

# A Machine Learning Model for Predicting Diabetic Nephropathy Based on TG/Cys-C Ratio and Five Clinical Indicators

Dongmei Zhou<sup>1,\*</sup>, Lingyu Shao<sup>2,\*</sup>, Libo Yang<sup>3,\*</sup>, Yongkang Chen<sup>1,\*</sup>, Yue Zhang<sup>4,\*</sup>, Feng Yue<sup>3</sup>, Weipeng Gu<sup>3</sup>, Shuyi Li<sup>1</sup>, Shuyan Li<sup>2</sup>, Jing Wei<sup>3</sup>

<sup>1</sup>Department of Rheumatology and Immunology, Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, People's Republic of China;

<sup>2</sup>School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou, Jiangsu, People's Republic of China; <sup>3</sup>Department of Endocrinology, The Affiliated Taian City Central Hospital of Qingdao University, Taian, Shandong, People's Republic of China; <sup>4</sup>Department of Endocrinology, Xuzhou New Health Hospital, Xuzhou, Jiangsu, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Jing Wei, Department of Endocrinology, The Affiliated Taian City Central Hospital of Qingdao University, Taian, Shandong, 271000, People's Republic of China, Email [drw996223@163.com](mailto:drw996223@163.com); Dongmei Zhou, Department of Rheumatology and Immunology, Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, 221000, People's Republic of China, Email [zdm@xzhmu.edu.cn](mailto:zdm@xzhmu.edu.cn)

**Objective:** Distinguishing diabetic nephropathy (DN) from non-diabetic renal disease (NDRD) remains challenging. This study developed and validated a machine learning model for differential diagnosis of DN and NDRD.

**Methods:** We included 100 type 2 diabetes mellitus (T2DM) patients with proteinuria from four Xuzhou hospitals (2013–2021), divided into DN (n=50) and NDRD (n=50) groups based on renal biopsy. Clinical data were used to build a predictive model. External validation was performed on 55 patients from The Affiliated Taian City Central Hospital of Qingdao University (2019–2023). Models were constructed using Python's scikit-learn library (v1.4.2), with feature selection via Recursive Feature Elimination (RFE).

**Results:** Compared to NDRD, DN patients had lower TG/Cys-c ratio [1.45 (0.75, 1.99) vs 2.78 (1.81, 4.48)], higher systolic blood pressure ( $156.80 \pm 20.14$  vs  $137.66 \pm 17.67$ ), longer diabetes duration [78 (24, 120) vs 18 (6, 48) months], higher diabetic retinopathy prevalence (60% vs 40%), higher HbA1c [7.98 (6.50, 10.40) vs 7.10 (6.70, 7.90)], and lower hemoglobin ( $115.66 \pm 22.20$  vs  $135.64 \pm 18.59$ ). The logistic regression (LR) model, incorporating TG/Cys-c ratio, SBP, diabetes duration, DR, HbA1c, and Hb, achieved an AUC of 0.9305, accuracy of 0.8333, sensitivity of 0.8283, and specificity of 0.8701. External validation showed an AUC of 0.9642, accuracy of 0.9455, sensitivity of 0.9615, and specificity of 0.9310. We named this method PDN (Prediction of Diabetic Nephropathy) and developed an online platform: <http://cppdd.cn/service/PDN>.

**Conclusion:** This machine learning-based method effectively differentiates DN from NDRD, aiding clinicians in diagnosis and treatment planning.

**Keywords:** diabetic nephropathy, non-diabetic renal disease, discriminant model, machine learning, logistic Regression

## Introduction

Diabetes mellitus (DM) is one of the most important health problems in the world. The number of people with diabetes is estimated to be 537 million in 2021 and is expected to increase to 784 million by 2045.<sup>1</sup> Diabetic nephropathy is one of the most common microvascular complications in DM patients. Diabetic nephropathy is the main cause of ESRD, and about 30–50% of End Stage Renal Disease (ESRD) worldwide is caused by DN.<sup>2</sup> However, proteinuria in T2DM patients is not necessarily DN, and may also be NDRD. The typical pathological manifestations of DN include thickening of the glomerular basement membrane and the formation of K-W nodules.<sup>3</sup> In contrast, the pathological mechanisms of NDRD are complex and are generally not induced by a hyperglycemic environment.<sup>4</sup> The incidence of NDRD in renal biopsy of T2DM patients is about 45%–80%.<sup>5,6</sup> Studies have found that, compared to DN patients, NDRD patients show significant improvement in

proteinuria and renal function after treatment with glucocorticoids and immunosuppressants, with an overall better prognosis than DN patients.<sup>7,8</sup> If these patients are misclassified as NDRD, their treatment may be limited, potentially missing opportunities for specific therapies that could improve renal outcomes.<sup>9</sup> Therefore, distinguishing between DN and NDRD in patients with type 2 diabetes and renal impairment is crucial. Renal biopsy remains the gold standard for differentiating DN from NDRD.<sup>10</sup> However, due to its invasive nature, risk of complications such as bleeding and infection, high cost, and complexity, it is unsuitable for routine screening. Biopsy is typically performed only when NDRD is suspected.<sup>11</sup> Studies have found that a short duration of diabetes, absence of DR, hematuria, or unexplained acute kidney injury are common clinical features used to predict NDKD in patients with type 2 diabetes.<sup>12,13</sup> However, the predictive value of these factors varies across studies. Therefore, there is an urgent need to develop non-invasive methods to assist in distinguishing DN from NDRD.

Previous studies have employed various conventional statistical methods to develop predictive models for DN.<sup>6,9,14</sup> Currently, machine learning (ML) methods offer several advantages over traditional regression approaches.<sup>12</sup> Research has demonstrated that ML has achieved notable success in the prediction and diagnosis of diabetes and DN.<sup>15–18</sup> ML models, combined with feature selection, have fewer restrictive statistical assumptions than traditional methods, thereby improving clinical decision-making. ML is regarded as an objective and reproducible approach that integrates multiple quantitative variables to enhance diagnostic accuracy.<sup>15,19,20</sup>

This study employed machine learning methods to develop a novel predictive model for distinguishing DN from NDRD. Using renal biopsy as the gold standard, we established a prediction model based on six non-invasive indicators: TG/Cys-C ratio, SBP, diabetes duration, DR, HbA1c, and Hb. Our goal is to utilize easily collected clinical information to create a clinical decision support system for identifying NDRD patients and facilitating early intervention.

In this study, five machine learning methods, RF, GBM, LR, SVM and XGB, were compared, and combined with clinical features and hematological indicators, a differential diagnosis model of DN and NDRD was constructed. Finally, LR can choose the best model, the best model can be on the web site, <http://cppdd.cn/service/PDN>, is convenient quickly identify early in clinical DN and NDRD.

