

# Serum Lipoprotein(a) and High-Density Lipoprotein Cholesterol Associate with Diabetic Nephropathy: Evidence from Machine Learning Perspectives

Rui-Huan Gao<sup>1,\*</sup>, Boyang Liu<sup>2,\*</sup>, Ying Yang<sup>1</sup>, Ruoxi Ran<sup>1</sup>, Yidan Zhou<sup>1,\*</sup>, Song-Mei Liu<sup>1,\*</sup>

<sup>1</sup>Department of Clinical Laboratory, Zhongnan Hospital of Wuhan University, Wuhan, Hubei Province, 430071, People's Republic of China;

<sup>2</sup>Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA

\*These authors contributed equally to this work

Correspondence: Song-Mei Liu; Yidan Zhou, Zhongnan Hospital of Wuhan University, 169 Donghu Road, Wuhan, Hubei Province, 430071, People's Republic of China, Email [smliu@whu.edu.cn](mailto:smliu@whu.edu.cn); [Yzhou12@whu.edu.cn](mailto:Yzhou12@whu.edu.cn)

**Purpose:** Diabetic nephropathy (DN) is a common complication of type 2 diabetes mellitus (T2DM) that significantly impacts the quality of life for affected patients. Dyslipidemia is a known risk factor for developing cardiovascular complications in T2DM patients. However, the association between serum lipoprotein(a) (Lp(a)) and high-density lipoprotein cholesterol (HDL-C) with DN requires further investigation.

**Patients and Methods:** For this cross-sectional study, we randomly selected T2DM patients with nephropathy (DN,  $n = 211$ ) and T2DM patients without nephropathy (T2DM,  $n = 217$ ) from a cohort of 142,611 patients based on predefined inclusion and exclusion criteria. We collected clinical data from the patients to identify potential risk factors for DN using binary logistic regression and machine learning. After obtaining the feature importance score of clinical indicators by building a random forest classifier, we examined the correlations between Lp(a), HDL-C and the top 10 indicators. Finally, we trained decision tree models with top 10 features using training data and evaluated their performance with independent testing data.

**Results:** Compared to the T2DM group, the DN group had significantly higher serum levels of Lp(a) ( $p < 0.001$ ) and lower levels of HDL-C ( $p = 0.028$ ). Lp(a) was identified as a risk factor for DN, while HDL-C was found to be protective. We identified the top 10 indicators that were associated with Lp(a) and/or HDL-C, including urinary albumin (uALB), uALB to creatinine ratio (uACR), cystatin C, creatinine, urinary  $\alpha$ 1-microglobulin, estimated glomerular filtration rate (eGFR), urinary  $\beta$ 2-microglobulin, urea nitrogen, superoxide dismutase and fibrinogen. The decision tree models trained using the top 10 features and with uALB at a cut-off value of 31.1 mg/L showed an average area under the receiver operating characteristic curve (AUC) of 0.874, with an AUC range of 0.870 to 0.890.

**Conclusion:** Our findings indicate that serum Lp(a) and HDL-C are associated with DN and we have provided a decision tree model with uALB as a predictor for DN.

**Keywords:** type 2 diabetes mellitus, diabetic nephropathy, high density lipoprotein cholesterol, lipoprotein(a), machine learning

## Introduction

Diabetic nephropathy (DN) is the most common complication of diabetes, characterized by renal microangiopathy, and it occurs in approximately 40% of type 2 diabetes mellitus (T2DM) patients.<sup>1</sup> Currently, DN has become the major contributor of chronic renal disease (CKD), renal failure<sup>2</sup> and end-stage renal disease.<sup>3</sup> DN could also lead to an increase of cardiovascular mortality.<sup>4</sup> The main manifestations of DN are the changes in renal morphology and function, such as decreased estimated glomerular filtration rate (eGFR) and renal hypertrophy at an early stage, proteinuria, decreased renal function, and even glomerulosclerosis at late stage.<sup>5</sup> The American Diabetes Association recommends using eGFR as an indicator for the progression of CKD and divides CKD into five stages (eGFR <15, 15–30, 30–60, 60–90, >90 mL/

min/1.73m<sup>2</sup>).<sup>6</sup> Importantly, the presence of proteinuria, reduced eGFR and renal pathological damage can be used for clinical diagnosis of DN.<sup>6</sup> Determining urinary albumin (uALB) is useful for identifying diabetic patients at risk for developing DN.<sup>7</sup>

Glucose metabolic disorders can cause renal microangiopathy, leading to changes in glomerular filtration rate and renal blood flow.<sup>8</sup> Dyslipidemia promotes the subcutaneous deposition of lipids in blood vessels, leading to atherosclerosis and cardio-cerebrovascular diseases (ASCVD).<sup>9</sup> Long-term hyperglycemia and lipid disorders in T2DM could lead to ASCVD.<sup>10,11</sup> Patients with concomitant DM and ASCVD tend to have a higher risk of future cardiovascular events.<sup>12</sup> Particularly, lipoprotein(a) (Lp(a)) is an independent risk factor for ASCVD.<sup>13</sup> Lowering LDL cholesterol (<100 mg/dl) is an effective strategy for preventing the development of microalbuminuria and delaying the progression of nephropathy.<sup>14</sup> A prospective study in Korean population demonstrated that Lp(a) is an independent risk factor for the progression of DN in T2DM patients with overt proteinuria.<sup>15</sup> In addition, in patients with type 1 diabetes, abnormal lipids are closely related to the occurrence of proteinuria, affecting the occurrence of macrovascular and microvascular complications.<sup>16</sup> These studies suggest that abnormal blood lipids are involved in the occurrence and development of DN. Lowering Lp(a) and increasing HDL-C therapy could prevent T2DM patients from developing DN.

Machine learning has been used for disease diagnosis/prediction, risk assessment and gene diagnosis. Olufemi Aromolaran et al applied machine learning for gene importance prediction.<sup>17</sup> As a common classifier in machine learning, random forest has a good classification performance, and its built-in feature selection method could be used for molecular marker prediction.<sup>18</sup> Early diagnosis of DN will be beneficial to the improvement of the life quality of T2DM patients. In this study, we aimed to obtain the feature importance score of clinical data from T2DM patients with or without nephropathy by building a random forest classifier, evaluate the correlations of the top 10 indicators with Lp(a) and HDL-C, and train a decision tree model for DN prediction.

## Materials and Methods

### Study Population

This cross-sectional study received approval from the Ethics Committee of Zhongnan Hospital, and it adhered to the provisions set forth in the Declaration of Helsinki. Informed consent was waived due to the following reasons: (1) participants' information was de-identified to safeguard their privacy; (2) no direct contact was made with the participants to collect samples, therefore no additional risk was posed to them; (3) all participants were discharged, and the data were not required to be shared with them; and (4) all findings were intended for scientific exploration only and not for diagnosis or commercial purposes. Initially, we obtained data from 142,611 patients with available blood lipids or urinary microalbumin from Zhongnan Hospital of Wuhan University between August 2016 and November 2022. The inclusion criteria were age >18 years and referral for T2DM patients with or without nephropathy. The clinical diagnosis of the enrolled patients was based on the Standards of Medical Care in Diabetes-2020.<sup>6,19</sup> The exclusion criteria included type 1 diabetes, gestational diabetes, other types of diabetes; acute complications of diabetes (such as diabetic ketoacidosis, etc.); other types of nephropathy, malignant tumors, and patients with missing clinical data.

