

Construction of risk prediction models for chronic kidney disease complicated with renal hypertension based on machine learning algorithms

Qin Zhu

Chongqing Medical University

Zhiyin Du (✉ duzhiyin@cqmu.edu.cn)

Chongqing Medical University

Ting Liu

Chongqing Medical University

Research Article

Keywords: chronic kidney disease, renal hypertension, machine learning, prediction model, electronic health record

Posted Date: April 22nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1563211/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Chronic kidney disease (CKD) is a causal relationship with hypertension. Renal hypertension, the main complication of CKD, is a traditional risk factor for cardiovascular events.

Objective: This retrospective study aimed to establish a risk prediction model for CKD with renal hypertension (RH) using machine learning algorithms.

Methods: Using the electronic medical record database of seven large hospitals in Chongqing, 1572 patients with CKD were selected. Based on the presence of RH, they were divided into RH (n = 400) and non-RH (n = 1172) groups. Data from 70% of patients were randomly allocated to the training set to construct the prediction model. The remaining 30% was used as the test set for internal verification. Single-factor logistic regression and correlation analysis were used to screen input indicators. Prediction models were constructed using these machine learning algorithms: support vector machine, random forest, XGBoost, LightGBM, GBDT, and CatBoost. The optimal parameters of these algorithms were determined using the grid search algorithm. The predictive values of the models constructed for predicting CKD with RH were compared.

Results: Urinary protein, urinary occult blood, creatinine, cystatin C, age, creatine kinase-MB, and $\beta 2$ microglobulin were predictors of CKD with RH. The XGBoost model performed best with a sensitivity of 0.820, a specificity of 0.945, an F1 score of 0.840, and an area under the relative operating characteristic curve of 0.935.

Conclusion: The clinical prediction model constructed by the XGBoost algorithm had the potential to predict CKD with RH.

1 Background

Chronic kidney disease (CKD) is a common disease worldwide. Its incidence rate has been increasing, and it is closely associated with diabetes mellitus and hypertension. Renal hypertension (RH) that complicates CKD is a common cause of secondary hypertension, is mainly caused by renal vascular and/or renal parenchymal lesions, and increases the risk of renal insufficiency. Hypertension (HT) and CKD have a unique association. They have a cause-and-effect relationship with each other, that is, HT is the main cause and complication of CKD. Additionally, HT is a traditional risk factor for cardiovascular events. In the absence of interventions, a vicious circle of heart and kidney disease may develop in patients with CKD^[1-2]. According to relevant studies, up to 50% of patients with CKD have RH as a complication^[3-5], which significantly increases mortality. Thus, it is particularly important that clinicians formulate targeted prevention and control strategies, improve the prognosis of patients, and reduce the risk of death by finding and addressing the potential risk factors of CKD with RH.

Significant effort has been made toward the early diagnosis and treatment of RH in patients with CKD. In 2014, relevant research on CKD complicated with renal hypertension in children found that hypertension

is one of the most common cardiovascular risk factors for CKD in children^[6]. Hypertension in patients with CKD can be diagnosed early through careful measurements of blood pressure. In 2019, Xiao, Liu, and others analyzed the risk factors of CKD complicated with hypertension^[7-8]. In 2020, Wang, Wan, and others used traditional statistical methods to analyze the risk factors of CKD complicated with renal hypertension^[9]. In 2021, Zhang, Xing, and others proposed that if renal hypertension is diagnosed early and corresponding interventional measures are performed, further development of nephropathy can be delayed and morbidity can be reduced^[10]. However, a convenient and rapid method to evaluate the risk of CKD complicated with RH has not yet been found.

Machine learning algorithms are powerful tools for prediction that use large amounts of data^[11]. In recent years, with the development of cloud computing and big data, machine learning has gradually become an important tool and research object for scientific research and practical application. As a classical prediction model, machine learning has good prediction performance and is widely used in the research of chronic diseases^[12]. In the 21st century, to better carry out evidence-based medicine, medical research began shifting from disease prevention and treatment to health maintenance, and the medical model has been shifting from simple disease treatment to including prevention, prediction, personalization, and patient participation^[13]. Thus, considering the multiple comorbidities in patients with CKD, reasonable and effective interventions to prevent RH should particularly be performed.

In this context, six machine learning algorithms (support vector machine, random forest, XGBoost, LightGBM, GBDT, and CatBoost) were used to construct and evaluate risk prediction models for chronic kidney disease complicated with renal hypertension to lay the foundation for RH risk assessment in patients with CKD and to provide references for clinical prevention and treatment.

2 Materials And Methods

Research Object

This study is a retrospective study. Between January 2019 and January 2021, 1572 patients with CKD were included from seven large medical institutions in Chongqing. Patients over 18 years of age with CKD were included. Patients with essential hypertension, those with CKD with a malignant tumor, fracture, or mental illness, and those with incomplete clinical data were excluded.

