BMC Nephrology

RESEARCH

Open Access

Risk prediction of cardiovascular events in peritoneal dialysis patients

Liang Liu[1†], Liu Zhang[2†], Daohai Zhang[1], Tao Guan[1], Ting He[1], Bo Liang[1*] and Jinghong Zhao[1*]

Abstract

Background Cardiovascular events (CVEs), which refer to a spectrum of conditions including heart attacks, stroke and peripheral vascular disease, are the primary cause of death among peritoneal dialysis (PD) patients, accounting for nearly 40% of deaths. Early identification of high-risk individuals is essential to lessen this burden. Machine learning is particularly suited for this task due to its ability to discern complex, non-linear relationships between various clinical variables, which is essential for accurately predicting CVEs in the context of PD. Our study aimed to develop a predictive machine learning model to identify PD patients at risk of CVEs, offering healthcare providers a tool for proactive intervention.

Methods A total of 251 PD patients were enrolled in the study, with an additional 42 patients included for external validation. Initially, 37 variables were collected but reduced to 25 via Lasso regression. Six supervised machine learning algorithms were evaluated, and XGBoost was chosen as the optimal model based on AUC. Both internal and external validation confirmed the model's efficacy, and a web application was developed using the final XGBoost model, which utilized 12 selected variables.

Results Among the 251 patients, 40 (15.94%) developed CVEs. The XGBoost model demonstrated an AUC of 0.94 in 5-fold cross-validation. A simplified XGBoost model using 12 variables demonstrated robust prediction capabilities with an AUC of 0.88 in 5-fold cross-validation and 0.78 in external validation. The top five predictors of CVEs were age at catheterization, height, HDL, gender and hemoglobin. According to the SHAP summary plot, older age at catheterization, shorter height, male gender, higher serum HDL and lower hemoglobin levels correlated with increased CVEs risk in PD patients.

Conclusions The machine learning model, based on 12 key variables, offers an effective tool for predicting CVEs in PD patients, enabling early identification of high-risk cases. This model has been integrated into a web application.

Keywords Peritoneal dialysis, Cardiovascular events, Machine learning

[†]Liang Liu and Liu Zhang contributed equally to this work.

*Correspondence:
Bo Liang
liangbo1993@tmmu.edu.cn
Jinghong Zhao
zhaojh@tmmu.edu.cn

[1]Department of Nephrology, Chongqing Key Laboratory of Prevention and Treatment of Kidney Disease, Chongqing Clinical Research Center of Kidney and Urology Diseases, Xinqiao Hospital, Army Medical University (Third Military Medical University), Chongqing 400037, China
[2]Department of Nephrology, Chongqing General Hospital, Chongqing University, Chongqing 401147, China

Introduction

Peritoneal dialysis (PD), a widely used form of kidney replacement therapy, is essential for extending the lives of patients with end-stage renal disease worldwide [1]. While PD has made significant strides in treating chronic end-stage renal disease, enhancing patients' quality of life and extending their lifespan, those relying on long-term PD face challenges due to a range of complex and severe complications. These include peritonitis, catheter malfunction, infections, hernias, pleuroperitoneal leaks, poor ultrafiltration, and cardiovascular events (CVEs). These complications are diverse, persistent, and often progressive, significantly hindering patient recovery [2–5].

Among the many complications related to PD, cardiovascular pathology and CVEs stand out as major concerns. Advanced stages of chronic kidney disease, particularly stages G4 and G5, are characterized by a high prevalence of cardiovascular disease, affecting approximately 50% of patients [6]. This elevated cardiovascular risk is similarly reflected in PD patients, among whom CVEs account for nearly 40% of all deaths [7, 8]. CVEs significantly affect the long-term survival and quality of life for PD patients, impacting them both physically and mentally. Several factors contribute to cardiovascular issues in PD patients, including common risk factors such as alcoholism, smoking, obesity, hyperlipidemia, diabetes, and hypertension [9]. Hypertension is very common in PD patients, and it affects about 90% of PD patients [10]. Hypertension in PD patients often correlates with fluid overload [11]. Factors linked to end-stage renal disease include inflammation, malnutrition, protein-energy wasting, endothelial dysfunction, oxidative stress, calcification, and anemia [9]. Low mean corpusular hemoglobin concentration is an independent predictor of non-atherosclerotic cardiovascular disease in new dialysis patients [12]. Unique abnormal lipid metabolism and inflammation in the context of uremia are associated with the development of coronary artery disease [13]. Factors associated with peritoneal dialysis include advanced glycation end products, residual renal function, and ultrafiltration failure [9]. Residual renal function is inversely associated with left ventricular hypertrophy in peritoneal dialysis patients [14]. Biocompatible dialysate may reduce peritoneal membrane damage [15]. Given this multifactorial pathogenesis, identifying key risks is crucial.

In this study, we developed a machine learning model to predict CVEs risk in PD patients, aiming to enable early intervention.

Materials and methods

Study population

In this study, we used data from 251 PD patients to establish and internally validate the model. Inclusion criteria: patients on peritoneal dialysis regularly followed up at Xinqiao Hospital's Nephrology Department from January 2018 to December 2020; All patients had complete clinical data and biological samples, including medical history, physical exams, lab tests, imaging, and other relevant information; Age ≥ 18 years old. Exclusion criteria: patients who had a prior diagnosis of any condition classified as CVEs before the initiation of peritoneal dialysis, as defined in our study; patients who had been on peritoneal dialysis for less than 3 months. The ethical review board approved the study, and individual patient consent was waived due to data de-identification. An additional 42 PD patients served as an external validation group. The external validation cohort was selected from the same hospital, but a different time period, specifically from January to December 2017. The external cohort was representative of a different temporal population, which can help assess the model's robustness over time. Inclusion criteria for the external validation group were similar to those of the primary cohort. For the external validation group, we specifically selected 21 PD patients who experienced CVEs and another 21 PD patients who did not experience CVEs. This balanced approach helped ensure that the predictive performance of the model could be adequately evaluated across a range of outcomes.

