



OPEN ACCESS

EDITED BY

Martina Tomić,
Merkur University Hospital, Croatia

REVIEWED BY

Mohd Anul Haq,
Majmaah University, Saudi Arabia
Babu Dhanendra Chaurasiya,
University of Southern California, United States

*CORRESPONDENCE

Ling Qin

✉ mhhkjwds@163.com

Fei Li

✉ li_fei@jlu.edu.cn

[†]These authors share first authorship

RECEIVED 20 May 2025

ACCEPTED 27 October 2025

PUBLISHED 10 November 2025

CITATION

Liu X, Wu S, Yang Y, Li Y, Zhang X, Liu R, Qin L
and Li F (2025) Development of machine
learning predictive model for type
2 diabetic retinopathy using the triglyceride-
glucose index explained by SHAP method.
Front. Endocrinol. 16:1631647.
doi: 10.3389/fendo.2025.1631647

COPYRIGHT

© 2025 Liu, Wu, Yang, Li, Zhang, Liu, Qin and
Li. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Development of machine learning predictive model for type 2 diabetic retinopathy using the triglyceride-glucose index explained by SHAP method

Xiaoqin Liu^{1†}, Shuying Wu^{1†}, Yue Yang¹, Yang Li¹,
Xinting Zhang¹, Rihui Liu², Ling Qin^{3*} and Fei Li^{1*}

¹The First Hospital of Jilin University, Changchun, Jilin, China, ²Guangdong Mental Health Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China, ³Meihekou Central Hospital, Meihekou, Jilin, China

Introduction: This study aimed to develop a diabetic retinopathy (DR) Prediction model using various machine learning algorithms incorporating the novel predictor Triglyceride-glucose index (TyG). Furthermore, the model was interpreted using the SHapley Additive exPlanations (SHAP) method.

Method: Real-world data were collected from a general hospital in a major city and a county clinic, then divided into the DR Group (1392) and non-DR group (2358). Baseline data were collected, and variables were selected using Recursive Feature Elimination with Cross-Validation (RFECV). The performance of five machine learning algorithms, including Logistic Regression model (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and XGBoost (XGB), was assessed based on accuracy, sensitivity, specificity, and Area Under the Curve (AUC) of the Receiver Operating characteristic Curve (ROC). The optimal model was interpreted using SHAP.

Result: LVM and LR demonstrated superior performance in both the test set and training set (ROC, 0.85 and 0.82, respectively). The top five predictors identified by SHAP analysis included TyG, Insulin therapy, HbA1c, Diabetes Course, HDL. HDL was identified as a protective factor, while the remaining factors were associated with retinopathy.

Conclusion: LR and SVM demonstrated the best performance. To our knowledge, this is the first machine learning-based DR prediction model integrating the triglyceride-glucose index (TyG) as a core predictor, overcoming limitations of insulin resistance (IR) assessment in resource-limited settings. TyG provides a cost-effective alternative to conventional IR biomarkers (e.g., HOMA-IR), enabling practical DR risk stratification in primary care.

KEYWORDS

TyG-index, diabetic retinopathy, machine learning, predictive model, SHAP

1 Introduction

Diabetes has a high incidence and mortality rate, thus emerging as a major global health challenge. The global prevalence of diabetes in people aged 20–79 years may increase to 12.2% by 2045 (1). Diabetic retinopathy (DR), the most common and serious ocular complication of type 2 diabetes, can result in visual impairment and even permanent vision loss (2). To date, DR is the leading cause of blindness among individuals aged 20–74, contributing to new cases of preventable blindness in developing countries (3, 4). A meta-analysis showed that the worldwide prevalence of DR is about 22.27% (5). Notably, China has the largest number of diabetic patients in the world (1), with about 27.9% of diabetes cases being DR (6), of which 34% are in rural areas. Most DR patients are asymptomatic or mild in the early stage and are only detected when great damage has occurred or at the late stage (2). Therefore, early screening of DR is crucial. However, the detection rate of DR remains suboptimal (7), particularly in certain primary hospitals and remote regions due to the constraints in medical conditions and limited social resources. Therefore, more accessible methods are needed to aid healthcare professionals in diagnosis and screening of DR.

To date, factors such as diabetes duration, age, BMI, smoking, blood pressure, HbA1c levels, and cholesterol have been identified as risk factors for diabetic retinopathy (DR) (8, 9). However, only a limited number of studies have incorporated insulin resistance (IR) into DR prediction models, despite substantial evidence (10, 11) demonstrating a strong association between IR and DR. The hyperinsulinemic-euglycemic clamp (HIEC) is the gold standard for detecting IR (12). However, it is expensive and complex, limiting its use. HOMA-IR (13) is the most commonly utilized method for estimating insulin resistance; however, it requires the measurement of fasting insulin levels, which imposes specific technical demands on laboratory capabilities. Additionally, this method is not suitable for patients undergoing insulin therapy and is not widely adopted in primary care settings or resource-limited regions. Therefore, HOMA-IR cannot be widely promoted in primary hospitals and poor areas (14, 15). A new index has been recently identified for detecting IR: Triglyceride-glucose index (TyG) (16). TyG is calculated by fasting triglyceride (TG) and fasting blood glucose (FBG), providing a simple, reliable, and cheap detection tool (17). Besides, this index tool has shown better results than HOMA-IR (18). Elsewhere, Srinivasan (19) and Yao (20) have shown that TyG and DR are closely related, while others not.

Abbreviations: TyG-index, Triglyceride-glucose index; DR, Diabetic retinopathy; RI, insulin resistance; HOMA-IR, Homeostatic Model Assessment for Insulin Resistance; TC, Total cholesterol; TG, Triglyceride; LDL, Low-density lipoprotein; HDL, High-density lipoprotein; FBG, Fasting blood glucose; BMI, Body mass index; LR, Logistic regression; DT, Decisiontree; RF, Randomforest; XGBoost, eXtremegradient boosting; SVM, Support vector machine; RFECV, Recursive Feature Elimination with Cross-Validation; AUC, Area under the curves; ROC, Receiver operating characteristic curve; SHAP, Shapley additive explanation.

In recent years, artificial intelligence (AI) has demonstrated significant advantages in disease prediction. Machine learning (ML), encompassing both deep learning (DL) and traditional ML approaches represents the most widely adopted technique (21). While DL has proven highly effective in medical image analysis—as evidenced by studies on brain tumor classification using DCNN (22), U-Net-based MRI segmentation (23), and mammography interpretation *via* XAI-RACapsNet (24)—its superior performance is primarily confined to high-dimensional (imaging) data. Furthermore, DL imposes higher computational demands than traditional ML methods. Due to the inherent opacity of its multiple processing layers (the so-called "black box" problem), DL poses greater challenges to model interpretability and accountability (25). In contrast, traditional ML algorithms (e.g., SVM, LR) exhibit distinct advantages when processing tabular data: superior interpretability through SHAP, critical for clinical trust (26); reduced computational demands, enabling deployment in resource-limited settings (27); and comparable performance to DL on tabular biomedical data with limited samples (27). Given our goal of developing a primary-care-suitable tool, we prioritized lightweight, interpretable models over complex DL architectures. In the absence of imaging inputs, ML outperforms DL in tabular data prediction. The newly discovered predictors can improve the predictive value for the prediction model (28, 29). TyG, as a new index for measuring IR, has shown great value in many studies. While integrating the TyG as a novel predictor further enhances accessibility without compromising performance. However, no prediction model uses TyG to predict the DR among the DM2 population.