Grouping

Based on the diagnosis of renal hypertension [5], as well as clinical manifestations and laboratory examination, the 1572 patients were divided into the RH group (n = 400) and the non-RH group (n = 1172). Patients in the RH and non-RH groups were randomly divided further between the training set (n = 1100) and test set (n = 472) in a 7:3 ratio. The training set was used to train the models, and the test set was used for internal verification.

Research Indicators

The general information of the patient (age, gender, past medical history, history of smoking, drinking, and diabetes mellitus) was collected, and laboratory examinations (routine blood tests, routine urine tests, renal function tests, liver function tests, blood electrolyte levels, and coagulation tests) were performed.

Statistical Processing

SPSS 25.0 software was used for statistical analysis. For variables with less than 30% data missing, the mean was calculated. Measurement data that followed a normal distribution were expressed as $\bar{X} \pm S$, and the t-test was used to compare between groups. Measurement data that did not conform to the normal distribution were represented by $M (P_{25}, P_{75})$, and the Mann–Whitney U test was used to compare between groups. Count data were expressed as rates (%), and to compare between groups, the χ^2 test was used. Univariate logistic regression analysis was used for internal validation, and a P value less than 0.05 was considered the standard for the inclusion of variables in multivariate analysis. Correlation analysis was carried out on the variables selected, and the variables with correlation coefficients greater than 0.15 were retained. Finally, after the multicollinearity testing of the remaining variables, provided the variance expansion factor between variables (variance inflation factor value) was within 10, it was considered that no collinearity was found between variables^[14], and those variables were included for statistical tests and feature screening.

Model Building

After determining the input variables of the model, 70% of the cases were randomly selected to constitute the training data set, and 30% of the cases constituted the test set. Taking the occurrence of RH as the outcome variable, six machine learning prediction models (support vector machine, random forest, XGBoost, LightGBM, GBDT, and CatBoost) were trained using the training set. Then, the optimal parameters of the model are searched by grid search algorithm (see Table 1).

All models were validated with 10-fold cross-validation.

Model Evaluation

Once the model was constructed, it must be evaluated to determine whether it is suitable to predict disease. In this study, the sensitivity and specificity of the model on the test set were calculated using Python language, and the area under the curve (AUC) of the relative operating characteristic curve was drawn to evaluate the prediction model. The 10-fold cross-validation method was used to verify the generalizability of the model.

3 Result

Baseline Data

Univariate logistic regression analysis showed that age, past medical history, history of drinking, history of dialysis, presence of diabetes mellitus, systolic blood pressure, diastolic blood pressure (DBP), pulse, β 2 microglobulin, urea, total cholesterol, total bilirubin, direct bilirubin, indirect bilirubin, lymphocyte count, urinary occult blood, urinary protein, serum creatinine, cystatin C, platelet count, blood sodium, phosphorus, and potassium levels, stage of CKD, alanine transaminase levels, aspartate transaminase levels, C-reactive protein, creatine kinase myocardial band (CK-MB), and neutrophil-to-lymphocyte ratio were significantly associated with renal hypertension (had P values of less than 0.05) (Table 2).

The results showed that treatment with dialysis, age, urea, urinary occult blood, CK-MB, β 2 microglobulin, urinary protein, DBP, stage of CKD, cystatin C, indirect bilirubin, and creatinine had high correlations with renal hypertension ($r > 0.15$) (shown in Figure 1).

Comparison of the Predictive Performance of Each Model

After optimization using the grid search algorithm, the support vector machine, random

forest, XGBoost, LightGBM, GBDT, and CatBoost models were internally verified in the test set. The results showed that the AUCs of each model were high: 0.831, 0.928, 0.935, 0.932, 0.929, and 0.929, respectively. The XGBoost model had the best comprehensive prediction efficiency and the highest AUC (0.935). (Figure 2, Table 3).

Analysis of the Important Influencing Factors of CKD Complicated with Renal Hypertension

All five models showed that urinary protein, urinary occult blood, creatinine, cystatin C, age, β 2 microglobulin, and CK-MB significantly influenced CKD with renal hypertension. Hence, these indicators could be used as the important influencing factors of RH in patients with CKD (shown in Figure 3).

4 Discussion

In this study, we established a machine learning prediction model for renal hypertension after CKD. Based on the single-factor analysis and correlation analysis, 12 variables were included for evaluation. The final evaluation of the models showed that the XGBoost model prediction effect on CKD with RH is the best. XGBoost showed advantages in processing many aspects of the data. It has been widely studied in the fields of genetics, proteomics, pharmacology, pathology, and so on. In this study, the AUC value of the XGBoost model was 0.935, the specificity was 0.945, and the sensitivity was 0.820, which showed that

the established XGBoost model had the best predictive ability. Simultaneously, this study compared five other commonly used machine learning models.