Sample size

The sample size was determined based on both statistical considerations and practical clinical constraints. The traditional rule of 10 events per predictor variable (EPV) was originally developed for logistic regression models, and recent machine learning studies have shown that this rule may be overly conservative for modern predictive modeling approaches. Several recent methodological studies have demonstrated that machine learning models can achieve reliable performance with lower EPV ratios when appropriate validation and optimization strategies are employed [16, 17]. To ensure model reliability, we implemented comprehensive methodological strategies: First, we applied dimensionality reduction through LASSO regression, which effectively reduced the number of variables from 37 to 25. Subsequently, we further refined the model by reducing the variables to 12 by AUC increment curves, thereby optimizing the model's complexity for the available data. Additionally, we utilized the 5-fold cross-validation to assess and ensure stability of the model. Finally, we appropriately addressed class imbalance. Therefore, our sample size of 251 should be appropriate for machine learning approach.

Data collection and data pre-processing

Demographic characteristics, clinical conditions, clinical examination data and laboratory test data were collected from the electronic medical record system of Xinqiao

Hospital at the first peritoneal equilibration test after PD treatment. A total of 37 variables were included, including demographic data, laboratory test data, and peritoneal equilibration test related parameters. Data on CVEs that occurred during the entire duration of peritoneal dialysis treatment were collected. Two investigators (D.Z. and L.Z.) manually collected all data, with an overall missing rate of 2.2%. Missing values were imputed using multiple imputation [18], and density plots were utilized to compare data distributions pre- and post-imputation.

Definition of CVEs in PD patients

CVEs in PD patients are defined as coronary artery events, cerebrovascular diseases, peripheral arterial diseases, and other unexplained cardiovascular abnormalities that occur after the initiation of PD treatment. Specifically, these encompass myocardial infarction, acute coronary syndromes, congestive heart failure, both hemorrhagic and non-hemorrhagic cerebrovascular disease, and peripheral vascular conditions like arterial embolism, diabetic foot, and non-traumatic amputations [19–21]. PD patients received hospital follow-ups every six months and monthly outpatient checks until CVEs occurred, PD ceased, they died from any cause, or follow-up ended on December 31, 2023. For patients not attending hospital follow-ups, phone contact was used.

Machine learning

Machine learning is a subset of artificial intelligence that enables computers to learn from data and make predictions [22]. In this study, we used machine learning algorithms to identify PD patients who developed CVEs during treatment. The machine learning model underwent both internal and external validation. Eighty percent of the dataset served as the internal training set, with the remaining 20% used for internal validation. Additionally, 42 PD patients were included as the external validation set [23]. In 5-fold cross-validation, the entire dataset was randomly divided into five equal parts. The model was trained using four parts and tested on the remaining part, which served as the internal validation set. This process repeated five times, ensuring each part was used for validation once. The results from the five runs were averaged to obtain an overall assessment of the model's performance. This method improves evaluation accuracy and reduces the impact of random data partitioning [24]. The Synthetic Minority Over-sampling Technique (SMOTE) was used to address data class imbalance. Unlike simple over-sampling or under-sampling, SMOTE created synthetic minority class examples. This method selected minority class samples and generated new instances near them, preserving the original dataset's complexity and distribution, thereby enhancing the model's reliability when handling imbalanced data [25]. Lasso regression was used for variable selection. Specifically, L1 regularization introduces a penalty term to the loss function that is proportional to the absolute values of the variables, which effectively reduces the importance of less relevant variables. This causes many of the variable coefficients to tend towards zero, effectively performing variable selection. By compressing the regression coefficients of unnecessary variables to zero, it retains only the most impactful features, reducing model complexity and enhancing generalization capability [26]. We use the following six supervised machine learning methods to build predictive models: random forest (RF), Gradient Boosting Decision Tree (GBDT), extreme gradient boosting (XGBoost), ensemble learning (RF + XGBoost), categorical boosting (CatBoost) and Light Gradient Boosting Machine (LightGBM). To strike a balance between model performance, simplicity, and clinical interpretability, we ranked the 25 variables based on their importance and sequentially incorporated them into the model that yielded the highest area under the curve (AUC) value, monitoring the change in AUC with each variable's inclusion. As the number of variables in the model increased, the performance enhancement effect of each additional variable showed a diminishing trend. Based on these observations, we constrained the number of variables in the model to 12, thereby constructing a prediction model that is both efficient and concise.

To evaluate the model's effectiveness, we utilized the area under the receiver operating characteristic (ROC) curve. For interpretability, we employed the ROC curve, a feature importance map, SHapley Additive exPlanations (SHAP) plots, and other graphical aids. SHAP plots provided insights into the prediction mechanisms by illustrating the impact of individual features and elucidating the model's decision process. SHAP summary plots offered an overview of feature importance, while SHAP dependence plots highlighted relationships between features. After external validation, we developed a web application for the predictive model using Streamlit, an open-source framework for creating and sharing data science applications. All data analyses were conducted using Python (version 3.11.5) and R (version 4.3.3).