### Data Collection and Description

For this cross-sectional study, we randomly selected 211 cases of T2DM patients with nephropathy (DN group) and 217 cases of T2DM patients without nephropathy (T2DM group) from a cohort of 142,611 patients. We ensured gender and age matching between the two groups, and the process of selecting the study population is outlined in [Supplementary Figure S1](#). We collected clinical information of 428 patients from their medical records using a standard protocol ([Supplementary Table S1](#)). The data were divided into training data (DN = 161, T2DM = 167; 57 features) and testing data (DN = 50, T2DM = 50; 57 features). The rate of missing values was 10.19%, with most of the missing data concentrated in height, weight, BMI, serum amyloid protein, urinary creatinine, urinary  $\beta$ 2-microglobulin (ug/L) and urinary  $\alpha$ 1-microglobulin (mg/L), which accounted for about 62% of the total missing data. The results of this study were based on training data, while the test data were only used for model testing.

## Calculations of uACR and eGFR

The equations for calculating uACR:

$$uACR = \frac{uALB}{uCREA} \times 8840$$

uALB is urinary microalbumin (mg/L). uCREA is urinary creatinine ( $\mu\text{mol/L}$ ).

eGFR were calculated from the 2012 Collaborative Equation for the Epidemiology of Chronic Kidney Disease (CKD-EPICYS):

$$eGFR = 133 \times \min\left(\frac{Scys}{\kappa}, 1\right)^{-0.499} \times \max\left(\frac{Scys}{\kappa}, 1\right)^{-1.328} \times 0.966^{Age} \times 0.932(\text{if female})$$

Scys is serum cystatin C (mg/dL),  $\kappa = 0.8$ , “min” indicates the minimum of Scys/ $\kappa$  or 1, and “max” indicates the maximum of Scys/ $\kappa$  or 1.<sup>20</sup>

## Data Preprocessing

As there are missing values present in the dataset, we employ the K-nearest neighbor<sup>21</sup> imputation method to fill in any gaps. This method utilizes the mean value of the top-k nearest neighbors to impute the missing value, ensuring that there are no missing values remaining in the dataset. Once the imputation process is complete, we divide the data into training and testing sets. We treat the value of K in K-nearest neighbor as a hyperparameter and conduct a hyperparameter search to identify the optimal value for our model. Our experiment finds that setting K = 5 yields the most accurate model predictions on the validation data. Thus, we will use K = 5 as our choice for K.

## Building Predictive Decision Tree Models

In this study, our pipeline of building models was as follows:

1. Extract the feature set X and corresponding label set Y from the training data, and construct a random forest classifier<sup>22</sup> using the default parameters. Subsequently, calculate the feature importance scores.
2. Sort the feature importance scores in descending order and choose the top 10 features based on their importance scores.
3. Train the decision tree model using the top 10 features selected by random forest for better interpretability. We employed 5-fold cross-validation<sup>23</sup> to partition the training data, with 80% of the data serving as the training set to build the tree model and 20% as the validation set to test the tree model performance. Successive halving search<sup>24,25</sup> was used to tune the hyperparameters, and we selected the optimal hyperparameter based on the lowest averaged validation loss.
4. Build the tree model on the entire training set using the best hyperparameters identified in step 3.

## Evaluation Tree Model

We obtained the prediction of the tree model on the testing set, compared the predicted label and ground-truth label, and calculated the AUC score. We repeated the above process 10 times to ensure the stability of our results.

## Statistical Analysis

All statistical analyses were performed by SPSS 25.0. The continuous variables of normal distribution were expressed as mean  $\pm$  standard deviation ( $\bar{X} \pm SD$ ), and the continuous variables of skewed distribution were presented as median (25th-75th percentile). Student's *t*-test, Mann-Whitney *U*-test or Kruskal-Wallis test was used to compare clinical parameters among groups. The Spearman's rank correlation was applied for correlation evaluation.  $\chi^2$  tests were used to compare frequencies between two groups. All statistical analysis was two-sided, and p value < 0.05 was considered statistically significant.

## Results

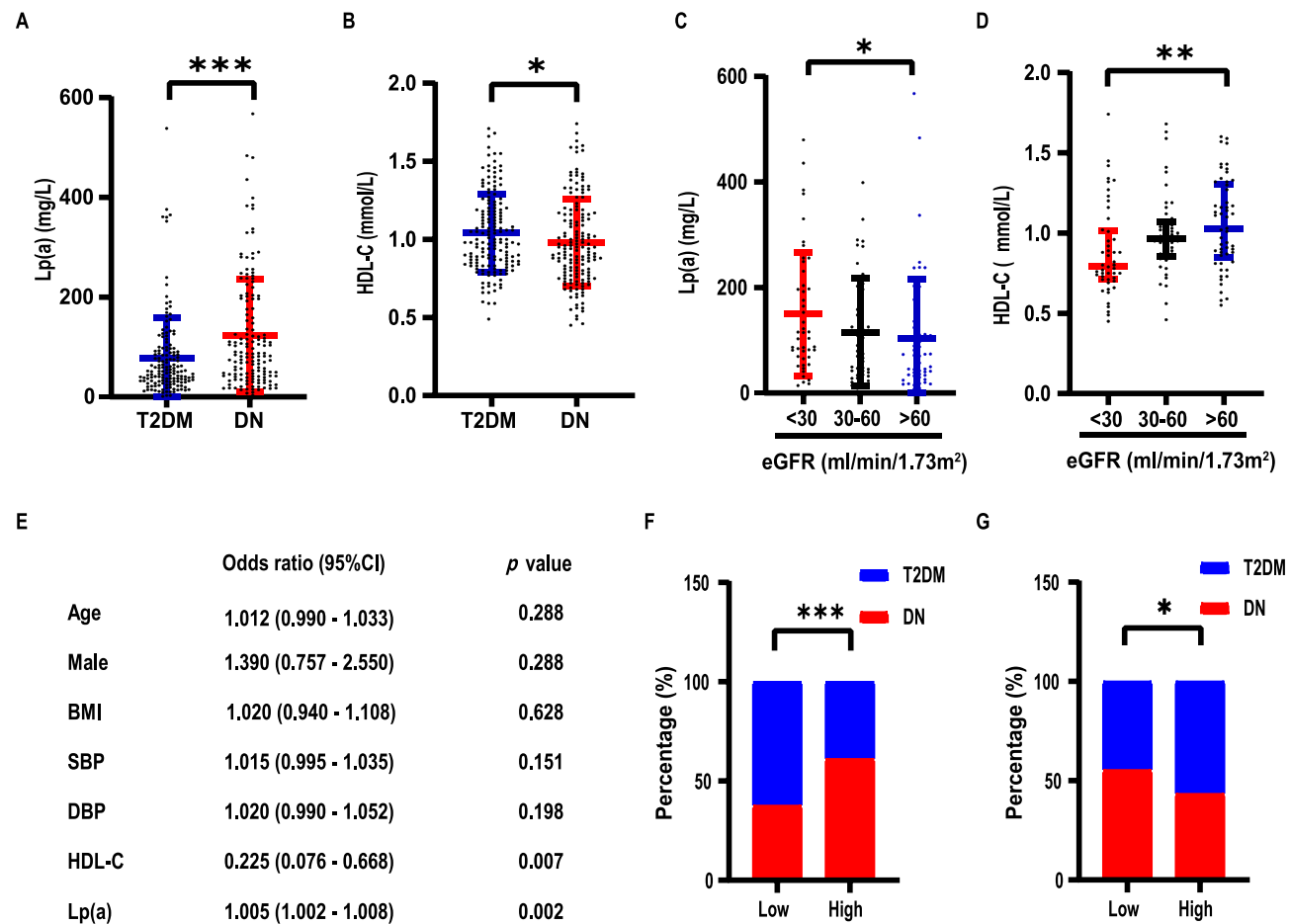
### Serum Lp(a) is Increased and HDL-C is Decreased in T2DM Patients with Nephropathy