## Methods

### Research Object

This is a multicenter retrospective observational study involving 4 nephropathy departments of Xuzhou Medical University Affiliated Hospital, Xuzhou Central Hospital, Xuzhou Municipal Hospital, and Xuzhou Hospital of Traditional Chinese Medicine. Patients with type 2 diabetes who had completed a kidney needle biopsy between January 2013 and October 2021 were collected. According to the pathological classification of kidney, 50 cases were divided into DN group and NDRD group, 90% of which were used as training set and the other 10% as test set. A total of 28 indicators including basic clinical features, blood routine, kidney function, diabetes related indicators were collected. Pearson correlation coefficient method was used to eliminate high correlation indicators, and finally 27 indicators were left. The missing values in the sample are filled in by the median or average. In addition, 26 patients with DN and 29 patients with NDRD diagnosed by kidney biopsy from Taian City Central Hospital from January 2019 to December 2023 were collected for external verification.

Inclusion criteria: (1) Age at which renal biopsy was performed (male or female) and gt; 18 years old with clear pathological findings; (2) The clinical diagnosis was type 2 diabetes, and the diagnosis of T2DM met the criteria established by the American Diabetes Association.<sup>21</sup> Exclusion criteria: (1) incomplete data or unclear medical history; (2) other acute complications of diabetes mellitus; (3) Patients who had taken fibrate lipid-lowering drugs within 2 months before renal biopsy; (4) Severe infection of other systems, failure of important organs, systemic immune system diseases and malignant tumors; (5) The pathological manifestations were DN combined with NDRD.

### Clinical and Laboratory Data

General Information: Data collected included the participants' sex, age, height, weight, and calculated body mass index (BMI). Information on hypertension history, SBP, diastolic blood pressure (DBP), duration of diabetes (in months), presence or absence of DR, and smoking or alcohol consumption history was also recorded.

**Laboratory Parameters:** Blood test indicators collected during the hospital admission for renal biopsy included Hb, fasting plasma glucose (FPG), HbA1c, triglycerides (TG), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), albumin (ALB), uric acid (UA), cystatin C (Cys-C), serum creatinine (Scr), and fibrinogen (FIB). All patients provided 24-hour urine specimens following standard procedures. The total urinary protein was measured using the turbidimetric method on a Roche Cobas c 701 fully automated analyzer. TG/Cys-C ratio was calculated by dividing the TG value by the Cys-C value, with the result rounded to two significant figures.

**Renal Biopsy:** Ultrasound-guided fine-needle renal biopsy was performed during hospitalization. All pathological tissues were examined using light microscopy (with HE, PAS, PASM, and Masson staining), electron microscopy, and immunofluorescence pathology. The analysis was conducted by two experienced pathologists.

This study was approved by the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University (approval number:XYFY2021-KL073-01). All patients and authorized signatories provided written informed consent. This study was conducted in accordance with the tenets of the Declaration of Helsinki.

## Comparison of Clinical Features

Measurement data were tested for normality. Data conforming to a normal distribution are expressed as  $\bar{x} \pm s$ , while non-normally distributed data are presented as  $M$  (Q1, Q3). Intergroup comparisons were performed using the Kruskal–Wallis  $H$ -test. Categorical data are expressed as counts (percentages), and intergroup comparisons were conducted using the chi-square test.

## Key Feature Selection Method

### Introduction to the RFE Method

Recursive Feature Elimination (RFE) is a feature selection method that selects the optimal feature subset by constructing a model recursively and gradually eliminating features with small weight coefficients. The process includes: The initial model is trained on all features, the importance of the features is calculated, the least important features are removed, and the model is retrained on the remaining features, iterating until a predetermined number of features is reached. By gradually eliminating low-weight features, RFE can effectively optimize model performance and improve interpretability.

### Introduction to the RF Method

Random Forest (RF) is an ensemble learning algorithm that improves the accuracy and robustness of classification or regression by building multiple decision trees and combining their predictions. It adopts Bootstrap Sampling method to extract multiple subsample sets from the original training data, and each subsample set is used to construct a decision tree. In the construction process, each node randomly selects some features to split, thus generating a diversified tree model. Finally, the stochastic forest integrates the prediction results of each tree by means of majority voting or average, which improves the robustness and anti-overfitting ability of the model.

### Introduction to the LR Method

Logistic Regression (LR) is a generalized linear model mainly used to solve binary classification problems. It classifies the logarithmic odds of the input variables by predicting them in a linear combination. In the training process, LR reduces the difference between the observed data and the prediction by adjusting the model parameters, and generates the final classification model. LR builds the model using a series of features, gradually adjusting the weight of each feature to the classification result. Model coefficients represent the degree to which features influence the output, and the training process involves optimizing the algorithm to maximize the likelihood function or minimize the loss function, ensuring efficient classification prediction on the data set.

All models are built using the scikit-learn library (1.4.2) in Python (3.10.10) and feature selection using the RFE method.

## Model Training and Feature Selection

The entire process is divided into three stages, as shown in [Figure 1](#).

- (1) Divide the data into the training set and the test set according to the ratio of 9:1. Random Forest, Gradient Boosting Machine, Logistic Regression, Support Vector Machine and Extreme Gradient were used respectively. Five algorithms train the data. On the basis of 5-fold cross-validation, the prediction model is established.
- (2) For feature selection and model parameter adjustment, Random Forest Classifier is mainly used as the basic model, and cross-validation is combined to evaluate the model performance and the effect of feature selection. Based on the random forest model, the average accuracy under different number of trees was calculated through 5-fold cross-validation, and the number of optimal trees, namely  $n\_estimators$ , was determined to be 30, as shown in Figure 2. Then, recursive feature elimination (RFE) was used to select features iteratively, gradually increase the number of features, observe the average accuracy of cross-validation, and draw a learning curve, as shown in Figure 3. Finally, 11 features were extracted.
- (3) According to the importance of each feature, forward feature selection is used to gradually select the feature with the greatest weight from the feature set obtained in the second stage for model construction and cross-validation to evaluate its prediction performance, as shown in Figure 4. According to the evaluation results, the five features

