Am J Transl Res 2023;15(10):6083-6094
www.ajtr.org /ISSN:1943-8141/AJTR0152913

*Original Article*

# Construction and clinical validation of a nomogram-based predictive model for diabetic retinopathy in type 2 diabetes

Yonghong Liu[1*], Zhaoling Zhou[2*], Zhiyong Wang[3], Huan Yang[4], Feifei Zhang[4], Qinhu Wang[4]

*[1]Department of Ophthalmology, Gansu Provincial Hospital of TCM, No. 418 Guazhou Road, Qilihe District, Lanzhou 730050, Gansu, China; [2]Department of Electrocardiogram, Gansu Provincial Hospital of TCM, No. 418 Guazhou Road, Qilihe District, Lanzhou 730050, Gansu, China; [3]Trauma Disease Diagnosis and Treatment Center, Gansu Provincial Hospital of TCM, No. 418 Guazhou Road, Qilihe District, Lanzhou 730050, Gansu, China; [4]Department of Endocrinology, The Third People's Hospital of Gansu Province, No. 763 Duanjiatan Road, Chengguan District, Lanzhou 730020, Gansu, China. [*]Equal contributors.*

Received August 13, 2023; Accepted September 14, 2023; Epub October 15, 2023; Published October 30, 2023

**Abstract:** Objective: This study aimed to identify risk factors for diabetic retinopathy (DR) in patients with type 2 diabetes mellitus (T2DM) and construct a nomogram prediction model for DR. Methods: T2DM patients (n = 520) who underwent funduscopic examinations from June 2020 to June 2022 were included. Of these patients, 220 had DR, yielding a disease rate of 40.38%. Patients were divided into a training set (n = 364) and a validation set (n = 156) at a 7:3 ratio. Feature variables were selected using LASSO regression, random forests, and decision trees. Venn diagrams identified common DR feature variables. The prediction model's validity was assessed using the C-index, decision curve analysis (DCA), receiver operating characteristic (ROC) curves, and calibration curves. Results: Factors influencing DR were age, Diabetic Peripheral Neuropathy (DPN), Hemoglobin A1C (HbA1C) levels, High-Density Lipoprotein (HDL) cholesterol, Low-Density Lipoprotein (LDL) cholesterol, Neutrophil-to-Lymphocyte Ratio (NLR), Triglycerides (TG), Blood Urea Nitrogen (BUN), and disease duration. Univariate analysis excluded LDL as being unrelated to DR. The DR prediction model, constructed using the remaining eight variables, showed internal validation metrics with a C-index of 0.937, area under the ROC curve (AUC) of 0.773, and DCA net benefit of 11%-95%. The external validation metrics demonstrated a C-index of 0.916, AUC of 0.735, and DCA net benefit of 17%-93%. Calibration curves indicated high consistency. Conclusion: This study developed a nomogram prediction model to assess the risk of DR in patients with T2DM. The model demonstrated high precision through internal validation.

**Keywords:** Type 2 diabetes mellitus, diabetic retinopathy, risk factors, nomogram, predictive model, LASSO

## Introduction

With the global improvement in quality of life, particularly in developing countries, the incidence of diabetes is steadily rising [1]. China, accounting for a third of the global diabetic population, has a diabetes prevalence rate of about 10% [2]. Type 2 diabetes mellitus (T2DM) is a chronic hyperglycemia disease that can exert stress on the microvasculature of multiple organ systems including the eyes [3]. Diabetic Retinopathy (DR), a complication of diabetes primarily due to metabolic dysregulation leading to abnormal ocular vasculature, is one of the leading causes of adult vision decline and visual impairment, and it has the potential to culminate in irreversible blindness [4]. Early DR has inconspicuous symptoms, but the onset of blindness is often acute. Studies have shown that in patients with T2DM, 60% are likely to develop DR within 20 years [5]. In China, nearly 28% of T2DM patients develop DR, among which approximately 45% of DR patients' vision is threatened [6], and there are currently no effective methods to reverse vision loss caused by DR.

Persistent hyperglycemia can lead to abnormal polyol metabolism, enhance vascular inflammatory response, and induce oxidative stress.

These factors collectively modify the biological properties and hemodynamics of the retinal microvasculature, thus leading to retinopathy [7]. The severe visual impairment caused by DR not only reduces patients' quality of life and hampers their ability to manage their disease, but also diminishes their productivity and expected lifespan. Therefore, effective therapies that can prevent or inhibit the progression of DR is of great value for both patient and society [8]. The idea of ophthalmologic screening in all diabetic patients for DR is unrealistic as many diabetes patients are treated in endocrinology, where there is often no ophthalmological examination equipment, such as wide-field retinal photography and optical coherence tomography, especially in developing countries like China [9, 10]. Still, the screening of DR should be given high priority. Existing studies have reported the risk factors for DR and the construction of risk models [11]. Although the results vary, age and duration of diabetes are the primary factors. Nowadays, increasingly researchers are putting tremendous efforts into constructing predictive models or improving existing ones [12, 13]. Among them, the construction of the nomogram models has become one of the hottest research areas. While it is acknowledged that the latest research has shown excellent prediction performance of nomograms, it's important to note that the excellence of model performance can be relative and context-dependent.

The rising prevalence of T2DM and its complications necessitates the construction and validation of predictive models tailored to different population groups. Our study, set in this context, aims to provide a tool that can help early risk identification of patients. While age and duration of diabetes are primary risk factors for DR, our study delved deeper into the potential connections between DR and other clinical indicators. By incorporating these indicators into the nomogram model, a more comprehensive understanding of DR risk factors and their interplay can be realized, and it serves as a practical tool for clinicians to predict the risk DR is provided. We believe that our study brings value to the ongoing research in this area by offering a nuanced understanding of DR risk factors and providing a practical predictive tool tailored to the needs of developing countries.

## Methods and materials

### *Sample source*

We collected data from 648 patients with T2DM who were diagnosed and underwent funduscopic examinations at Gansu Provincial Hospital of TCM between June 2020 and June 2022 for a retrospective study. The flow chart of this study is presented in **Figure 1**. This study was conducted with the approval of the medical ethics committee of Gansu Provincial Hospital of TCM.

### *Inclusion and exclusion criteria*

Inclusion criteria: patients with confirmed diagnosis of T2DM [14]; patients who underwent funduscopic examinations, which confirmed the diagnosis of DR; patients with age between 18 and 80 years; patients with available urine and blood sample data.