This study ranked the importance of variables on the basis of their predictive abilities. The variables with high predictive abilities included urinary protein, urinary occult blood, and creatinine. For patients with CKD, renal damage mainly manifested with increased creatinine, decreased glomerular filtration rate, or increased urinary albumin excretion, resulting in increased urinary protein and urinary occult blood^[15]. Some researchers^[16] suggest that patients with CKD develop increased nocturia, limb edema, and mild anemia. With the progression of the disease, the renal function continues to deteriorate, and abnormalities of urinary components, such as hematuria and proteinuria, develop^[9]. These variables were followed in predictive ability by cystatin C, age, CK-MB, and β 2 microglobulin. When CKD progresses to a certain extent, it causes renal hypertension. Renal function further reduces, water and sodium retention increase the capacity of extracellular fluid space, and the activation of the local sympathetic nervous system and renin–angiotensin system increase the risk of concurrent hypertension. Cystatin C accounts for a high proportion in the model of CKD complicated with RH. Cystatin C is an inhibitor of cysteine protease, an endogenous marker reflecting the changes of glomerular filtration, and a relatively stable index for detecting renal function. Shi, Zhang^[17], and others found that cystatin C and CKD are closely related to renal function injury, and cystatin C has high specificity and accuracy and can be used as a reliable index for the assessment of renal function in patients with CKD^[18]. With an increase in age, the body's immunity decreases, worsening the morbidity in patients with CKD while increasing the risk of renal hypertension^[8,19]. CK-MB is a marker of myocardial damage. There is a close causal relationship between blood pressure and the risk of cardiovascular and cerebrovascular disease. Thus, an increase in the creatine kinase isozyme suggests that attention should be paid to preventing the increase of blood pressure in advance^[20]. β 2 microglobulin and CKD are closely related to hypertension. The increase of β 2 microglobulin indicates that the renal tubular function has been damaged. Hence, the blood pressure should be controlled actively and β 2 microglobulin measured regularly.

A large number of studies^[8,10,21-23] have shown that proteinuria, hematuria, creatinine, and cystatin C are closely related to renal hypertension. Thus, the prediction model reflects the actual clinical situation and has a good predictive ability for CKD complicated with RH. Thus, we should strengthen the risk assessment among older patients with CKD, patients with CKD stage 4–5, and patients with CKD and cardiovascular disease. Timely screening must be performed, and early disease warning must be given.

Our study had several limitations. First, we included many variables. The optimal number of variables obtained by single-factor analysis and correlation analysis may cause a lack of practicability in clinical practice. Perhaps, the number of variables can be optimized by other methods in the future. Second, this study is a single-center study. The included sample size is small, and some variables had to be deleted as they were missing values. Increasing the sample size and improving the collection of variables will provide data closer to the real results. Presently, this study is only an exploratory study and must be verified using larger studies.

5 Conclusion

In conclusion, the XGBoost model had a good predictive ability for patients with CKD complicated with RH. It can help clinicians classify patients based on risk to provide an effective reference in clinical practice. Simultaneously, in patients with CKD and proteinuria, hematuria, and high creatinine, active interventions to reduce the risk of RH can be performed. However, before the XGBoost model is applied to clinical practice, external validation research must be conducted.

6 List Of Abbreviations

CKD: Chronic kidney disease

RH: Renal hypertension

SVM support vector machine model

RF random forest model

SBP systolic blood pressure

DBP diastolic blood pressure;

PP pulse pressure

BT temperature;

ALT alanine aminotransferase;

AST aspartate aminotransferase;

CRP C-reactive protein;

CK-MB creatine kinase myocardial band isoenzyme;

NLR neutrophil-to-lymphocyte ratio

Declarations

Ethics approval and consent to participate

The data we obtained comes from the medical big data platform. The data in the platform has been desensitized and does not contain any personal privacy data. The medical research ethics committee of Chongqing Medical University approved this study, and all data was de-identified and informed consent was waived for the retrospective data. All methods of this study were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Conflict of Interest

The authors declare that no potential conflict of interest exists.

Funding

Not applicable.

Authors' Contributions

Qin Zhu and Zhiyin Du contributed to the concept and design of the research. Qin Zhu and Ting Liu participated in data collection and data processing. Qin Zhu contributed to statistical analysis and data interpretation. Zhu Qin and Zhiyin Du completed the drafting and revision of the manuscript. All authors have made critical changes to the important knowledge content of the manuscript.

Acknowledgement

We thank seven hospitals for providing electronic medical record data in Chongqing.