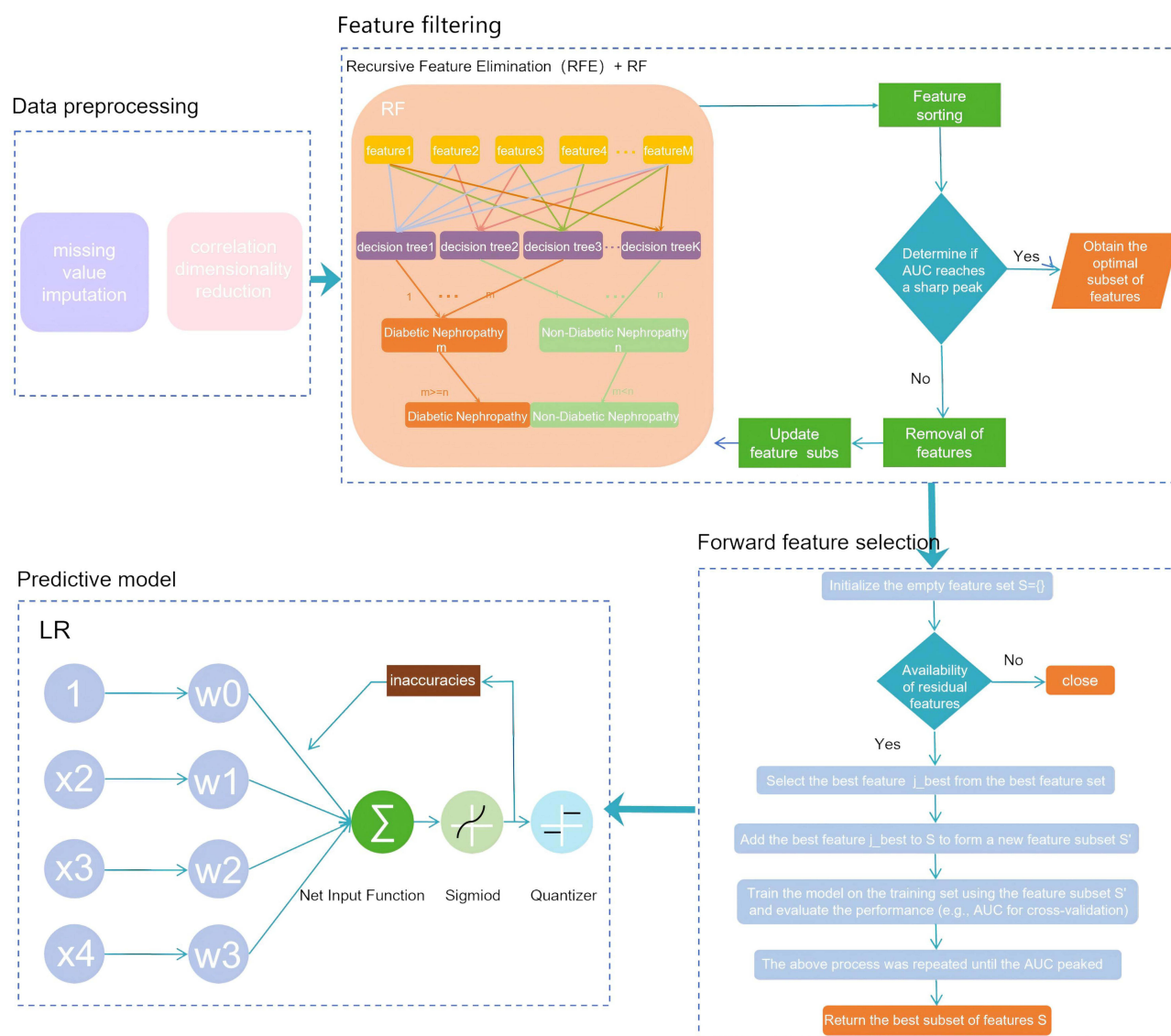
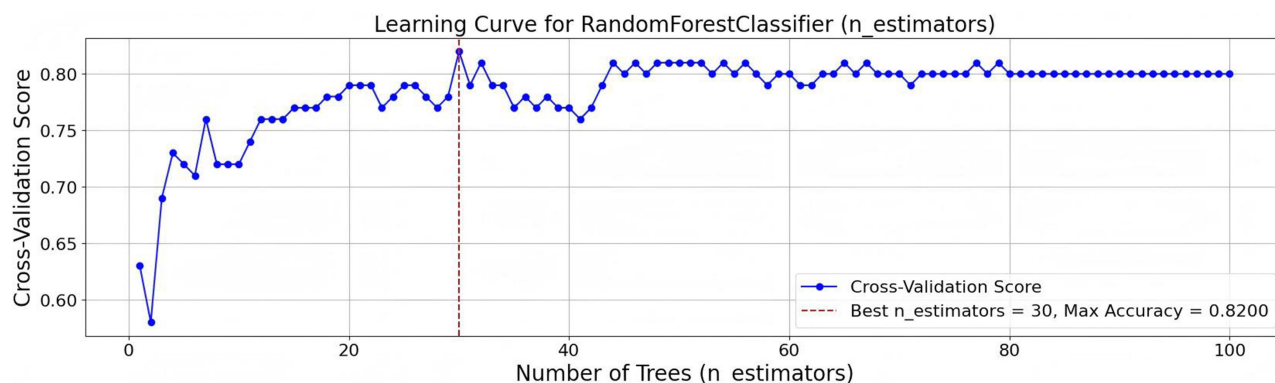
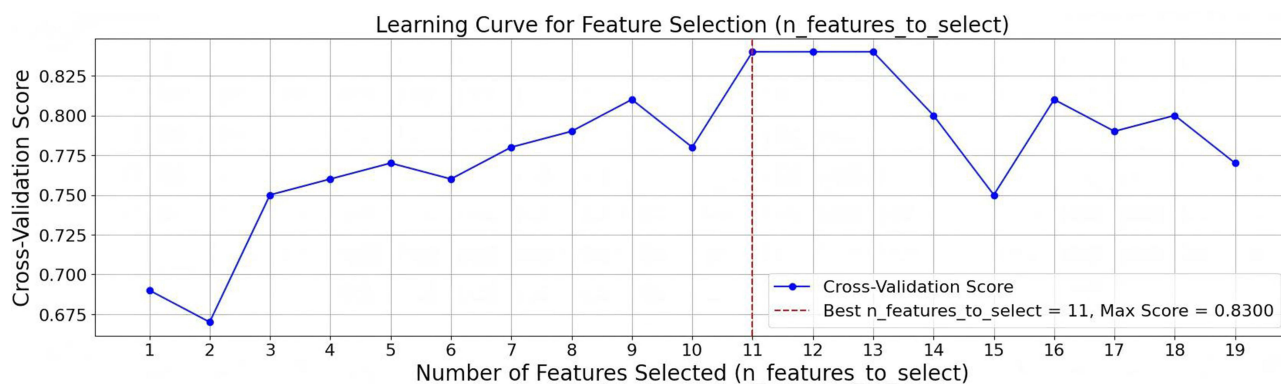


Figure 1 Flowchart of machine learning model construction.



**Figure 2** Determines the number of  $n\_estimators$ .



**Figure 3** Determines the number of  $n\_features\_to\_select$ .

with the best performance are selected, as shown in Figure 5. In addition, based on the previous research results of our center, “TG/Cys-C ratio” plays an important role in DN and NDRD, so it is specially included in this study, and these six indicators ultimately constitute the optimal feature set.<sup>22,23</sup> These characteristics include TG/Cys-C ratio, SBP, diabetes course, DR, HbA1c, and Hb.

## Methods of Validation

In this study, validation sets and external validation were used to evaluate the model. In order to ensure the accuracy of the algorithm and the reliability of the model, considering the small sample size, the five-times cross-validation method was used to verify the model. The data set was randomly divided into five non-overlapping parts, four of which served as the training set and one as the validation set. Repeating this process five times is called five-pass cross-validation, so each sample can be used as a validation set. Finally, the performance of the model is reevaluated using an external validation set.