Exclusion criteria: patients with tumors; patients with systolic blood pressure > 200 mmHg or diastolic blood pressure > 120 mmHg; patients with severe lipid abnormalities: total cholesterol (TC) > 10 mmol/L or triglycerides (TG) > 15 mmol/L; patients with incomplete clinical data; pregnant women.

### *Diagnostic criteria*

In patients diagnosed with T2DM, funduscopic examinations were performed using both direct and indirect ophthalmoscopy. The diagnosis of DR was indicated by the presence of one or more of the following retinal changes: 1. Microaneurysms: small red dots on the retina; 2. Hard exudates: yellow lipid deposits from leaking blood vessels; 3. Cotton-wool spots: pale and fluffy areas on the retina resulting from nerve fiber layer infarctions; 4. Venous beading: changes in the caliber of retinal veins; 5. Intraretinal microvascular abnormalities; 6. Neovascularization: growth of new blood vessels on the retina or optic disc; 7. Fluorescein angiography (if required) showing areas of leakage, non-perfusion, or neovascularization; 8. Optical coherence tomography (OCT) showing retinal thickening or cystoid macular edema (if applicable) [15].

Patients without DR were confirmed through the absence of any retinal changes indicative of

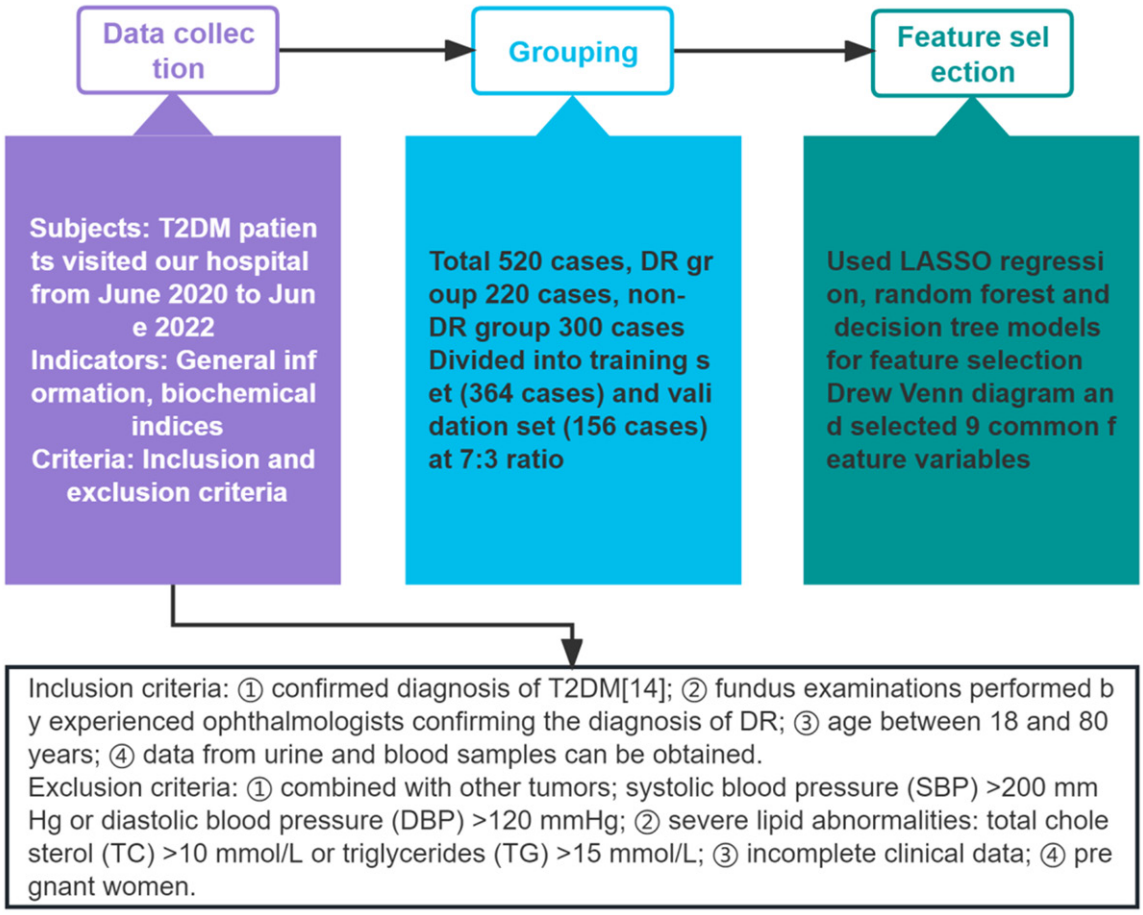


**Figure 1.** Study flow chart. T2DM, type 2 diabetes mellitus; DR, Diabetic Retinopathy.

DR as mentioned above, no evidence of leakage, non-perfusion, or neovascularization on fluorescein angiography (if performed), and no signs of retinal thickening or cystoid macular edema found in OCT (if applicable).

*Sample selection*

According to the inclusion and exclusion criteria, a total of 520 samples were collected for this study. Among them, 210 patients were diagnosed with DR, and 310 patients were not, with a morbidity rate of 40.38%. Subsequently, patients were divided into a training group (n = 364) and a validation group (n = 156) at a ratio of 7:3.

*Clinical data collection*

All of the patients were admitted between June 2020 and June 2022. The general data and biochemical data were collected through electronic medical records and outpatient review records. General data included gender, age, Diabetic Peripheral Neuropathy (DPN), hypertension, systolic blood pressure, diastolic blood pressure, and disease course. For biochemical data, an automatic blood analyzer (Sysmex, pocH-100i) was used to test patient blood routine indicators including White Blood Cell (WBC) count, Eosinophil count (EOS), Neutrophil count (NE), Lymphocyte count (LY), and the Neutrophil to Lymphocyte Ratio (NLR). An automatic biochemical analyzer (Beckman Coulter, AU5800) was used to test TC, Aspartate Aminotransferase (AST), High-Density Lipoprotein (HDL), Low-Density Lipoprotein (LDL), Alanine Aminotransferase (ALT), Glycated Serum Protein (GSP), Glucose (GLC), Hemoglobin A1c (HbA1C), Apolipoprotein A1 (ApoA1), Apolipoprotein B (ApoB), TG, and Blood Urea Nitrogen (BUN).