References

1. Charles C, Ferris A H. Chronic Kidney Disease. Primary Care: Clinics in Office Practice. 2020; 47:585–595.
2. Kono K, Fujii H, Nakai K, Goto S, Watanabe S, Watanabe K, et al. Relationship between type of hypertension and renal arteriosclerosis in chronic glomerular disease. Kidney & Blood Pressure Research. 2016; 41:374–83. doi: 10.1159/000443440.
3. Diwan V, Brown L, Gobe GC. Adenine-induced chronic kidney disease in rats. Nephrology. 2017; 23:5–11.

4. Hanqing Wang, Minghe Wang. Diagnosis and treatment of renovascular hypertension. *World clinical medicine*. 2012; 33: 15-17.
5. Kirkendall W M, Fitz A E, Lawrence M S. Renal hypertension. Diagnosis and surgical treatment. *New England Journal of Medicine*. 1967; 276:479.
6. Peco-Antic A, Paripovic D. Renal hypertension and cardiovascular disorder in children with chronic kidney disease. *Srpski arhiv za celokupno lekarstvo*. 2014; 142:113–117.
7. Yangyang Xiao, Qiuyue Li, Qinkai Chen. Risk factors of hypertension in patients with chronic kidney disease stage 5. *Journal of Huazhong University of science and Technology (Medical Edition)*. 2015; 6:696–699.
8. Yang Liu, Yu Wenjuan Yu, Xiuzhen Li, Bei Zhao, Gang Yao, et al. Risk factors of chronic kidney disease in elderly patients with essential hypertension. *Chinese Journal of geriatric multiple organ diseases*. 2019; 18:1–5.
9. Jing Wang, Changliang Wan, Xiaonan Li. Analysis of risk factors and preventive measures of chronic kidney disease complicated with renal hypertension. *Guizhou medicine*. 2020; 44:1058–1059.
10. Yanan Zhang, Yan Xing, Lina Yao, Wenming Niu, Ziman Lai, et al. Plasma ET-1, no and aldosterone levels in patients with renal hypertension and their correlation with renal function and blood pressure. *PLA Journal of medicine*. 2021; 33:54–58.
11. Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and validation of machine learning models in prediction of remission in patients with moderate to severe crohn disease. *JAMA Network Open*. 2019; 2: e193721.
12. Yafei Wu, Ya Fang. Application progress of machine learning methods in chronic disease research. *China health statistics*. 2020; 37:624–628.
13. Yansheng Li, Houwu Gong, Yichao Li, Mingliang Su. Research on disease risk prediction based on real world data. *Medical information*. 2020; 33:17–19.
14. Yuehua Hu, Shicheng Yu, Xiao Qi, Wenjing Zheng, Qiqi Wang, Hongyan Mei. Multiple linear regression model and its application. *Chinese Journal of preventive medicine*. 2019; 6:653–656.
15. Cravedi P, Remuzzi G. Pathophysiology of proteinuria and its value as an outcome measure in CKD. *British Journal of Clinical Pharmacology*. 2013; 76:516–23.
16. Shuang Li, Wei Huang, Xiaodong Han, Yajun Chang, Zhongkai Yan, Xiaoning Cao. Clinical study of Yiqi Huoxue Recipe on urinary protein of chronic kidney disease. *Chinese Journal of traditional Chinese medicine*. 2021:1–9.
17. Wuqi Shi, Zhiya Zhang, Bing Li. Application of serum creatinine and cystatin C in the diagnosis of chronic kidney disease and evaluation of renal function injury. *PLA Journal of medicine*. 2017; 29:89–92.
18. YuLi L, IChen C, HungHsiang L, ChihHsien W, YuHsien L, ChiuHuang K, et al. Serum indices based on creatinine and cystatin C predict mortality in patients with non-dialysis chronic kidney disease. *Scientific reports*. 2021; 11:16863.

19. McClelland RL, Jorgensen NW, Budoff M, Blaha MJ, Post WS, Kronmal RA, et al. 10-year coronary heart disease risk prediction using coronary artery calcium and traditional risk factors. *Journal of the American College of Cardiology*. 2015; 66:1643–53.
20. Chinese guidelines for the prevention and treatment of hypertension (2018 Revision). *Chinese Journal of Cardiology*. 2019; 24:24–56
21. Mingxiang Weng, Chunqin Yang, Min Huang. Correlation between serum omentin-1, PAPP-A, vWF levels and AS in patients with renal hypertension during hemodialysis. *Modern practical medicine*. 2021; 33:471–472.
22. Lizhen Liu. Effects of ambovide combined with amlodipine besylate on urinary protein, serum creatinine and serum uric acid levels in patients with renal hypertension. *International medical and health Herald*. 2020; 26:404–405.
23. Ruggenenti P, Cravedi P, Remuzzi G. Mechanisms and treatment of CKD. *Journal of the American Society of Nephrology*. 2012; 23:1917–28.