Sensitivity (Sens), specificity (Spec), accuracy (ACC), Matthews correlation coefficient (MCC) and Area under curve (AUC) were used to comprehensively evaluate the model, and the indicators were calculated as follows:

$$Sens = TP / (TP + FN)$$

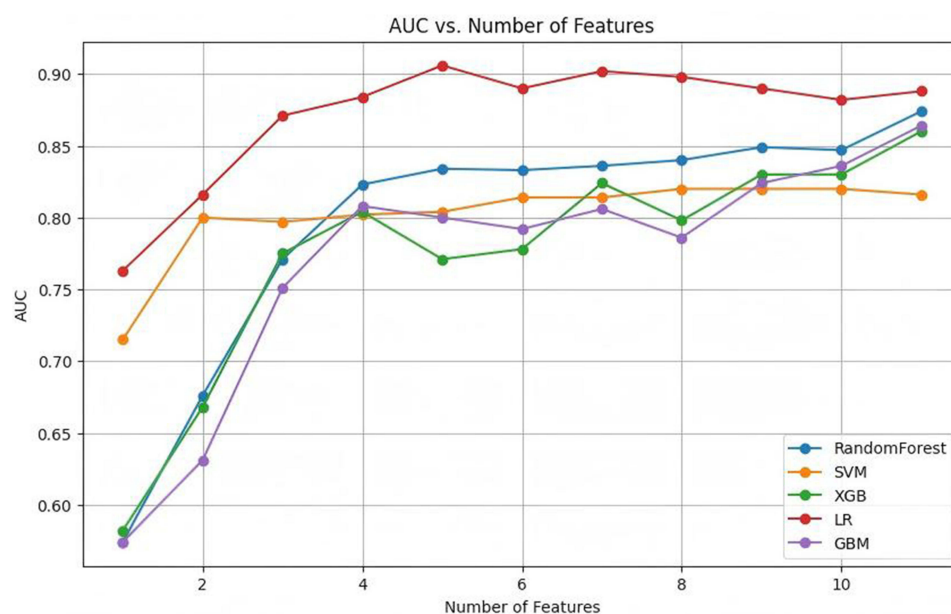
$$Spec = TN / (TN + TP)$$

$$ACC = (TP + TN) / (TP + TN + FP + FN)$$

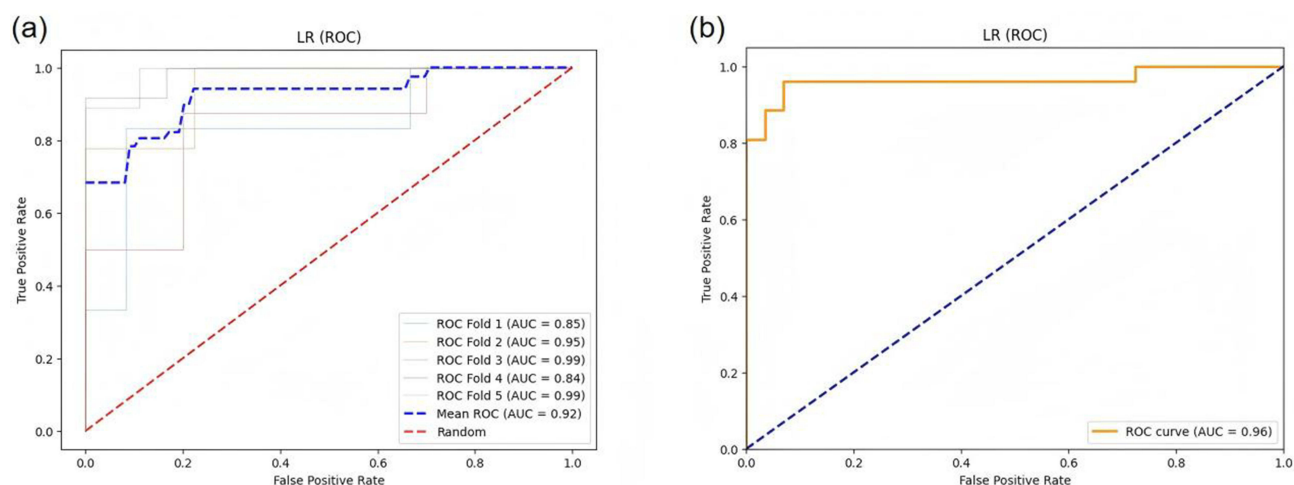
$$MCC = (TP \times TN - FP \times FN) / \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$$

TP is true positive, TN is true negative, FP is false positive, FN is false negative.





**Figure 4** Forward feature selection diagram.



**Figure 5** (a) ROC curve of a 5-fold cross-validated LR model. (b) ROC curve of the LR model external validation set.

Generally, Receiver Operating Characteristic (ROC) is used to evaluate the diagnostic value of the model. Generally speaking, the larger the area under the ROC curve (AUC), the higher the diagnostic value of the model and the better the prediction performance.

## Results

### Trial Population

This study included 100 patients, divided into DN (50 cases) and NDRD (50 cases) groups based on renal biopsy pathology. The comparison of general characteristics between the two groups is presented in Table 1.

DN patients were younger on average compared to NDRD patients. The DN group had higher proportions of hypertension and DR, along with elevated levels of SBP, DBP, diabetes duration, HbA1c, FPG, Cys-C, and Scr. In contrast, the DN group showed lower levels of Hb, eGFR, TC, TG, LDL-C, and TG/Cys-C ratio, with all differences being statistically significant ( $P < 0.05$ ).

**Table 1** Comparison of Clinical Characteristics Between the Two Groups

Clinical Parameters	DN Group (n=50)	NDRD Group (n=50)	Z/t/ $\chi^2$	P value
Age (years, $\bar{x} \pm s$ )	47.98 $\pm$ 11.02	54.16 $\pm$ 10.57	-2.861	0.005
Gender (%)			0.694	0.405
Male	34 (68)	30 (60)		
Female	16 (32)	20 (40)		
Hypertension (%)			10.928	<0.001
Yes	43 (86)	28 (56)		
No	7 (14)	22 (44)		
Tobacco smoking (%)			2.250	0.134
Yes	13 (26)	7 (14)		
No	37 (74)	43 (86)		
Alcohol drinking (%)			0.796	0.372
Yes	5 (10)	8 (16)		
No	45 (90)	42 (84)		
DR (%)			30.125	<0.001
Yes	30 (60)	4 (8)		
No	20 (40)	46 (92)		
Hematuria (%)			1.999	0.157
Yes	32 (64)	25 (50)		
No	18 (36)	25 (50)		
Urine sediment (%)			1.099	0.295
Yes	35 (70)	30 (60)		
No	15 (30)	20 (40)		
BMI [kg/m <sup>2</sup> , M (Q1, Q3)]	23.71 (21.48, 26.73)	24.07 (22.04, 27.34)	-0.925	0.355
SBP (mmHg, $\bar{x} \pm s$ )	156.80 $\pm$ 20.14	137.66 $\pm$ 17.67	5.052	<0.001
DBP [mmHg, M (Q1, Q3)]	90 (80, 100)	80 (76, 93)	-2.728	0.006
Diabetes duration [month, M (Q1, Q3)]	78 (24, 120)	18 (6, 48)	-3.882	<0.001
HbA1c [%; M (Q1, Q3)]	7.98 (6.50, 10.40)	7.10 (6.70, 7.90)	-2.410	0.016
Hb [g/L, M (Q1, Q3)]	115.66 $\pm$ 22.20	135.64 $\pm$ 18.59	-4.879	<0.001
FPG [mmol/L, M (Q1, Q3)]	7.40 (6.51, 10.55)	6.84 (5.59, 7.46)	-2.306	0.021
ALB [g/L, $\bar{x} \pm s$ ]	33.32 $\pm$ 8.22	30.53 $\pm$ 9.23	1.597	0.113
TC [mmol/L, M (Q1, Q3)]	6.18 (4.64, 7.37)	7.43 (5.84, 10.07)	-3.178	0.001
TG [mmol/L, M (Q1, Q3)]	1.89 (1.28, 2.50)	2.90 (1.72, 4.23)	-3.347	0.001
HDL-C [mmol/L, M (Q1, Q3)]	1.19 (0.95, 1.52)	1.38 (1.07, 1.64)	-1.693	0.091
LDL-C [mmol/L, M (Q1, Q3)]	3.70 (2.62, 4.69)	4.32 (3.18, 6.35)	-2.072	0.038
Cys-C [mg/L, M (Q1, Q3)]	1.26 (0.93, 1.71)	0.89 (0.71, 1.20)	-4.472	<0.001
TG/Cys-C ratio [M (Q1, Q3)]	1.45 (0.75, 1.99)	2.78 (1.81, 4.48)	-5.088	<0.001
Scr [ $\mu$ mol/L, M (Q1, Q3)]	93.50 (67.00, 139.00)	68.50 (55.00, 86.00)	-3.823	<0.001
UA ( $\mu$ mol/L, $\bar{x} \pm s$ )	337.34 $\pm$ 81.59	355.57 $\pm$ 82.42	-1.111	0.269
eGFR [mL/min/1.73m <sup>2</sup> , M (Q1, Q3)]	68.29 (40.54, 92.15)	98.90 (73.22, 119.54)	-3.547	<0.001
UPE, [g/24 h, M (Q1, Q3)]	4.55 (3.49, 7.60)	3.80 (2.04, 7.36)	-1.624	0.104
FIB [g/L, $\bar{x} \pm s$ ]	4.41 $\pm$ 1.33	4.19 $\pm$ 1.10	0.918	0.361

