

Original Paper

# Predicting End-Stage Renal Disease and Mortality in Chronic Kidney Disease Using Machine Learning: Retrospective Cohort Study

Tz-Heng Chen<sup>1,2,3</sup>, MD; Kuan-Hsun Lin<sup>4,5</sup>; Yang Ho<sup>1,2</sup>, MD; Wei-Cheng Tseng<sup>1,2</sup>, MD, PhD; Yuan-Chia Chu<sup>4,5,6\*</sup>, PhD; Der-Cherng Tarn<sup>2,7\*</sup>, MD, PhD

<sup>1</sup>Division of Nephrology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>2</sup>School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

<sup>3</sup>Institute of Emergency and Critical Care Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

<sup>4</sup>Department of Information Management, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>5</sup>Department of Information Management, National Taipei University of Nursing and Health Science, Taipei, Taiwan

<sup>6</sup>Big Data Center, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>7</sup>Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

\*these authors contributed equally

## Corresponding Author:

Yuan-Chia Chu, PhD

Department of Information Management

Taipei Veterans General Hospital

No 201, Sec 2, Shipai Rd, Beitou District

Taipei 11217

Taiwan

Phone: +886 986-680623

Fax: 886 2-28757841

Email: [xd.yuanchia@gmail.com](mailto:xd.yuanchia@gmail.com)

## Abstract

**Background:** Chronic kidney disease (CKD) is a global health burden characterized by heterogeneous progression trajectories. Without timely and appropriate management, CKD can lead to increased morbidity and mortality and a reduced quality of life. Therefore, early identification of patients at high risk of developing end-stage renal disease (ESRD) or mortality is essential to facilitate timely intervention and improve patient outcomes.

**Objective:** This study aimed to develop and validate machine learning models to predict ESRD and all-cause mortality in patients with CKD.

**Methods:** We developed and validated machine learning models using data from patients with CKD and an estimated glomerular filtration rate of  $<60$  mL/min/1.73 m<sup>2</sup>, who were treated at Taipei Veterans General Hospital between 2011 and 2021. Predictors included 69 routinely available demographic, clinical, medication, laboratory, and echocardiographic variables. The outcomes were ESRD and all-cause mortality. The cohort was randomly divided into training (n=23,741, 80%) and testing (n=5936, 20%) sets. The evaluated models included extreme gradient boosting, light gradient boosting machine, categorical boosting, random forest, and a stacking classifier. Model performance was assessed using the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve, calibration, and decision-curve analysis. Supplementary time-to-event analyses were performed using the kidney failure risk equation and survival-based machine learning models.

**Results:** A total of 29,677 patients were included in the study. The median age was 79 (IQR 70.0–88.0) years, and 16,359 (55.1%) were male. Among these patients, 14,993 (50.5%) had hypertension, 7908 (26.6%) had diabetes mellitus, and 1768 (6%) had cancer. During follow-up, 649 patients (2.2%) developed ESRD and 3631 (12.2%) died. The models demonstrated high predictive performance for ESRD, with AUROCs ranging from 0.839 to 0.894. For all-cause mortality, the predictive performance was more modest, with AUROCs ranging from 0.752 to 0.774. Given the low incidence of ESRD in this cohort, model performance was additionally evaluated using precision-recall curves. The area under the precision-recall curve ranged

from 0.172 to 0.216 for ESRD prediction and from 0.330 to 0.356 for all-cause mortality across models. Calibration and decision-curve analyses supported model reliability and clinical utility.

**Conclusions:** Machine learning algorithms may serve as useful tools for risk stratification of ESRD and all-cause mortality in patients with CKD, with the potential to support more individualized clinical management.

*JMIR Med Inform* 2026;14:e81152; doi: [10.2196/81152](https://doi.org/10.2196/81152)

**Keywords:** chronic kidney disease; end-stage renal disease; machine learning; mortality; Shapley additive explanations; SHAP

## Introduction

Chronic kidney disease (CKD) is recognized as a major global public health challenge, with substantial impacts on morbidity and health care utilization worldwide [1,2]. In Taiwan, CKD affects approximately 10% of the adult population. The consistently high end-stage renal disease (ESRD) incidence imposes a substantial burden on the health care system and is accompanied by steadily increasing financial pressure [3,4]. This local epidemiologic context supports the need to improve risk stratification for kidney failure and mortality using Taiwan-based real-world data. Despite advancements in CKD research and management, accurately predicting progression to ESRD remains challenging. CKD is heterogeneous, with diverse etiologies, clinical presentations, and pathological features, and progression rates can vary substantially even among patients with identical primary kidney diseases [5]. Numerous clinical factors, including baseline estimated glomerular filtration rate (eGFR), proteinuria levels, age, and comorbidities, have been identified as contributors to CKD progression [6,7]. However, these factors and traditional risk equations, such as the kidney failure risk equation (KFRE), may not fully capture the complex, multidimensional nature of CKD, in which clinical, laboratory, and demographic variables can interact nonlinearly.