Tables

Table1 Comparison of parameters before and after grid search optimization

Models	Default parameters	Optimal parameters
SVM	C=1.0 kernel='linear' probability=True	C = 10 kernel = 'linear' probability = True
RF	max_depth=None min_samples_leaf=1	max_depth = 9 min_samples_leaf = 2
XGboost	max_depth=6 reg_alpha=0 subsample=1 colsample_bytree=1	max_depth = 3 reg_alpha = 3 subsample = 1 colsample_bytree = 0.7
LightGBM	max_depth=-1 reg_alpha=0 subsample=1 colsample_bytree=1	max_depth=3 reg_alpha=2 subsample=0.1 colsample_bytree=0.8
GBDT	max_depth=3 subsample=1	max_depth=3 subsample=0.6
Catboost	max_depth=6 subsample=None iterations = 1000	max_depth=5 subsample=0.7 iterations = 100

Abbreviations: SVM, support vector machine model; RF, random forest model

Table 2. Comparison of clinical data between the two groups

Variable	Non-RH (n = 1172)	RH n = 400	Z/ χ^2	P
Age [years, M (P ₂₅ , P ₇₅)]	67.00 (55.00, 77.00)	53.50(42.00, 66.00)	-11.932	<0001
Sex [n (%)]			0.000	0.988
Female	480 (40.96%)	164 (41.00%)		
Male	692 (59.04%)	236 (59.0%)		
SBP [mmHg, M (P ₂₅ , P ₇₅)]	142.50(126.00, 161.00)	150.00(132.25,165.00)	-3.429	0.001
DBP [mmHg, M (P ₂₅ , P ₇₅)]	81.00 (70.00, 93.00)	88.00 (76.25, 99.00)	-6.675	<0001
PP [mmHg, M (P ₂₅ , P ₇₅)]	61.00 (48, 76)	60.00 (46.00, 73.00)	-1.351	0.177
BT [°C, M (P ₂₅ , P ₇₅)]	36.50 (36.30, 36.60)	36.50 (36.40, 36.60)	-1.189	0.234
Pulse [M (P ₂₅ , P ₇₅)]	84.00 (74.00, 94.00)	86.00 (78.00, 97.00)	-3.500	<0.001
Past medical history [n (%)]			4.636	0.031
No	180 (15.36%)	44 (11.00%)		
Yes	992 (84.64%)	356 (89.00%)		
Smoking status [n (%)]			0.922	0.337
No	799 (68.17%)	283 (70.75%)		
Yes	373 (31.83%)	117 (29.25%)		
Drinking status [n (%)]			10.999	0.001
No	878 (74.91%)	332 (83.00%)		
Yes	294 (25.09%)	68 (17.00%)		
Dialysis therapy [n (%)]			57.731	<0.001
No dialysis	1015 (86.60%)	280 (70.00%)		
Hemodialysis	145 (12.40%)	107 (26.75%)		
Peritoneal dialysis	12 (1.00%)	13 (3.25%)		
Diabetes [n (%)]			9.804	0.002
No	838 (71.50%)	318 (79.50%)		
Yes	334 (28.50%)	82 (20.50%)		
Urinary occult blood [n (%)]			301.687	<0.001

-	825 (70.40%)	112 (28.00%)		
+	182 (15.53%)	235 (58.75%)		
++	86 (7.33%)	33 (8.25%)		
+++	79 (6.74%)	20 (0.5%)		
Urine protein [n (%)]			413.637	<0.001
-	429 (36.60%)	23 (5.75%)		
+	374 (31.90%)	48 (12.00%)		
++	190 (16.20%)	271 (67.75%)		
+++	164 (14.00%)	55 (13.75%)		
++++	15 (1.30%)	3 (0.75%)		
CKD stage [n (%)]			140.328	<0.001
1~3 stage	476 (40.61%)	34 (8.50%)		
4~5 stage	696 (59.39%)	366 (91.50%)		
Length of stay [days, M (P ₂₅ , P ₇₅)]	9.00 (6.00, 14.00)	9.00 (9.00, 15.00)	-1.149	0.250
b ₂ microglobulin [mg/L, M (P ₂₅ , P ₇₅)]	15.93 (5.44, 15.93)	18.22 (15.38, 26.26)	-12.874	<0.001
Neutrophils [$\times 10^{12}$ /L, M (P ₂₅ , P ₇₅)]	4.72 (3.42, 6.20)	4.68 (3.53, 6.00)	-0.061	0.951
Urea [mmol/L, M (P ₂₅ , P ₇₅)]	13.60 (8.80, 20.68)	17.02 (13.22, 25.53)	-7.662	<0.001
Uric acid [μ mol/L, M (P ₂₅ , P ₇₅)]	425.90(334.19, 524.78)	421.91(330.13,504.86)	-0.802	0.422
Total cholesterol [mmol/L, M (P ₂₅ , P ₇₅)]	4.35 (3.65, 4.80)	4.35 (3.56, 4.35)	-2.434	0.015
Total bilirubin [μ mol/L, M (P ₂₅ , P ₇₅)]	7.81 (5.10, 10.85)	6.38 (4.49, 8.63)	-5.350	<0.001
Total protein [g/L, M (P ₂₅ , P ₇₅)]	65.87 (60.56, 71.20)	65.87 (60.35, 71.73)	-0.291	0.771
	2.40 (1.50, 3.50)	2.10 (1.30, 2.90)	-4.354	<0.001