The comparison of clinical data between T2DM group and DN group was presented in [Supplementary Table S1](#). There was no difference in the distribution of age and sex, indicating that the effects of age and sex on lipid changes could be negligible. However, compared with T2DM group, the level of Lp(a) was increased and HDL-C was decreased in DN group ([Figure 1A](#) and [B](#)). There was no difference in cholesterol, triglyceride, low-density lipoprotein cholesterol, small dense low-density lipoprotein cholesterol, free fatty acid or phospholipid levels ([Table 1](#)). Taken together, the distinct features of lipid disorders between DN group and T2DM group are the changes of Lp(a), and HDL-C levels.

To investigate the difference of Lp(a) and HDL-C across various stages of CKD, T2DM patients with nephropathy were divided into three groups based on eGFR (<30, 30–60, >60 mL/min/1.73m<sup>2</sup>), due to the small number of patients with eGFR < 15 and eGFR >90. By Kruskal–Wallis *H*-test, we found that an increase of eGFR was associated with a decline in Lp(a) levels ([Figure 1C](#)), whereas an increase of eGFR was associated with elevation of HDL-C ([Figure 1D](#)).

### Lp(a) is a Risk Factor, While HDL-C is a Protective Factor for DN

Binary logistic regression analysis (0 for T2DM group and 1 for DN group) was used to evaluate whether Lp(a) and HDL-C were the risk factors of DN. After adjusting age, sex, BMI, systolic blood pressure (SBP) and diastolic blood



**Figure 1** Increased Lp(a) and decreased HDL-C presenting a risk factor for DN. **(A)** The level of serum lipoprotein(a) in DN group was higher than that in T2DM group. **(B)** The level of serum HDL-C in DN group was lower than that in T2DM group. **(C)** The increase of eGFR was associated with the decline of Lp(a) level. **(D)** The increase of eGFR was associated with the elevation of HDL-C level. **(E)** Binary logistic regression analysis showed Lp(a) was a risk factor, while HDL-C was a protective factor for DN. **(F)**  $\chi^2$  test showed the incidence of DN in the high-Lp(a) group was higher than that in the low-Lp(a) group (61.18% vs 37.75%). **(G)**  $\chi^2$  test showed the incidence of DN in the high-HDL-C group was lower than that in the low-HDL-C group (43.48% vs 55.48%). \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001.

**Table 1** Serum Lipids Comparison of Type 2 Diabetes Mellitus Patients with or without Nephropathy

Parameter	T2DM Group (n=167)	DN Group (n=161)	p value
Age (years)	60 ± 1	61 ± 1	0.784
Male, n (%)	167 (66.47%)	161 (65.84%)	0.904
SBP (mmHg)	131 (120, 140)	138 (126, 150)	<0.001
DBP (mmHg)	77.0 (70.0, 84.0)	80.0 (74.5, 90.0)	0.002
Height (cm)	169 (162, 174)	168 (162, 172)	0.290
Weight (kg)	69.5 (61.0, 77.0)	71.9 ± 1.2	0.114
BMI (kg/m <sup>2</sup> )	24.11 (22.34, 27.36)	25.76 ± 0.32	0.010
Cholesterol (mmol/L)	4.52 ± 1.11	4.37 (3.50, 5.20)	0.286
Triglyceride (mmol/L)	1.49 (1.09, 2.31)	1.62 (1.11, 2.43)	0.472
HDL-C (mmol/L)	1.00 (0.85, 1.20)	0.94 (0.78, 1.16)	0.028
LDL-C (mmol/L)	2.70 ± 0.870	2.46 (1.73, 3.24)	0.056
sdLDL-C (mmol/L)	1.03 (0.73, 1.34)	0.90 (0.70, 1.26)	0.214
Lp(a) (mg/L)	51.20 (31.15, 92.35)	85.25 (41.55, 193.4)	<0.001
NEFA (μmol/L)	428 (313, 575)	419 ± 213	0.171
Phospholipid (mmol/L)	2.42 ± 0.471	2.37 (1.99, 2.71)	0.622

**Abbreviations:** T2DM, type 2 diabetes mellitus without nephropathy; DN, type 2 diabetes mellitus with nephropathy; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body Mass Index; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; sdLDL-C, small dense low-density lipoprotein cholesterol; Lp(a), lipoprotein(a); NEFA, free fatty acid.

pressure (DBP), Lp(a) was found to be a risk factor for DN (OR = 1.005; 95% CI, 1.002–1.008), while HDL-C was a protective factor (OR = 0.225; 95% CI, 0.076–0.668) (Figure 1E). Then, based on the median of Lp(a), patients were divided into low-Lp(a) group (Lp(a) <64.30mg/L) and high-Lp(a) group (Lp(a) ≥64.30mg/L). Similarly, patients were divided into low-HDL-C group (HDL-C <0.97mmol/L) and high-HDL-C group (HDL-C ≥0.97mmol/L).  $\chi^2$  test showed that the incidence of DN in the high-Lp(a) group (61.18%) was higher than that of the low-Lp(a) group (37.75%) ( $p < 0.001$ , Figure 1F), while the incidence of high-HDL-C group (43.48%) was lower than that of the low-HDL-C group (55.48%) ( $p = 0.033$ , Figure 1G).

## Top 10 Indicators Ranking by the Feature Importance Score

As shown in Table 2, the feature importance score of 57 clinical parameters were obtained by building a random forest classifier. After ranking the feature importance score, the top 10 indicators were urinary albumin (uALB), uALB to creatinine ratio (uACR), cystatin C, creatinine, urinary  $\alpha$ 1-microglobulin, eGFR, urinary  $\beta$ 2-microglobulin, urea nitrogen, superoxide dismutase and fibrinogen. Of which, the top 8 feature indicators are well known for renal function evaluation in clinical practice.

