

# Design of Machine Learning Algorithms and Internal Validation of a Kidney Risk Prediction Model for Type 2 Diabetes Mellitus

Ying Wang<sup>1</sup>, Han-Xin Yao<sup>1</sup>, Zhen-Yi Liu<sup>1</sup>, Yi-Ting Wang<sup>1</sup>, Si-Wen Zhang<sup>2</sup>, Yuan-Yuan Song<sup>1</sup>, Qin Zhang<sup>1</sup>, Hai-Di Gao<sup>1</sup>, Jian-Cheng Xu<sup>1</sup> 

<sup>1</sup>Department of Laboratory Medicine, First Hospital of Jilin University, Changchun, 130021, People's Republic of China; <sup>2</sup>Department of Endocrinology & Metabolism, First Hospital of Jilin University, Changchun, 130021, People's Republic of China

Correspondence: Jian-Cheng Xu, Department of Laboratory Medicine, First Hospital of Jilin University, Changchun, 130021, People's Republic of China, Tel/Fax +86-0431 88782595, Email xjc@jlu.edu.cn

**Objective:** This study aimed to explore specific biochemical indicators and construct a risk prediction model for diabetic kidney disease (DKD) in patients with type 2 diabetes (T2D).

**Methods:** This study included 234 T2D patients, of whom 166 had DKD, at the First Hospital of Jilin University from January 2021 to July 2022. Clinical characteristics, such as age, gender, and typical hematological parameters, were collected and used for modeling. Five machine learning algorithms [Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF)] were used to identify critical clinical and pathological features and to build a risk prediction model for DKD. Additionally, clinical data from 70 patients ( $n_{T2D} = 20$ ,  $n_{DKD} = 50$ ) were collected for external validation from the Third Hospital of Jilin University.

**Results:** The RF algorithm demonstrated the best performance in predicting progression to DKD, identifying five major indicators: estimated glomerular filtration rate (eGFR), glycosylated albumin (GA), Uric acid, HbA1c, and Zinc (Zn). The prediction model showed sufficient predictive accuracy with area under the curve (AUC) values of 0.960 (95% CI: 0.936–0.984) and 0.9326 (95% CI: 0.8747–0.9885) in the internal validation set and external validation set, respectively. The diagnostic efficacy of the RF model (AUC = 0.960) was significantly higher than each of the five features screened with the highest feature importance in the RF model.

**Conclusion:** The online DKD risk prediction model constructed using the RF algorithm was selected based on its strong performance in the internal validation.

**Keywords:** diabetic kidney disease, type 2 diabetes, machine learning model, random forest algorithm

## Introduction

The prevalence and incidence of type 2 diabetes (T2D), representing more than 90% of all diabetes cases, are rapidly increasing worldwide.<sup>1</sup> Diabetic kidney disease (DKD) is a critical microvascular complication of diabetes mellitus with a high prevalence, mortality rate, and substantial financial burden.<sup>2</sup> Approximately one-third of T2D patients can progress to DKD,<sup>3</sup> subsequently leading to end-stage kidney disease (ESKD) and death.<sup>4</sup> Identifying risk factors and early diagnosis of DKD are crucial for the prevention and treatment of ESKD.

Chronic kidney disease leading to ESKD in T2D includes DKD, nondiabetic kidney disease (NDKD), or a combination of DKD and NDKD. Given the different treatment options, renal biopsy histopathology remains the gold standard for distinguishing between DKD and NDKD.<sup>5</sup> However, renal biopsy is cautiously used in clinical diagnosis. In contrast to histopathological biopsy, diagnosing DKD based on clinical, physiological, and biochemical indicators can shorten the diagnostic time, reduce patient discomfort, and minimize medical risks. Therefore, many risk factors involved in the initiation and progression of DKD, such as increasing age, family history, hyperglycemia, hypertension, dyslipidemia, dietary patterns, and lifestyles have been reported.<sup>6</sup> Physiological indicators, such as urinary albumin-to-creatinine ratio and estimated

glomerular filtration rate (eGFR), have been included as screening indicators for the occurrence and grading of DKD in clinical diagnostic guidelines for DKD.

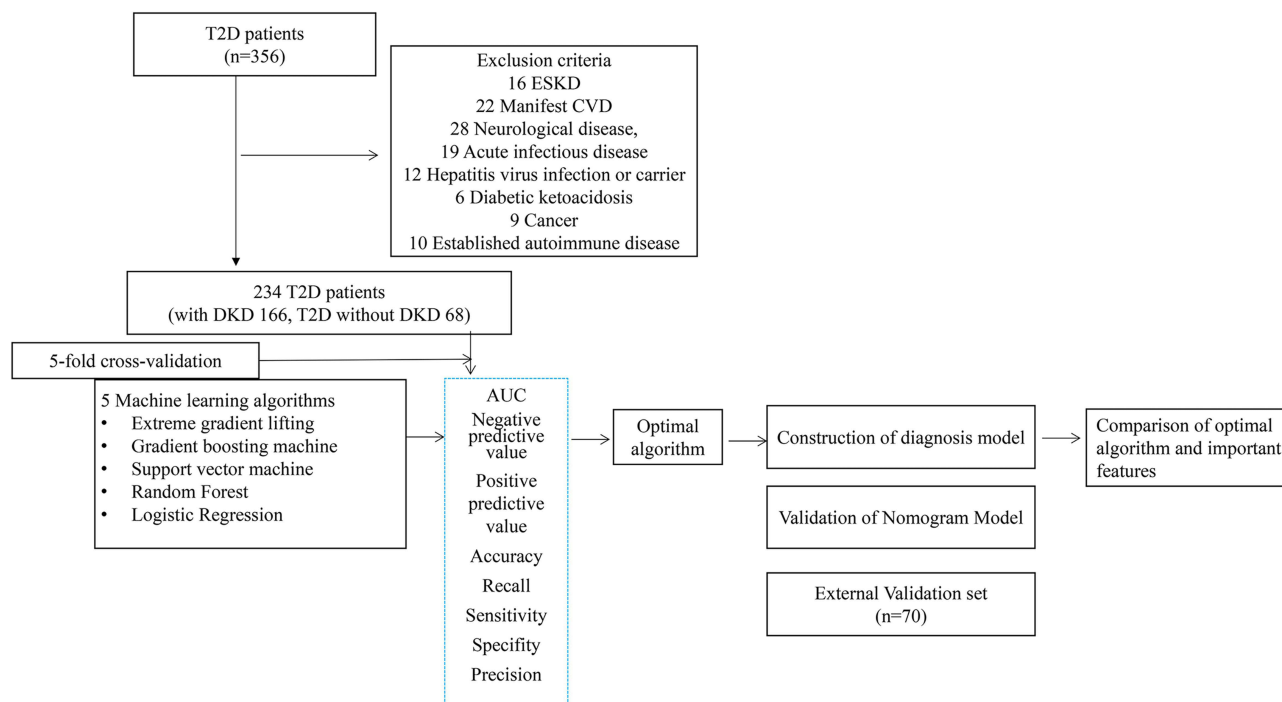
Although researchers have studied early predictive markers of DKD, including proteomics and genomics, these new biomarkers are challenging to popularize in clinical setting due to their low sensitivity and specificity in evaluating early DKD. Even the effectiveness of microalbuminuria as a traditional DKD marker and the best opportunity for intervention has recently been questioned. In fact, some biochemical indicators of T2D patients have been considered to provide guidance for the early diagnosis of DKD. The varying presentation of DKD poses challenges for clinicians in terms of accurate detection and selecting appropriate individualized interventions in high-risk individuals.<sup>7</sup> Therefore, it is necessary to build a simple and practical DKD risk prediction model to help clinicians identify patients' risks early, provide early treatment, and prevent T2D patients from progressing to DKD, rather than blindly pursuing new indicators.