*Feature selection process*

Decision trees, LASSO, and Random Forests were used for feature selection from the training dataset, and a Venn diagram was used to determine the final feature variables. These feature variables were validated for comparability in both training and validation datasets,

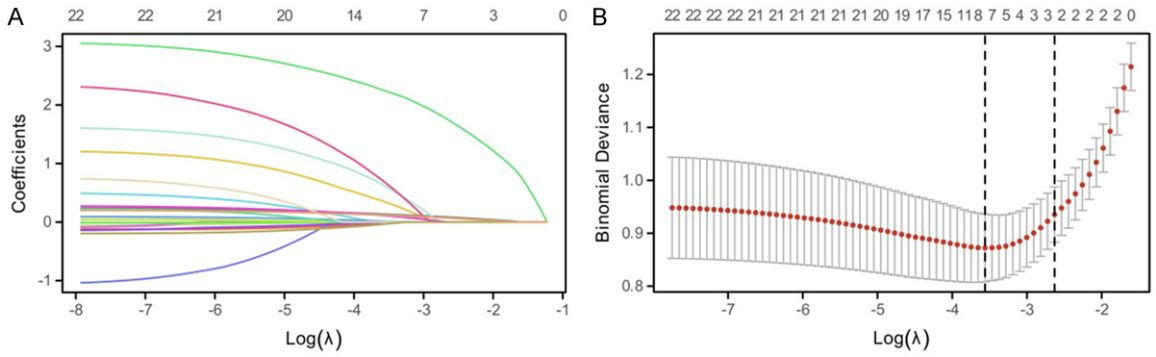


**Figure 2.** Lasso Regression for the screening of feature variables. A. Selection of genes with non-zero coefficients for model construction. B. Log lambda values of genes corresponding to the point with minimal cross-validation error.

and a DR risk prediction nomogram was constructed based on these features.

*Model evaluation and validation*

Several statistical methods were employed to evaluate the performance and validity of the constructed risk prediction model. 1. C-index: This index reflects the proportion of consistent case pairs between the predicted and actual results. It indicates the consistency of the predicted results with the actual observations. A higher C-index suggests better model performance. 2. Calibration curve: This curve represents the fit line between the predicted risk and actual risk. The closer the predicted values are to the actual values, the higher the model accuracy. 3. Area Under the Receiver Operating Characteristic (ROC) Curve (AUC): AUC indicates the ability of the model to distinguish between positive and negative outcomes. The larger the AUC value, the higher the accuracy of the model. 4. Decision Curve Analysis (DCA): Although not a conventional validation method, DCA assumes significance as a vital tool for evaluating the clinical utility of a model. It assesses the net benefits of using the model in decision-making, thereby helping high-risk patients to receive necessary interventions and low-risk patients to avoid unnecessary treatments. This ensures that the model meets practical clinical decision needs.

*Statistical analysis*

Initial data processing was performed using SPSS26.0. This included tasks such as data cleaning (e.g., deletion of missing or irrelevant data), data imputation (filling in missing values using appropriate statistical methods), and random grouping for subsequent analyses. Decision tree and random forest analyses were conducted through the SPSSAU platform. The decision tree was constructed to identify the primary splits in the data, while the random forest model was used to assess the importance of each variable and improve prediction accuracy. Advanced statistical analyses were performed using R software. The "glmnet" package was employed to construct the LASSO regression model, which aids in feature selection by shrinking some regression coefficients to zero. The "rms" package was used to plot the nomogram, a graphical representation of the prediction model. The "rmda" package was utilized for DCA plotting, and the "rocr" package for ROC curve plotting. Both are essential for assessing the model's predictive performance. The "rms" package was also used to plot the calibration curve, which assesses the agreement between predicted and observed outcomes, and to calculate the C-index, a measure of the model's discriminative ability. In addition, normally distributed measured data were described by mean ± standard deviation (mean ± SD) and compared by t-test. Intergroup comparisons were performed using independent samples t-test. The $\chi^2$ test was used to compare the counting data, expressed by percentage (%). All statistical tests were two-sided, and a *P*-value less than 0.05 was considered statistically significant.

## Results

*Lasso feature selection*

Using the LASSO algorithm and performing 10-fold cross-validation, we selected the lambda.1se parameter as 0.028664 (**Figure 2A**, **2B**). This resulted in a total of 12 relevant variables: age, ApoA1, DPN, HbA1C, HDL,

**Table 1.** The random forest feature variable screening for weight value

| Nape | Weight value |
|---|---|
| DPN | 0.189 |
| Age | 0.13 |
| BUN (mmol/L) | 0.104 |
| HbA1C (%) | 0.054 |
| AST (mmol/L) | 0.048 |
| NE ($10^9$/L) | 0.04 |
| HDL (mmol/L) | 0.039 |
| Disease course (year) | 0.037 |
| ALT (mmol/L) | 0.036 |
| EOS ($10^9$/L) | 0.034 |
| GSP (%) | 0.033 |
| ApoA1 (g/L) | 0.032 |
| LDL (mmol/L) | 0.028 |
| NLR | 0.027 |
| GLC (mmol/L) | 0.027 |
| LY ($10^9$/L) | 0.026 |
| TC (mmol/L) | 0.026 |
| WBC ($10^9$/L) | 0.025 |
| ApoB (g/L) | 0.024 |
| TG (mmol/L) | 0.024 |
| Hypertension | 0.014 |

Note: DPN, Diabetic Peripheral Neuropathy; WBC, White Blood Cell Count; EOS, Eosinophil Fraction; NE, Neutrophil Count; LY, Lymphocyte Count; NLR, Neutrophil to Lymphocyte Ratio; TC, Total Cholesterol; AST, Aspartate Aminotransferase; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; ALT, Alanine Aminotransferase; GSP, Glycated Serum Protein; GLC, Glucose; HbA1C, Hemoglobin A1c; ApoA1, Apolipoprotein A1; ApoB, Apolipoprotein B; TG, Triglycerides; BUN, Blood Urea Nitrogen.

hypertension, LDL, NE ($10^9$/L), NLR, TG, BUN, and duration of disease.