## Correlations Between Lp(a) and Top 10 Indicators

To evaluate whether Lp(a) was related to the top 10 indicators in DN, we performed Spearman's rank correlation analysis. As illustrated in Figure 2, there were positive correlations between serum Lp(a) and uALB ( $r = 0.207$ ,  $p = 0.001$ ), uACR ( $r = 0.214$ ,  $p = 0.006$ ), cystatin C ( $r = 0.186$ ,  $p = 0.001$ ), urinary  $\alpha$ 1-microglobulin ( $r = 0.209$ ,  $p = 0.033$ ), urinary  $\beta$ 2-microglobulin ( $r = 0.208$ ,  $p = 0.031$ ), fibrinogen ( $r = 0.414$ ,  $p < 0.001$ ) and negative correlations between Lp(a) and eGFR ( $r = -0.201$ ,  $p = 0.001$ ) and superoxide dismutase ( $r = -0.286$ ,  $p < 0.001$ ). However, Lp(a) was not related to creatinine and urea nitrogen. In short, Lp(a) was related to 8 out of 10 indicators, suggesting that higher level of Lp(a) implicated poor renal function.

**Table 2** Ranking of Index Importance

Index	Importance Score	Standard Deviation
<b>uALB (mg/L)</b>	<b>0.147639962</b>	<b>0.006891206</b>
<b>uACR (mg/g)</b>	<b>0.102045601</b>	<b>0.003800898</b>
<b>Cystatin C (mg/L)</b>	<b>0.059223046</b>	<b>0.003244178</b>
<b>Creatinine (μmol/L)</b>	<b>0.049524732</b>	<b>0.00194059</b>
<b>Urinary α1-microglobulin (mg/L)</b>	<b>0.045712685</b>	<b>0.00290581</b>
<b>eGFR (mL/min/1.73m<sup>2</sup>)</b>	<b>0.045155647</b>	<b>0.002332287</b>
<b>Urinary β2-microglobulin (ug/L)</b>	<b>0.03748505</b>	<b>0.002362612</b>
<b>Urea nitrogen (mmol/L)</b>	<b>0.034022249</b>	<b>0.002582486</b>
<b>Superoxide dismutase (U/L)</b>	<b>0.019634366</b>	<b>0.000966312</b>
<b>Fibrinogen (mg/dL)</b>	<b>0.018179109</b>	<b>0.00113984</b>
Carbon dioxide (mmol/L)	0.016206725	0.001016206
Serum amyloid protein (mg/L)	0.015993248	0.000712355
Globulin (g/L)	0.013687041	0.000593917
Sialic acid (mg/L)	0.013529818	0.0006511
Uric acid (μmol/L)	0.012908594	0.000778313
A/G	0.01284098	0.000595353
Calcium (mmol/L)	0.01260451	0.000488119
BMI (kg/m <sup>2</sup> )	0.012526778	0.000681684
D-dimer (ng/mL)	0.011876261	0.000750182
K <sup>+</sup> (mmol/L)	0.011870701	0.000699801
Indirect bilirubin (μmol/L)	0.011353344	0.000810906
NEFA (μmol/L)	0.010846996	0.000273531
Urinary creatinine (μmol/L)	0.010737078	0.000569962
Total protein (g/L)	0.010736913	0.000467672
DBP (mmHg)	0.010734574	0.000707376
Glucose (mmol/L)	0.010620109	0.000633742
SBP (mmHg)	0.010597897	0.000662161
Albumin (g/L)	0.01044024	0.000721211
Age (years)	0.010135166	0.000374472
Phosphate (mmol/L)	0.009667562	0.000397528
APTT (s)	0.009401446	0.00055728
Weight (kg)	0.009323374	0.000446333
Cl <sup>-</sup> (mmol/L)	0.009303612	0.000287252
Triglyceride (mmol/L)	0.009100923	0.000515982
Total bilirubin (μmol/L)	0.008891763	0.000335444
Complement C1q (mg/L)	0.008770598	0.000406383
AST/ALT	0.008614687	0.000359756
Thrombin time (s)	0.008480246	0.000327303
Na <sup>+</sup> (mmol/L)	0.008469606	0.000289718
Mg <sup>2+</sup> (mmol/L)	0.008138087	0.000343414
HDL-C (mmol/L)	0.008009633	0.000404038
Total bile acid (μmol/L)	0.007937504	0.000534054
PTTA (%)	0.007768872	0.00030398
Direct bilirubin (μmol/L)	0.007696741	0.000478416
Lp(a) (mg/L)	0.007608078	0.000479737
Height (cm)	0.007342507	0.000430207
γ-glutamyl transpeptidase (U/L)	0.00719477	0.000460904
Phospholipid (mmol/L)	0.007146312	0.000340156
Alkaline phosphatase (U/L)	0.006849449	0.000289154
ALT (U/L)	0.006845283	0.0005644
LDL-C (mmol/L)	0.006786628	0.000308364

(Continued)

Table 2 (Continued).

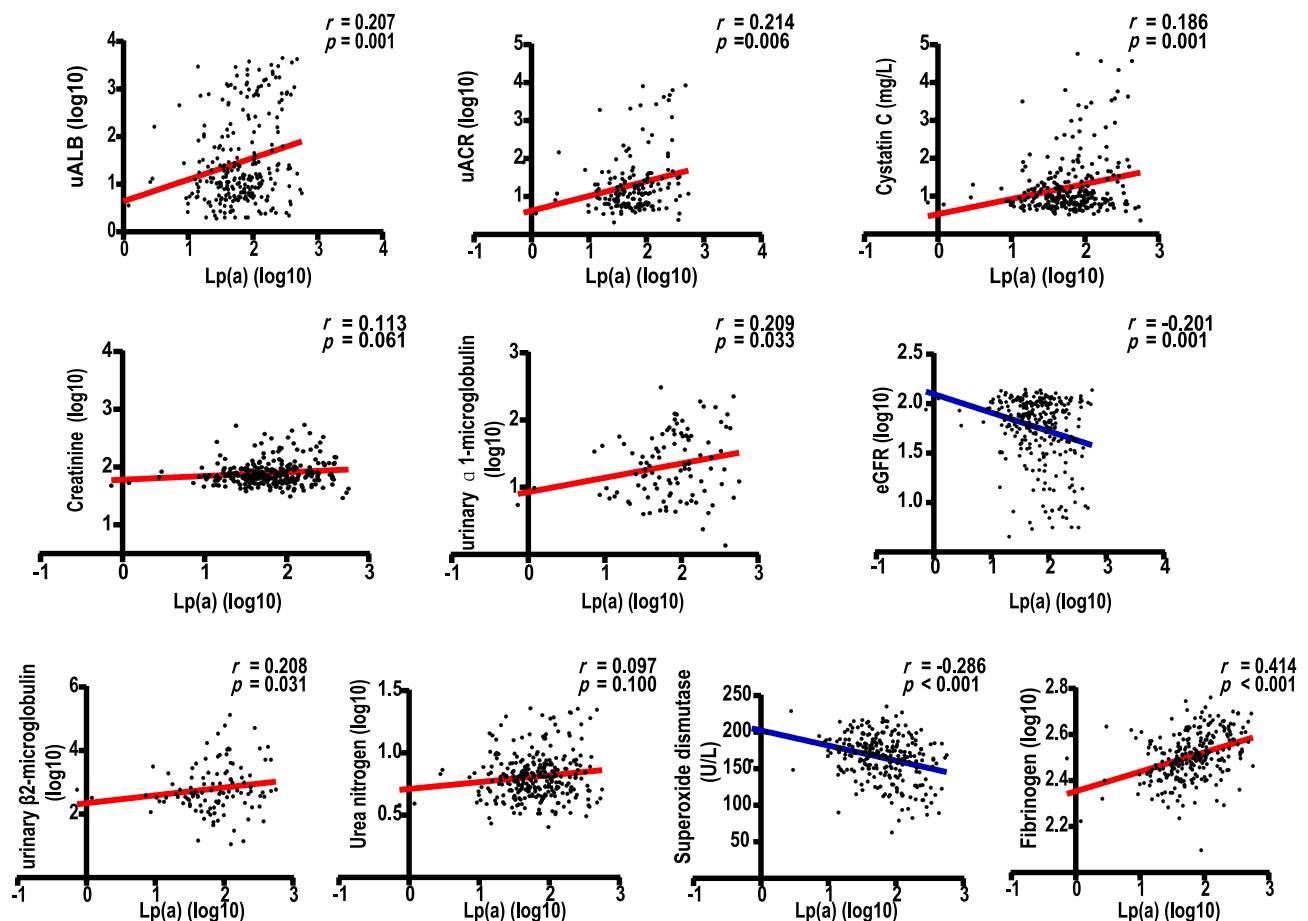
Index	Importance Score	Standard Deviation
Prothrombin time (s)	0.006541392	0.000557753
INR (%)	0.006455686	0.000303502
sdLDL-C (mmol/L)	0.006176813	0.000465997
Cholesterol (mmol/L)	0.005758407	0.000413009
AST (U/L)	0.005700455	0.000392745
Sex	0.001150145	0.0001483