In recent years, machine learning (ML) algorithms have made significant progress in disease prediction and prognosis forecasting. Previous studies have successfully applied ML algorithms for the prediction of ESRD in DKD participants.<sup>8</sup> As a novel statistical modeling method, based on conventional laboratory test data, we will apply multiple ML algorithms to develop a real-world DKD artificial intelligence early warning model. Our research will provide effective guidance for early diagnosis and risk prediction of DKD.

## Methods

### Study Populations and Design

T2D hospitalized patients in this study were enrolled from January 2021 to July 2022 at the First Hospital of Jilin University and the Third Hospital of Jilin University. Only patients with a clinical diagnosis of T2D and a definite pathological diagnosis were included. This study does not include ESKD patients (defined as those receiving basic dialysis, kidney transplantation, or with eGFR < 15 mL/min/1.73 m<sup>2</sup>), severe cardiovascular disease (CVD), nervous system disease, acute infectious disease, hepatitis virus infection, or carrying a history of such infection, diabetic ketoacidosis, any type of cancer, or identified autoimmune diseases. T2D was diagnosed according to the American Diabetes Association standard.<sup>9</sup> Nephropathological Classification of DKD is based on the Society of Renal Pathology (RPS) criteria from 2010.<sup>10</sup> A schematic diagram of patient selection is shown in Figure 1. Our research complies with the Declaration of Helsinki and was approved by the Ethics



**Figure 1** Flowchart showing the selection of patients and construction of the predictive model.

Committee of the First Hospital of Jilin University (No. 2023–333) and the Ethics Committee of the Third Hospital of Jilin University (No. 2023033021).

## Blood Sampling and Laboratory Testing

Blood samples from patients were collected into metal-free tubes after an overnight fast. After centrifugation at 1500 g for 10 min, serum was aliquoted into metal-free Eppendorf tubes and stored at  $-80^{\circ}\text{C}$ . A total of 50 hematological and biochemical parameters were tested at baseline, including fasting plasma glucose, hemoglobin A1c (HbA1c), glycated albumin (GA), and aspartate aminotransferase (AST). Detailed indicators are shown in Table 1. These indicators were examined using the AU5800 (Sunto-gun Bakman, Japan, 454–32 Higashino, Nagaizumi-cho, Sunto-gun, Japan), XN-9000 (Sysmex, Hyogo Prefecture, Japan), UniCel DXI800 (Chaska Bakman, USA), and HLC-723G8 glycosylated hemoglobin analyzer (Tosoh,

**Table 1** Clinical Characteristics of the Study Participants

	<b>DKD group (n=166)</b>	<b>T2D group (n=68)</b>	<b>T /W</b>	<b>P</b>
Age, year	53.54±14.44	60.98±14.71	-3.53	0.001 **
Sex				
Male	102 (61.4%)	41 (60.3%)	0.163	0.870
Female	64 (38.6%)	27 (39.7%)		
FPG, g/L	8.99 (7.48–11.50)	8.98 (7.63–11.91)	5227	0.376
HbA1c, %	7.20 (6.20–8.23)	8.30 (7.43–9.48)	3470	0.000 **
GA, %	22.60 (18.00–24.62)	27.39 (24.29–33.13)	2070	0.000 **
GA/HbA1c	2.98 (2.57–3.32)	3.49 (3.04–3.94)	3516	0.000 **
AST, U/L	18.55 (14.20–23.53)	15.00 (12.60–22.15)	6550	0.054
ALT, U/L	15.00 (8.85–22.33)	10.30 (7.00–18.20)	6795	0.014 *
ALB, g/L	36.60 (32.88–39.23)	34.00 (28.63–38.26)	7054.5	0.003 **
$\gamma$ -GGT, U/L	35.25 (21.50–52.08)	30.70 (20.83–51.03)	6105.5	0.327
TBil, $\mu\text{mol/L}$	11.50 (8.70–15.70)	8.25 (6.43–13.08)	7388.5	0.000 **
DBil, $\mu\text{mol/L}$	2.10 (1.48–3.08)	1.65 (1.10–2.58)	6819	0.012 *
IBil, $\mu\text{mol/L}$	8.65 (7.20–12.73)	6.95 (5.10–9.88)	7395.5	0.000 **
Ferritin, $\mu\text{g/L}$	253.55 (116.58–503.25)	246.00 (95.98–537.05)	6094.5	0.339
UIBC, $\mu\text{mol/L}$	31.69±15.52	32.84±13.08	-0.58	0.562
BUN, mmol/L	5.65 (4.32–6.63)	10.91 (6.98–21.80)	2367	0.000 **
Cre, $\mu\text{mol/L}$	64.90 (52.13–72.40)	140.95 (78.50–366.08)	2102	0.000 **
Uric acid, $\mu\text{mol/L}$	276.50 (185.75–360.50)	390.50 (305.75–489.50)	2938	0.000 **
RBP, mg/L	44.55 (32.85–58.95)	56.50 (40.53–82.03)	4061.5	0.001 **
Cystatin C, mg/L	0.83 (0.76–0.99)	1.98 (0.96–3.50)	2485	0.000 **
Egfr, mL/min	108.00 (97.00–113.00)	41.00 (14.00–92.25)	9331	0.000 **
TG, mmol/L	1.76 (1.15–2.84)	1.72 (1.11–2.70)	5702.5	0.902
CHOL, mmol/L	4.67 (3.80–5.30)	4.12 (3.25–5.37)	6434	0.093
HDL, mmol/L	0.96 (0.79–1.18)	0.87 (0.71–1.06)	6502.5	0.068
LDL, mmol/L	2.84 (2.25–3.47)	2.62 (2.03–3.38)	6159	0.274
NEFA, mmol/L	0.67 (0.43–0.83)	0.60 (0.38–0.82)	6276	0.179
sd-LDL-C, mmol/L	0.91 (0.72–1.13)	0.82 (0.59–1.13)	6368.5	0.124
Mg, mmol/L	0.77 (0.63–0.86)	0.76 (0.60–0.86)	5847	0.667
Zn, $\mu\text{mol/L}$	12.52±3.50	11.39±3.64	2.189	0.030 *
Cu, $\mu\text{mol/L}$	15.00 (12.93–17.83)	16.30 (13.95–19.25)	4546	0.020 *
Fe, $\mu\text{mol/L}$	14.75 (10.63–19.80)	11.45 (7.33–15.90)	7338	0.000 **
NE#, $\times 10^9/\text{L}$	4.32 (2.68–6.03)	4.48 (3.33–6.34)	4940	0.134
NE%, %	0.65 (0.54–0.71)	0.66 (0.59–0.76)	4643.5	0.033 *
LY#, $\times 10^9/\text{L}$	1.54 (1.09–2.14)	1.54 (1.04–1.92)	5862.5	0.643
LY%, %	0.25 (0.20–0.34)	0.22 (0.13–0.29)	6785.5	0.015 *