**Abbreviations:** DR, diabetic retinopathy; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HbA1c, hemoglobin A1c; Hb, hemoglobin; FPG, fasting plasma glucose; ALB, albumin; TC, total cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; Cys-C, cystatin C; TG/Cys-C ratio, triglyceride/cystatin C ratio; Scr, serum creatinine; UA, uric acid; eGFR, estimated glomerular filtration rate; UPE, urinary protein excretion obtained over 24 h; FIB, fibrinogen.

## Feature Selection

In this study, recursive feature elimination (RFE) method is used to screen important features. At the initial stage, the model was trained using all features, and the results showed that the prediction model built by random forest (RF) was the best (AUC: 0.8997, ACC: 0.8000, Sens: 0.8313, Spec: 0.8111). See Table 2 for details. In the second stage, RFE was

**Table 2** Model Comparison of Five Machine Learning Methods in the First Stage

	RF		GBM		LR		SVM		XGB	
	Test	Validation	Test	Validation	Test	Validation	Test	Validation	Test	Validation
AUC	0.8095	0.8997	0.8333	0.8871	0.8333	0.8855	0.7143	0.8393	0.8333	0.8497
ACC	0.8000	0.8000	0.8000	0.8000	0.6000	0.8111	0.7000	0.7667	0.7000	0.7333
Sens	0.6667	0.8313	0.7500	0.8733	0.6667	0.8620	0.6667	0.8106	0.7500	0.8123
Spec	0.8571	0.8111	0.8333	0.7364	0.5000	0.7490	0.7143	0.7492	0.6667	0.7112
MCC	0.5238	0.6388	0.5833	0.6230	0.1667	0.6166	0.3563	0.5539	0.4082	0.5057

**Table 3** Model Comparison of Five Machine Learning Methods in the Second Stage

	RF		GBM		LR		SVM		XGB	
	Test	Validation	Test	Validation	Test	Validation	Test	Validation	Test	Validation
AUC	0.9523	0.9121	0.8750	0.8930	0.8571	0.8612	0.8095	0.8458	0.9200	0.9041
ACC	0.9000	0.8222	0.7000	0.8000	0.6000	0.7667	0.7000	0.7778	0.7000	0.7667
Sens	0.6666	0.8328	0.7500	0.8629	0.6667	0.8335	0.6667	0.7383	0.6666	0.9202
Spec	0.7142	0.8333	0.6667	0.7252	0.5714	0.7031	0.7143	0.8189	0.8000	0.8441
MCC	0.7637	0.6643	0.4082	0.5925	0.5345	0.5315	0.3563	0.5687	0.4082	0.5704

used for feature extraction to screen out 11 features, including SBP, diabetes course, DR, HbA1c, Hb, FPG, TC, TG, LDL-C, TG/ Cys-C ratio and urinary protein quantification. As shown in Table 3, the effect of the model built by RF was improved (AUC: 0.9121, ACC: 0.8222, Sens: 0.8328, Spec: 0.8333). Finally, through forward feature selection, six optimal features were identified: TG/Cys-C ratio, SBP, diabetes course, DR, HbA1c and Hb. The predictive model based on logistic regression (LR) performed best (AUC: 0.9305, ACC: 0.8333, Sens: 0.8283, Spec: 0.8701).

## Model Comparison and Validation

The predictive models generated at each stage were compared using five machine learning methods: RF, GBM, LR, SVM, and XGB. Table 4 provides a comparison of the models generated in the third stage using the five machine learning methods. In the third stage, the prediction model built by LR algorithm had the best performance (AUC: 0.9305, ACC: 0.8333, Sens: 0.8283, Spec: 0.8701). We also performed external validation and finally achieved good predictive performance (AUC: 0.9642, ACC: 0.9455, Sens: 0.9615, Spec: 0.9310). ROC curves of LR prediction model and external validation are shown in Figure 5.

## Prediction Tool Development

Finally, in order to facilitate clinical use and promotion, we established a prediction model of DN using six indexes: TG/ Cys-C ratio, SBP, diabetes course, DR, HbA1c and Hb. It is named PDN (prediction of diabetic nephropathy). At the same time, we design a web site, <http://cppdd.cn/service/PDN>, as shown in Figure 6. The site is quick and easy to use,

**Table 4** Model Comparison of Five Machine Learning Methods in the Third Stage

	RF		GBM		LR		SVM		XGB	
	Test	Validation	Test	Validation	Test	Validation	Test	Validation	Test	Validation
AUC	0.9048	0.9000	0.7500	0.8799	0.9200	0.9305	0.7143	0.8704	0.7917	0.8831
ACC	0.8000	0.7889	0.8000	0.8000	0.9000	0.8333	0.8000	0.8000	0.7000	0.7778
Sens	0.6667	0.8480	0.7500	0.7883	0.9000	0.8283	0.6667	0.7717	0.8333	0.7577
Spec	0.8571	0.7778	0.8333	0.7683	0.8000	0.8701	0.8571	0.8444	0.5000	0.8317
MCC	0.5238	0.6264	0.5833	0.5799	0.8165	0.6914	0.5238	0.6127	0.3563	0.5857



CPPDD HOME SERVICES RESOURCES login register

*Prediction of Diabetic Nephropathy*

HOME DOCUMENT RESEARCH CONTACT US

Please input single sample data in the text form:

Serum:

\* Triglycerides to Cystatin C Ratio, TG: 1.9898

\* Systolic Blood Pressure, SBP: 168 mmHg

\* Diabetes Duration: 48 months

\* Diabetic Retinopathy, DR: 0

\* Hemoglobin A1c, HbA1c: 11.6 %

\* Hemoglobin, Hb: 128 g/L

Or upload multiple samples (Please strictly adherence to the csv format as [sample file](#)):

Drop file here or click to upload

limit 1 file, must be csv format

Submit Reset

**Figure 6** Web page diagram of DN prediction model.

and users simply enter the model indicators into the specified location in order and click “Submit”. During the input process, ensure that the input unit is the same as the unit on the interface. After calculation and analysis, our model will provide the percentage probability that the sample is DN on the results page.