*Random forest feature selection*

Random forest model was used on the training dataset to determine the weights between the relevant variables and the model, as well as rank the important variables (**Table 1**). Variables with weights less than 0.01 were eliminated, and we finally obtained 21 relevant variables: DPN, age, BUN, HbA1C, AST, NE, HDL, duration of disease, ALT, EOS, GSP, ApoA1, LDL, NLR, GLC, LY, TC, WBC, ApoB, TG, and hypertension.

*Decision tree feature selection*

We used the decision tree model on the training dataset to determine the weight relationships between the relevant variables and the model, and rank the variables by importance (**Table 2**). Variables with weights less than 0.01 were eliminated, and we finally obtained 15 relevant variables: DPN, age, BUN, GSP, EOS, GLC, HbA1C, HDL, ALT, duration of disease, ApoB, LDL, NLR, AST, and TG.

**Table 2.** Decision tree feature variable screening weight value

| Nape | Weight value |
|---|---|
| DPN | 0.348 |
| Age | 0.152 |
| BUN (mmol/L) | 0.099 |
| GSP (%) | 0.082 |
| EOS ($10^9$/L) | 0.052 |
| GLC (mmol/L) | 0.049 |
| HbA1C (%) | 0.04 |
| HDL (mmol/L) | 0.039 |
| ALT (mmol/L) | 0.039 |
| Disease course (year) | 0.022 |
| ApoB (g/L) | 0.021 |
| LDL (mmol/L) | 0.016 |
| NLR | 0.015 |
| AST (mmol/L) | 0.011 |
| TG (mmol/L) | 0.011 |

Note: DPN, Diabetic Peripheral Neuropathy; EOS, Eosinophil Fraction; NLR, Neutrophil to Lymphocyte Ratio; AST, Aspartate Aminotransferase; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; ALT, Alanine Aminotransferase; GSP, Glycated Serum Protein; GLC, Glucose; HbA1C, Hemoglobin A1c; ApoB, Apolipoprotein B; TG, Triglycerides; BUN, Blood Urea Nitrogen.

*Common feature selection*

Using a Venn diagram, we further screened the feature variables selected by the above three methods. As a result, we found 9 common feature variables: age, DPN, HbA1C, HDL, LDL, NLR, TG, BUN, and duration of disease (**Figure 3**).

*Univariate analysis of the 9 feature variables*

Initially, values were assigned to the nine variables, and subsequently, the measurement data were grouped based on the descending ROC cut-off value (**Table 3**). This was followed by conducting a univariate analysis. The results showed statistical differences in 8 feature variables (age, DPN, NLR, HDL, duration of disease, HbA1C, TG, and BUN) between DR and non-DR groups (**Table 4**, $P < 0.05$), but no statistical difference in LDL between the two groups ($P >$

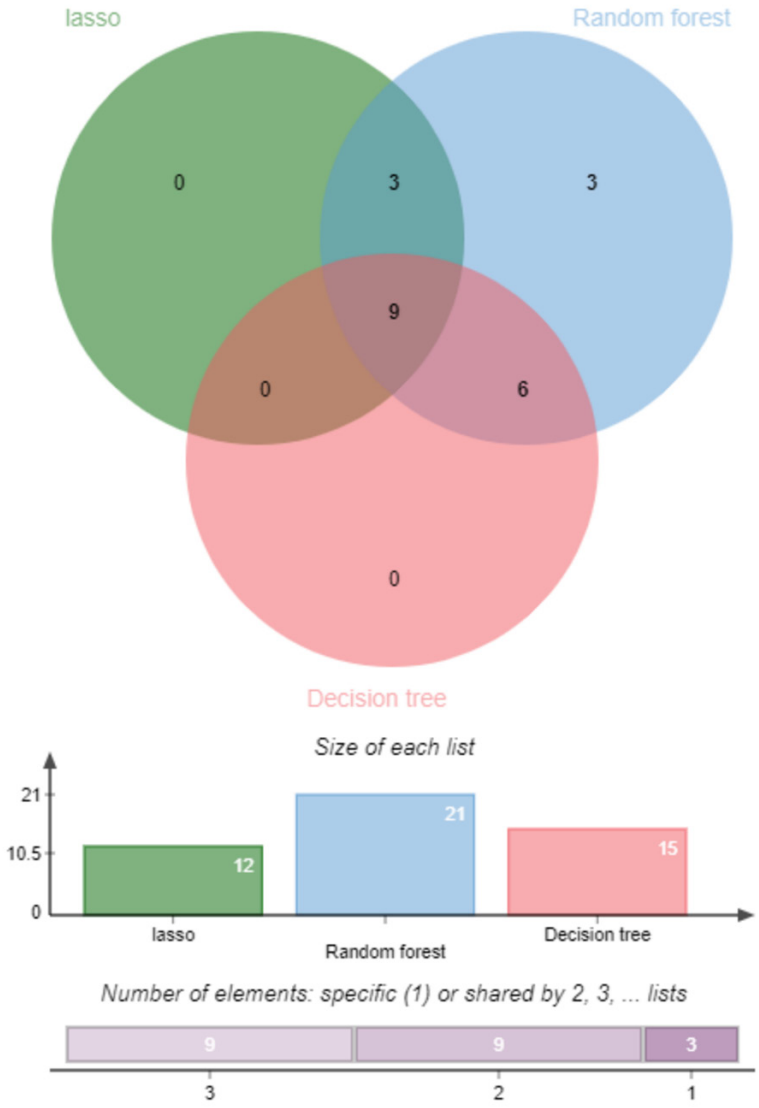


**Figure 3.** Wayn diagram for the screening of feature variables.

**Table 3.** Assignment table

| Factor | Assignment |
|---|---|
| Age | ≥ 50 year = 1, < 50 year = 0 |
| DPN | Yes = 1, No = 0 |
| NLR | ≥ 2.357 = 1, < 2.357 = 0 |
| HDL (mmol/L) | ≥ 0.945 = 1, < 0.945 = 0 |
| LDL (mmol/L) | ≥ 2.395 = 1, < 2.395 = 0 |
| Disease course (year) | ≥ 9.89 = 1, < 9.89 = 0 |
| HbA1C (%) | ≥ 8.835 = 1, < 8.835 = 0 |
| TG (mmol/L) | ≥ 2.515 = 1, < 2.515 = 0 |
| BUN (mmol/L) | ≥ 46.436 = 1, < 46.436 = 0 |
| DR | Yes = 1, No = 0 |

Note: DPN, Diabetic Peripheral Neuropathy; NLR, Neutrophil to Lymphocyte Ratio; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; HbA1C, Hemoglobin A1c; TG, Triglycerides; BUN, Blood Urea Nitrogen; DR, Diabetic Retinopathy.