To bridge this critical gap, we present the first machine learning framework integrating TyG for DR prediction, which simultaneously addresses three key limitations of prior approaches: 1) Methodological Barrier: Existing models (30–33) neglect insulin resistance (IR) or rely on impractical biomarkers [e.g., HOMA-IR requiring fasting insulin (11)]; 2) Clinical Applicability: TyG leverages routine lipid/glucose tests, enabling IR assessment in resource-limited settings where specialized assays are unavailable (17, 18); 3) Model Interpretability: We employ SHAP to decode the "black box" nature of ML models, quantifying TyG's non-linear contribution to DR risk. By validating this approach across multi-tier healthcare centers (urban tertiary vs. rural primary hospitals), we establish TyG as a novel, deployable biomarker for scalable DR screening.

2 Methods

2.1 Research design

2.1.1 Population

This is a retrospective study, where patient data were extracted from the real-world database of The First Hospital of Jilin University (a general hospital in a major city) and Meihokou Central Hospital (a primary health care institution in the county) from January 1, 2010 to December 31, 2023.

2.1.2 Inclusion criteria

a. Patients diagnosed with type 2 diabetes (T2DM) following the criteria of the 2024 American Diabetes Association (34); b. Patients aged ≥ 18 years; c. Patients with complete indicators (triglyceride and fasting glucose).

2.1.3 Exclusion criteria

a. Patients diagnosed with retinopathy at admission; b. Patients suffering from other retinal diseases, glaucoma, optic neuropathy, or eye diseases caused by systemic diseases; c. Patients with a history of eye surgery; d. Patients with severe systemic diseases (cancer, myocardial infarction, and dialysis history); e. Patients with data loss exceeding 20%.

2.1.4 Outcome

The extraction variable was the first measurement record at first admission. Diagnosis information was extracted from the discharge diagnosis, and the follow-up was conducted until the first DR diagnosis, otherwise the last visit time was selected as the follow-up endpoint. Diagnostic criteria of DR included: A spectrum of retinal microvascular lesions on retinal examination with diabetic patients (35), including mild, moderate, and severe non-proliferative DR and proliferative DR. DR was diagnosed using 45° photos of macular center and indirect ophthalmoscopy when pupils were dilated. DR diagnosis was mainly achieved by endocrinologists and ophthalmologists. Patients diagnosed with retinopathy in other hospitals during the follow-up process were marked as DR, and the follow-up time was based on the earliest diagnosis time. Positive and negative samples were defined as DR and non-DR patients, respectively.

2.1.5 Ethical approval

This study was conducted following the Helsinki Declaration and was approved by the Research Ethics Committee of The First Hospital of Jilin University (approval number: 2024-918). Each participant provided signed written informed consent. This study was reported based on TRIPOD (28).

2.1.6 Baseline data collection

Data indicators, including sex, age, height, weight, smoking, drinking, course of T2DM, insulin therapy, hypertension history, and laboratory parameters, were mainly obtained from literature reports (2, 36, 37).

Laboratory parameters included glycated hemoglobin (HbA1c, g/dL), total cholesterol (TC, mmol/L), high-density lipoprotein (HDL, mmol/L), low-density lipoprotein (LDL, mmol/L), triglyceride (TG, mg/dL), fasting blood glucose (FPG, mg/dL), fasting C-Peptide (C-PE, ng/ml), fasting insulin (FINS, μ U/ml) and C-reactive protein (CRP, μ g/L).

Body mass index (BMI) was calculated using the formula as follows: $\text{weight (kg)} / \text{height (m)}^2$. TyG index: $\text{LN} [\text{triglyceride (mg/dl)} \times \text{plasma glucose (mg/dl)} / 2]$. Three graduate students collated and cross-checked the collected data. A unified training for the data collection and collation personnel was conducted to ensure the accuracy and consistency of the data.

2.1.7 Sample size calculation

The sample size was mainly based on Riley (38) standard for accurate estimation. Four sample sizes (at least 878) (predicted value with small average error, intercept model only, ensuring shrinkage coefficient of 0.9, and ensuring small optimism of apparent model) were calculated. The total sample size was at least 966 [$878 + 10\% (878)$], considering that some medical record information was incomplete and 10% contingency.

2.1.8 Model construction

Three single model algorithms (LR, DT, and SVM) and two integrated methods (RF and XGBoost) were used to train the model. LR is a machine learning method for solving binary classification problems, and is used to estimate the possibility of something. DT is a simple and easy-to-operate tree-type classification prediction model, providing intuitive and easy-to-understand results. However, DT can easily lead to overfitting during the classification process. SVM is widely used to construct a hyperplane concept to classify the observed values and can be used to deal with classification and regression problems. Compared with the single DT model, the integrated algorithms have higher accuracy but present more complicated and difficult results to explain.

2.1.9 Statistical analysis

2.1.9.1 Statistical interpretation

R4.2.1 software was used for all data analysis. The variables missing more than 20%, including fasting C-Peptide, fasting insulin, and CRP were deleted to improve the utilization rate of the data. For individual missing values, the average interpolation method and mode interpolation method were used for counting data and measuring data, respectively. The normally distributed data were expressed as $\bar{X} \pm S$, and compared using two independent samples T-test. The non-normally distributed data were expressed as P50 (P25, P75) and compared using Mann-Whitney U test. The counting data were expressed as frequency (%) and compared using χ^2 test. The correlation between the predictive indicators and retinopathy was assessed using univariate and multivariate logistic regression models. Variables with univariate analysis $P < .05$ were included in multivariate logistic regression analysis. The odds ratio (OR) and corresponding 95% confidence interval (CI) were used to indicate the trend of correlation.

2.1.9.2 Variable selection

The model was trained and verified using Python3.9.15 software and tool kits Sklearn1.0.2, XGBoost1.7.4, and shap0.41.0. The recursive feature elimination with cross-validation (RFECV) was used to screen the optimal index. The least important features were continuously eliminated by training the model. The performance of the model was evaluated until the optimal performance index was reached. In this study, the variables with $P < 0.1$ in the comparison between groups were included in the RFECV model, and the best indicators were selected for subsequent prediction.

2.1.9.3 Model performance assessment

The prediction efficiency of LR, DT, SVM, RF, and XGBoost 5 models was evaluated based on accuracy, sensitivity, specificity, and Area Under the Curve (AUC) of receiver operating characteristic curve. Accuracy, precision, recall rate, and F1 score were used as the indicators. SHAP was used to visually explain the machine learning model. The Summary Plot, Shap Heat map, and importance ranking diagram were also drawn. The actual application of a single sample model was visualized *via* Force Plot. $P < .05$ was considered a statistically significant difference.

To further assess model stability, we performed 5-fold cross-validation on the entire dataset. Performance metrics (AUC, accuracy, sensitivity, specificity) were averaged over 5 iterations with standard deviations calculated.

2.1.9.4 Model interpretation

We employed SHapley Additive exPlanations (SHAP) (26), to interpret model predictions using Python's shap package (v0.41.0). Given computational advantages for tree-based models, SHAP analysis was applied to the XGBoost classifier *via* TreeExplainer. SHAP values – quantifying each feature's marginal contribution to predictions – were calculated for every sample in the training dataset. Four interpretability visualizations were generated: Global Feature Importance: Features ranked by mean absolute SHAP

value. Summary Plot: Dot-plot distribution of SHAP values per feature. Horizontal position indicates impact direction (positive/negative SHAP); color encodes feature value (red=high, blue=low). Force Plot: Illustrates additive feature contributions driving an individual prediction from the base value (dataset average) to the model output $f(x)$. Partial Dependence Plot: Generated using scikit-learn's PartialDependenceDisplay, depicting the marginal effect of TyG on predicted DR risk while averaging other features.