**Notes:** Bold font represents the top 10 indicators after ranking the feature importance score.

**Abbreviations:** SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body Mass Index; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; sdLDL-C, small dense low-density lipoprotein cholesterol; Lp(a), lipoprotein(a); NEFA, free fatty acid; eGFR, estimated glomerular filtration rate; uALB, urinary albumin; uCREA, urinary creatinine; uACR, urinary albumin to creatinine ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; INR, International standardized ratio; PTTA, prothrombin time activity; APTT, activated partial thromboplastin time.

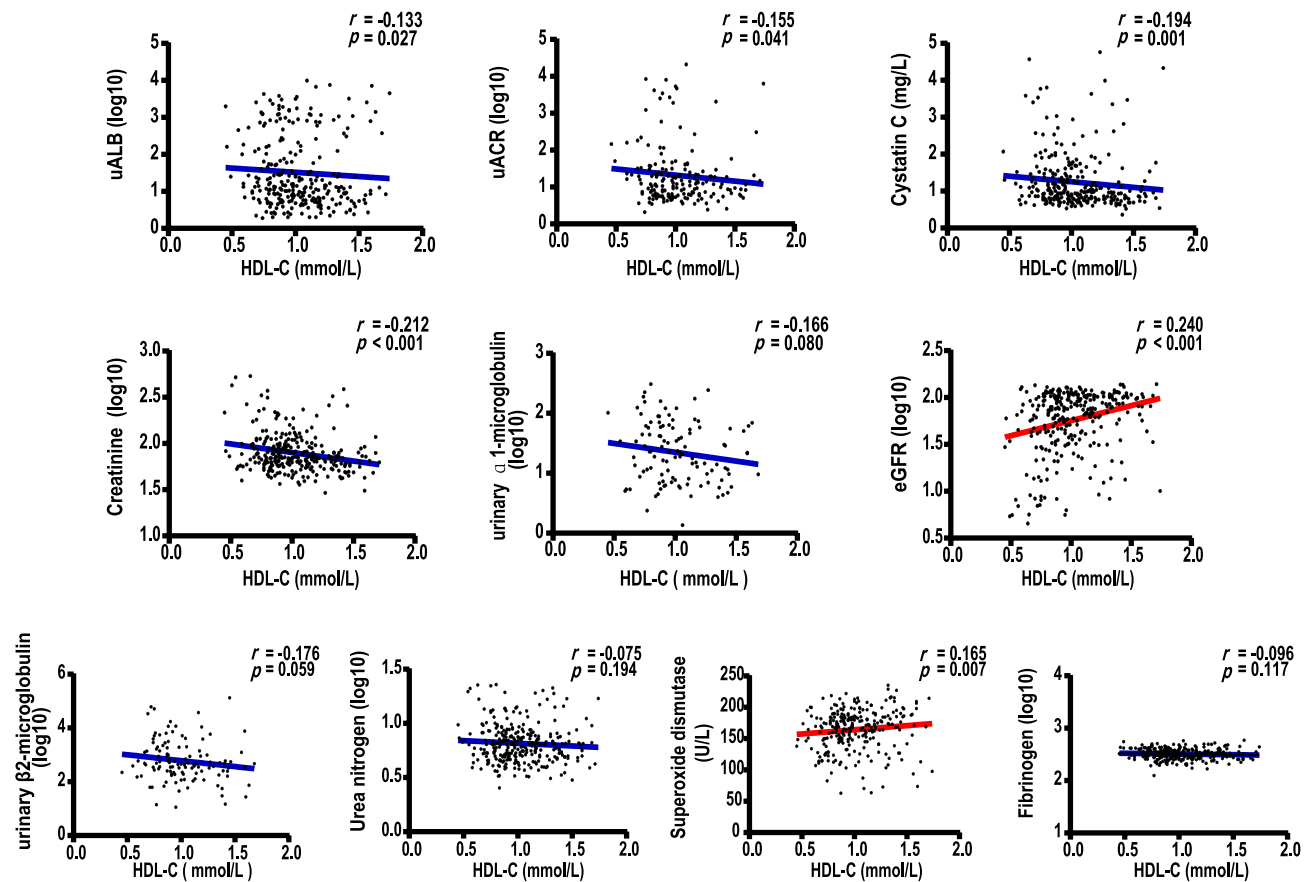
## Correlations Between HDL-C and Top 10 Future Indicators

We then performed Spearman's rank correlation analysis to explore the correlations between HDL-C and the top 10 future indicators. As shown in Figure 3, HDL-C was negatively correlated with uALB ( $r = -0.133$ ,  $p = 0.027$ ), uACR ( $r = -0.155$ ,  $p = 0.041$ ), cystatin C ( $r = -0.194$ ,  $p = 0.001$ ), and creatinine ( $r = -0.212$ ,  $p < 0.001$ ). In addition, HDL-C



**Figure 2** Correlations between Lp(a) and top 10 indicators. Serum Lp(a) was positively correlated with uALB, uACR, cystatin C, urinary  $\alpha$ 1-microglobulin, urinary  $\beta$ 2-microglobulin and fibrinogen, while negatively correlated with eGFR and superoxide dismutase. Lp(a) was not related to creatinine and urea nitrogen.