In recent years, advancements in artificial intelligence have significantly impacted the medical field. The integration of machine learning and deep learning with big data has found extensive applications across various medical disciplines, including image interpretation, disease diagnosis, and risk prediction [8]. Previous research has highlighted the potential of artificial intelligence for predicting the risk of patients with diabetes developing diabetic kidney disease or progressing to ESRD [9,10]. By incorporating a comprehensive range of clinical, demographic, and laboratory variables, these machine learning models have demonstrated considerable promise for achieving more precise risk stratification compared to traditional methods. Despite these advantages, several challenges remain in translating machine learning models into clinical practice. Many models have limited transparency, which can reduce interpretability and clinician trust [11]. In addition, differences in data sources, preprocessing strategies, and modeling pipelines may affect reproducibility across studies [12,13]. Furthermore, a substantial proportion of published machine learning prediction models lack external validation, which raises concerns about their generalizability and real-world clinical usefulness [14]. To bridge this gap between predictive modeling and clinical

application, there is a critical need for transparent algorithms that can accurately handle the complex, time-to-event nature of chronic disease progression. Integrating tools such as Shapley additive explanations (SHAP) and survival-based machine learning techniques can demystify these models, providing clinicians with actionable insights into individual risk trajectories over time.

The aim of this study was to develop and validate machine learning-based models to predict the risks of ESRD and all-cause mortality in patients with CKD using a large-scale single-center cohort. We hypothesized that these models would enhance the precision of risk stratification, identify clinically meaningful predictors to inform clinical decision-making, and ultimately facilitate timely interventions to improve patient outcomes.

## Methods

### Data Sources and Study Population

This retrospective cohort study involved patients with CKD, defined as persistent eGFR  $<60$  mL/min/1.73 m<sup>2</sup> for at least 3 months. Data were comprehensively obtained from the Taipei Veterans General Hospital Big Data Center, covering the period from January 2011 to December 2021. Patient information included demographic data, comorbidities, medication prescriptions, laboratory results, and echocardiography findings. Exclusion criteria were prior renal replacement therapy (dialysis or kidney transplantation) before study enrollment and individuals younger than 20 years at the time of CKD diagnosis. eGFR calculations used the 2021 Chronic Kidney Disease Epidemiology Collaboration equation [15].

### Feature Selection

Our machine learning models incorporated 69 features across multiple clinical domains, comprising a broad set of routinely available electronic health record (EHR) variables to capture multidimensional risk domains relevant to CKD progression and all-cause mortality. Demographic and lifestyle characteristics included age, sex, body weight, BMI, systolic and diastolic blood pressure, alcohol consumption, betel nut chewing, and cigarette smoking. Comorbidity profiles included hypertension, diabetes mellitus, and cancer. Medication history encompassed antidiabetic agents (metformin, sodium-glucose cotransporter 2 inhibitors, glucagon-like peptide-1 receptor agonists, sulfonylureas,  $\alpha$ -glucosidase inhibitors, thiazolidinediones, dipeptidyl peptidase-4 inhibitors, insulin), cardiovascular drugs (statins, aspirin, clopidogrel, direct oral anticoagulants, warfarin,

angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, beta-blockers, calcium channel blockers, thiazide diuretics, loop diuretics, mineralocorticoid receptor antagonists, nitrates, sacubitril or valsartan), and anti-inflammatory medications (nonsteroidal anti-inflammatory drugs, cyclooxygenase-2 inhibitors). Laboratory parameters included serum creatinine, blood urea nitrogen, sodium, potassium, chloride, bicarbonate, uric acid, calcium, phosphate, serum albumin, total protein, creatine kinase, glucose, hemoglobin A<sub>1c</sub>, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, alanine aminotransferase, alkaline phosphatase, gamma-glutamyl transferase, total bilirubin, international normalized ratio, free thyroxine, thyroid-stimulating hormone, white blood cell count, hemoglobin, platelet count, urine protein-to-creatinine ratio (UPCR), and urine albumin-to-creatinine ratio (UACR). Cardiac function was evaluated using left ventricular ejection fraction via echocardiography.