2.1.9.5 Multicollinearity assessment

To evaluate potential multicollinearity among predictors, we calculated the variance inflation factor (VIF) for all variables included in both the machine learning models and multivariate logistic regression. VIF values < 5 were considered indicative of no significant multicollinearity.

3 Result

3.1 Social demographic and clinical characteristics

The distribution of study participants is shown in Figure 1. A total of 1014 positive cases were extracted from real-world data, and

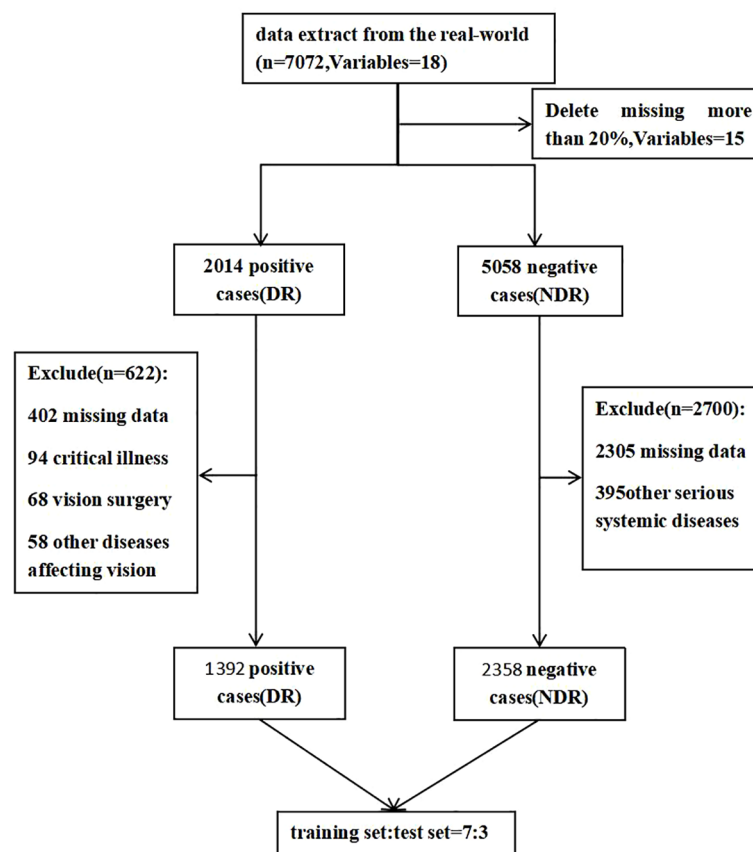


FIGURE 1
Flowchart of participant selection. DR, diabetic retinopathy; NDR, No DR.

TABLE 1 Comparison of baseline data between the two groups.

Variables		Overall	Control	Case	P
		(n=3750)	(n=2358)	(n=1392)	
Sex	Female	1384 (36.91)	800 (33.93)	584 (41.95)	<.001
	Male	2366 (63.09)	1558 (66.07)	808 (58.05)	
Age		54.8 ± 12.3	53.3 ± 12.8	57.2 ± 11.1	<.001
BMI		25.9 ± 3.7	26.0 ± 3.8	25.7 ± 3.4	.021
Smoking	No	2798 (74.61)	1724 (73.11)	1074 (77.16)	.007
	Yes	952 (25.39)	634 (26.89)	318 (22.84)	
Drinking	No	2912 (77.65)	1803 (76.46)	1109 (79.67)	.025
	Yes	838 (22.35)	555 (23.54)	283 (20.33)	
Hypertension	No	1922 (51.25)	1313 (55.68)	609 (43.75)	<.001
	Yes	1828 (48.75)	1045 (44.32)	783 (56.25)	
Insulin therapy	No	1484 (39.57)	1248 (52.93)	236 (16.95)	<.001
	Yes	2266 (60.43)	1110 (47.07)	1156 (83.05)	
Course of diabetes		10.0 (4.0, 15.0)	6.0 (3.0, 10.1)	12.0 (8.0, 20.0)	<.001
HbA1c		8.7 (7.3, 10.2)	8.2 (7.0, 9.8)	9.3 (8.2, 10.7)	<.001
TC		4.9 (4.2, 5.7)	4.9 (4.3, 5.7)	4.8 (4.1, 5.6)	.001
HDL		1.1 (1.1, 1.3)	1.1 (1.1, 1.2)	1.1 (0.9, 1.3)	<.001
LDL		3.0 (2.4, 3.5)	3.0 (2.5, 3.6)	2.9 (2.3, 3.4)	<.001
TG		2.0 (1.3, 3.4)	2.0 (1.3, 3.2)	2.0 (1.4, 4.0)	.072
FBG		8.9 (6.9, 12.0)	8.5 (6.7, 11.6)	9.7 (7.3, 12.6)	<.001
TyG		2.2 (1.7, 2.9)	2.2 (1.6, 2.8)	2.3 (1.7, 3.0)	<.001

5058 negative cases were extracted *via* the computer random sampling method. Finally, 622 cases were excluded from the positive group (including 402 with missing data, 108 with other diseases affecting vision, 108 with vision surgery, and 94 with critical illness) and 2700 cases were excluded from the negative group (2305 with missing data and 395 with other serious systemic diseases), leaving 3755 cases (1392 cases in the positive group and 2358 cases in the negative group) About 63.09% of the remaining cases were males and 36.91% were females. The subjects were aged 18~91 years, with an average age of 54.8 ± 12.3 (Table 1).

The variables with univariate $P < 0.05$ were included in multivariate logistics regression analysis. The results showed that HDL ($P < .001$, OR = 0.176, 95% CI:0.144~0.215), smoking ($P = .002$, OR = 0.759, 95% CI:0.620~0.92), LDL ($P = .021$, OR = 0.811, 95% CI:0.678~0.969), BMI ($P = .001$, OR = 0.815, 95% CI:0.726~0.915), TyG ($P = .038$, OR = 1.091, 95% CI:1.005, 1.185), age ($P = .014$, OR = 1.149, 95% CI:1.028~1.284), hypertension history ($P < .001$, OR = 1.506, 95% CI:1.273~1.781), course of T2DM ($P < .001$, OR = 1.798, 95% CI:1.615~2.002), insulin therapy ($P < .001$, OR = 3.166, 95% CI:2.621~3.824), HbA1c ($P < .001$, OR = 3.443, 95% CI:2.879~4.116), were the independent influencing factors of DR (Table 2).

3.2 RFECV screening

The independent variables were selected *via* RFECV method, and recursive features were eliminated. The results showed that nine indexes, including age, BMI, Diabetes course, Insulin therapy, Hypertension, HbA1c, TC, HDL, and TyG, were retained, and five indexes (sex, smoking, drinking, FBG, and LDL) were excluded (Figure 2).

3.3 Training and verification of the model

The data were grouped into the training set and testing set (7:3). The last nine indexes were used as the optimal solution for model training. Moreover, SVM and LR showed the best performance in the test set and training set. The AUC curves of the five models are shown in Figure 3. The comparison of various prediction indexes (Table 3).