Direct bilirubin [$\mu\text{mol/L}$, M (P ₂₅ , P ₇₅)]	6.02 (3.80, 7.72)	4.20 (2.83, 6.94)	-8.054	<0.001
Indirect bilirubin [$\mu\text{mol/L}$, M (P ₂₅ , P ₇₅)]	1.17 (0.79, 1.54)	0.98 (0.71, 1.31)	-5.333	<0.001
lymphocyte [$\times 10^{12}/\text{L}$, M (P ₂₅ , P ₇₅)]	37.30 (33.90, 41.10)	37.22 (34.03, 41.60)	-0.795	0.427
Albumin [g/L, M (P ₂₅ , P ₇₅)]	1.77 (1.13, 1.86)	1.77 (1.32, 1.77)	-1.476	0.140
Triglycerides [mmol/L, M (P ₂₅ , P ₇₅)]	236.30(127.68, 607.14)	608.85(438.94,879.00)	-13.773	<0.001
Creatinine [$\mu\text{mol/L}$, M (P ₂₅ , P ₇₅)]	3.41 (1.95, 4.99)	4.13 (4.03, 6.27)	-11.710	<0.001
Cystatin C [mg/L, M (P ₂₅ , P ₇₅)]	181.00(138.00, 221.75)	171.00(132.50,212.00)	-2.268	0.230
Platelet count [$\times 10^9 /\text{L}$, M (P ₂₅ , P ₇₅)]	140.10(137.81, 142.44)	139.29(137.10,141.50)	-4.129	<0.001
Sodium [mmol/L, M (P ₂₅ , P ₇₅)]	1.44 (1.10, 1.63)	1.53 (1.25, 1.98)	-5.748	<0.001
Phosphorus [mmol/L, M (P ₂₅ , P ₇₅)]	4.41 (3.92, 4.90)	4.51 (4.06, 5.03)	-2.914	0.004
Potassium [mmol/L, M (P ₂₅ , P ₇₅)]	2.18 (2.04, 2.32)	2.16 (1.98, 2.32)	-1.528	0.126
Calcium [mmol/L, M (P ₂₅ , P ₇₅)]	2.18 (2.04, 2.32)	15.33 (9.45, 29.09)	-3.274	0.001
ALT [U/L, M (P ₂₅ , P ₇₅)]	17.00 (11.60, 29.07)	17.50 (13.00, 26.00)	-4.718	<0.001
AST [U/L, M (P ₂₅ , P ₇₅)]	20.00 (15.00, 28.00)	4.73 (1.80, 6.53)	-6.074	<0.001
CK-MB [ng/L, M (P ₂₅ , P ₇₅)]	6.53 (2.30, 6.53)	10.06 (2.26, 22.00)	-2.073	0.042
CRP [mg/L, M (P ₂₅ , P ₇₅)]	17.66 (1.31, 22.00)	4.53 (3.23, 6.97)	-4.136	<0.001

Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; PP, pulse pressure; BT, temperature; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; CK-MB, creatine kinase myocardial band isoenzyme; NLR, neutrophil-to-lymphocyte ratio.

Table 3. Comparison of the predictive performance of each model in the test set

Model	Sensitivity	Specificity	AUC	F1 score
SVM	0.670	0.965	0.831	0.690
RF	0.790	0.965	0.924	0.820
XGBoost	0.820	0.945	0.935	0.840
LightGBM	0.830	0.948	0.932	0.850
GBDT	0.800	0.945	0.927	0.820
CatBoost	0.810	0.963	0.927	0.830