(Continued)

**Table 1** (Continued).

	<b>DKD group (n=166)</b>	<b>T2D group (n=68)</b>	<b>T /W</b>	<b>P</b>
MO#, ×10 <sup>9</sup> /L	0.50 (0.42–0.73)	0.50 (0.40–0.67)	5887.5	0.605
MO%, %	0.07 (0.06–0.09)	0.07 (0.06–0.09)	5958	0.500
EO#, ×10 <sup>9</sup> /L	0.11 (0.05–0.14)	0.11 (0.07–0.20)	4757.5	0.058
EO%, %	0.018 (0.008–0.021)	0.018 (0.011–0.03)	4762	0.060
BA#, ×10 <sup>9</sup> /L	0.03 (0.03–0.04)	0.03 (0.02–0.04)	6151.5	0.269
BA%, %	0.01 (0.01–0.01)	0.01 (0.01–0.01)	6008	0.027 *
RBC, ×10 <sup>12</sup> /L	3.99 (2.93–4.90)	3.93 (3.15–4.50)	5921.5	0.556
HCT, L/L	0.36 (0.25–0.43)	0.35 (0.29–0.40)	5900	0.587
Hb, g/L	119.50 (86.75–146.25)	118.25 (95.25–134.00)	5965	0.495
MCV, fL	89.75 (87.10–93.10)	89.75 (86.30–92.88)	5994	0.457
MCH, pg	30.55 (29.35–32.08)	30.30 (29.20–31.10)	6371.5	0.122
MCHC, g/L	338.00 (329.75–348.00)	338.00 (328.00–345.00)	6239	0.206
RDW, %	11.75 (10.73–13.43)	11.75 (10.70–12.80)	6068	0.367
PLT, ×10 <sup>9</sup> /L	191.50 (91.25–239.50)	199.75 (159.25–243.00)	4846	0.090
PCT, %	0.21 (0.16–0.25)	0.21 (0.18–0.25)	5217	0.363
MPV, fL	10.40 (10.10–11.50)	10.40 (10.00–10.98)	6177	0.256
PDW, %	11.75 (10.73–13.43)	11.75 (10.70–12.80)	6068	0.367

Notes: \*p<0.05, \*\*p<0.01.

Japan). All laboratory data have passed external quality evaluation and certification by the Jilin Clinical Trial Center. These instruments underwent strict quality control inspection before use. Demographic data of these patients, including age and sex, were extracted from their medical records. The requirement for written informed consent was waived by the ethics committee because the data used in the research were anonymous.

## Data Cleaning and Normalization

To improve the data quality and ensure data accuracy, consistency, and availability, we performed data cleaning and standardization. The patient's ID number was used as the unique identifier.<sup>1</sup>

Data elimination and reporting: we removed data points with a missed detection rate exceeding 30%. Abnormal values, defined as those greater than three times the interquartile range, were also removed. For variables with missing data, we utilize appropriate filling methods such as median substitution, mean substitution, or complex substitution based on the data distribution characteristics.<sup>2</sup>

eGFR reporting: we assessed the renal function periodically using the CKD-EPI method.<sup>11</sup>

## Statistical Analysis

The frequency difference of categorical variables was estimated using either a block test or Fisher's precision test. For the comparison between continuous variables with normal distribution, we used *t*-test and expressed them as the mean ± standard deviation. Variables that did not follow a normal distribution were compared using the Mann–Whitney *U*-test or Kruskal–Wallis test, with results expressed as the median and interquartile range (IQR). Categorical variables were presented as frequency with proportion and analyzed using the chi-square test. To explore correlations among the parameters, Spearman correlation analysis was conducted.

Feature selection includes two steps as follows. First, we calculated the Pearson correlation coefficient (PCC) among all features. If the PCC value between a pair of features exceeded 0.8, one of the features was randomly removed. The remaining features underwent an F-test, and 50% of them were further selected to establish the model.

## Model Development

The Deepwise & Beckman Coulter DxAI platform was used for feature selection and model building. Five commonly used ML algorithms, namely Extreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Logistic Regression (LR), were used to select key features and construct risk prediction models.

A 5-fold cross-validation approach was applied, where the entire dataset was divided into five subsets. In each iteration, four of these subsets were used as a training set to train the model, while the remaining one served as a validation set to evaluate the model. This process was repeated five times, and the results were averaged.

The maximum likelihood model was constructed using the training set, and the diagnostic performance of the optimal likelihood model was evaluated. A comparison was made between the best-performing ML model and each major risk factor (Figure 1).

## Assessment of Model Performance

To assess the performance of the diagnostic model, receiver operating characteristic (ROC) curves were employed. We calculated the 95% CI and the area under the curve (AUC). A model with an AUC greater than 0.75 was considered to demonstrate good performance. A *P*-value of less than 0.05 was considered statistically significant. The calibration curve was used to evaluate the goodness-of-fit of the model. A *P*-value exceeding 0.05 was considered indicative of a well-fitting model. Moreover, decision curve analysis (DCA) was used to evaluate the clinical efficacy of the model. The data analysis software used in this study includes SPSS 24.0 (IBM, Armonk, NY), Stata 15.0 (Stata Corp LLC, Texas, USA), and GraphPad Prism 8.0 (GraphPad Software Corp, San Diego, USA).