Five-fold cross-validation confirmed model stability: Logistic Regression: AUC 0.824 ± 0.012, Accuracy 0.760 ± 0.007; SVM: AUC 0.815 ± 0.007, Accuracy 0.762 ± 0.016; The minimal standard deviations (<0.02) indicate robust performance across data subsets.

TABLE 2 Results of multivariate logistic regression analysis.

Variable	B	Wald	P	OR (95%CI)
HDL	-1.736	291.731	<.001	0.176(0.144, 0.215)
Smoking	-0.275	7.136	.008	0.759(0.620, 0.929)
LDL	-0.21	5.3	.021	0.811(0.678, 0.969)
BMI	-0.205	12.035	.001	0.815(0.726, 0.915)
TyG	0.087	4.319	.038	1.091(1.005, 1.185)
Age	0.139	6.005	.014	1.149(1.028, 1.284)
Hypertension	0.409	22.879	<.001	1.506(1.273, 1.781)
Course of diabetes	0.587	115.088	<.001	1.798(1.615, 2.002)
Insulin_therapy	1.152	143.06	<.001	3.166(2.621, 3.824)
HbA1c	1.236	183.72	<.001	3.443(2.879, 4.116)

3.4 SHAP model analysis

3.4.1 Ranking of feature importance

The top 5 variables based on SHAP values in the ML model are shown in Figure 4. A summary plot is shown in Figure 4B, where “red” and “blue” represent higher and lower eigenvalues, respectively. SHAP value<0 indicates negative influence, SHAP value>0 indicates positive influence. The dispersion of feature distribution (located above the Y axis) was directly related to the importance of the feature. Furthermore, TyG, Insulin therapy, HbA1c, and Diabetes course showed positive effects on retinopathy, while HDL showed negative effects.

3.4.2 Practical application of the model

The force Plot is shown in Figure 5. The blue and red arrows indicate that this factor reduces and increases the risk of retinopathy, respectively. The reference value represents the average SHAP value of all samples. $F(x)$ represents the comprehensive SHAP value of each patient. The model can only predict the patient's retinopathy if the value of $f(x)$ is greater than the base value. The 1045th case (Figure 5) was randomly predicted using the test set. Notably, $f(x)$ was less than the base value, which was accurately predicted as the control group.

3.4.3 The effect of key features on outcomes

A partial dependence diagram of the influence of the first three indicators on infection was drawn, showing the marginal effect relationship between important characteristics and outcome variables (how important influencing factors affect retinopathy). TyG > 4 showed significant impact (Figure 6). Multicollinearity assessment showed all VIF values were below 1.4 for both machine learning and logistic regression models, well under the threshold of 5, indicating no significant multicollinearity concerns.

4 Discussion

Early screening and identification of people at high risk of developing DR is important for prevention and treatment.

However, Li et al. (39) showed that only 17.48% of T2DM undergo routine DR Screening yearly. Doctors in the diabetes department, especially in primary hospitals that lack equipments and professionals, often overlook the early signs of retinopathy in their patients (36, 37). This study introduces a paradigm shift in diabetic retinopathy (DR) risk stratification by integrating the triglyceride-glucose index (TyG) – an easily measurable, low-cost surrogate of insulin resistance (IR) – into machine learning prediction models. Unlike prior models that either overlook IR (40–42) or rely on impractical biomarkers like HOMA-IR (requiring specialized insulin assays) (11), our approach leverages routinely available lipid/glucose data, addressing a critical barrier in primary care settings. The prominent contribution of TyG in SHAP analysis (Figures 4–6) further establishes its role as a novel, interpretable predictor for scalable DR screening.

Results showed that logistic regression and SVM had the best performance among the five models, with area under ROC curve of the testing set and training set of 0.85 and 0.82 respectively. Similarly, Tsao et al. (43) showed that the SVM model has good predictive performance. Jiang et al. (37) also showed that logistic regression has good predictive performance. LR is a linear model suitable for small sample analysis, and it is sensitive to outliers. SVM finds the optimal hyperplane by optimizing the objective function, which is suitable for the analysis of complex data (37). Different algorithms have their own advantages in the modeling process, and no algorithm can be applied to all models. Therefore, the corresponding model should be selected according to the research design. This study was conducted based on the computational minimum sample size. Therefore, a study with a larger sample size is needed to compare the performance between LR and SVM. Our SVM model demonstrated test set AUC (0.85) comparable to Jiang et al. (37) (AUC = 0.89) and superior to Wang et al. (40) (AUC = 0.709). The integration of TyG provided predictive power comparable to HOMA-IR-based models (11) without requiring specialized tests. Unlike Roşu et al. (42) (AUC = 0.72) and Yang et al. (32) (AUC = 0.78) which excluded insulin resistance indicators, our TyG-incorporated model achieved higher AUC (0.85) with clinically accessible variables. This balances performance and feasibility for primary care settings. While our initial hold-out test set showed excellent performance (AUC >0.83), the additional 5-fold cross-validation provides stronger evidence of model robustness. The cross-validated AUC remained above 0.81 with standard deviations <0.02, confirming reliable performance across diverse data partitions.

Herein, the variables included in LR and SVM were consistent. The remaining variables were described in previous studies except for the significant predictive ability of TyG (9, 31, 37, 44, 45). Previous studies showed that the interpretability of ML models is challenging (46). In the present study, SHAP was used to improve model interpretability. SHAP is used to interpret a “black box” model that calculates a Shapley value for each feature in the prediction model to assess the importance of all feature combinations, reflecting their contribution to the predictive power of the overall model (26). SHAP results showed that the top 5 variables included TyG, HDL, Insulin therapy, Diabetes course, and HbA1c.

This study observed discordant results for Smoking and LDL between univariate and multivariate analyses, with a notable

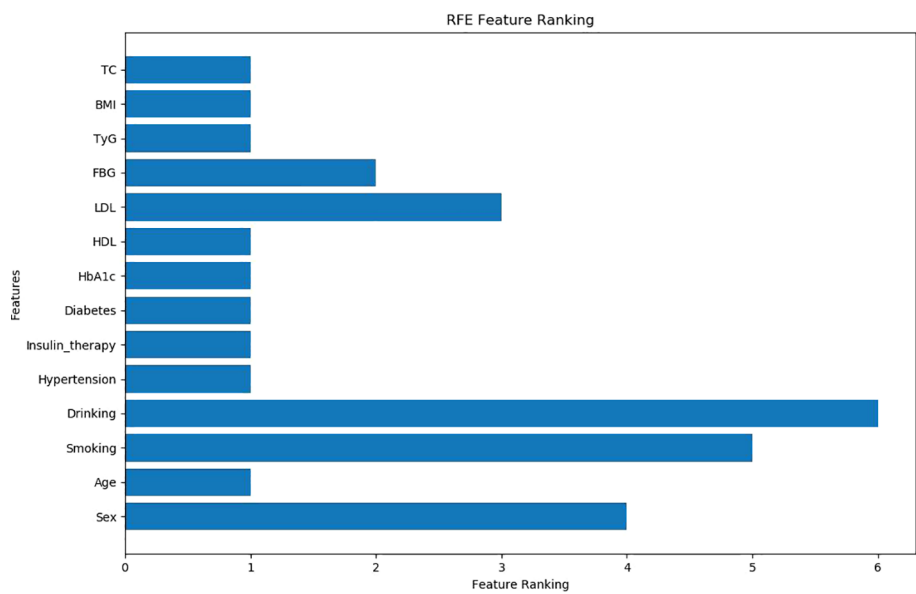


FIGURE 2
Selection results of RFECV method (9 indexes were retained, 5 indexes were excluded).