## Discussion

With the continuous development of medical information technology, more and more machine learning models have been applied to the diagnosis and prediction of medical diseases. There are few similar studies on the correlation between DN and NDRD when learning models are applied. In a single-center retrospective study, elevated serum creatinine and urea, decreased eGFR, and the presence of retinopathy and peripheral neuropathy were identified as key biomarkers associated with DN.<sup>18</sup> Therefore, this study intends to collect multi-center data for retrospective analysis, and build a differential prediction model for DN and NDRD based on machine learning model. The purpose is to early screen out patients with higher probability of complicated NDRD among people with type 2 diabetes with proteinuria, encourage kidney puncture biopsy, identify pathological types, and actively adjust treatment plan to improve patient prognosis. Improve patient survival rate.

The establishment of the model is a process of gradual exploration. At the initial stage, the model was trained using all features, and the results showed that the prediction model built by random forest (RF) was the best (AUC: 0.8997, ACC: 0.8000, Sens: 0.8313, Spec: 0.8111). At the initial stage, there are many clinical indicators in the prediction model, so we want to reduce the number of indicators in the model and improve the prediction ability of the model. Therefore, in the second stage, RFE was used for feature extraction to screen out 11 features, including SBP, duration of diabetes, DR, HbA1c, Hb, fasting blood glucose, cholesterol, triglyceride, low density lipoprotein cholesterol, TG/ Cys-C ratio and urinary protein quantification. The results of the RF model were improved (AUC: 0.9121, ACC: 0.8222, Sens: 0.8328,

Spec: 0.8333). In the second stage, compared with the first stage, the prediction index is more simplified and the prediction ability is improved. Finally, based on the previous research results, the high level (TG/Cys-C ratio) has good predictive performance for predicting NDRD.<sup>22,23</sup> Therefore, in the third stage, after the inclusion of TG/Cys-C ratio, we determined six optimal features through forward feature selection: TG/Cys-C ratio, SBP, diabetes course, DR, HbA1c and Hb. The predictive model based on logistic regression (LR) performed best (AUC: 0.9305, ACC: 0.8333, Sens: 0.8283, Spec: 0.8701). Finally, 55 cases from Tai 'an Central Hospital were used for external validation, and the results showed that the model achieved good predictive performance (AUC: 0.9642, ACC: 0.9455, Sens: 0.9615, Spec: 0.9310).

TG/Cys-C ratio may be a good predictor of differential diagnosis of DN and NDRD. At present, there are few studies on the correlation of TG/Cys-C ratio to identify DN and NDRD. Our previous study found that the greater the serum TG/Cys-C ratio in T2DM patients with kidney injury, the greater the probability of predicting NDRD.<sup>21,22</sup> Previous studies of our center have found that the lipid level of DN patients is lower than that of NDRD. Therefore, we speculate that there may be different pathophysiological mechanisms in lipid metabolism of DN and NDRD patients. Of course, this will be our next research focus. The TG/Cys-C ratio is an indicator that takes into account both the dyslipidemia of NDRD and the abnormal renal function of DN in patients with type 2 diabetes mellitus. As for the differential diagnostic efficacy of TG/Cys-C ratio in DN and NDRD patients, the results of the prediction model constructed in this study showed that TG/Cys-C ratio showed good predictive efficacy in the model.

Previous studies have shown that the SBP in DN patients is significantly higher than that in NDRD patients.<sup>11,24,25</sup> In the prediction model constructed in this study, SBP is also a clinical predictor for distinguishing DN from NDRD. The lower the SBP level in T2DM patients with renal injury, the greater the probability of predicting the diagnosis of NDRD. The pathogenesis of hypertension in T2DM patients with proteinuria is usually complicated. This may be due to the activation of the RASS system and the influence of the diabetic environment on the release of reactive oxygen species and inflammatory mediators.<sup>26</sup> A prospective study has shown that strict control of blood pressure in DN patients can reduce the incidence of macrovascular and microvascular complications.<sup>27,28</sup> In clinical work, diabetic patients with poor blood sugar control are more likely to be complicated with complications such as heart, brain and kidney. High glucose toxicity is associated with the glomerular arteries, which will cause a series of oxidative stress reactions, resulting in glomerular injury, proteinuria, and thus the occurrence of DN. The model established in this study shows that the duration of diabetes is a predictor of differential diagnosis of DN and NDRD. The shorter the course of diabetes, the greater the likelihood of NDRD in T2DM patients with kidney damage.

Diabetic retinopathy (DR) and DN are both diabetic microvascular complications, the pathogenesis is very similar, the physiological structure of the retina and the physiological structure of the kidney are homologous, so most people will think that the occurrence of DR Means that DN occurs at the same time.<sup>29–32</sup> However, with the continuous improvement of renal puncture biopsy techniques, studies have shown that the inconsistency between DR And DN also exists.<sup>33,34</sup> Meta-analyses have shown that DR Has high sensitivity and specificity in predicting the diagnosis of DN.<sup>33</sup> Other studies have also confirmed that the presence of DR strongly suggests DN, while the absence of retinopathy is a major predictor of NDRD.<sup>13,35</sup> In the prediction model constructed in this study, DR Is a good predictor for distinguishing patients with DN and NDRD. In patients with T2DM with proteinuria and DR Absence, we are more likely to consider NDRD, suggesting that further renal biopsy is more important.

HbA1c reflects the average blood sugar level of the past 2 to 3 months and is an indicator to monitor the level of blood sugar control in diabetes. Previous studies have shown that serum HbA1c is also considered to be an important predictor of DN.<sup>36,37</sup> Previous studies have suggested that a higher HbA1c level in T2DM patients with kidney injury means that DN is more likely to occur, while a lower HbA1c is more likely to be associated with NDRD.<sup>38</sup> The prediction model constructed in this study showed that HbA1c level also had good predictive performance for the differential diagnosis of DN and NDRD.