**Abbreviations:** uALB, urinary microalbumin; uACR, urinary albumin to creatinine ratio;  $r$ , related coefficient;  $p$ ,  $p$  value.



**Figure 3** Correlations between HDL-C and top 10 indicators. Serum HDL-C was negatively correlated with uALB, uACR, cystatin C and creatinine. HDL-C was positively correlated with eGFR and superoxide dismutase. HDL-C was not related to urinary  $\alpha$ 1-microglobulin, urinary  $\beta$ 2-microglobulin, urea nitrogen and fibrinogen.

**Abbreviations:** uALB, urinary microalbumin; uACR, urinary albumin to creatinine ratio;  $r$ , related coefficient;  $p$ ,  $p$  value.

was positively correlated with eGFR ( $r = 0.240$ ,  $p < 0.001$ ) and superoxide dismutase ( $r = 0.165$ ,  $p = 0.007$ ). However, HDL-C was not related to urinary  $\alpha$ 1-microglobulin, urinary  $\beta$ 2-microglobulin, urea nitrogen and fibrinogen. Our data indicated that higher HDL-C is associated with better renal function.

## Decision Tree Models for DN Prediction

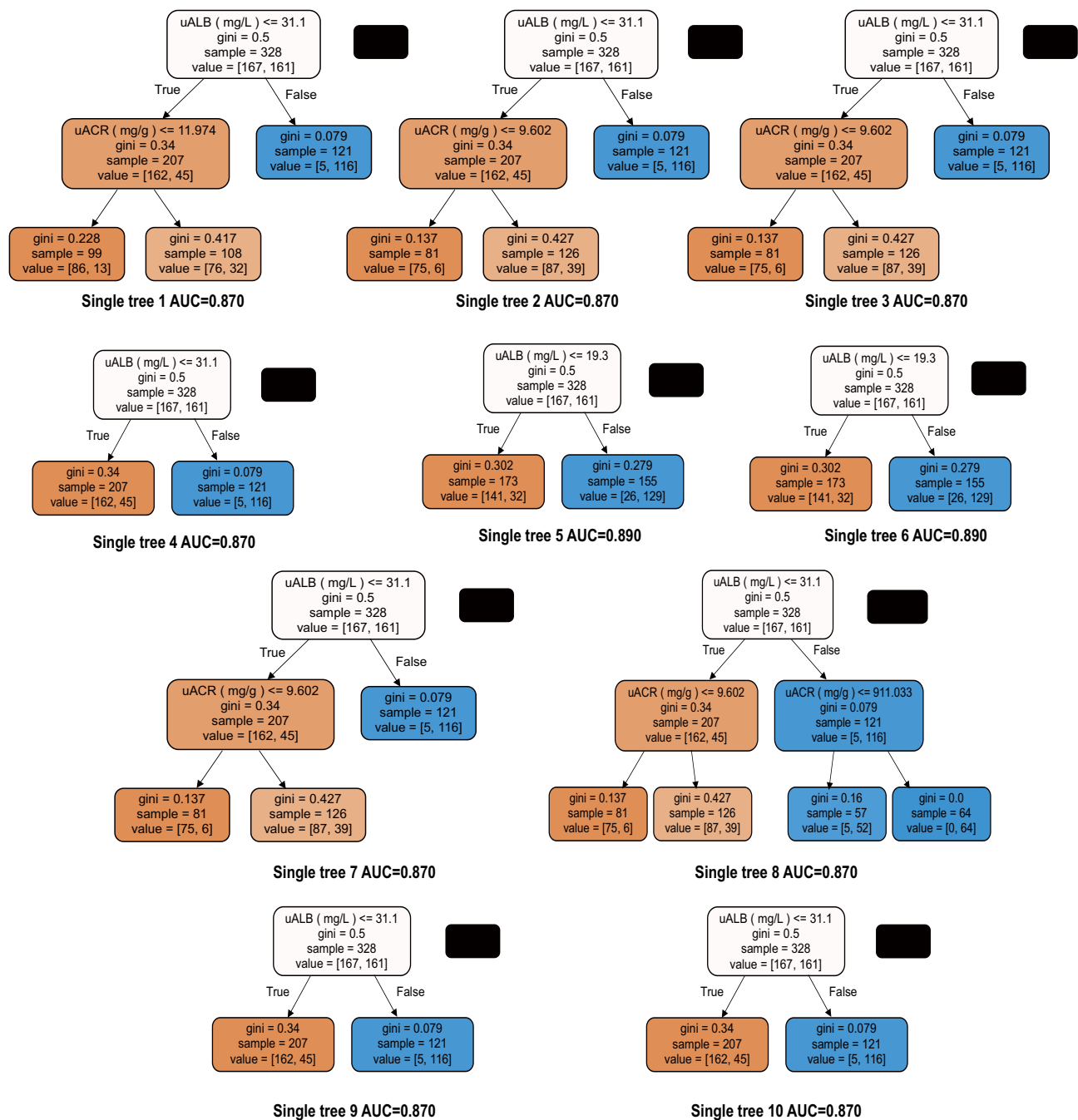
As shown in Figure 4, the top-ranked index uALB showed robust performance for DN predication. The area under the receiver operating characteristic curve (AUC) of the decision tree models at cut-off value of 31.1mg/L ranged from 0.870 to 0.890, with an average AUC of 0.874.

## Discussion

For the first time, we employed a combination of machine learning and correlation analysis to explore the links between Lp(a), HDL-C, and diabetic nephropathy. This led us to develop a decision tree model with uALB for predicting DN, which showed an AUC of 0.874 and a trained cut-off value that aligns with the clinical normal range (uALB  $\leq$  30 mg/L), providing a reliable and simple model for clinical DN prediction. Our data indicated that Lp(a) is a risk factor for DN, while HDL-C is a protective factor after controlling blood pressure. Furthermore, Lp(a) and HDL-C were associated with the top 10 indicators selected from 57 clinical parameters by feature selection with random forest classifier, emphasizing their clinical significance for assessing DN.