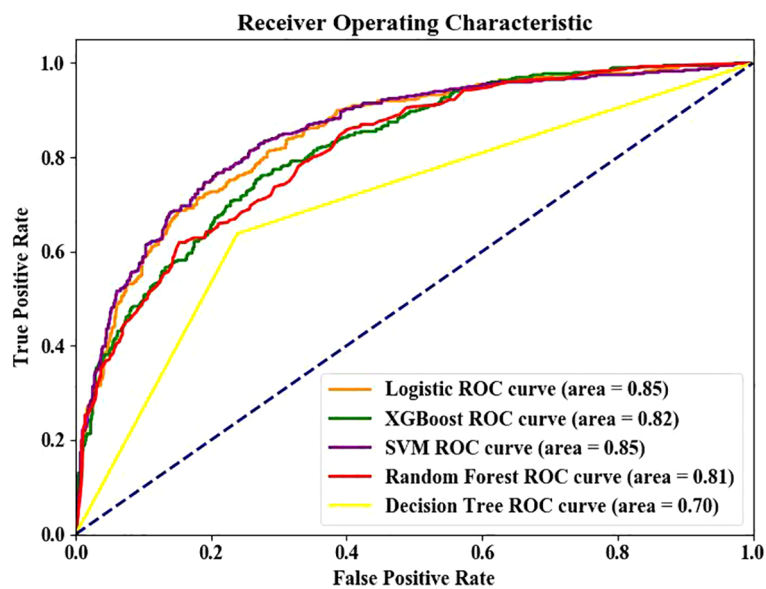


FIGURE 3
Receiver operating characteristic (ROC) curves based of five machine learning models.

TABLE 3 Prediction performance indicators of five models in the training and testing sets.

Model	Train set				Test set			
	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
LR	0.70	0.58	0.64	0.75	0.70	0.69	0.70	0.79
XGBoost	0.69	0.62	0.65	0.75	0.65	0.62	0.64	0.75
SVM	0.70	0.60	0.65	0.75	0.73	0.67	0.7	0.80
RF	0.68	0.61	0.64	0.74	0.64	0.64	0.64	0.75
DT	0.59	0.56	0.58	0.69	0.59	0.64	0.62	0.72

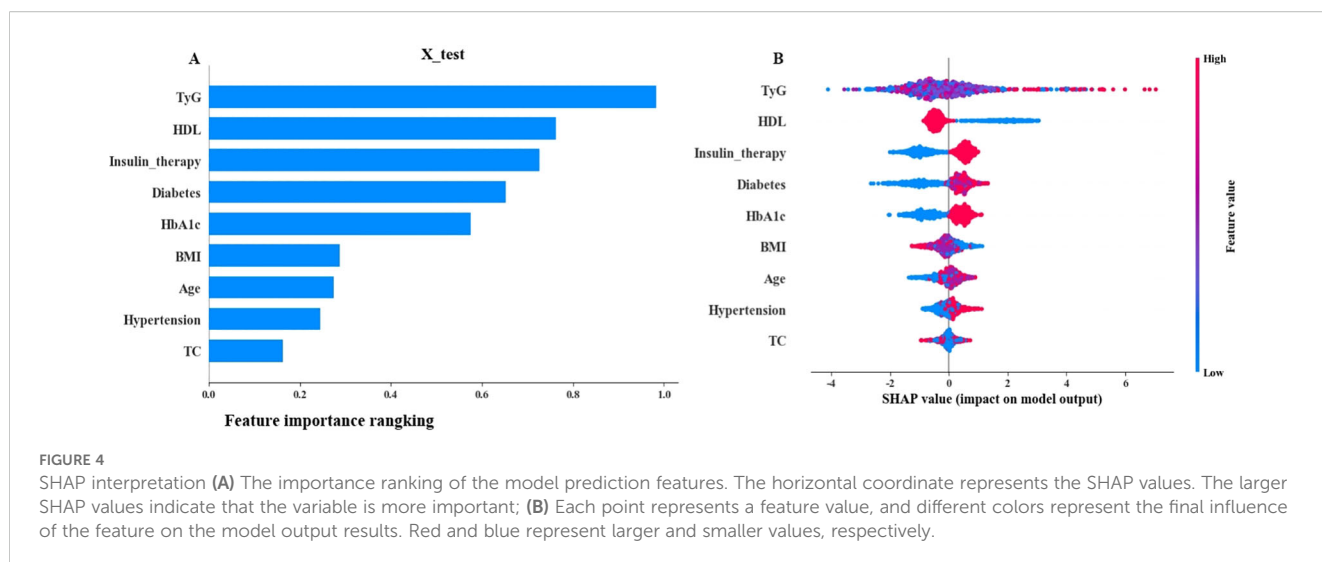


FIGURE 4

SHAP interpretation (A) The importance ranking of the model prediction features. The horizontal coordinate represents the SHAP values. The larger SHAP values indicate that the variable is more important; (B) Each point represents a feature value, and different colors represent the final influence of the feature on the model output results. Red and blue represent larger and smaller values, respectively.

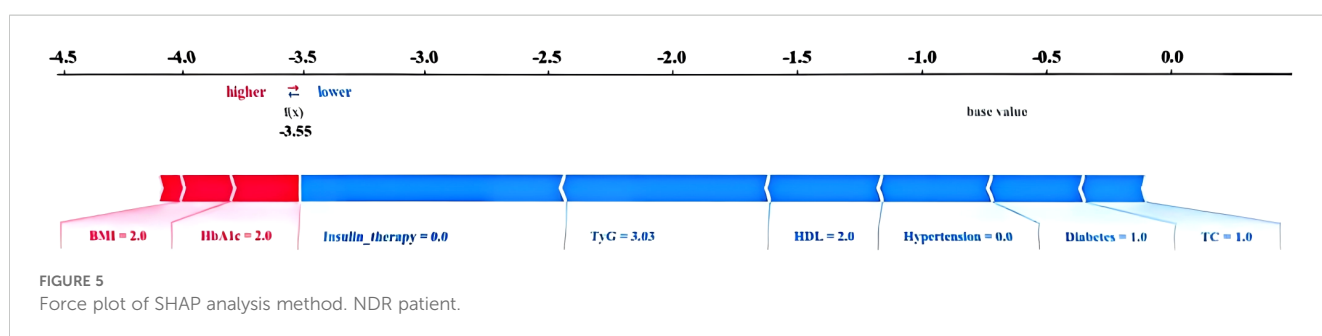


FIGURE 5

Force plot of SHAP analysis method. NDR patient.