The incidence of decreased Hb level is higher in DN patients. The results of our previous study showed that the prevalence of anemia in diabetic patients receiving kidney biopsy and biopsy-confirmed DN cases was higher than that in the NDRD group, which was consistent with the results of previous studies.<sup>39–41</sup> The mechanisms of anemia in DN patients are complex, including increased bleeding tendencies associated with antiplatelet action, decreased oxygen sensitivity due to autonomic neuropathy or use of inhibitors of the renin-angiotensin-aldosterone system, inhibition of

inflammatory cytokines, loss of erythropoietin (EPO) in the urine, and poor response to EPO.<sup>42</sup> Studies have shown that there is a significant correlation between anemia and DN, and the prevalence of anemia may be related to interstitial fibrosis and renal tubule atrophy in DN patients.<sup>43</sup> In the prediction model constructed in this study, Hb is a good predictor for distinguishing DN from NDRD.

The limitations of this study include the following: it is a retrospective study, and all patients' medical history data were obtained from the internal electronic medical record system of hospital, which may not be accurate in the collection and recording of medical history, so information bias cannot be avoided. The sample size of this study is relatively small, so more data from other hospitals can be included for model construction and comparison in the future, so as to select a better model for the differential diagnosis of DN and NDRD. The data in this study are all from Chinese cities, and considering ethnic differences, the research results are only of reference value for the global human race, and the data of different ethnic groups can be included in the future statistical analysis. In addition, there are many factors involved in the pathogenesis of DN and NDRD, and further exploration is needed in the future.

## Conclusion

In summary, we use LR to screen features and build a differential diagnosis model between DN and NDRD. The model consisted of TG/Cys-C ratio, SBP, duration of diabetes, DR, HbA1c and Hb. The results of this study show that the model can identify DN and NDRD quickly and easily. Personalized kidney biopsies in patients with type 2 diabetes and proteinuria require a comprehensive evaluation of the patient's clinical presentation, potential benefits of surgery, inherent risks, and more. The results of this study highlight the importance of conducting kidney biopsies in patients with type 2 diabetes who are highly suspected of having NDRD, as accurate diagnosis and appropriate treatment can greatly improve their prognosis.

Distinguishing between DN and NDRD remains a significant challenge. Compared to similar studies, the feature selection and model tuning methods in this research exhibit better robustness and interpretability, but there is still potential for improvement in computational complexity and parameter optimization. Additionally, based on our findings, we developed the first online calculator for predicting DN, enabling easier calculation of DN probability. Nevertheless, this study has limitations, and future efforts will focus on expanding the sample size and developing more sophisticated models and methods to enhance the accuracy of DN identification.

## Ethical Statement

Ethics approval and consent to participate Written informed consent for all data was obtained from patients during their hospitalization, and the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University approved the study. (XYFY2021-KL073-01).

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Disclosure

The authors declare that they have no conflicts of interest.

## References

1. Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2022 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney Int.* 2022;102(5S):S1–S127. doi:10.1016/j.kint.2022.06.013.
2. Ruiz-Ortega M, Rodrigues-Diez RR, Lavoiz C, Rayego-Mateos S. Special issue “diabetic nephropathy: diagnosis, prevention and treatment”. *J Clin Med.* 2020;9(3):813. doi:10.3390/jcm9030813