Results

Study population and study design

The overall workflow of the study and patient information are illustrated in Fig. 1. Specifically, the study included 251 PD patients. Among these, 40 (15.94%) PD patients experienced CVEs during PD, while 211 (84.06%) did not. Variables showing significant differences between the two groups included age at catheterization, albumin levels, and prealbumin levels (Table 1). To address group imbalance, SMOTE over-sampling was applied, followed by Lasso regression to select 25 variables from an initial

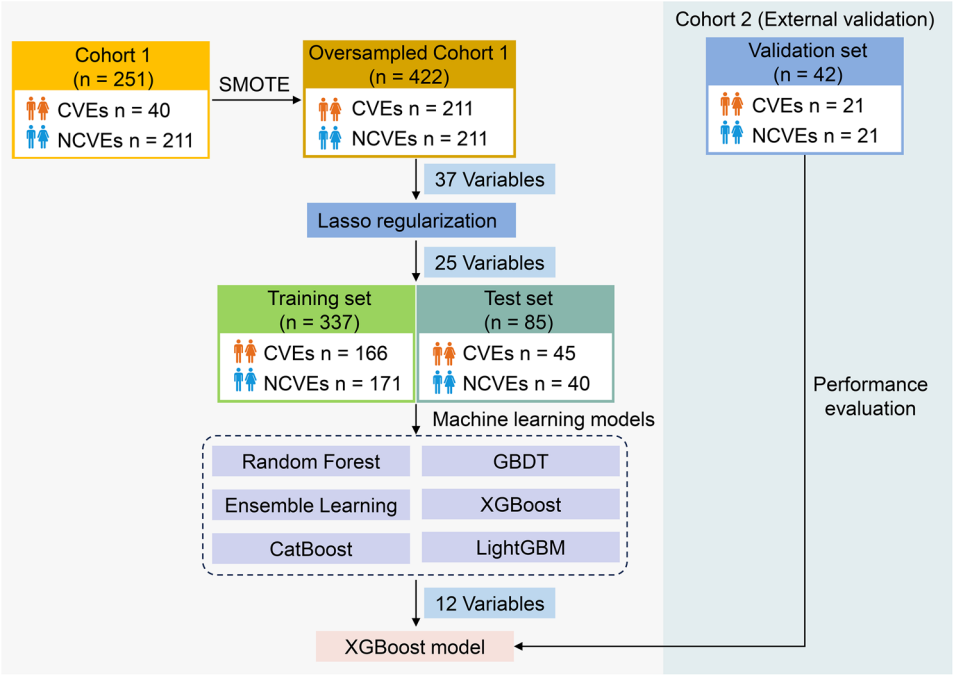

Fig. 1 Study schematic. In this study, 251 PD patients were included, and SMOTE was applied to address data imbalance. Lasso regression identified 25 variables from 37 in the oversampled Cohort 1. Further refinement led to an XGBoost model with 12 variables. Cohort 2 served for external validation. Abbreviations: CVEs, cardiovascular events; NCVEs, non-cardiovascular events; SMOTE, Synthetic Minority Over-sampling Technique; GBDT, Gradient Boosting Decision Tree; XGBoost, extreme gradient boosting; CatBoost, categorical boosting; LightGBM, Light Gradient Boosting Machine

set of 37 (Supplementary Figure 1). These 25 variables were used in six machine learning algorithms to examine the association between variables and CVEs. Eighty percent of the PD patients were randomly assigned to the training set, and the remaining 20% were allocated for the validation set (Supplementary Table 1). XGBoost achieved the best prediction performance among the six models. The variables were then included in the XGBoost model one by one based on their importance. Ultimately, a CVEs risk prediction model incorporating 12 variables was developed. Additionally, an external validation set was used to confirm the model's robustness.

Establishment and selection of machine learning models

We employed six machine learning methods, and XGBoost achieved the best AUC of 0.96 among them (Fig. 2A). Consequently, we selected the XGBoost model for our subsequent analyses. The XGBoost model achieved a threshold of 0.61, a Youden index of 78.33%, sensitivity of 93.33%, specificity of 85.00%, an F1 score of 0.88, a positive predictive value of 0.84, a negative predictive value of 0.91, and an accuracy of 0.87 (Supplementary Table 2). Figure 2B shows the 5-fold cross-validation average AUC value of XGBoost model, which is 0.940 ± 0.03. Figure 2C shows the ranking of feature importance for all 25 variables in the XGBoost model. Subsequently, we sequentially entered all 25 variables into the XGBoost model in order of feature importance and showed the growth of the model AUC value. It was found that the model could achieve good prediction performance after inputting the 12th variable (Fig. 2D). In order to further simplify the model and improve the generalization, we use these 12 variables to re-establish the XGBoost model.

XGBoost model based on 12 variables

We established a new XGBoost model based on the previously selected 12 variables, achieving an AUC value of 0.91 for the ROC curve (Fig. 3A). A 5-fold cross-validation of the XGBoost model yielded an average AUC of 0.88 ± 0.03 (Fig. 3B). Figure 3C demonstrated that both the training error and the test error gradually decreased and plateaued, indicating steady learning by the model. The SHAP summary plot revealed the importance of each feature in the model prediction and its influence trend (Fig. 3D). The top five contributing factors were age at catheterization, height, HDL, gender, and hemoglobin. Older age at catheterization, shorter height, male gender, higher serum HDL, and lower hemoglobin were associated with a higher risk of CVEs. We validated the XGBoost model using an external validation set of 42 patients, with an AUC value of 0.78 (Fig. 3E). Figure 3F displayed the confusion matrix for the model's external validation.