0.05). We also compared the differences in these 8 feature variables between the validation and training datasets, and found no statistical differences, demonstrating comparability (P > 0.05, **Table 5**).

*Construction of the DR nomogram*

We used the LASSO, random forest, and decision tree machine learning algorithms to select feature variables, and 8 common features were identified through the Wayn diagram (**Figure 3**). These variables were then used to construct the nomogram, which was set up based on the coefficients obtained from the multivariate logistic regression analysis. In logistic regression, the coefficients of each variable determine its weight in the nomogram, ensuring that the nomogram accurately reflects the relative importance of each variable in predicting outcomes (**Figure 4**). In the visualization of the risk prediction nomogram, the "Points" represent the scores corresponding to the variables. Different variable values correspond to different "Points" values. The summation of the scores corresponding to each variable yields the "TotalPoints", from which the corresponding DR risk for each patient can be ascertained from the "Risk of DR" below. This model can facilitate individualized prediction of DR in clinical practice.

*Risk prediction and model validation*

We validated the risk prediction model using four methods, AUC of ROC curve, C-index, DCA, and calibration curve. The effectiveness of the risk prediction model was obtained: (1) The internal validation AUC was 0.773, and the external validation AUC was 0.735, with no difference in AUC between internal and external AUC (Delong test, P > 0.05), indicating that the

**Table 4.** Univariate analysis

| Factor | DR group (n = 144) | Non-DR group (n = 220) | $\chi^2$ value | *P* value |
|---|---|---|---|---|
| Age | | | 131.039 | < 0.0001 |
| ≥ 50 | 125 | 56 | | |
| < 50 | 19 | 164 | | |
| DPN | | | 65.447 | < 0.0001 |
| Yes | 120 | 89 | | |
| No | 24 | 131 | | |
| NLR | | | 4.717 | 0.029 |
| ≥ 2.357 | 44 | 92 | | |
| < 2.357 | 100 | 128 | | |
| HDL (mmol/L) | | | 9.531 | 0.002 |
| ≥ 0.945 | 106 | 127 | | |
| < 0.945 | 38 | 93 | | |
| LDL (mmol/L) | | | 1.648 | 0.199 |
| ≥ 2.395 | 84 | 143 | | |
| < 2.395 | 60 | 77 | | |
| Disease course (year) | | | 10.271 | 0.001 |
| ≥ 9.89 | 72 | 73 | | |
| < 9.89 | 72 | 147 | | |
| HbA1C (%) | | | 13.655 | < 0.001 |
| ≥ 8.835 | 80 | 79 | | |
| < 8.835 | 64 | 141 | | |
| TG (mmol/L) | | | 6.157 | 0.013 |
| ≥ 2.515 | 64 | 127 | | |
| < 2.515 | 80 | 93 | | |
| BUN (mmol/L) | | | 50.932 | < 0.0001 |
| ≥ 46.436 | 113 | 89 | | |
| < 46.436 | 31 | 131 | | |

Note: DPN, Diabetic Peripheral Neuropathy; NLR, Neutrophil to Lymphocyte Ratio; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; HbA1C, Hemoglobin A1c; TG, Triglycerides; BUN, Blood Urea Nitrogen; DR, Diabetic Retinopathy.

prediction model has good discriminative ability (**Figure 5A**, **5B**). (2) The calibration curves of internal and external validations indicated that the predicted probability of DR aligns well with the actual situation, demonstrating the accuracy of the prediction model (**Figure 5C**, **5D**). (3) The C-index of internal and external validation was 0.937 (0.913-0.961) and 0.916 (0.874-0.959), respectively, indicating good consistency between the actual and predicted probabilities of DR. (4) The DCA of internal and external validations demonstrated a good clinical net benefit of the prediction model at different threshold probabilities (internal: 11%~95%; external: 17%~93%), confirming its practicality (**Figure 5E**, **5F**).

## Discussion

The development of DR is associated with several factors, including oxidative stress triggered by high blood sugar, accumulation of advanced glycation end-products, increased reactive oxygen species, abnormal activation of protein kinase C, and aberrant activation of the renin-angiotensin system [16]. These factors can also result in the concurrent of DR and diabetic nephropathy. However, despite the similar pathogenesis, pathological changes and influencing factors vary between the two [17]. Clinically, due to the lack of simplified DR screening methods, DR and diabetic nephropathy exhibit asynchrony in clinical manifestation [18]. Therefore, actively exploring risk factors for DR and intervening accordingly bears significant clinical relevance for early diagnosis and treatment of diabetic microvascular complications and even for the prevention of these complications.

Machine learning is multidisciplinary field that draws upon statistics and various academic

**Table 5.** Comparison of 9 characteristic variables between patients in the training group and the validation group

| Factor | Training groups (n = 364) | Validation groups (n = 156) | $\chi^2$ value | *P* value |
|---|---|---|---|---|
| Age | | | 2.029 | 1.543 |
| ≥ 50 | 181 | 94 | | |
| < 50 | 183 | 62 | | |
| DPN | | | 2.029 | 0.154 |
| Yes | 209 | 79 | | |
| No | 155 | 77 | | |
| NLR | | | 0.056 | 0.812 |
| ≥ 2.357 | 136 | 60 | | |
| < 2.357 | 228 | 96 | | |
| HDL (mmol/L) | | | 0.066 | 0.795 |
| ≥ 0.945 | 233 | 98 | | |
| < 0.945 | 131 | 58 | | |
| Disease course (year) | | | 1.793 | 0.180 |
| ≥ 9.89 | 145 | 72 | | |
| < 9.89 | 219 | 84 | | |
| HbA1C (%) | | | 1.757 | 0.184 |
| ≥ 8.835 | 159 | 78 | | |
| < 8.835 | 205 | 78 | | |
| TG (mmol/L) | | | 1.744 | 0.186 |
| ≥ 2.515 | 191 | 72 | | |
| < 2.515 | 173 | 84 | | |
| BUN (mmol/L) | | | 0.005 | 0.938 |
| ≥ 46.436 | 202 | 86 | | |
| < 46.436 | 162 | 70 | | |

Note: DPN, Diabetic Peripheral Neuropathy; NLR, Neutrophil to Lymphocyte Ratio; HDL, High-Density Lipoprotein; HbA1C, Hemoglobin A1c; TG, Triglycerides; BUN, Blood Urea Nitrogen; DR, Diabetic Retinopathy.

disciplines. It leverages computer technology to process big data and is a subset of artificial intelligence. By using machine learning algorithms, we can filter out required feature variables from large amounts of data, thereby effectively improving learning efficiency [19]. At present, machine learning has been widely applied in the medical field, especially in radiomics and pathological image segmentation, but further research is needed for risk prediction of clinical diseases [20, 21]. Thus, in this study, we used decision tree, LASSO, and random forest machine learning algorithms to screen risk factors for DR and constructed a risk prediction model [22, 23]. Furthermore, we performed internal and external validations of this model to verify its accuracy, reliability, applicability, and clinical application value, providing reliable data support for the prevention, treatment, and prognosis of DR.