3. Tervaert TW, Mooyaart AL, Amann K, et al. Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol.* 2010;21(4):556–563. doi:10.1681/ASN.2010010010
4. Santoro D, Torreggiani M, Pellicanò V, et al. Kidney biopsy in type 2 diabetic patients: critical reflections on present indications and diagnostic alternatives. *Int J Mol Sci.* 2021;22(11):5425. doi:10.3390/ijms22115425
5. Bermejo S, González E, López-Revuelta K, et al. Risk factors for non-diabetic renal disease in diabetic patients. *Clin Kidney J.* 2020;13(3):380–388. doi:10.1093/ckj/sfz177
6. Fontana F, Perrone R, Giaroni F, et al. Clinical predictors of nondiabetic kidney disease in patients with diabetes: a single-center study. *Int J Nephrol.* 2021;2021:9999621. doi:10.1155/2021/9999621
7. Lee YH, Kim KP, Kim YG, et al. Clinicopathological features of diabetic and nondiabetic renal diseases in type 2 diabetic patients with nephrotic-range proteinuria. *Medicine.* 2017;96(36):e8047. doi:10.1097/MD.00000000000008047
8. Li L, Zhang X, Li Z, et al. Renal pathological implications in type 2 diabetes mellitus patients with renal involvement. *J Diabetes Complications.* 2017;31(1):114–121. doi:10.1016/j.jdiacomp.2016.10.024
9. Li L, Yang Y, Zhu X, et al. Design and validation of a scoring model for differential diagnosis of diabetic nephropathy and nondiabetic renal diseases in type 2 diabetic patients. *J Diabetes.* 2020;12(3):237–246. doi:10.1111/1753-0407.12994
10. Yang YZ, Wang JW, Wang F, et al. Incidence, development, and prognosis of diabetic kidney disease in China: design and methods. *Chin Med J.* 2017;130(2):199–202. doi:10.4103/0366-6999.198002
11. Umanath K, Lewis JB. Update on diabetic nephropathy: core curriculum 2018. *Am J Kidney Dis.* 2018;71(6):884–895. doi:10.1053/j.ajkd.2017.10.026
12. Fiorentino M, Bolognino D, Tesar V, et al. Renal biopsy in patients with diabetes: a pooled meta-analysis of 48 studies. *Nephrol Dial Transplant.* 2017;32(1):97–110. doi:10.1093/ndt/gfw070
13. Dong Z, Wang Y, Qiu Q, et al. Clinical predictors differentiating non-diabetic renal diseases from diabetic nephropathy in a large population of type 2 diabetes patients. *Diabet Res Clin Pract.* 2016;121:112–118. doi:10.1016/j.diabres.2016.09.005
14. Yang Z, Feng L, Huang Y, Xia N. A differential diagnosis model for diabetic nephropathy and non-diabetic renal disease in patients with type 2 diabetes complicated with chronic kidney disease. *Diabetes Metab Syndr Obes.* 2019;12:1963–1972. doi:10.2147/DMSO.S223144
15. Sabanayagam C, He F, Nusinovic S, et al. Prediction of diabetic kidney disease risk using machine learning models: a population-based cohort study of Asian adults. *Elife.* 2023;12:e81878. doi:10.7554/eLife.81878
16. Zueger T, Schallmoser S, Kraus M, Saar-Tsechansky M, Feuerriegel S, Stettler C. Machine learning for predicting the risk of transition from prediabetes to diabetes. *Diabetes Technol Ther.* 2022;24(11):842–847. doi:10.1089/dia.2022.0210
17. Zou Y, Zhao L, Zhang J, et al. Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease. *Ren Fail.* 2022;44(1):562–570. doi:10.1080/0886022X.2022.2056053
18. Zhu Y, Zhang Y, Yang M, et al. Machine learning-based predictive modeling of diabetic nephropathy in type 2 diabetes using integrated biomarkers: a single-center retrospective study. *Diabetes Metab Syndr Obes.* 2024;17:1987–1997. doi:10.2147/DMSO.S458263
19. Kolker E, Özdemir V, Kolker E. How healthcare can refocus on its super-customers (patients, n=1) and customers (doctors and nurses) by leveraging lessons from Amazon, Uber, and Watson. *OMICS.* 2016;20(6):329–333. doi:10.1089/omi.2016.0077
20. Daimiel Naranjo I, Gibbs P, Reiner JS, et al. Radiomics and machine learning with multiparametric breast MRI for improved diagnostic accuracy in breast cancer diagnosis. *Diagnostics.* 2021;11(6):919. doi:10.3390/diagnostics11060919
21. American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. *Diabetes Care.* 2018;41(Suppl 1):S13–S27. doi:10.2337/dc18-S002.
22. Wei J, Wang B, Shen FJ, Zhang TT, Duan Z, Zhou DM. Diagnostic value of triglyceride and cystatin C ratio in diabetic kidney disease: a retrospective and prospective cohort study based on renal biopsy. *BMC Nephrol.* 2022;23(1):270. doi:10.1186/s12882-022-02888-3
23. Zhou DM, Wei J, Zhang TT, Shen FJ, Yang JK. Establishment and validation of a nomogram model for prediction of diabetic nephropathy in type 2 diabetic patients with proteinuria. *Diabetes Metab Syndr Obes.* 2022;15:1101–1110. doi:10.2147/DMSO.S357357
24. Zhang J, Wang Y, Zhang R, et al. Serum levels of immunoglobulin G and complement 3 differentiate non-diabetic renal disease from diabetic nephropathy in patients with type 2 diabetes mellitus. *Acta Diabetol.* 2019;56(8):873–881. doi:10.1007/s00592-019-01339-0
25. Hu F, Zhang T. Study on risk factors of diabetic nephropathy in obese patients with type 2 diabetes mellitus. *Int J Gen Med.* 2020;13:351–360. doi:10.2147/IJGM.S255858
26. Jiang S, Fang J, Yu T, et al. Novel model predicts diabetic nephropathy in type 2 diabetes. *Am J Nephrol.* 2020;51(2):130–138. doi:10.1159/000505145
27. Hsieh JT, Chang FP, Yang AH, Tarng DC, Yang CY. Timing of kidney biopsy in type 2 diabetic patients: a stepwise approach. *BMC Nephrol.* 2020;21(1):131. doi:10.1186/s12882-020-01794-w
28. Wang J, Han Q, Zhao L, et al. Identification of clinical predictors of diabetic nephropathy and non-diabetic renal disease in Chinese patients with type 2 diabetes, with reference to disease course and outcome. *Acta Diabetol.* 2019;56(8):939–946. doi:10.1007/s00592-019-01324-7
29. Farrah TE, Dhillon B, Keane PA, Webb DJ, Dhaun N. The eye, the kidney, and cardiovascular disease: old concepts, better tools, and new horizons. *Kidney Int.* 2020;98(2):323–342. doi:10.1016/j.kint.2020.01.039
30. Fursova AZ, Derbeneva AS, Vasilyeva MA, Tarasov MS, Nikulich IF, Galkina EV. Development, clinical manifestations and diagnosis of retinal changes in chronic kidney disease. *Vestn Oftalmol.* 2021;137(1):107–114. doi:10.17116/oftalma202113701107
31. Lovshin JA, Lytvyn Y, Lovblom LE, et al. Retinopathy and RAAS activation: results from the Canadian study of longevity in type 1 diabetes. *Diabetes Care.* 2019;42(2):273–280. doi:10.2337/dc18-1809
32. Wang Q, Cheng H, Jiang S, et al. The relationship between diabetic retinopathy and diabetic nephropathy in type 2 diabetes. *Front Endocrinol.* 2024;15:1292412. doi:10.3389/fendo.2024.1292412
33. He F, Xia X, Wu XF, Yu XQ, Huang FX. Diabetic retinopathy in predicting diabetic nephropathy in patients with type 2 diabetes and renal disease: a meta-analysis. *Diabetologia.* 2013;56(3):457–466. doi:10.1007/s00125-012-2796-6
34. Zhuo L, Ren W, Li W, Zou G, Lu J. Evaluation of renal biopsies in type 2 diabetic patients with kidney disease: a clinicopathological study of 216 cases. *Int Urol Nephrol.* 2013;45(1):173–179. doi:10.1007/s11255-012-0164-6
35. Wang X, Li J, Huo L, et al. Clinical characteristics of diabetic nephropathy in patients with type 2 diabetic mellitus manifesting heavy proteinuria: a retrospective analysis of 220 cases. *Diabet Res Clin Pract.* 2019;157:107874. doi:10.1016/j.diabres.2019.107874

36. Guo J, Liu C, Wang Y, et al. Dose-response association of diabetic kidney disease with routine clinical parameters in patients with type 2 diabetes mellitus: a systematic review and meta-analysis. *EClinicalMedicine*. 2024;69:102482. doi:10.1016/j.eclinm.2024.102482
37. Shah HS, McGill JB, Hirsch IB, et al. Poor glycemic control is associated with more rapid kidney function decline after the onset of diabetic kidney disease. *J Clin Endocrinol Metab*. 2024;109(8):2124–2135. doi:10.1210/clinem/dgae044
38. Demirelli B, Boztepe B, Şenol EG, et al. Non-diabetic nephropathy in diabetic patients: incidence, HbA1c variability and other predictive factors, and implications. *Int Urol Nephrol*. 2024;56(9):3091–3100. doi:10.1007/s11255-024-04066-w
39. Kim M, Lee SH, Park KS, Kim EJ, Yeo S, Ha IH. Association between diabetes mellitus and anemia among Korean adults according to sex: a cross-sectional analysis of data from the Korea National Health and Nutrition Examination Survey (2010–2016). *BMC Endocr Disord*. 2021;21(1):209. doi:10.1186/s12902-021-00873-9
40. Chen X, Xie J, Zhang Y, et al. Prognostic value of hemoglobin concentration on renal outcomes with diabetic kidney disease: a retrospective cohort study. *Diabetes Metab Syndr Obes*. 2024;17:1367–1381. doi:10.2147/DMSO.S471641
41. Okada A, Yamaguchi S, Imaizumi T, et al. Modification effects of albuminuria on the association between kidney function and development of anemia in diabetes. *J Clin Endocrinol Metab*. 2024;109(4):1012–1032. doi:10.1210/clinem/dgad660
42. Tsai SF, Tarng DC. Anemia in patients of diabetic kidney disease. *J Chin Med Assoc*. 2019;82(10):752–755. doi:10.1097/JCMA.000000000000175
43. Ito K, Yokota S, Watanabe M, et al. Anemia in diabetic patients reflects severe tubulointerstitial injury and aids in clinically predicting a diagnosis of diabetic nephropathy. *Intern Med*. 2021;60(9):1349–1357. doi:10.2169/internalmedicine.5455-20

## Diabetes, Metabolic Syndrome and Obesity

### Publish your work in this journal

Diabetes, Metabolic Syndrome and Obesity is an international, peer-reviewed open-access journal committed to the rapid publication of the latest laboratory and clinical findings in the fields of diabetes, metabolic syndrome and obesity research. Original research, review, case reports, hypothesis formation, expert opinion and commentaries are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/diabetes-metabolic-syndrome-and-obesity-journal>

**Dovepress**  
Taylor & Francis Group