Table 1 Patient characteristics

	Overall	non-CVEs	CVEs	*P* value
Number of patients, n	251	211 (84.06%)	40 (15.94%)	
Duration of PD (month)	21.00 [11.00, 35.00]	21.00 [11.00, 35.00]	19.50 [11.75, 36.25]	0.867
Age at catheterization (years)	38.29 (11.21)	37.34 (10.84)	43.33 (11.91)	0.002
Female, n (%)	77 (30.7)	64 (30.3)	13 (32.5)	0.852
Primary disease, n (%)				0.217
primary glomerular disease	204 (81.3)	175 (82.9)	29 (72.5)	
hypertensive nephropathy	22 (8.8)	18 (8.5)	4 (10.0)	
diabetic nephropathy	9 (3.6)	6 (2.8)	3 (7.5)	
Other	16 (6.4)	12 (5.7)	4 (10.0)	
Diabetes history, n (%)	14 (5.6)	9 (4.3)	5 (12.5)	0.053
Height (cm)	165.47 (7.38)	165.76 (7.28)	163.93 (7.79)	0.149
Weight (kg)	60.85 (11.02)	60.77 (10.91)	61.26 (11.68)	0.801
BMI (kg/m^2)	22.09 (3.16)	22.01 (3.16)	22.54 (3.13)	0.326
Marital status, n (%)				0.352
unmarried	40 (15.9)	36 (17.1)	4 (10.0)	
Married	203 (80.9)	169 (80.1)	34 (85.0)	
Divorce	8 (3.2)	6 (2.8)	2 (5.0)	
PD type, n (%)				0.350
CAPD	212 (84.5)	176 (83.4)	36 (90.0)	
APD	39 (15.5)	35 (16.6)	4 (10.0)	
SBP (mmHg)	148.42 (18.74)	148.08 (18.66)	150.25 (19.31)	0.502
DBP (mmHg)	96.83 (13.34)	97.04 (13.06)	95.72 (14.86)	0.569
Residual urine volume (ml/d)	869.67 (473.08)	874.91 (467.73)	842.02 (505.70)	0.688
Residual urine volume < 400 mL/d, n (%)	44 (17.5)	36 (17.1)	8 (20.0)	0.653
Albumin (g/L)	39.90 [36.85, 42.60]	40.20 [37.40, 42.95]	38.60 [35.43, 40.20]	0.003
Prealbumin (mg/L)	398.00 [348.00, 447.00]	408.00 [356.50, 451.50]	360.30 [313.70, 423.50]	0.010
Urea (mmol/L)	18.12 (5.46)	18.13 (5.48)	18.08 (5.45)	0.96
Creatinine (umol/L)	886.46 (285.59)	894.39 (284.91)	844.61 (289.15)	0.313
EGFR ($ml/min/1.73\ m^2$)	6.21 (2.22)	6.16 (2.19)	6.46 (2.39)	0.432
Uric acid (umol/L)	409.38 (92.17)	408.67 (90.04)	413.12 (103.88)	0.780
Calcium (mmol/L)	2.26 (0.22)	2.26 (0.22)	2.26 (0.22)	0.884
Phosphorus (mmol/L)	1.52 (0.39)	1.52 (0.40)	1.51 (0.34)	0.818
Hemoglobin (g/L)	110.98 (20.60)	111.68 (20.76)	107.33 (19.56)	0.221
Blood glucose (mmol/L)	5.02 (0.89)	4.99 (0.88)	5.20 (0.92)	0.166
Triglyceride (mmol/L)	1.73 (1.01)	1.78 (1.06)	1.46 (0.61)	0.070
Total cholesterol (mmol/L)	4.85 (1.11)	4.88 (1.14)	4.73 (0.96)	0.460
HDL (mmol/L)	1.20 (0.32)	1.18 (0.32)	1.29 (0.36)	0.052
LDL (mmol/L)	2.84 (0.86)	2.86 (0.90)	2.76 (0.63)	0.535
PTH (pg/ml)	402.62 (304.29)	410.00 (311.53)	363.68 (262.91)	0.378
Total Kt/V	2.20 (0.59)	2.18 (0.51)	2.33 (0.91)	0.159
Residual kidney Kt/V	0.68 (0.50)	0.68 (0.43)	0.72 (0.77)	0.601
Dialysis Kt/V	1.55 (0.38)	1.54 (0.36)	1.59 (0.48)	0.434
Total Ccr (mL/min)	73.88 (23.79)	73.17 (21.11)	77.65 (34.78)	0.275
Residual kidney Ccr (mL/min)	33.92 (23.74)	33.55 (21.32)	35.91 (34.06)	0.565
Dialysis Ccr (mL/min)	39.65 (7.95)	39.29 (7.91)	41.57 (7.98)	0.096
PET, n (%)				0.699
High transport	25 (10.0)	21 (10.0)	4 (10.0)	
High average transport	110 (43.8)	89 (42.2)	21 (52.5)	
Low average transport	96 (38.2)	83 (39.3)	13 (32.5)	

Table 1 (continued)

	Overall	non-CVEs	CVEs	*P* value
Low transport	20 (8.0)	18 (8.5)	2 (5.0)	
Peritonitis, n (%)	84 (33.5)	67 (31.8)	17 (42.5)	0.203

Continuous variables that conform to a normal distribution are expressed as mean (standard deviation), continuous variables that are not normally distributed are expressed as median [interquartile spacing], and categorical variables are expressed as number (%). Abbreviations: CVEs, cardiovascular events; PD, Peritoneal Dialysis; BMI, body mass index; CAPD, Continuous Ambulatory Peritoneal Dialysis; APD, Automated Peritoneal Dialysis; SBP, systolic blood pressure; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HDL, high-density lipoprotein; LDL, low-density lipoprotein; PTH, Parathyroid Hormone; Kt/V, Urea Reduction Ratio; Ccr, Creatinine Clearance; PET, Peritoneal Equilibration Test