## Results

### Baseline Characteristics of the Participants

In total, data from 166 T2D patients with DKD and 68 T2D patients without DKD were collected. All routine laboratory indices, age, and sex of the patients were selected as modeling characteristics. The Kolmogorov–Smirnov Normality Test results showed that the data for laboratory indicators were not normally distributed (except for  $P_{Age} = 0.168$ ,  $P_{Zn} = 0.582$ ,  $P_{UIBC} = 0.062$ , all other *P*-values were  $<0.01$ ). Some routine laboratory indicators had missing data, and median interpolation was used to address it. The baseline characteristics of the study participants are presented in Table 1.

### Construction of ML Model and Evaluation

The Pearson correlation coefficient (PCC) value for certain feature pairs exceeding 0.8, and as a result, the F-test was applied, leading to the removal of 50% of those characteristics. These features included DBil, IBil, BUN, Cystatin C, Cre, and LY%.

In the training set, the result of the internal 5-fold cross-validation for the five algorithms indicated that the RF algorithm method showed superior performance during internal validation. The RF algorithm identified five major factors: eGFR, GA, Uric acid, HbA1c, and Zn. The model's performance was assessed using various metrics, including AUC, negative predictive value, positive predictive value, accuracy (ACC), recall, sensitivity (SEN), specificity (SPE), and precision (Table 2) When compared to all evaluation techniques, the RF algorithm outperformed others by achieving a higher AUC.

**Table 2** 5-Fold Cross-Validation Effectiveness of Five ML Algorithms

	AUC (95% CI)	NPV	PPV	ACC	Recall	SEN	SPE	Precision
XGBoost	0.9548 (0.927–0.983)	0.8485	0.9286	0.906	0.9398	0.9398	0.8235	0.9286
RF	0.9600 (0.936–0.984)	0.8636	0.9345	0.9145	0.9458	0.9458	0.8382	0.9345
SVM	0.9478 (0.921–0.975)	0.8385	0.9281	0.9017	0.9337	0.9337	0.8235	0.9281
GBM	0.9569 (0.932–0.982)	0.8276	0.8864	0.8718	0.9398	0.9398	0.7059	0.8864
LR	0.9500 (0.923–0.977)	0.75	0.9667	0.8889	0.8735	0.8735	0.9265	0.9667

## Model Implementation in Clinical Settings

To understand the contribution of each predictor to the ML models, we computed the feature variables' importance in the RF algorithm model for each outcome. All risk factors (eGFR, GA, Uric acid, HbA1c, and Zn) were assigned scores based on their weightings, with the highest score being 38.01.

We encouraged doctors and professionals to test our optimal predictive ML model during the early stages of diagnostic testing for educational purposes. To facilitate this, we have developed a web-based risk calculation tool for predicting DKD. The Deepwise & Beckman Coulter DxAI platform was used for feature selection and model building. The system can be accessed by users for easy evaluation of DKD risk via the following link: DKD Risk Prediction Tool. (<https://dxonline.deepwise.com/prediction/index.html?baseUrl=%2Fapi%2F&id=34208&topicName=undefined&from=share&platformType=wisdom>)

## Internal Validation of the Model

The RF model demonstrated a high level of prediction accuracy, with the AUC value of 0.960 (95% CI: 0.936–0.984) during internal validation (Figure 2). The result of the Hosmer-Lemeshow goodness-of-fit test indicated that the *P*-value for each model in both the training and validation sets exceeded 0.05, indicating a good model fit (Figure 3). The DCA model has a significant net gain in both the training and validation sets (Figure 4)

## External Validation of the Model

To validate the ML model established in this study, data from patients at the Third Hospital of Jilin University were collected as an external test set. Among all the results, the RF algorithm demonstrated the best predictive performance and successfully passed external validation. In this analysis, the model's accuracy was assessed using the AUC (0.9316), sensitivity (0.9000) and overall accuracy (0.9000). An external dataset was used to ensure consistent analysis with the internal validation (Figure 5).

## Comparison of RF and the Five Features Screened

We compared the five features screened with the highest feature importance in the RF model and the RF algorithm method. Among these, the diagnostic effectiveness of the RF model (AUC = 0.960) was significantly higher than each of the five features. The performance of the model was evaluated by AUC, Cut-off, sensitivity, specificity, Youden's index (YI), standard error and 95% CI (Table 3).

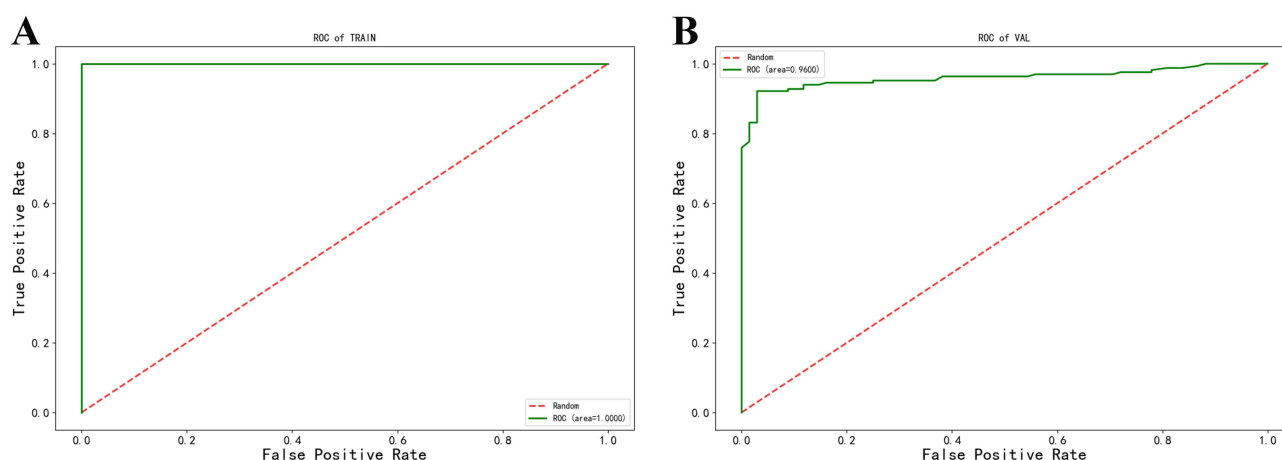


Figure 2 Receiver operating characteristic curve. (A) Training set. (B) Validation set.