Abbreviations: SVM, support vector machine model; RF, random forest model

Figures

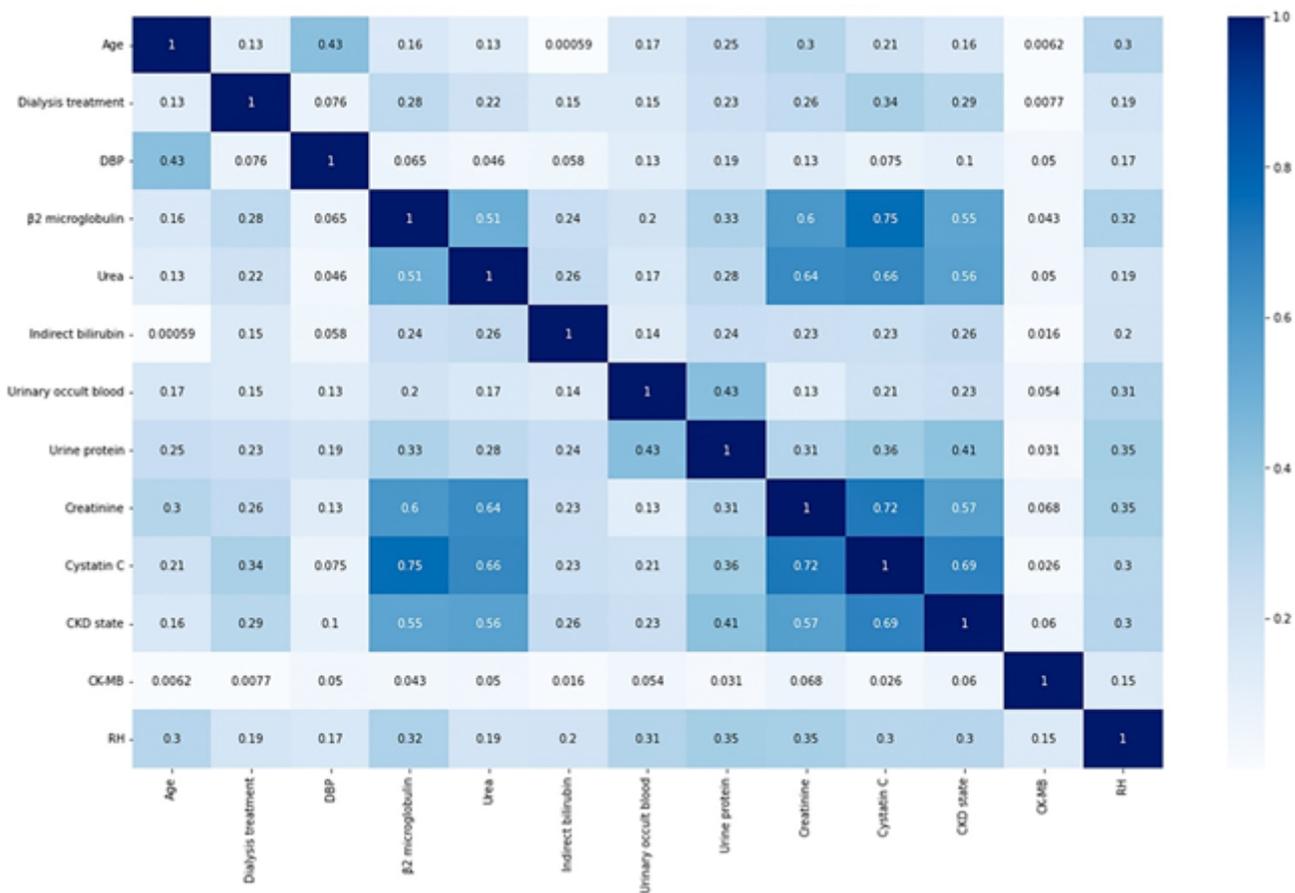


Figure 1

Variable correlation analysis

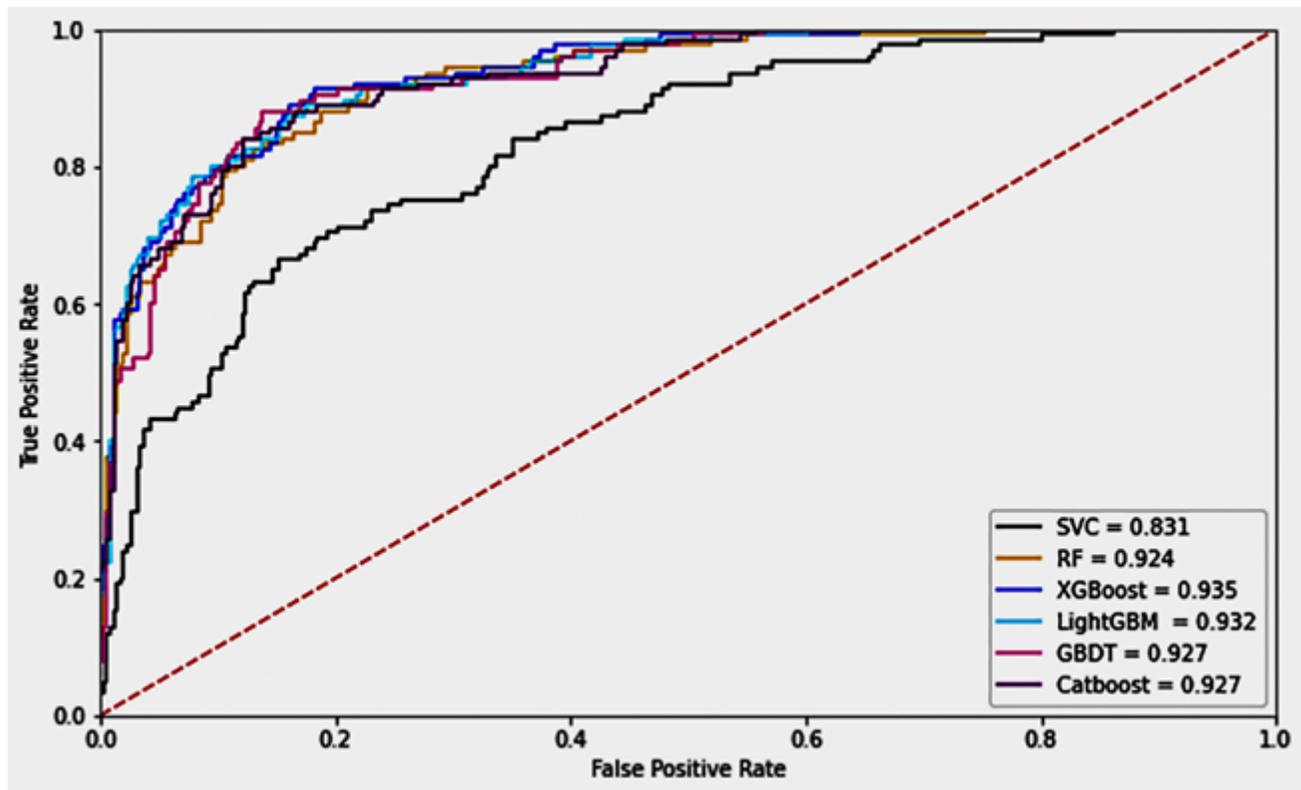


Figure 2

