song Y, Chen K, et al. Predictive model and risk analysis for diabetic retinopathy using machine learning: a retrospective cohort study in China. *BMJ Open*. 2021;11:e050989.
10. Ogunyemi O, Kermah D. Machine learning approaches for detecting diabetic retinopathy from clinical and public health records. *AMIA Annu Symp Proc*. 2015;2015:983.
11. Ogunyemi OI, Gandhi M, Tayek C. Predictive models for diabetic retinopathy from non-image teleretinal screening data. *AMIA Jt Summits Transl Sci Proc*. 2019;2019:472.
12. Haider S, Sadiq SN, Moore D, et al. Prognostic prediction models for diabetic retinopathy progression: a systematic review. *Eye*. 2019;33:702−713.
13. Nathan DM, Balkau B, Bonora E, et al. International expert committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care*. 2009;32:1327−1334.
14. Voigt M, Schmidt S, Lehmann T, et al. Prevalence and progression rate of diabetic retinopathy in type 2 diabetes patients in correlation with the duration of diabetes. *Exp Clin Endocrinol Diabetes*. 2018;126:570−576.
15. Schreur V, van Asten F, Ng H, et al. Risk factors for development and progression of diabetic retinopathy in Dutch patients with type 1 diabetes mellitus. *Acta Ophthalmol*. 2018;96:459−464.
16. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1−67.
17. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28:1−26.
18. Sabanayagam C, Banu R, Chee ML, et al. Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol*. 2019;7:140−149.
19. City of Philadelphia. Southeastern Pennsylvania community health needs assessment 2022: department of public health. https://www.phila.gov/media/20220628103151/6-rCHNA-2022-Philadelphia-pp225-291.pdf; 2022. Accessed August 23, 2024.
20. van der Heijden AA, Nijpels G, Badloe F, et al. Prediction models for development of retinopathy in people with type 2 diabetes: systematic review and external validation in a Dutch primary care setting. *Diabetologia*. 2020;63:1110−1119.
21. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical Research: a tutorial on multiple imputation. *Can J Cardiol*. 2021;37:1322−1331.