In this study, we collected T2DM patients who had already undergone funduscopic examination and divided them into DR and non-DR groups. We compared the relevant data between the two groups, screened out significant risk factors, and built a DR prediction model based on them. Our research found that age, DPN, NLR, HDL, disease duration, HbA1C, TG, and BUN were risk factors affecting the occurrence of DR in diabetic patients. Compared to older patients, it was found that patients of younger age and shorter disease duration had a higher risk of DR, suggesting that younger age might be a protective factor for DR. Yin et al. [24] also found that age was a risk factor for DR in T2DM patients, and the older the age, the greater the likelihood of disease worsening, and the higher the risk of DR, which is consistent with our research results. Our study also found that high expression of

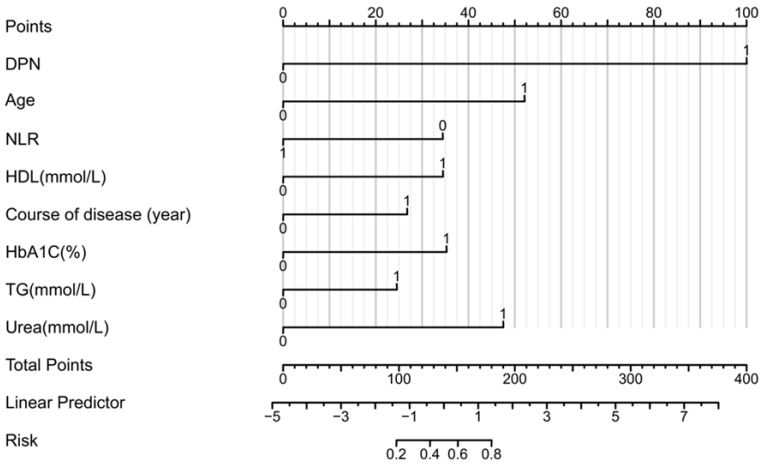


**Figure 4.** DR Risk prediction nomogram. Note: DPN, diabetic peripheral neuropathy; NLR, Lymphocyte to neutrophil ratio to value; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; HbA1C, Hemoglobin A1c; TG, Triglycerides; Urea, Blood Urea Nitrogen; DR, diabetic retinopathy.

NLR was a protective factor for DR. This might be related to the systemic inflammatory response and immune system response in diabetic patients. NLR is an indicator reflecting the body's inflammatory and immune status. Existing research has shown [25] that inflammatory and immune responses play a key role in the pathogenesis of diabetes. High blood sugar can lead to an enhanced systemic inflammatory response, which may lead to an increase in the number of neutrophils, while affecting the function of lymphocytes, which may cause a relative decrease in lymphocyte count, resulting in an increase of NLR in diabetic patients [26]. However, the occurrence of DR is related to the inflammatory response and angiogenesis. In the process of DR, some inflammatory mediators and growth factors, such as vascular endothelial growth factor, might promote angiogenesis and pathological changes [27]. This process is possibly related to the response of lymphocytes, potentially causing a relative increase in lymphocytes, thereby reducing NLR.

DPN is a type of diabetic microvascular complication, and its development and progression involve several interrelated pathophysiological mechanisms [28]. Foremost of these is the sustained hyperglycemic state caused by poor long-term blood sugar control, which triggers a series of biochemical reactions, such as nonenzymatic glycosylation, activation of the sorbitol pathway, oxidative stress, inflammatory response, and endothelial dysfunction [29]. A multicenter study by Peng et al. [30] revealed that the occurrence of DPN was related to retinopathy in diabetic patients, which is consistent with our research results. In addition, the cross-sectional study by Wang et al. [31] also found that DPN was a risk factor for the occurrence of DR. Currently, HbA1c, which can reflect blood glucose levels over the past 6-8 weeks, is widely used as a screening indicator for T2DM in clinical practice and is considered the "gold standard" for diagnosing T2DM. Our research found that abnormally elevated levels of HbA1c are also a risk factor for DR, which is consistent with the research results of Su et al. [32]. In addition, studies have shown that an increase in HbA1c mediates damage to vascular endothelial cells and triggers adhesion of leukocytes to the surface of vascular endothelial cells, thereby promoting thrombosis. At the same time, oxidative stress products and/or inflammation states induced by high glucose levels, through their effects on the vascular wall and the vascular matrix, intensify the harm to the vascular endothelium and tissues. They collaboratively contribute to the pathogenesis and progression of DR. Therefore, patients with elevated HbA1c levels have a significantly higher risk of developing DR. In addition, our results also showed that elevated HDL, TG, and BUN were also risk factors for DR, which is consistent with previous research results [33-35].