In line with previous reports,<sup>26</sup> we found that higher levels of serum Lp(a) are associated with renal insufficiency in diabetes patients. For instance, Lp(a) was related to renal function indicators (uALB, uACR, eGFR, urea nitrogen, creatinine and cystatin C). However, the role of Lp(a) in the pathogenesis of DN remains unclear. Lp(a) can mediate





**Figure 4** Decision tree models for DN prediction. Gini represents the purity of the node.  
**Abbreviation:** AUC, area under the receiver operating characteristic curve.

inflammation, promote thrombosis and atherosclerotic plaque formation.<sup>27</sup> Mechanically, Lp(a) induces endothelial cells to express adhesion molecules vascular cellular adhesion molecule-1, E-selectin and monocyte chemoattractant protein, promotes endothelial cell proliferation and foam cell formation, mediates plaque inflammation, inhibits the production of anti-atherosclerotic nitric oxide, promotes tissue factor-mediated thrombosis, and induces platelet activation.<sup>28</sup> Oxidative stress and glomerular atherosclerosis mediated by Lp(a)-induced reactive oxygen metabolites also play important roles in the pathogenesis of DN.<sup>29–31</sup> Interestingly, we observed a negative correlation between Lp(a) and superoxide dismutase activity, which is the first line of defense against reactive oxygen species-mediated damage.<sup>32</sup>

Lp(a) has been reported to inhibit the fibrinolytic system, leading to hypercoagulable state and renal damage.<sup>33</sup> DN is related to changes in the coagulation system,<sup>34–36</sup> and our results indicated that the hypercoagulable state in T2DM patients with nephropathy was related to Lp(a) ([Supplementary Table S2](#)). Lp(a) was positively correlated with blood fibrinogen and D-dimer, which significantly increased in T2DM patients with nephropathy. However, our data did not show any correlation between Lp(a) and creatinine. This can be attributed to the fact that the main determinant of Lp(a) is liver uptake, and the mechanism of regulating Lp(a) secretion by the kidney differs between individuals with impaired renal function and those with a normal to moderate decrease in renal clearance.<sup>37–39</sup> Even under physiological conditions, the kidney accounts for only 1.3% of Lp(a) clearance.<sup>38</sup>

Low levels of HDL-C have been considered an independent risk factor for the development and progression of DN,<sup>40</sup> and our results confirmed this. HDL-C has nephroprotective, anti-atherosclerotic, and anti-inflammatory effects.<sup>41,42</sup> Our data revealed correlations between HDL-C and several renal function indicators, suggesting that low levels of HDL-C in T2DM patients with nephropathy are associated with poor renal function. Furthermore, HDL-C is involved in oxidative stress and regulates endothelial cell apoptosis, vascular smooth muscle, and platelet function.<sup>43,44</sup> The regulation of HDL-C on endothelial cells is dependent on calcium homeostasis,<sup>43</sup> and our results showed that the level of HDL-C was positively correlated with antioxidant levels and calcium levels ([Supplementary Table S3](#)).

There were still several limitations in this study. First, due to various therapeutic requirements of the patients, their diets, exercise and drug treatments could not be normalized, which prevented us from taking these factors into account in the current study. Second, although our correlation and machine learning data found connections between Lp(a), HDL-C and diabetic nephropathy, more direct causal evidence was lacking. Third, this was a retrospective and single-center study, therefore, a multicenter prospective study and mechanism investigations will be necessary in the future.

## Conclusion

In summary, the increase of Lp(a) was a risk factor for DN, while the decrease of HDL-C was a protective factor. We reported serum Lp(a) and HDL-C related to the top 10 features of DN. We also provided a decision tree model for DN prediction, and uALB could be a reliable and easily detectable indicator.

## Abbreviations

T2DM, Type 2 diabetes mellitus; DN, diabetic nephropathy; CKD, chronic renal disease; ASCVD, atherosclerosis and cardio-cerebrovascular diseases; Lp(a), lipoprotein(a); HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; sdLDL-C, small dense low-density lipoprotein cholesterol; NEFA, free fatty acid; eGFR, estimated glomerular filtration rate; uALB, urinary albumin; uCREA, urinary creatinine; uACR, urinary albumin to creatinine ratio; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; ALT, alanine aminotransferase; AST, aspartate aminotransferase; INR, International standardized ratio; PTTA, Prothrombin time activity; APTT, activated partial Thromboplastin time.

## Acknowledgments

We are indebted to all the participants for their dedicated and conscientious collaboration. Rui-Huan Gao and Boyang Liu are co-first authors for this study. Yidan Zhou and Song-Mei Liu are co-senior authors for this study.

## 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.

## Funding

We thank the financial support from the National Natural Science Foundation of China (81972009, 82172359) and Health Commission of Hubei Province Scientific Research Project (2019CFA018).

## Disclosure

The authors declare no conflicts of interest in this work.