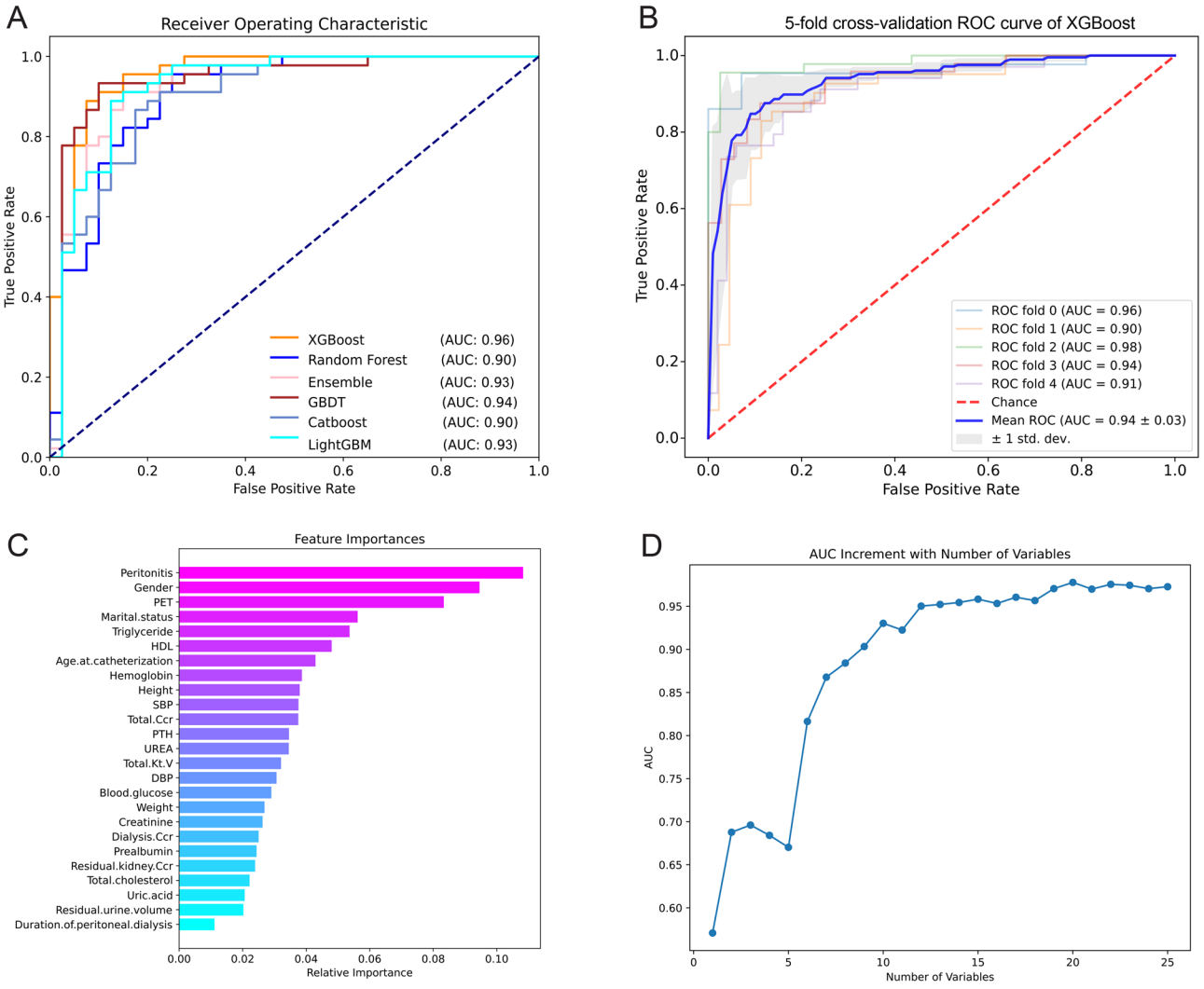

Fig. 2 Development of machine-learning models and selection of variables. **A** ROC plot of 6 machine learning; **B** 5-fold cross-validation of XGBoost; **C** Feature importance plot of XGBoost; **D** AUC increment plot of XGBoost. Abbreviations: XGBoost, extreme gradient boosting; CatBoost, categorical boosting; GBDT, Gradient Boosting Decision Tree; LightGBM, Light Gradient Boosting Machine; ROC, Receiver Operating Characteristic; AUC, Area Under the Curve; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL, high-density lipoprotein; PTH, Parathyroid Hormone; Kt/V, Urea Reduction Ratio; Ccr, Creatinine Clearance; PET, Peritoneal Equilibration Test

Web-app for predicting the risk of CVEs in PD patients

We built a web application using the XGBoost model with 12 variables and posted it online, enabling doctors and nurses to evaluate the risk of CVEs in PD patients (Supplementary Figure 2). The website is as follows: https://pd-cve-prediction-tool.streamlit.app/.

Discussion

In this study, we developed and validated a machine learning model to predict the risk of CVEs in PD patients. Using data from 251 patients, we applied six machine learning algorithms. The XGBoost model showed the highest predictive accuracy, achieving an average AUC of