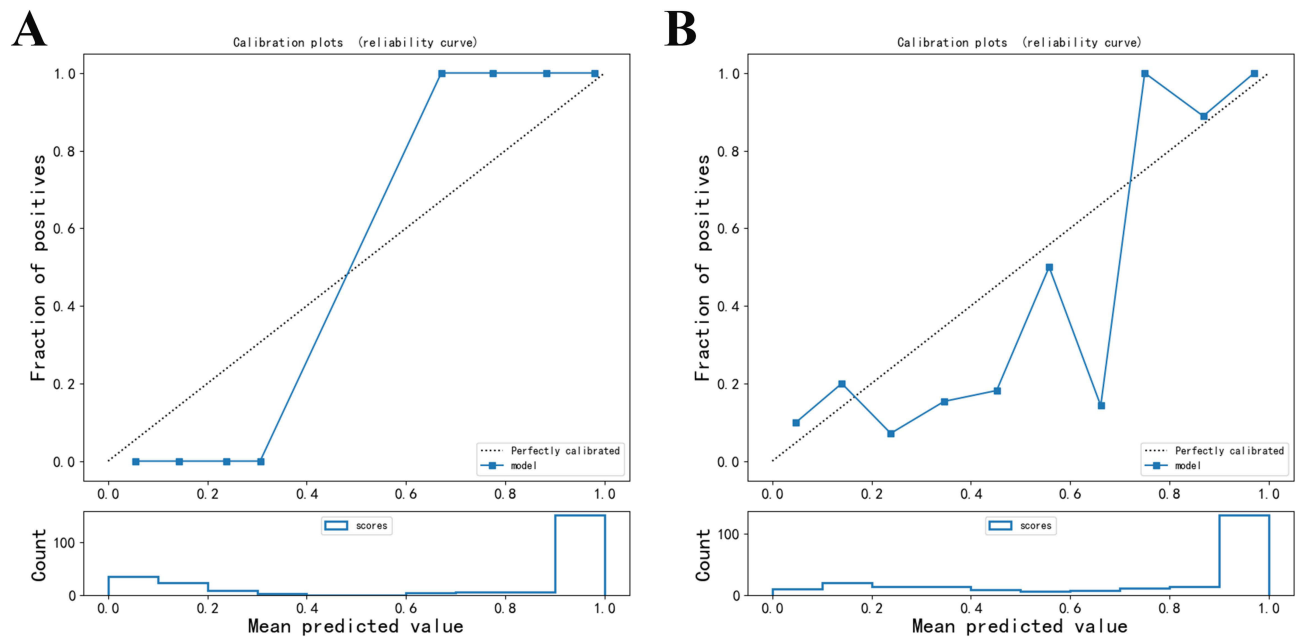


Figure 3 Calibration curve. (A) Training set. (B) Validation set.

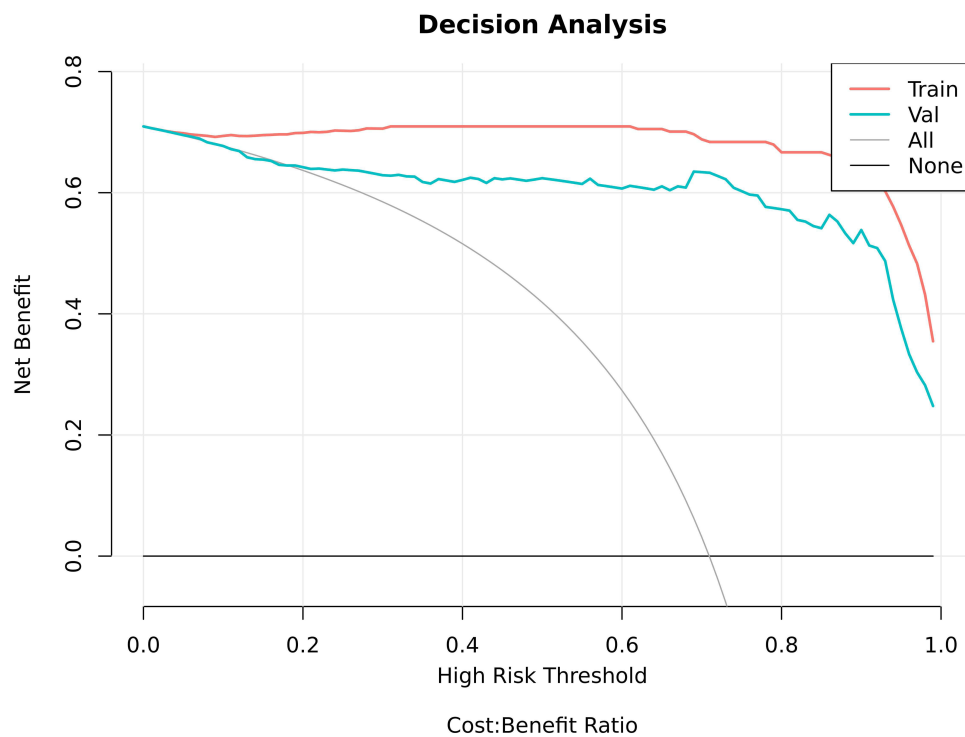


Figure 4 Decision curve analysis in the training set and validation set.

## Discussion

Despite the significance of DKD, few studies have established a differential diagnosis model for DKD, particularly when combined with laboratory parameters. The importance of developing a simple and practical diagnostic model lies in its potential to enable accurate and rapid clinical diagnosis of DKD, rather than blindly adding more indicators. Our study aimed to construct a risk prediction model for patients with DKD using artificial intelligence (AI), a method whose performance has been extensively tested and validated.