## References

1. Umanath K, Lewis JB. Update on diabetic nephropathy: core curriculum 2018. *Am J Kidn Dis*. 2018;71(6):884–895. doi:10.1053/ajkd.2017.10.026
2. Rayego-Mateos S, Morgado-Pascual JL, Opazo-Ríos L, et al. Pathogenic pathways and therapeutic approaches targeting inflammation in diabetic nephropathy. *Int J Mol Sci*. 2020;21(11):3798. doi:10.3390/ijms21113798
3. Selby NM, Taal MW. An updated overview of diabetic nephropathy: diagnosis, prognosis, treatment goals and latest guidelines. *Diabetes Obes Metab*. 2020;22(Suppl 1):3–15. doi:10.1111/dom.14007
4. Li J, Albajrami O, Zhuo M, Hawley CE, Paik JM. Decision algorithm for prescribing SGLT2 Inhibitors and GLP-1 receptor agonists for diabetic kidney disease. *Clin J Am Soc Nephrol*. 2020;15(11):1678–1688. doi:10.2215/CJN.02690320
5. Alicic RZ, Rooney MT, Tuttle KR. Diabetic kidney disease: challenges, progress, and possibilities. *Clin J Am Soc Nephrol*. 2017;12(12):2032–2045. doi:10.2215/CJN.11491116
6. American Diabetes Association. Microvascular complications and foot care: standards of medical care in diabetes-2019. *Diabe Care*. 2019;42(Suppl 1):S124–s138. doi:10.2337/dc19-S011
7. Wheeler DC, Stefánsson BV, Jongs N, et al. Effects of dapagliflozin on major adverse kidney and cardiovascular events in patients with diabetic and non-diabetic chronic kidney disease: a prespecified analysis from the DAPA-CKD trial. *Lancet Diabet Endocrinol*. 2021;9(1):22–31. doi:10.1016/S2213-8587(20)30369-7
8. Tu C, Wang L, Wei L. The role of PKM2 in diabetic microangiopathy. *Diabetes Meta Syndr Obesity*. 2022;15:1405–1412. doi:10.2147/DMSO.S366403
9. Michos ED, McEvoy JW, Blumenthal RS. Lipid management for the prevention of atherosclerotic cardiovascular disease. *N Engl J Med*. 2019;381(16):1557–1567. doi:10.1056/NEJMra1806939
10. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American heart association task force on clinical practice guidelines. *Circulation*. 2019;140(11):e563–e595. doi:10.1161/CIR.0000000000000677
11. Deng X, Qin S, Chen Y, et al. B-RCA revealed circulating miR-33a/b associates with serum cholesterol in type 2 diabetes patients at high risk of ASCVD. *Diabetes Res Clin Pract*. 2018;140:191–199. doi:10.1016/j.diabres.2018.03.024
12. Lee YJ, Cho JY, You SC, et al. Moderate-intensity statin with ezetimibe vs. high-intensity statin in patients with diabetes and atherosclerotic cardiovascular disease in the RACING trial. *Eur Heart J*. 2023;44(11):972–983. doi:10.1093/eurheartj/ehac709
13. Mohammadi-Shemirani P, Chong M, Narula S, et al. Elevated Lipoprotein(a) and risk of atrial fibrillation: an observational and Mendelian randomization study. *J Am Coll Cardiol*. 2022;79(16):1579–1590. doi:10.1016/j.jacc.2022.02.018
14. Gross JL, de Azevedo MJ, Silveiro SP, Canani LH, Caramori ML, Zelmanovitz T. Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabe Care*. 2005;28(1):164–176.
15. Chandni R, Ramamoorthy KP. Lipoprotein(a) in type 2 diabetic subjects and its relationship to diabetic microvascular complications. *World J Diabetes*. 2012;3(5):105–109. doi:10.4239/wjd.v3.i5.105
16. Chaturvedi N, Fuller JH, Taskinen MR. Differing associations of lipid and lipoprotein disturbances with the macrovascular and microvascular complications of type 1 diabetes. *Diabe Care*. 2001;24(12):2071–2077. doi:10.2337/diacare.24.12.2071
17. Aromolaran O, Aromolaran D, Isewon I, Oyelade J. Machine learning approach to gene essentiality prediction: a review. *Brief Bioinform*. 2021;22(5):bbab128.
18. Lin E, Lin CH, Lane HY. Prediction of functional outcomes of schizophrenia with genetic biomarkers using a bagging ensemble machine learning method with feature selection. *Sci Rep*. 2021;11(1):10179. doi:10.1038/s41598-021-89540-6
19. Care D. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020. *Diabe Care*. 2020;43(Suppl 1):S14–s31. doi:10.2337/dc20-S002
20. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20–29. doi:10.1056/NEJMoa1114248
21. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001;17(6):520–525. doi:10.1093/bioinformatics/17.6.520
22. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. doi:10.1023/A:1010933404324
23. Stone MA. Cross-validated choice and assessment of statistical predictions. *J Royal Stat Soc*. 1974;36(2):111–147.
24. Jamieson K, Talwalkar A. Non-stochastic best arm identification and hyperparameter optimization. *Artif Intell Stat*. 2016;51:240–248.
25. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: a novel bandit-based approach to hyperparameter optimization. *J Mach Learn Res*. 2017;18(1):6765–6816.
26. Hermans MP, Ahn SA, Rousseau MF. The mixed benefit of low lipoprotein(a) in type 2 diabetes. *Lipids Health Dis*. 2017;16(1):171. doi:10.1186/s12944-017-0564-9
27. Ugovšek S, Šebešljen M. Lipoprotein(a)-the crossroads of atherosclerosis, atherothrombosis and inflammation. *Biomolecules*. 2021;12(1):26. doi:10.3390/biom12010026
28. Boffa MB. Beyond fibrinolysis: the confounding role of Lp(a) in thrombosis. *Atherosclerosis*. 2022;349:72–81. doi:10.1016/j.atherosclerosis.2022.04.009
29. Hung CC, Tsai JC, Kuo HT, Chang JM, Hwang SJ, Chen HC. Dyslipoproteinemia and impairment of renal function in diabetic kidney disease: an analysis of animal studies, observational studies, and clinical trials. *Rev Diabet Stud*. 2013;10(2–3):110–120. doi:10.1900/RDS.2013.10.110
30. Piani F, Melena I, Tommerdahl KL, et al. Sex-related differences in diabetic kidney disease: a review on the mechanisms and potential therapeutic implications. *J Diabetes Complications*. 2021;35(4):107841. doi:10.1016/j.jdiacomp.2020.107841
31. Greiber S, Kreuzel M, Pavenstädt H, Schollmeyer P, Wanner C. Lipoprotein(a) induces glomerular superoxide anion production. *Nephrol Dialysis Tran*. 1997;12(7):1330–1335. doi:10.1093/ndt/12.7.1330
32. Zhao H, Zhang R, Yan X, Fan K. Superoxide dismutase nanozymes: an emerging star for anti-oxidation. *J Materials Chem B*. 2021;9(35):6939–6957. doi:10.1039/D1TB00720C

33. Berg K. A new serum type system in man--The Ld system. *Vox Sang.* 1965;10(5):513–527. doi:10.1111/j.1423-0410.1965.tb01404.x
34. Alsharidah AS. Diabetes mellitus and diabetic nephropathy: a review of the literature on hemostatic changes in coagulation and thrombosis. *Blood Res.* 2022;57(2):101–105. doi:10.5045/br.2022.2021204
35. Goldberg RB. Cytokine and cytokine-like inflammation markers, endothelial dysfunction, and imbalanced coagulation in development of diabetes and its complications. *J Clin Endocrinol Metab.* 2009;94(9):3171–3182. doi:10.1210/jc.2008-2534
36. Pan L, Ye Y, Wo M, et al. Clinical significance of hemostatic parameters in the prediction for type 2 diabetes mellitus and diabetic nephropathy. *Dis Markers.* 2018;2018:5214376. doi:10.1155/2018/5214376
37. Cain WJ, Millar JS, Himebauch AS, et al. Lipoprotein [a] is cleared from the plasma primarily by the liver in a process mediated by apolipoprotein [a]. *J Lipid Res.* 2005;46(12):2681–2691. doi:10.1194/jlr.M500249-JLR200
38. Doucet C, Mooser V, Gonbert S, et al. Lipoprotein(a) in the nephrotic syndrome: molecular analysis of lipoprotein(a) and apolipoprotein(a) fragments in plasma and urine. *J Am Soc Nephrol.* 2000;11(3):507–513. doi:10.1681/ASN.V113507
39. Lippi G, Salvagno GL, Montagnana M, Targher G, Franchini M, Guidi GC. No correlation between lipoprotein(a) and biochemical markers of renal function in the general population. *Arch Pathol Lab Med.* 2008;132(9):1436–1438. doi:10.5858/2008-132-1436-NCBLAB
40. Molitch ME, DeFronzo RA, Franz MJ, Keane WF, Mogensen CE, Parving HH. Diabetic nephropathy. *Diabe Care.* 2003;26(1):S94–98.
41. Li X, Su T, Xiao H, et al. Association of the HDL-c Level with HsCRP, IL-6, U-NAG, RBP and Cys-C in Type 2 diabetes mellitus, hypertension, and chronic kidney disease: an epidemiological survey. *Diabetes Meta Syndr Obesity.* 2020;13:3645–3654. doi:10.2147/DMSO.S265735
42. Lüscher TF, Landmesser U, von Eckardstein A, Fogelman AM. High-density lipoprotein: vascular protective effects, dysfunction, and potential as therapeutic target. *Circ Res.* 2014;114(1):171–182. doi:10.1161/CIRCRESAHA.114.300935
43. Mineo C, Shaul PW. Novel biological functions of high-density lipoprotein cholesterol. *Circ Res.* 2012;111(8):1079–1090. doi:10.1161/CIRCRESAHA.111.258673
44. Zewinger S, Drechsler C, Kleber ME, et al. Serum amyloid A: high-density lipoproteins interaction and cardiovascular risk. *Eur Heart J.* 2015;36(43):3007–3016. doi:10.1093/eurheartj/ehv352

## Diabetes, Metabolic Syndrome and Obesity

Dovepress

### 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>