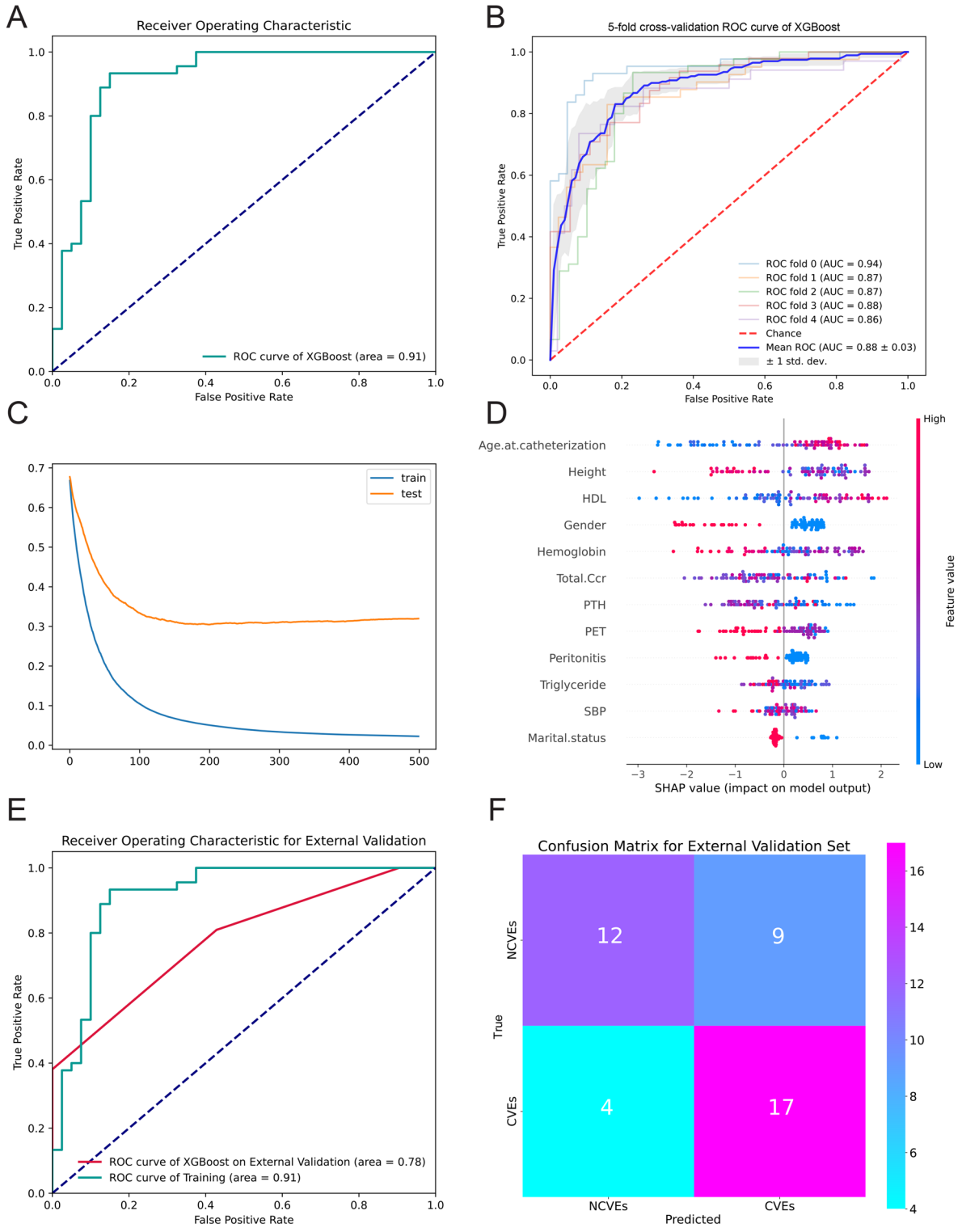

Fig. 3 XGBoost model with 12 variables. **A** ROC plot; **B** 5-fold cross-validation; **C** Iterative learning curve; **D** SHAP summary plot. Each point represents a sample; its color indicates the magnitude of the feature value, and its position on the X-axis shows the SHAP value. The SHAP value signifies the magnitude and direction of the feature's influence on the model prediction; **E** ROC plots containing external validation; **F**. Confusion matrix plot. Abbreviations: ROC, Receiver Operating Characteristic; AUC, Area Under the Curve; SBP, systolic blood pressure; HDL, high-density lipoprotein; PTH, Parathyroid Hormone; Ccr, Creatinine Clearance; PET, Peritoneal Equilibration Test

0.94 in 5-fold cross-validation. We further simplified the model to include 12 variables without sacrificing predictive performance, maintaining an average AUC of 0.88 in 5-fold cross-validation and an AUC of 0.78 in external validation. Additionally, we created a web-based tool to facilitate real-time CVE risk assessment in clinical practice. These results suggest that machine learning can be a valuable tool for predicting CVEs in PD patients, potentially improving patient care through earlier risk identification and intervention.

The XGBoost model, based on 12 variables, identified age at catheterization, height, HDL, gender, and hemoglobin as the top five predictive factors. Among these factors, age, gender, and height at catheterization are non-modifiable. The SHAP summary plot indicates that older age at catheterization, shorter height, and male gender are linked to a higher risk of CVEs according to our model. Age has been recognized as a traditional factor contributing to CVEs [27]. Roderburg et al. analyzed 657,310 outpatients in Germany and found that lower height was associated with a higher prevalence of hypertension, coronary heart disease, heart failure, and ischemic stroke, regardless of gender [28]. This population-level evidence suggests height may serve as a possible risk stratification tool for CVEs, which may also be useful in PD patients. Our research indicates that male patients undergoing peritoneal dialysis have a higher risk of CVEs compared to females, consistent with the trend of gender impact on CVEs risk in the general population. A review article delved into the role of gender in CVEs risk, highlighting how gender may increase CVEs risk in men by influencing traditional cardiovascular risk factors such as hypertension, smoking, diabetes, physical inactivity, and obesity [29]. Dev R et al. analyzed gender differences in cardiovascular health in six South Asian countries and found that women had better cardiovascular health. Sex-related factors such as marital status and family size are associated with poorer cardiovascular health and higher CVEs risk, and the effect is more pronounced among men [30]. Our study suggests that HDL can be used as a tool to predict the risk of CVEs in PD patients. The SHAP summary plot revealed that higher HDL levels correlated with an increased risk of CVEs, challenging the conventional wisdom that higher HDL levels are protective. Current studies suggest a U-shaped relationship between HDL levels and CVEs mortality, where both low and high levels of HDL can increase the risk [31–33]. In PD patients, HDL often exhibits dysfunction, shifting towards a more proinflammatory phenotype [34], which correlates with elevated mortality risk [35]. The SHAP summary plot suggested that lower hemoglobin would increase the risk of CVEs in PD patients. Anemia is associated with risk for coronary artery disease in a Mendelian randomization study [36]. Previous studies have shown that recombinant erythropoietin can reduce left ventricular mass index and improve left ventricular hypertrophy by correcting anemia in patients with severe renal failure [37].