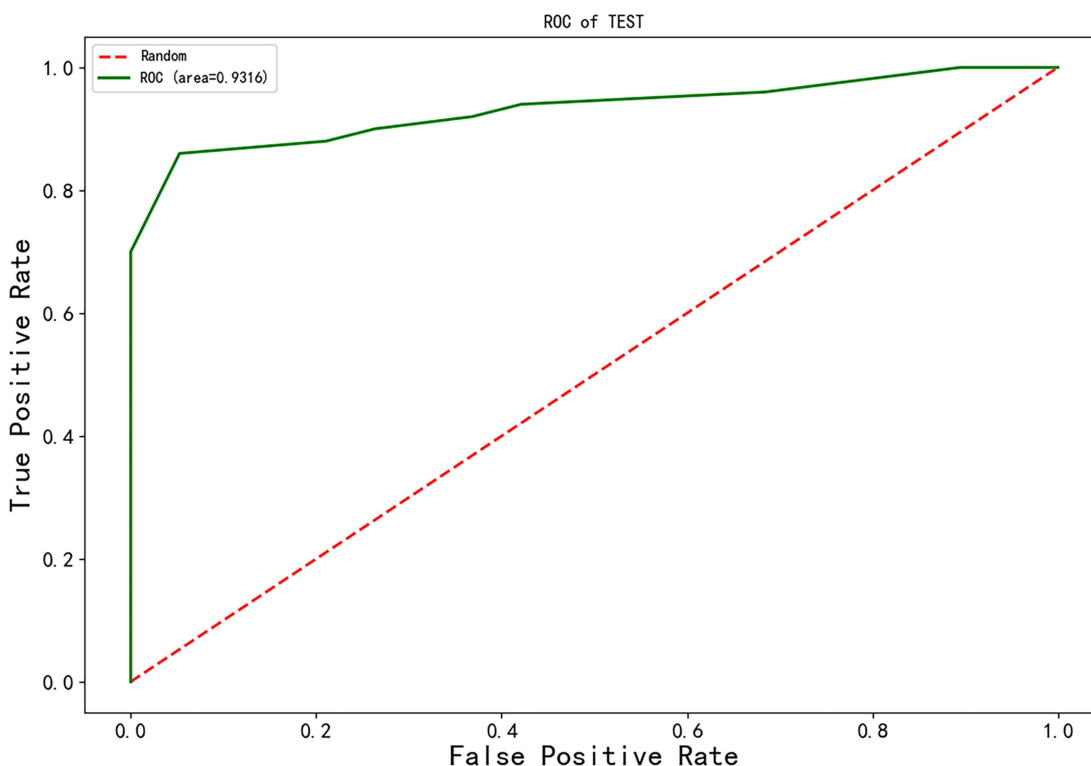


Figure 5 Receiver operating characteristic curve in the external validation set.

Furthermore, this model could play a crucial role in the prevention and management of T2D, especially in individuals with diabetes at an increased risk of developing DKD. To be more specific, if high-risk T2D patients exhibit abnormalities in the parameters outlined in the diagnostic model during routine physical examinations, this could serve as an early alert to the possibility of progressing to DKD, thus enabling timely intervention.

The results of the ML model can be influenced by the characteristics of the participants. Due to the imbalance between the groups studied, there may be qualitative differences, such as gender, and age, between the two groups. Significant correlations between the factors under investigation may lead to deviations in the results. In this study, the correlation of indicators was analyzed using Spearman correlation. Highly correlated metrics were excluded to ensure that other causes or confounding factors did not affect the study results.

The efficient predictive model established in this study offers several advantages. First, the simple model has been uploaded to a website, making it easy to understand and use. The system developed as part of the research is based on an RF model on a server, allowing for high-speed analysis of patient data. Users can easily access this data through the Internet using PCs and smartphones.

Table 3 Comparison of RF and the Five Features Screened in the Validation Set

	AUC	Cut-off	SEN	SPE	YI	SE	95% CI	P
RF	0.960	0.705	0.922	0.971	0.892	0.012	0.936–0.984	0.000 **
Egfr	0.827	90.500	0.747	0.853	0.600	0.030	0.768–0.885	0.000 **
GA	0.817	23.890	0.789	0.735	0.524	0.030	0.758–0.875	0.000 **
HbA1c	0.693	7.550	0.735	0.574	0.308	0.038	0.618–0.767	0.000 **
Uric acid	0.740	327.000	0.729	0.662	0.391	0.037	0.667–0.813	0.000 **
Zn	0.590	11.245	0.524	0.676	0.201	0.039	0.513–0.667	0.023 *

Notes: \* $p < 0.05$ , \*\*  $p < 0.01$ .

Second, the model comprehensively assesses representative laboratory test indicators to distinguish between DKD and T2D, including hematological and biochemical parameters, with a special focus on the novel indicator GA to improve diagnostic accuracy. To reduce the workload of measuring and inputting patient data, the number of variables in our system model has been limited. AI and communication technology have been used to implement this system.

Nevertheless, the performance of the diagnosis model was excellent, highlighting its potential utility as an adjunct test for a wide range of DKD patients. In this study, the diagnostic model, which combines eGFR, GA, Uric acid, HbA1c, and Zn to assess kidney function damage in T2D, was established using ML algorithms. A progressive decline in eGFR is a clear indication of DKD and is part of the diagnostic criteria.<sup>12</sup> Poor and continuous control of blood glucose is one of the most significant risk factors for DKD. GA is the final product of non-enzymatic glycosylation resulting from the reaction between serum albumin and glucose. GA is associated with blood glucose concentration and albumin half-life, reflecting the average blood glucose status of patients in the first 2–3 weeks.<sup>13</sup> The GA concentration can be determined based on the quantitative measurement of glycosylated serum albumin and the percentage of serum albumin. HbA1c is currently recognized as a key marker for the development of long-term diabetic complications in patients with diabetes. Additionally, HbA1c is an important index for preventing the occurrence and progression of diabetic complications by achieving good blood glucose control.<sup>14</sup> Previous studies have indicated that the concentration of HbA1c, unrelated to blood sugar, may be affected by various conditions, such as iron deficiency, red blood cell lifespan, hemolytic anemia, renal anemia, hemoglobin mutations, chronic kidney disease, and more.<sup>15–17</sup> GA detects glycemic fluctuations more effectively than HbA1c in the context of DKD.

Moreover, the measurement of GA serves as an additional tool for the clinical diagnosis of DKD, facilitating an index for evaluating the severity of kidney function disorder to guide treatment.<sup>17–19</sup> GA contributes to the progression of DKD by promoting the production of acidic molecules by membrane cells and epithelial cells after being absorbed by cells. GA also accelerates insulin resistance by increasing the production of intracellular reactive oxygen species, which inhibits glucose uptake by muscle and fat cells.<sup>20</sup> Moreover, we performed the study to obtain data from participants to establish a novel model, that highlights the possibility for clinical assessment of blood glucose control and the identification of kidney function damage in the diagnosis and screening of DKD.