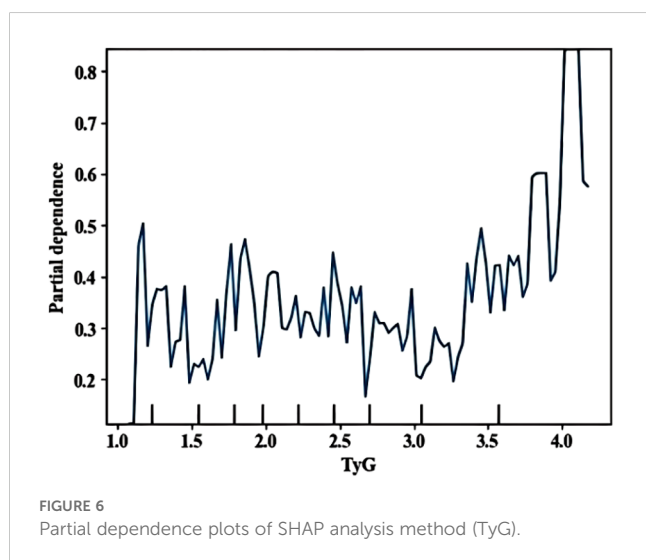


FIGURE 6

Partial dependence plots of SHAP analysis method (TyG).

reversal in smoking's effect direction (univariate: lower prevalence in DR group, $P = 0.007$; multivariate: protective association, OR = 0.759, $P = 0.008$). This reversal is explained by two synergistic mechanisms: First, hypertension mediation—hypertension was more prevalent in the DR group (56.25% vs 44.32%, $P < 0.001$) and strongly predicted DR (OR = 1.506, $P < 0.001$); since smoking

promotes hypertension (47), adjusting for it in multivariate models revealed smoking's residual protective effect *via* metabolic pathways. Second, exclusion of severe comorbidities may favor resilient smokers, consistent with the "smoker's paradox" in chronic disease cohorts (48, 49). Regarding LDL and HDL, their reversals reflect collinearity within lipid metabolism (univariate analysis failed to disentangle HDL/LDL effects from TyG, which correlates with triglycerides) and a synergistic risk interaction where low HDL combined with high TyG increases DR risk—captured only by multivariate models. These reversals underscore that risk factors' net effects depend on comorbidities: Heavy smokers with hypertension remain high-risk despite smoking's statistical "protection," while LDL's reversal highlights complex lipid interactions in DR pathogenesis.

TyG emerged as a key predictor with dual significance. Notably, increased TyG levels were associated with a high risk of DR due to the excessive production of mitochondrial superoxide in microvascular endothelial cells. This production is caused by pathway-specific insulin resistance, meanwhile, triggering intracellular hyperglycemia and vascular damage (50). As a biochemical indicator reflecting synergistic dysregulation of glucose/lipid metabolism (51), TyG directly links to DR pathogenesis. Our SHAP analysis revealed a "non-linear risk surge" at $TyG > 4$, aligning with the point where dysregulated glucose/lipid metabolism triggers mitochondrial superoxide

overproduction in retinal microvessels – a key driver of early vascular damage. Based on multivariate regression, each unit increase in TyG elevates DR risk by 9.1% (OR = 1.091, 95% CI: 1.005–1.185). The threshold effect at TyG > 4 reflects amplified risk accumulation consistent with Zhou et al.'s (52) report. This threshold is calculable from routine lipid/glucose tests. We recommend annual fundus screening for patients exceeding TyG > 4. Crucially, TyG overcomes limitations of traditional IR metrics. Our study is the first to operationalize TyG within a predictive algorithm, with its threshold effect (TyG > 4) providing actionable clinical guidance. Though TyG cut-offs vary by disease context [e.g., 4.65 for IR (16), 8.88 for diabetes progression (53)], our threshold specifically targets DR risk stratification in primary care settings where standardized values remain undefined. A meta-analysis (54) showed that there is no standardized threshold value for TyG and clear threshold value related to TyG and DR. Therefore, more high-quality studies are needed to determine the optimal cut-off point for TyG in the future.

Insulin therapy is another key predictor of DR. Wang, Li et al (40, 45) found that the insulin treatment is associated with a high risk of DR development, possibly because insulin therapy indicates worse islet function. While Ricard found it possibly related to the rapid reduction of blood glucose (55). Nonetheless, more basic research is needed to confirm the findings. This study, along with others, has demonstrated that the duration of diabetes mellitus is closely associated with the development of DR (30, 41, 44, 56), possibly due to the prolonged exposure of blood vessels to risk factors.

Herein, DR was linked to increased HbA1c levels. This may be due to the continuous elevation of blood glucose, which can lead to dysfunction of the retinal vascular endothelium and cause retinal ischemia and increased vascular permeability (2). Similar findings have been reported previously (11, 30, 45).

In addition, HDL was identified as a key predictor of DR. Low HDL levels may indicate a higher risk of DR, consistent with Roşu, Liu, Li et al (42, 57, 58).

By addressing three critical gaps – 1) IR assessment simplification *via* TyG, 2) cross-tier validation (urban tertiary/rural primary hospitals), and 3) SHAP-driven interpretability – our model enables scalable DR screening. Our model requires only 9 routinely collected variables, enabling seamless integration into electronic health records (EHR) for real-time risk stratification during diabetic outpatient visits. Based on SHAP probability thresholds (Figures 4–6), we propose a tiered screening pathway: High-risk patients (TyG > 4 and other key predictors elevated): Prioritize immediate ophthalmologist referral. Low-to-intermediate-risk patients: Maintain routine annual/biannual retinal screening. This approach is particularly valuable for resource-limited settings (e.g., rural clinics), where the model can optimize specialist resource allocation by focusing on high-risk individuals identified through universally available tests like TyG (calculated from routine lipid/glucose tests). Besides, Our model exhibits inherent generalizability potential through its reliance on universal predictors requiring only routine clinical data, demonstrated demographic resilience across diverse populations, and established scalability pathways including a simplified screening protocol and ongoing multi-region validation.

This study bridges a translational gap: while TyG-DR associations were reported, we are the first to validate its clinical utility within a predictive algorithm. The integration with SHAP interpretation further distinguishes our model from conventional approaches. While retrospective data from two hospitals with different resource levels (urban tertiary vs. rural primary) may introduce selection bias, this heterogeneity enhances real-world generalizability.

Limitation: First, only internal verification was conducted, thus effective external verification is needed. Secondly, excluding high-missingness variables may omit useful information, and while mean/mode imputation is practical, advanced techniques (e.g., multiple imputation) might reduce potential bias. Last, this study did not differentiate early vs. late-stage DR in predictions due to ungraded diagnostic records. However, our model prioritizes biomarkers linked to DR progression (TyG, diabetes course). Future work should incorporate standardized severity staging to enable stage-specific optimization.

5 Conclusion and recommendations

In this study, a DR prediction model was built using TyG and other easily available clinical data. LR and SVM models had the best performance. SHAP showed that the most important predictors of DR were TyG, insulin therapy, diabetes course, HbA1c, and HDL. These findings suggest that baseline characteristics of patients can be used to screen for high-risk DR to improve early diagnosis and implement timely referral for patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Research Ethics Committee of The First Hospital of Jilin University (approval number: 2024-918). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

XL: Investigation, Writing – original draft, Writing – review & editing. SW: Data curation, Investigation, Writing – original draft, Writing – review & editing. YY: Data curation, Writing – review & editing. YL: Data curation, Writing – review & editing. XZ: Data curation, Writing – review & editing. RL: Software, Writing – review & editing. LQ: Data curation, Writing – original draft.

FL: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. The study was supported by the grants from Jilin Medical and Health Talents Special Project (JLSWSRCZX2023-29).