This study possesses several strengths. It utilized a comprehensive dataset derived from actual PD patients, ensuring the results are clinically applicable. The study incorporated standard clinical examination metrics as variables, resulting in an XGBoost model incorporating 12 clinically relevant factors, thereby simplifying the model without compromising predictive accuracy. The model's robustness was further confirmed through external validation. A web application was developed to facilitate real-time clinical deployment, integrating predictions into routine care for early identification of high-risk individuals and targeted interventions. This approach may improve patient outcomes by reducing adverse CVEs. Despite these strengths, the study does have limitations. It was conducted within a single institution with a relatively limited external validation, which may limit generalizability. The observed AUC value was lower in the external validation, possibly due to the small sample size amplifying outliers' effects. The external validation cohort, from January to December 2017, represents a different temporal population, which could affect model performance. Despite the drop in AUC, the external validation still demonstrates acceptable discriminatory power, suggesting that the model can generalize reasonably well to similar populations. Further validation in larger, more diverse cohorts is needed to fully assess the robustness and generalizability of the model. Additionally, unmeasured variables or confounders may influence the model's predictions, and the web application's integration into clinical practices requires further validation.

Conclusion

Our machine learning model provides a practical tool for CVE risk stratification in PD patients using routinely collected parameters. Clinical implementation could enable targeted interventions for high-risk individuals, though broader validation remains essential.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12882-025-04091-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

None.

Author contributions
J.Z. and L.L. designed the study. D.Z. and L.Z. conducted data collection. L.L. and B.L. performed the data analysis. L.L. and T.G. drafted the manuscript. J.Z. and B.L. revised it. L.L. and L.Z. contributed equally to this research.

Funding
This study was supported by the Joint Fund of National Natural Science Foundation of China (Grant No. U22A20279) and National Natural Science Foundation of China (Grant No. 82100783).

Data availability
Due to ethical considerations and participant confidentiality, requests for access to the raw data should be directed to the corresponding author.

Declarations

Ethics approval and consent to participate
Prior to the commencement of this study, approval was granted by the Medical Ethics Committee of the Second Affiliated Hospital of the Army Medical University (Approval No. 2024-140-01), confirming that the risk to patients would not exceed a minimal level. As a result, the Medical Ethics Committee of the Second Affiliated Hospital of the Army Medical University gave waiver for informed consent. Additionally, it was confirmed that participation or non-participation would not negatively impact the personal interests of the individuals involved. Our study was conducted in strict accordance with the ethical principles outlined in the Declaration of Helsinki, ensuring that all aspects of the research were compliant with international standards for medical research involving human participants.

Consent for publication
Not applicable.

Clinical trial number
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 November 2024 / Accepted: 24 March 2025
Published online: 05 April 2025