In recent years, there has been a focus on developing predictive models for DR in patients with T2DM. For instance, Wang et al. [36] developed a DR risk nomogram for a Chinese population with T2DM. Their model, based on data from 213 patients, identified 8 prediction variables and achieved a C-index of 0.848. Pan et al. [37], utilizing machine learning techniques, created a DR risk prediction model based on a larger sample size of 2,385 T2DM patients. The AUC of their model was 0.703. In comparison to these studies, our study employed three machine learning algorithms (LASSO, random forest, and decision tree) and identified 8 key

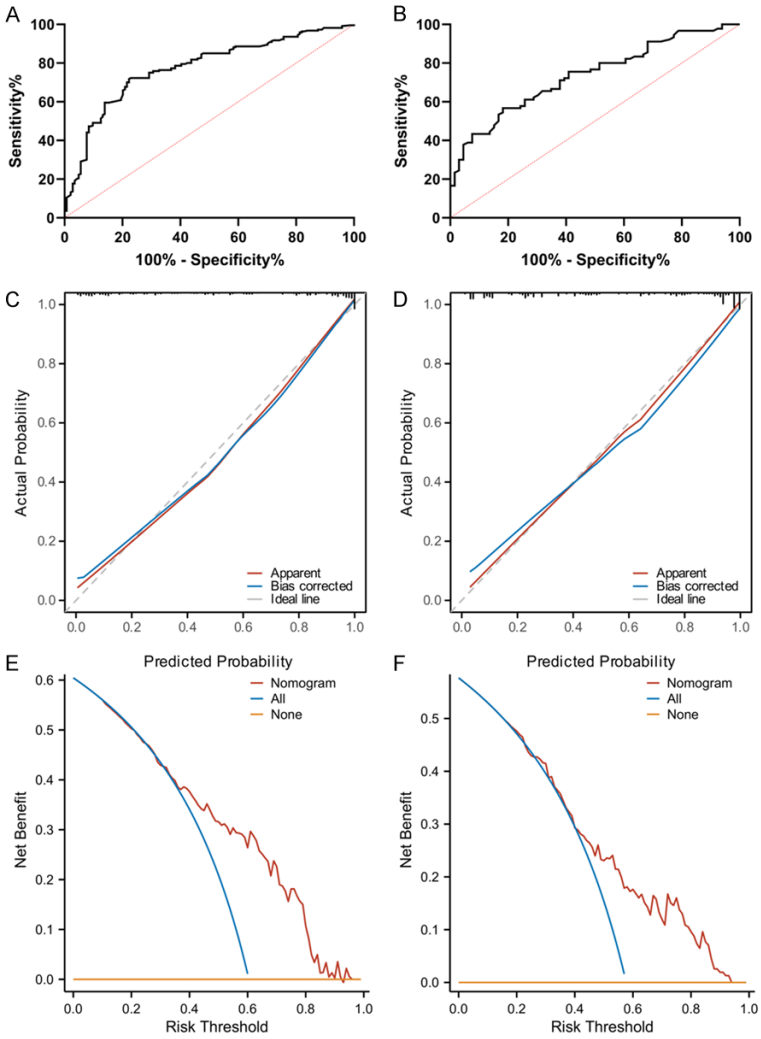


**Figure 5.** Risk forecast and model verification. A. ROC curve analysis of the model effectiveness in the training group. B. ROC curve analysis of the model effectiveness in the validation groups. C. Calibration curve of the model in the training group. D. Calibration curve of the model in the validation group. E. DCA curve of the model in the training group. F. DCA curve of the model in the validation group. Note: DR, diabetic retinopathy; ROC, subject working curve; DCA, decision curve analysis.

characteristic variables. We successfully constructed a DR risk prediction nomogram, which demonstrated an AUC of 0.773 for internal validation and 0.735 for external validation, indicating a high accuracy and discriminative power of our model. Additionally, our model performed well in terms of C-index, calibration curve, and DCA, further showcasing its significance in clinical practice. This research offers a comprehensive understanding of risk factors for DR and their interactions, as well as provides clinicians with a practical tool. It is important to note that various models exist for predicting the risk of DR in patients with T2DM, and each model offers its own unique perspective. Our study contributes to the expanding knowledge in this field by providing a detailed understanding of DR risk factors and practical predictive tools that are tailored to specific patient populations and settings.

However, there are certain limitations to this study. First, as a single-center study, whether our results can be generalized requires validation with more data. Secondly, this study only used data from the same medical center for validation and did not carry out validation with data from other medical centers. Therefore, in subsequent research, it is necessary to validate with data from other medical centers to reduce selection bias, calibrate the prediction model, and optimize this model, so as to provide a more advantageous reference for clinical determination of the risk of DR in T2DM patients.

In conclusion, this study developed an innovative, intuitive, objective, and accurate Nomogram prediction model that can predict the risk of retinopathy in patients with T2DM, and proved through internal validation that it has high precision.

## Disclosure of conflict of interest

None.

**Address correspondence to:** Qinhu Wang, Department of Endocrinology, The Third People's Hospital of Gansu Province, No. 763 Duanjiatan Road, Chengguan District, Lanzhou 730020, Gansu, China. Tel: +86-0931-8497337; E-mail: 86531311@qq.com

## References


[1] Tomkins M, Lawless S, Martin-Grace J, Sherlock M and Thompson CJ. Diagnosis and management of central diabetes insipidus in adults. J Clin Endocrinol Metab 2022; 107: 2701-2715.

[2] Christ-Crain M and Gaisl O. Diabetes insipidus. Presse Med 2021; 50: 104093.

[3] Christ-Crain M, Winzeler B and Refardt J. Diagnosis and management of diabetes insipidus for the internist: an update. J Intern Med 2021; 290: 73-87.

[4] Refardt J. Diagnosis and differential diagnosis of diabetes insipidus: update. Best Pract Res Clin Endocrinol Metab 2020; 34: 101398.

[5] Roikjer J and Ejskjaer N. Diabetic peripheral neuropathy. Handb Exp Pharmacol 2022; 274: 309-328.

[6] Barrell K and Smith AG. Peripheral neuropathy. Med Clin North Am 2019; 103: 383-397.

[7] Selvarajah D, Kar D, Khunti K, Davies MJ, Scott AR, Walker J and Tesfaye S. Diabetic peripheral neuropathy: advances in diagnosis and strategies for screening and early intervention. Lancet Diabetes Endocrinol 2019; 7: 938-948.

[8] Yang K, Wang Y, Li YW, Chen YG, Xing N, Lin HB, Zhou P and Yu XP. Progress in the treatment of diabetic peripheral neuropathy. Biomed Pharmacother 2022; 148: 112717.

[9] Javed S, Hayat T, Menon L, Alam U and Malik RA. Diabetic peripheral neuropathy in people with type 2 diabetes: too little too late. Diabet Med 2020; 37: 573-579.

[10] Seyedizadeh SH, Cheragh-Birjandi S and Hamedi Nia MR. The effects of combined exercise training (resistance-aerobic) on serum kinesin and physical function in type 2 diabetes patients with diabetic peripheral neuropathy (randomized controlled trials). J Diabetes Res 2020; 2020: 6978128.