Uric acid is the end product of purine base metabolism, generated in the exchange process between nucleotides and adenosine triphosphate (ATP). Increasing epidemiological evidence indicates that hyperuricemia, acting as an oxidant, serves as an independent risk factor for DKD. It is closely associated with the onset and progression of DKD. Small proof-of-concept clinical trials have suggested that reducing uric acid levels with allopurinol may lead to a decrease in the rate of eGFR decline, though whether it benefits DKD progression remains controversial.<sup>21–26</sup>

The potential mechanisms behind uric acid-induced kidney damage may involve mediating insulin resistance, inactivation through oxidative reactions that lower nitric oxide (NO) levels, pancreatic  $\beta$ -cell dysfunction, oxidative stress, stimulation of the renin-angiotensin system activation, the inflammatory response, and more.<sup>27–30</sup>

Patients in this study received treatment with hypoglycemic, and anti-inflammatory drugs, as well as traditional Chinese medicine and other medications. The results do not entirely rule out the potential effects of some of these drugs on renal function. However, the predictive model was constructed using data from the initial laboratory study at admission, a point at which most patients had not yet started taking medication. Therefore, the impact of drugs on renal function can be effectively disregarded.

The model's performance was evaluated through ROC curve analysis and DCA, focusing on discrimination, calibration, and clinical application value. The model demonstrated good results across all parameters. Therefore, we believe that this model can be used in routine clinical practice to assist doctors in identifying renal dysfunction in T2D patients. The significance of zinc in preventing and slowing the progression of DKD has been highly regarded in experimental studies. This underscores the importance of paying attention to this trace element and understanding how it can protect against kidney damage caused by diabetes.

Zinc has a negative correlation with the risk of DKD, and is considered a vital factor in reducing the incidence of diabetic kidney damage. This can be attributed to its ability to alleviate oxidative stress and inflammation.<sup>31,32</sup>

Studies have confirmed that Zn plays a protective role in DKD by mediating inflammation, regulating insulin receptors, addressing glycoxidative stress, countering pancreatic  $\beta$ -cell impairment, promoting healthy lipid profiles, and activating important proteins related to cell signaling, which contribute to maintaining glucose homeostasis.<sup>33–35</sup> We observed a notable reduction in serum Zn concentration in DKD compared to T2D subjects ( $p = 0.056$ ), which is

consistent with previous studies.<sup>36,37</sup> Zn appears to have a potential role in assessing glycemic control and managing renal impairment in the course of diabetes. Zn antagonizes diabetes-related risk factors and complications, such as insulin accumulation in the pancreas, improving blood glucose control, and stabilizing insulin hexamers.<sup>38</sup>

The study showed that abnormal Zn metabolism may be associated with diabetes complications.<sup>39</sup> Achieving and maintaining appropriate serum Zn levels may be a valuable goal for alleviating symptoms related to T2D complications and slowing the progression of renal damage in patients at high risk of end-stage renal disease. In summary, analyzing the beneficial effects of Zn in our study will contribute to a better understanding of potential treatment strategies for diabetes and DKD.

According to the ROC analysis, the AUC for the training set was 0.960 (95% CI: 0.936–0.984), and for the validation set, it was 0.9316 (95% CI: 0.8747–0.9885), respectively. When comparing the model's results with the five features screened with the highest feature importance in the RF model, it shows good results across all parameters. Therefore, this model can be used in daily clinical practice to assist doctors judge renal function damage in DKD patients.

This study still has some limitations. First of all, the statistical significance of the data may be affected by the relatively small sample size of the patients included in this study. Second, the data used in this study was derived from the initial laboratory examination conducted after each patient's admission to the hospital, rather than from laboratory results at the onset of symptoms. Obtaining data at the onset of symptoms can be challenging unless patients seek immediate medical attention after symptom onset. Finally, although this study suggests that the diagnostic model could serve as a potential indicator for predicting and preventing the progression of DKD, it remains unclear what the appropriate cutoff values are and the prognostic value of the indicators included in this model. Further longitudinal studies may provide additional clarification in the future.

## Conclusions

In conclusion, the online DKD risk prediction model constructed using the RF algorithm was selected based on its strong performance in the internal validation. In contrast to previous risk prediction models, our model emphasized the significance of the novel indicator GA. The RF algorithm exhibited a high level of accuracy in predicting the progression of kidney disease, surpassing the predictive ability of the major risk factors associated with the occurrence of DKD. Therefore, it may serve as a valuable alternative or complementary tool for enhancing the diagnosis of DKD. To further validate the clinical applicability of our model, large-scale studies involving diverse patient cohorts across multiple sites should be conducted in the future.

## Funding

This study was funded by Establishment and Validation of a Diagnostic Model to Identify Diabetic Kidney Disease in Patients with Type 2 diabetes mellitus (2022LC106).

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Laakso M. Biomarkers for type 2 diabetes. *Mol Metabol.* 2019;27s:S139–s46. doi:10.1016/j.molmet.2019.06.016
2. Kim KS, Lee JS, Park JH, et al. Identification of novel biomarker for early detection of diabetic nephropathy. *Biomedicines.* 2021;9(5).
3. Pugliese G, Penno G, Natali A, et al. Diabetic kidney disease: new clinical and therapeutic issues. Joint position statement of the Italian diabetes society and the Italian society of nephrology on “The natural history of diabetic kidney disease and treatment of hyperglycemia in patients with type 2 diabetes and impaired renal function”. *J Nephrol.* 2020;33(1):9–35. doi:10.1007/s40620-019-00650-x
4. Barutta F, Bellini S, Canepa S, Durazzo M, Gruden G. Novel biomarkers of diabetic kidney disease: current status and potential clinical application. *Acta diabetologica.* 2021;58(7):819–830. doi:10.1007/s00592-020-01656-9
5. 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
6. Wang JS, Yen FS, Lin KD, Shin SJ, Hsu YH, Hsu CC. Epidemiological characteristics of diabetic kidney disease in Taiwan. *J Diabetes Invest.* 2021;12(12):2112–2123. doi:10.1111/jdi.13668
7. Jiang W, Wang J, Shen X, et al. Establishment and validation of a risk prediction model for early diabetic kidney disease based on a systematic review and meta-analysis of 20 cohorts. *Diabetes Care.* 2020;43(4):925–933. doi:10.2337/dc19-1897
8. 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. *Renal Failure.* 2022;44(1):562–570. doi:10.1080/0886022X.2022.2056053
9. Care, Diabetes. Microvascular Complications and Foot Care: standards of Medical Care in Diabetes-2019. *Diabetes Care.* 2019;42(Suppl 1).
10. Tervaert TW, Mooyaart AL, Amann K, et al. Pathologic classification of diabetic nephropathy. *JASN.* 2010;21(4):556–563. doi:10.1681/ASN.2010010010

11. Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. *Official j Nation Kid Found.* **2022**;79(2):268–88.e1. doi:10.1053/j.ajkd.2021.08.003
12. Wang K, Xu W, Zha B, Shi J, Wu G, Ding H. Fibrinogen to albumin ratio as an independent risk factor for type 2 diabetic kidney disease. *Targets Therap.* **2021**;14:4557–4567. doi:10.2147/DMSO.S337986
13. Copur S, Siritopol D, Afsar B, et al. Serum glycated albumin predicts all-cause mortality in dialysis patients with diabetes mellitus: meta-analysis and systematic review of a predictive biomarker. *Acta diabetologica.* **2021**;58(1):81–91. doi:10.1007/s00592-020-01581-x
14. Wang M, Hng TM. HbA1c: More than just a number. *Australian j Gene Pract.* **2021**;50(9):628–632. doi:10.31128/AJGP-03-21-5866
15. Altuntaş S Ç, Evran M, Gürkan E, Sert M, Tetiker T. HbA1c level decreases in iron deficiency anemia. *Wiener klinische Wochenschrift.* **2021**;133(3–4):102–106. doi:10.1007/s00508-020-01661-6
16. American Diabetes Association. Classification and Diagnosis of Diabetes: standards of Medical Care in Diabetes-2021. *Diabetes Care.* **2021**;44(Suppl 1).
17. Ohigashi M, Osugi K, Kusunoki Y, et al. Association of time in range with hemoglobin A1c, glycated albumin and 1,5-anhydro-d-glucitol. *J Diabetes Invest.* **2021**;12(6):940–949. doi:10.1111/jdi.13437
18. Fang M, Daya N, Coresh J, Christenson RH, Selvin E. Glycated albumin for the diagnosis of diabetes in US adults. *Clin Chem.* **2022**;68(3):413–421. doi:10.1093/clinchem/hvab231
19. Gan T, Liao B, Xu G. The clinical usefulness of glycated albumin in patients with diabetes and chronic kidney disease: progress and challenges. *J diabet complicat.* **2018**;32(9):876–884. doi:10.1016/j.jdiacomp.2018.07.004
20. Zendjabil M. Glycated albumin. *clinica chimica acta. Int J Clin Chem.* **2020**;502:240–244. doi:10.1016/j.cca.2019.11.007
21. Doria A, Galecki AT, Spino C, et al. Serum urate lowering with allopurinol and kidney function in type 1 diabetes. *New Engl J Med.* **2020**;382(26):2493–2503. doi:10.1056/NEJMoa1916624
22. Badve SV, Pascoe EM, Tiku A, et al. Effects of allopurinol on the progression of chronic kidney disease. *New Engl J Med.* **2020**;382(26):2504–2513. doi:10.1056/NEJMoa1915833
23. Jordan DM, Choi HK, Verbanck M, et al. No causal effects of serum urate levels on the risk of chronic kidney disease: a Mendelian randomization study. *PLoS Med.* **2019**;16(1):e1002725. doi:10.1371/journal.pmed.1002725
24. Zou Y, Zhao L, Zhang J, et al. Association between serum uric acid and renal outcome in patients with biopsy-confirmed diabetic nephropathy. *Endocr Connections.* **2021**;10(10):1299–1306. doi:10.1530/EC-21-0307
25. Mauer M, Doria A. Uric acid and diabetic nephropathy risk. *Contrib Nephrol.* **2018**;192:103–109.
26. Mauer M, Doria A. Uric acid and risk of diabetic kidney disease. *J Nephrol.* **2020**;33(5):995–999. doi:10.1007/s40620-020-00796-z
27. Volpe A, Ye C, Hanley AJ, Connelly PW, Zinman B, Retnakaran R. Changes over time in uric acid in relation to changes in insulin sensitivity, beta-cell function, and glycemia. *J Clin Endocrinol Metab.* **2020**;105(3):e651–9. doi:10.1210/clinem/dgz199
28. King C, Lanaspas MA, Jensen T, Tolan DR, Sánchez-Lozada LG, Johnson RJ. Uric acid as a cause of the metabolic syndrome. *Contrib Nephrol.* **2018**;192:88–102.
29. Lu J, He Y, Cui L, et al. Hyperuricemia predisposes to the onset of diabetes via promoting pancreatic  $\beta$ -cell death in uricase-deficient male mice. *Diabetes.* **2020**;69(6):1149–1163. doi:10.2337/db19-0704
30. Yang X, Gu J, Lv H, et al. Uric acid induced inflammatory responses in endothelial cells via up-regulating(pro)renin receptor. *Biomed Pharmacoth.* **2019**;109:1163–1170. doi:10.1016/j.biopha.2018.10.129
31. Shen Y, Yin Z, Lv Y, et al. Plasma element levels and risk of chronic kidney disease in elderly populations ( $\geq 90$  Years old). *Chemosphere.* **2020**;254:126809. doi:10.1016/j.chemosphere.2020.126809
32. Tokuyama A, Kanda E, Itano S, et al. Effect of zinc deficiency on chronic kidney disease progression and effect modification by hypoalbuminemia. *PLoS One.* **2021**;16(5):e0251554. doi:10.1371/journal.pone.0251554
33. Grădinaru D, Margină D, Ungurianu A, et al. Zinc status, insulin resistance and glycoxidative stress in elderly subjects with type 2 diabetes mellitus. *Exp Ther Med.* **2021**;22(6):1393. doi:10.3892/etm.2021.10829
34. Prasad AS, Bao B. Molecular mechanisms of zinc as a pro-antioxidant mediator: Clinical Therapeutic Implications. *Antioxidants.* **2019**;8(6):164. doi:10.3390/antiox8060164
35. Gerardo Yanowsky-Escatell F, Andrade-Sierra J, Pazarín-Villaseñor L, et al. The role of dietary antioxidants on oxidative stress in diabetic nephropathy. *Iranian j Kidney Dis.* **2020**;14(2):81–94.
36. Tavares A, Mafta D, Leal VO, et al. Zinc plasma status and sensory perception in nondialysis chronic kidney disease patients. *J Ren Nutr.* **2021**;31(3):257–262. doi:10.1053/j.jrn.2020.05.012
37. Damianaki K, Lourenco JM, Braconnier P, et al. Renal handling of zinc in chronic kidney disease patients and the role of circulating zinc levels in renal function decline. *Nephrol Dial Transplant.* **2020**;35(7):1163–1170. doi:10.1093/ndt/gfz065
38. Gembillo G, Visconti L, Giuffrida AE, et al. Role of zinc in diabetic kidney disease. *Nutrients.* **2022**;14(7):1353. doi:10.3390/nu14071353
39. Björklund G, Dadar M, Pivina L, Doşa MD, Semenova Y, Aaseth J. The role of zinc and copper in insulin resistance and diabetes mellitus. *Curr Med Chem.* **2020**;27(39):6643–6657. doi:10.2174/0929867326666190902122155

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. 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.