Acknowledgments

We express our gratitude to all the participants of the study, including patients and researchers. We thank Core Facility of the First Hospital of Jilin University for their help and valuable comments on this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract.* (2022) 183:109119. doi: 10.1016/j.diabres.2021.109119
- Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet.* (2010) 376:124–36. doi: 10.1016/S0140-6736(09)62124-3
- Congdon NG, Friedman DS, Lietman T. Important causes of visual impairment in the world today. *JAMA-J. Am Med Assoc.* (2003) 290:2057–60. doi: 10.1001/jama.290.15.2057
- Maniadakis N, Konstantakopoulou E. Cost effectiveness of treatments for diabetic retinopathy: A systematic literature review. *Pharmacoeconomics.* (2019) 995–1010. doi: 10.1007/s40273-019-00800-w
- Teo ZL, Tham Y-C, Yu M, Chee ML, Rim TH, Cheung N, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology.* (2021) 128:1580–91. doi: 10.1016/j.ophtha.2021.04.027
- Zhang G, Chen H, Chen W, Zhang M. Prevalence and risk factors for diabetic retinopathy in China: a multi-hospital-based cross-sectional study. *Br J Ophthalmol.* (2017) 101:1591–5. doi: 10.1136/bjophthalmol-2017-310316
- Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T, et al. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol.* (2020) 8:337–47. doi: 10.1016/S2213-8587(19)30411-5
- Usman TM, Saheed YK, Nsang A, Ajibesin A, Rakshit S. A systematic literature review of machine learning based risk prediction models for diabetic retinopathy progression. *Artif Intell Med.* (2023) 143:102617. doi: 10.1016/j.artmed.2023.102617
- Guo H, Han F, Qu J-R, Pan C-Q, Sun B, Chen L-M. Scoring and validation of a simple model for predicting diabetic retinopathy in patients with type 2 diabetes based on a meta-analysis approach of 21 cohorts. *Ann Med.* (2024) 56:2413920. doi: 10.1080/07853890.2024.2413920
- Amoaku WM, Ghanchi F, Bailey C, Banerjee S, Banerjee S, Downey L, et al. Diabetic retinopathy and diabetic macular oedema pathways and management: UK Consensus Working Group. *Eye (Lond).* (2020) 34:1–51.
- Wang Q, Zeng N, Tang H, Yang X, Yao Q, Zhang L, et al. Diabetic retinopathy risk prediction in patients with type 2 diabetes mellitus using a nomogram model. *Front Endocrinol (Lausanne).* (2022) 13:993423. doi: 10.3389/fendo.2022.993423
- Gastaldelli A. Measuring and estimating insulin resistance in clinical and research settings. *Obes (Silver Spring Md.).* (2022) 30:1549–63.
- Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia.* (1985) 28:412–9. doi: 10.1007/BF00280883
- Zhao Q, Zhang T-Y, Cheng Y-J, Ma Y, Xu Y-K, Yang J-Q, et al. Impacts of triglyceride-glucose index on prognosis of patients with type 2 diabetes mellitus and non-ST-segment elevation acute coronary syndrome: results from an observational cohort study in China. *Cardiovasc Diabetol.* (2020) 19:108. doi: 10.1186/s12933-020-01086-5
- Wang A, Tian X, Zuo Y, Chen S, Meng X, Wu S, et al. Change in triglyceride-glucose index predicts the risk of cardiovascular disease in the general population: a prospective cohort study. *Cardiovasc Diabetol.* (2021) 20:113. doi: 10.1186/s12933-021-01305-7
- Simental-Mendía LE, Rodríguez-Morán M, Guerrero-Romero F. The product of fasting glucose and triglycerides as surrogate for identifying insulin resistance in apparently healthy subjects. *Metab Syndrome Related Disord.* (2008) 6:299–304. doi: 10.1089/met.2008.0034
- Abbasi F, Reaven GM. Comparison of two methods using plasma triglyceride concentration as a surrogate estimate of insulin action in nondiabetic subjects: triglycerides $\times$ glucose versus triglyceride/high-density lipoprotein cholesterol. *Metabolism: Clin Exp.* (2011) 60:1673–6. doi: 10.1016/j.metabol.2011.04.006
- Park HM, Lee HS, Lee Y-J, Lee J-H. The triglyceride-glucose index is a more powerful surrogate marker for predicting the prevalence and incidence of type 2 diabetes mellitus than the homeostatic model assessment of insulin resistance. *Diabetes Res Clin Pract.* (2021) 180:109042. doi: 10.1016/j.diabres.2021.109042
- Srinivasan S, Singh P, Kulothungan V, Sharma T, Raman R. Relationship between triglyceride glucose index, retinopathy and nephropathy in Type 2 diabetes. *Endocrinology Diabetes Metab.* (2021) 4:e00151. doi: 10.1002/edmd.151
- Yao L, Wang X, Zhong Y, Wang Y, Wu J, Geng J, et al. The triglyceride-glucose index is associated with diabetic retinopathy in chinese patients with type 2 diabetes: A hospital-based, nested, case-control study. *Diabetes Metab Syndrome Obesity: Targets Ther.* (2021) 14:1547–55. doi: 10.2147/DMSO.S294408

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2025.1631647/full#supplementary-material>