References

1. Himmelfarb J, Vanholder R, Mehrotra R, Tonelli M. The current and future landscape of dialysis. Nat Rev Nephrol. 2020;16(10):573–85. https://doi.org/10.1038/s41581-020-0315-4.
2. Li PK, Chow KM, Cho Y, et al. ISPD peritonitis guideline recommendations: 2022 update on prevention and treatment. Peritoneal Dial Int. 2022;42(2):110–53. https://doi.org/10.1177/08968608221080586.
3. Cheng XBJ, Bargman J. Complications of peritoneal dialysis part II: nonmechanical complications. Clin J Am Soc Nephrol. 2024;19(6):791–99. https://doi.org/10.2215/CJN.0000000000000418.
4. Scalamogna A, Nardelli L, Cicero E, Castellano G. Analysis of mechanical complications in urgent-start peritoneal dialysis. J Nephrol. 2022;35(5):1489–96. https://doi.org/10.1007/s40620-022-01294-0.
5. Lluesa JH, Lopez-Romero LC, Monzo JJB, et al. Lipidic profiles of patients starting peritoneal dialysis suggest an increased cardiovascular risk beyond classical dyslipidemia biomarkers. Sci Rep. 2022;12(1):16394. https://doi.org/10.1038/s41598-022-20757-9.
6. Stevens PE, O'Donoghue DJ, de Lusignan S, et al. Chronic kidney disease management in the United Kingdom: NEOERICA project results. Kidney Int. 2007;72(1):92–99. https://doi.org/10.1038/sj.ki.5002273.
7. Mehrotra R, Devuyst O, Davies SJ, Johnson DW. The current state of peritoneal dialysis. J Am Soc Nephrol. 2016;27(11):3238–52. https://doi.org/10.1681/ASN.2016010112.
8. Modi ZJ, Lu Y, Ji N, et al. Risk of cardiovascular disease and mortality in young adults with end-stage renal disease: an analysis of the US renal data system. JAMA Cardiol. 2019;4(4):353–62. https://doi.org/10.1001/jamacardio.2019.0375.
9. Krediet RT, Balafa O. Cardiovascular risk in the peritoneal dialysis patient. Nat Rev Nephrol. 2010;62010(8):451–60. https://doi.org/10.1038/nrneph.2010.68.
10. Vaios V, Georgianos PI, Liakopoulos V, Agarwal R. Assessment and management of hypertension among patients on peritoneal dialysis. Clin J Am Soc Nephrol. 2019;14(2):297–305. https://doi.org/10.2215/CJN.07480618.
11. Chen W, Cheng LT, Wang T. Salt and fluid intake in the development of hypertension in peritoneal dialysis patients. Renal Failure. 2007;29(4):427–32. https://doi.org/10.1080/08860220701260461.
12. Yamaguchi S, Hamano T, Oka T, et al. Mean corpuscular hemoglobin concentration: an anemia parameter predicting cardiovascular disease in incident dialysis patients. J Nephrol. 2022;35(2):535–44. https://doi.org/10.1007/s40620-021-01107-w.
13. Stompor T, Pasowicz M, Sullowicz W, et al. An association between coronary artery calcification score, lipid profile, and selected markers of chronic inflammation in ESRD patients treated with peritoneal dialysis. Am J Kidney Dis. 2003;41(1):203–11. https://doi.org/10.1053/ajkd.2003.50005.
14. Wang AY, Wang M, Woo J, et al. A novel association between residual renal function and left ventricular hypertrophy in peritoneal dialysis patients. Kidney Int. 2002;62(2):639–47. https://doi.org/10.1046/j.1523-1755.2002.00471.x.
15. van Westrhenen R, Zweers MM, Kunne C, de Waart DR, van der Wal AC, Krediet RT.A pyruvate-buffered dialysis fluid induces less peritoneal angiogenesis and fibrosis than a conventional solution. Peritoneal Dial Int. 2008;28(5):487–96.
16. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. BMC Bioinf. 2023;24(1):48. https://doi.org/10.1186/s12859-023-05156-9.
17. Infante G, Miceli R, Ambrogi F. Sample size and predictive performance of machine learning methods with survival data: a simulation study. Stat Med. 2023;42(30):5657–75. https://doi.org/10.1002/sim.9931.
18. Blazek K, van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. Kidney Int. 2021;99(1):68–74. https://doi.org/10.1016/j.kint.2020.07.035.
19. Regan JA, Mentz RJ, Nguyen M, et al. Mitochondrial metabolites predict adverse cardiovascular events in individuals with diabetes. JCI Insight. 2023;8(17). https://doi.org/10.1172/jci.insight.168563.
20. Zhang L, Guan X, Liu L, Huang Y, Xiong J, Zhao J. Risk factors and outcomes in patients who switched from peritoneal dialysis to physician-oriented or patient-oriented kidney replacement therapy. Renal Failure. 2024;46(1):2337286. https://doi.org/10.1080/0886022X.2024.2337286.
21. Dai L, Yang Y, Liu L, et al. The association of left ventricular fraction shortening with cardiovascular events in peritoneal dialysis patients. Renal Failure. 2023;45(2):2261786. https://doi.org/10.1080/0886022X.2023.2261786.
22. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40–55. https://doi.org/10.1038/s41580-021-00407-0.
23. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14(1):49–58. https://doi.org/10.1093/ckj/sfaa188.
24. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. JAMA Psychiatry. 2020;77(5):534–40. https://doi.org/10.1001/jamapsychiatry.2019.3671.
25. Fernández A, García S, Herrera F, Chawla NV.SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Int Res. 2018;61(1):863–905.
26. Tibshirani R.Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Ser B (Stat Methodol). 2011;73(3):273–82.
27. Cozzolino M, Mangano M, Stucchi A, Ciceri P, Conte F, Galassi A. Cardiovascular disease in dialysis patients. Nephrol Dial Transplant. 2018;33(suppl_3):iii28–iii34. https://doi.org/10.1093/ndt/gfy174.
28. Krieg S, Kostev K, Luedde M, et al. The association between the body height and cardiovascular diseases: a retrospective analysis of 657,310 outpatients in Germany. Eur J Med Res. 2022;27(1):240. https://doi.org/10.1186/s40001-022-00881-y.
29. Connelly PJ, Azizi Z, Alipour P, Delles C, Pilote L, Raparelli V. The importance of gender to understand sex differences in cardiovascular disease. Can J Cardiol. 2021;37(5):699–710. https://doi.org/10.1016/j.cjca.2021.02.005.
30. Dev R, Raparelli V, Pilote L, et al. Cardiovascular health through a sex and gender lens in six South Asian countries: findings from the WHO STEPS surveillance. J Glob Health. 2022;12:04020.
31. Lu J, Han G, Liu X, et al. Association of high-density lipoprotein cholesterol with all-cause and cause-specific mortality in a Chinese population of 3.3

million adults: a prospective cohort study. Lancet Reg Health West Pac. 2024;42:100874.
32. Endo Y, Fujita M, Ikewaki K. HDL functions-current status and future perspectives. Biomolecules. 2023;13(1). https://doi.org/10.3390/biom13010105.
33. Moradi H, Streja E, Kashyap ML, Vaziri ND, Fonarow GC, Kalantar-Zadeh K. Elevated high-density lipoprotein cholesterol and cardiovascular mortality in maintenance hemodialysis patients. Nephrol Dial Transplant. 2014;29(8):1554–62. https://doi.org/10.1093/ndt/gfu022.
34. Holzer M, Schilcher G, Curcic S, et al. Dialysis modalities and HDL composition and function. J Am Soc Nephrol. 2015;26(9):2267–76. https://doi.org/10.1681/ASN.2014030309.
35. Sanguankeo A, Upala S. Metabolic syndrome increases mortality risk in dialysis patients: a systematic review and meta-analysis. Int J Endocrinol Metab. 2018;16(2):e61201. https://doi.org/10.5812/ijem.61201.
36. Gan T, Hu J, Liu W, et al. Causal association between anemia and cardiovascular disease: a 2-sample bidirectional mendelian randomization study. J Am Heart Assoc. 2023;12(12):e029689. https://doi.org/10.1161/JAHA.123.029689.
37. Ayus JC, Go AS, Valderrabano F, et al. Effects of erythropoietin on left ventricular hypertrophy in adults with severe chronic renal failure and hemoglobin <10 g/dL. Kidney Int. 2005;68(2):788–95. https://doi.org/10.1111/j.1523-1755.2005.00458.x.

Publisher's Note