[11] Tang Y, Xu L, Ren Y, Li Y, Yuan F, Cao M, Zhang Y, Deng M and Yao Z. Identification and validation of a prognostic model based on three MVI-related genes in hepatocellular carcinoma. Int J Biol Sci 2022; 18: 261-275.

[12] Zuo D, Li C, Liu T, Yue M, Zhang J and Ning G. Construction and validation of a metabolic risk model predicting prognosis of colon cancer. Sci Rep 2021; 11: 6837.

[13] Zheng Y, Liu B, Deng X, Chen Y, Huang Y, Zhang Y, Xu Y, Sang L, Liu X and Li Y. Construction and validation of a robust prognostic model based on immune features in sepsis. Front Immunol 2022; 13: 994295.

[14] Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z and Zhao Y. From pre-diabetes to diabetes: diagnosis, treatments and translational research. Medicina (Kaunas) 2019; 55: 546.

[15] Genuth SM, Palmer JP and Nathan DM. Classification and diagnosis of diabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, Gregg EW, Knowler WC, Barrett-Connor E, Becker DJ, Brancati FL, Boyko EJ, Herman WH, Howard BV, Narayan KMV, Rewers M, Fradkin JE, editors. Diabetes in America. 3rd edition. Bethesda (MD): Interest.; 2018.

[16] Elafros MA, Andersen H, Bennett DL, Savelieff MG, Viswanathan V, Callaghan BC and Feldman EL. Towards prevention of diabetic peripheral neuropathy: clinical presentation, pathogenesis, and new treatments. Lancet Neurol 2022; 21: 922-936.

[17] Castelli G, Desai KM and Cantone RE. Peripheral neuropathy: evaluation and differential diagnosis. Am Fam Physician 2020; 102: 732-739.

[18] Sloan G, Selvarajah D and Tesfaye S. Pathogenesis, diagnosis and clinical management of diabetic sensorimotor peripheral neuropathy. Nat Rev Endocrinol 2021; 17: 400-420.

[19] Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ and Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med 2018; 284: 603-619.

[20] Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF and Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 2020; 9: 14.

[21] Greener JG, Kandathil SM, Moffat L and Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022; 23: 40-55.

[22] Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr and Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019; 23: 64.

[23] Yang C, Ren J, Li B, Jin C, Ma C, Cheng C, Sun Y and Shi X. Identification of gene biomarkers in patients with postmenopausal osteoporosis. Mol Med Rep 2019; 19: 1065-1073.

[24] Yin L, Zhang D, Ren Q, Su X and Sun Z. Prevalence and risk factors of diabetic retinopathy in diabetic patients: a community based cross-sectional study. Medicine (Baltimore) 2020; 99: e19236.

[25] Yalinbas Yeter D, Eroglu S, Sariakcali B, Bozali E, Vural Ozec A and Erdogan H. The usefulness of monocyte-to-high density lipoprotein and neutrophil-to-lymphocyte ratio in diabetic macular edema prediction and early anti-VEGF treatment response. Ocul Immunol Inflamm 2022; 30: 901-906.

[26] Hu Y, Cheng Y, Xu X, Yang B, Mei F, Zhou Q, Yan L, Wang J and Wu X. Pretreatment neutrophil-to-lymphocyte ratio predicts prognosis in patients with diabetic macular edema treated

with ranibizumab. BMC Ophthalmol 2019; 19: 194.
[27] Gokce SE, Celik A and Baskan C. The role of blood neutrophil lymphocyte ratio in predicting the initial response to anti-VEGF treatment in neovascular AMD patients. Ir J Med Sci 2023; [Epub ahead of print].
[28] Rasheed R, Pillai GS, Kumar H, Shajan AT, Radhakrishnan N and Ravindran GC. Relationship between diabetic retinopathy and diabetic peripheral neuropathy - neurodegenerative and microvascular changes. Indian J Ophthalmol 2021; 69: 3370-3375.
[29] Preguica I, Alves A, Nunes S, Gomes P, Fernandes R, Viana SD and Reis F. Diet-induced rodent models of diabetic peripheral neuropathy, retinopathy and nephropathy. Nutrients 2020; 12: 250.
[30] Peng Y, Guo X, Liu J, Yao Y, Guo H, Wang Y, Xiao X and Ye L. Incidence and risk factors for diabetic retinopathy in the communities of Shenzhen. Ann Palliat Med 2021; 10: 615-624.
[31] Wang W, Ji Q, Ran X, Li C, Kuang H, Yu X, Fang H, Yang J, Liu J, Xue Y, Feng B, Lei M and Zhu D. Prevalence and risk factors of diabetic peripheral neuropathy: a population-based cross-sectional study in China. Diabetes Metab Res Rev 2023; e3702.
[32] Sun Q, Jing Y, Zhang B, Gu T, Meng R, Sun J, Zhu D and Wang Y. The risk factors for diabetic retinopathy in a Chinese population: a cross-sectional study. J Diabetes Res 2021; 2021: 5340453.
[33] Chen X, Zhao J, You Y, Li Z and Chen S. The ratio of fibrinogen to albumin is related to the occurrence of retinopathy in type 2 diabetic patients. Diabetes Metab Syndr Obes 2023; 16: 1859-1867.
[34] Yang QP, Chen YY, Li Z and Xu M. Major risk factors analysis of pruritus complicated by type 2 diabetes mellitus and the effect of comprehensive nursing intervention. Front Surg 2022; 9: 842884.
[35] Xing L, Peng F, Liang Q, Dai X, Ren J, Wu H, Yang S, Zhu Y, Jia L and Zhao S. Clinical characteristics and risk of diabetic complications in data-driven clusters among type 2 diabetes. Front Endocrinol (Lausanne) 2021; 12: 617628.
[36] Wang Q, Zeng N, Tang H, Yang X, Yao Q, Zhang L, Zhang H, Zhang Y, Nie X, Liao X and Jiang F. Diabetic retinopathy risk prediction in patients with type 2 diabetes mellitus using a nomogram model. Front Endocrinol (Lausanne) 2022; 13: 993423.
[37] Pan H, Sun J, Luo X, Ai H, Zeng J, Shi R and Zhang A. A risk prediction model for type 2 diabetes mellitus complicated with retinopathy based on machine learning and its application in health management. Front Med (Lausanne) 2023; 10: 1136653.