21. Haq MA, Khan I, Ahmed A, Eldin SM, Alshehri A, Ghamry NA. DCNNBT: A NOVEL DEEP CONVOLUTION NEURAL NETWORK-BASED BRAIN TUMOR CLASSIFICATION MODEL. *Fractals*. (2023) 31:2340102. doi: 10.1142/S0218348X23401023
22. Ganesh S, Gomathi R, Kannadhasan S. Brain tumor segmentation and detection in MRI using convolutional neural networks and VGG16. *Cancer biomark.* (2025) 42:18758592241311184. doi: 10.1177/18758592241311184
23. Yousef R, Khan S, Gupta G, Siddiqui T, Albahall BM, Alajlan SA, et al. U-net-based models towards optimal MR brain image segmentation. *Diagnostics (Basel Switzerland)*. (2023) 13:1624. doi: 10.3390/diagnostics13091624
24. Alhussen A, Anul Haq M, Ahmad Khan A, Mahendran RK, Kadry S. XAI-RACapsNet: Relevance aware capsule network-based breast cancer detection using mammography images via explainability O-net ROI segmentation. *Expert Syst Appl.* (2025) 261:125461. doi: 10.1016/j.eswa.2024.125461
25. Theodosiou AA, Read RC. Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician. *J Infect.* (2023) 87:287–94. doi: 10.1016/j.jinf.2023.07.006
26. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: *Proceedings of the Thirty-first Annual Conference on Neural Information Processing Systems (NeurIPS)* (2017). Long Beach, California, USA: Curran Associates Inc. 4768–77
27. Somvanshi S, Das S, Aaqib Javed S, Antariksh G, Hossain A. A survey on deep tabular learning. (2024). 2410.12034
28. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. (2015) 350:g7594. doi: 10.1161/CIRCULATIONAHA.114.014508
29. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest.* (2012) 42:216–28. doi: 10.1111/j.1365-2362.2011.02562.x
30. Mulat Tebeje T, Kindie Yenit M, Gedlu Nigatu S, Bizuneh Mengistu S, Kidie Tesfie T, Byadgie Gelaw N, et al. Prediction of diabetic retinopathy among type 2 diabetic patients in University of Gondar Comprehensive Specialized Hospital, 2006–2021: A prognostic model. *Int J Med Inform.* (2024) 190:105536. doi: 10.1016/j.jiminf.2024.105536
31. Jin S, Zhang X, Liu H, Hao J, Cao K, Lin C, et al. Identification of the optimal model for the prediction of diabetic retinopathy in chinese rural population: handan eye study. *J Diabetes Res.* (2022) 2022:4282953. doi: 10.1155/2022/4282953
32. Yang J, Jiang S. Development and validation of a model that predicts the risk of diabetic retinopathy in type 2 diabetes mellitus patients. *Acta Diabetol.* (2023) 60:43–51. doi: 10.1007/s00592-022-01973-1
33. Li H-Y, Dong L, Zhou W-D, Wu H-T, Zhang R-H, Li Y-T, et al. Development and validation of medical record-based logistic regression and machine learning models to diagnose diabetic retinopathy. *Graefes Arch Clin Exp Ophthalmol.* (2023) 261:681–9. doi: 10.1007/s00417-022-05854-9
34. American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: standards of care in diabetes-2024. *Diabetes Care.* (2024) 47:S20–42. doi: 10.2337/dc24-S002
35. Wong TY, Cheung CMG, Larsen M, Sharma S, Simó R. Diabetic retinopathy. *Nat Rev Dis Primers.* (2016) 2:16012. doi: 10.1038/nrdp.2016.12
36. Flaxel CJ, Adelman RA, Bailey ST, Fawzi A, Lim JJ, Vemulakonda GA, et al. Diabetic retinopathy preferred practice pattern®. *Ophthalmology.* (2020) 127:P66–P145. doi: 10.1016/j.ophtha.2019.09.025
37. Jiang W, Li Z. Comparison of machine learning algorithms and nomogram construction for diabetic retinopathy prediction in type 2 diabetes mellitus patients. *Ophthalmic Res.* (2024) 67:537–48. doi: 10.1159/000541294
38. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ.* (2020) 368:m441. doi: 10.1136/bmj.m441
39. Li W-Y, Yang M, Song Y-N, Luo L, Nie C, Zhang M-N. An online diabetic retinopathy screening tool for patients with type 2 diabetes. *Int J Ophthalmol.* (2021) 14:1748–55. doi: 10.18240/ijo.2021.11.15
40. Wang G-X, Hu X-Y, Zhao H-X, Li H-L, Chu S-F, Liu D-L. Development and validation of a diabetic retinopathy risk prediction model for middle-aged patients with type 2 diabetes mellitus. *Front Endocrinol (Lausanne).* (2023) 14:1132036. doi: 10.3389/fendo.2023.1132036
41. Yang Y, Tan J, He Y, Huang H, Wang T, Gong J, et al. Predictive model for diabetic retinopathy under limited medical resources: A multicenter diagnostic study. *Front Endocrinol (Lausanne).* (2022) 13:1099302. doi: 10.3389/fendo.2022.1099302
42. Roşu CD, Bratu ML, Stoicescu ER, Iacob R, Haegan OA, Ghenciu LA, et al. Cardiovascular risk factors as independent predictors of diabetic retinopathy in type II diabetes mellitus: the development of a predictive model. *Med (Kaunas Lithuania).* (2024) 60:1617.
43. Tsao H-Y, Chan P-Y, Su EC-Y. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinf.* (2018) 19:283. doi: 10.1186/s12859-018-2277-0
44. Chen X, Xie Q, Zhang X, Lv Q, Liu X, Rao H. Nomogram prediction model for diabetic retinopathy development in type 2 diabetes mellitus patients: A retrospective cohort study. *J Diabetes Res.* (2021) 2021:3825155. doi: 10.1155/2021/3825155
45. Li W, Song Y, Chen K, Ying J, Zheng Z, Qiao S, et al. Predictive model and risk analysis for diabetic retinopathy using machine learning: a retrospective cohort study in China. *BMJ Open.* (2021) 11:e050989. doi: 10.1136/bmjopen-2021-050989
46. Luo H, Xiang C, Zeng L, Li S, Mei X, Xiong L, et al. SHAP based predictive modeling for 1 year all-cause readmission risk in elderly heart failure patients: feature selection and model interpretation. *Sci Rep.* (2024) 14:17728. doi: 10.1038/s41598-024-67844-7
47. Lauder L, Mahfoud F, Azizi M, Bhatt DL, Ewen S, Kario K, et al. Hypertension management in patients with cardiovascular comorbidities. *Eur Heart J.* (2023) 44:2066–77. doi: 10.1093/eurheartj/ehac395
48. Hu G, Gu H, Jiang Y, Yang X, Jiang Y, Wang C, et al. Revisiting the smoking paradox in acute ischemic stroke patients: findings from the chinese stroke center alliance study. *J Am Heart Assoc.* (2023) 12:e029963. doi: 10.1161/JAHA.123.029963
49. Gupta T, Kolte D, Khera S, Harikrishnan P, Mujib M, Aronow WS, et al. Smoker's paradox in patients with ST-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention. *J Am Heart Assoc.* (2016) 5:e003370. doi: 10.1161/JAHA.116.003370
50. Giacco F, Brownlee M. Oxidative stress and diabetic complications. *Circ Res.* (2010) 107:1058–70. doi: 10.1161/CIRCRESAHA.110.223545
51. Augustine J, Troendle EP, Barabas P, McAleese CA, Friedel T, Stitt AW, et al. The role of lipoxidation in the pathogenesis of diabetic retinopathy. *Front Endocrinol (Lausanne).* (2020) 11:621938. doi: 10.3389/fendo.2020.621938
52. Zhou J, Zhu L, Li Y. Association between the triglyceride glucose index and diabetic retinopathy in type 2 diabetes: a meta-analysis. *Front Endocrinol (Lausanne).* (2023) 14:1302127. doi: 10.3389/fendo.2023.1302127
53. Chen X, Liu D, He W, Hu H, Wang W. Predictive performance of triglyceride glucose index (TyG index) to identify glucose status conversion: a 5-year longitudinal cohort study in Chinese pre-diabetes people. *J Transl Med.* (2023) 21:624. doi: 10.1186/s12967-023-04402-1
54. Sánchez-García A, Rodríguez-Gutiérrez R, Mancillas-Adame L, González-Nava V, Díaz González-Colmenero A, Solís RC, et al. Diagnostic accuracy of the triglyceride and glucose index for insulin resistance: A systematic review. *Int J Endocrinol.* (2020) 2020:4678526. doi: 10.1155/2020/4678526
55. Ricard N, Bailly S, Guignabert C, Simons M. The quiescent endothelium: signalling pathways regulating organ-specific endothelial normalcy. *Nat Rev Cardiol.* (2021) 18:565–80. doi: 10.1038/s41569-021-00517-4
56. Semeraro F, Parrinello G, Cancarini A, Pasquini L, Zarra E, Cimino A, et al. Predicting the risk of diabetic retinopathy in type 2 diabetic patients. *J Diabetes its Complications.* (2011) 25:292–7. doi: 10.1016/j.jdiacomp.2010.12.002
57. Liu Y, Zhou Z, Wang Z, Yang H, Zhang F, Wang Q. Construction and clinical validation of a nomogram-based predictive model for diabetic retinopathy in type 2 diabetes. *Am J Transl Res.* (2023) 15:6083–94.
58. Li Y, Hu B, Lu L, Li Y, Caika S, Song Z, et al. Development and external validation of a predictive model for type 2 diabetic retinopathy. *Sci Rep.* (2024) 14:16741. doi: 10.1038/s41598-024-67533-5