Relative operating characteristic curve analysis of each model in the test set

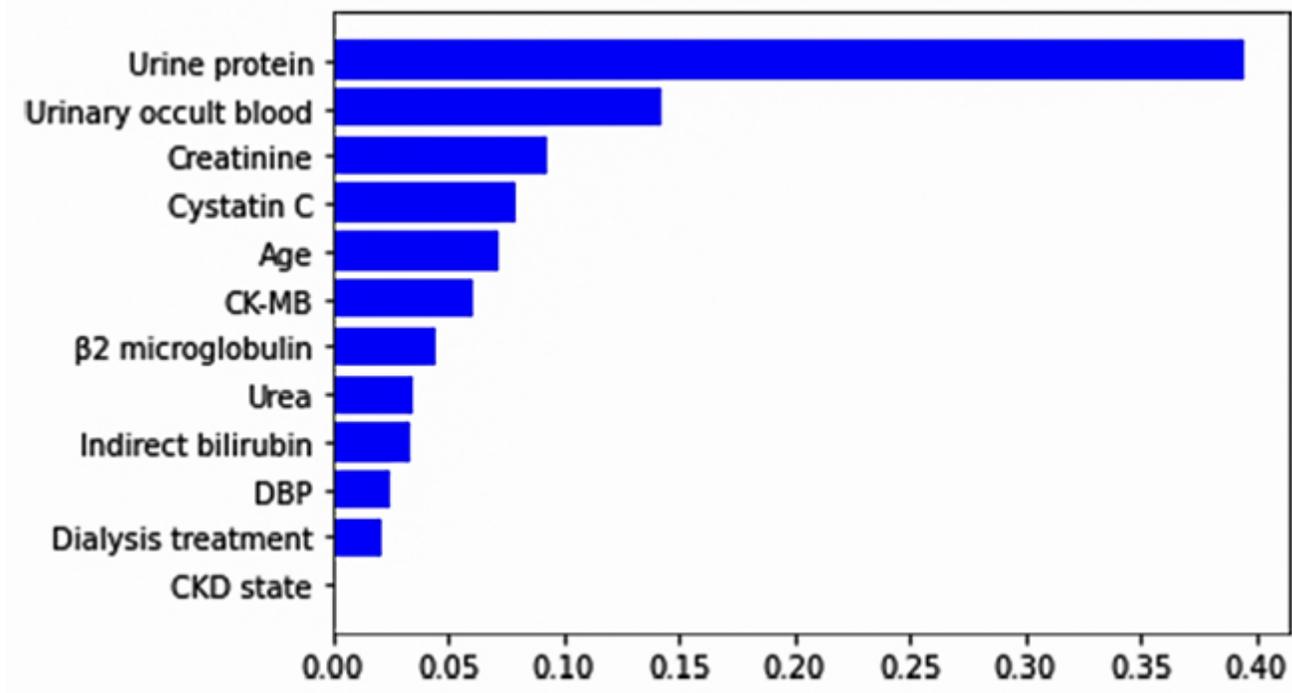


Figure 3

Importance ranking of XGBoost model variables

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [RawData.zip](#)