### Outcome Definition

The primary clinical outcomes were progression to ESRD and all-cause mortality. Both outcomes were ascertained using electronic health records from the Taipei Veterans General Hospital. To prevent the misclassification of temporary acute dialysis, ESRD was strictly defined as the presence of dialysis-related procedure codes recorded for at least 3 consecutive months, or the occurrence of kidney transplantation identified by the *International Classification of Diseases, Ninth or Tenth Revision*, Clinical Modification codes. Outcomes were represented as binary variables, with 1 indicating occurrence and 0 indicating nonoccurrence during follow-up.

### Data Preprocessing

Categorical variables were summarized as frequencies (percentages) and continuous variables as medians with IQRs. The dataset was randomly divided into training (23,741/29,677, 80%) and testing (5936/29,677, 20%) sets using stratified sampling with a fixed random seed to preserve the outcome distribution. Infinite values were converted to missing values, and only numeric variables were retained for model development. In the primary analysis, missing values were imputed using K-nearest-neighbor imputation ( $k=5$ ), fitted on the training set, and then applied to the testing set. Continuous variables were standardized using a StandardScaler fitted on the training set. To address significant class imbalance for ESRD, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data only with a sampling strategy of 0.9.

### Machine Learning Models

The analysis used five advanced machine learning models: extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), categorical boosting (CatBoost), random forest, and a stacking classifier integrating these base models. Each tree-based algorithm applied ensemble learning with distinct optimization strategies. The stacking classifier used logistic regression as a meta-learner

to leverage complementary strengths, reduce variance, and enhance overall predictive accuracy.

### Hyperparameter Optimization

Hyperparameter optimization was conducted using randomized search combined with 3-fold cross-validation, targeting the maximization of the area under the precision-recall curve (AUPRC), which is more appropriate for imbalanced outcomes. The optimization explored 30 random combinations within predefined parameter ranges, identifying optimal configurations efficiently. Key hyperparameters included learning rates, tree depths, regularization parameters, and specific configurations for each algorithm. The stacking classifier used logistic regression as the final meta-learner.

### Model Evaluation and Feature Importance

Model performance was comprehensively evaluated using multiple metrics, with the area under the receiver operating characteristic curve (AUROC) as the primary measure of discrimination. Additional evaluation metrics included the AUPRC, accuracy, precision, recall,  $F_1$  score, specificity, and log loss. Confusion matrices were also generated using the optimal cutoff determined by the Youden index. Calibration and clinical utility were further assessed using calibration curves and decision-curve analysis. Feature importance was evaluated using SHAP, which quantified the contribution of individual variables to model predictions and facilitated interpretation of clinically relevant predictors.

### Survival Analyses

As time-to-event analyses for ESRD, we evaluated the KFRE and random survival forest (RSF) at 2-year and 5-year horizons. To ensure a fair comparison, these models were evaluated using a unified horizon-specific test mask requiring landmark eligibility, complete age, sex, eGFR, and UACR inputs and nonmissing model predictions. Time-dependent discrimination was assessed using inverse probability of censoring weighting-based AUROC.

### Sensitivity Analyses

Sensitivity analyses were performed to assess the robustness of model performance. First, an age-stratified analysis was conducted using the XGBoost model to evaluate discrimination separately in patients aged  $\geq 60$  years and  $< 60$  years in the testing set. Second, to examine the influence of missing-data handling, we repeated model evaluation without imputation using complete-case data only. Model discrimination in these sensitivity analyses was assessed using AUROC.

### Implementation Details

Analyses were implemented using Python 3.9 with specialized libraries. Scikit-learn provided core machine learning functionality, supplemented by XGBoost, LightGBM, and CatBoost for gradient boosting. The stacking classifier was implemented using `sklearn.ensemble.StackingClassifier` with logistic regression as the final estimator. Pandas and NumPy facilitated data manipulation, while visualization

was achieved using Matplotlib and Seaborn. Missing value imputation used `sklearn.impute.KNNImputer`, dataset partitioning used `sklearn.model_selection.train_test_split`, and class balancing used `imblearn.over_sampling.SMOTE`. Hyperparameter optimization applied `sklearn.model_selection.RandomizedSearchCV`, and stratified k-fold cross-validation used `sklearn.model_selection.StratifiedKFold`. Complementary statistical analyses were performed with SAS version 9.4 (SAS Institute, Cary, NC), and statistical significance was defined as  $P < .05$ . Complementary survival analyses were implemented using `scikit-survival` and `PyTorch` for time-to-event modeling and inverse probability of censoring weighting-based evaluation.

## Ethical Considerations

This study protocol was approved by the institutional review board of Taipei Veterans General Hospital (approval number 2023-07-001CC), with informed consent waived due to the use of deidentified data.

## Results

### Characteristics and Distribution of Patients

A total of 29,677 patients with CKD were enrolled during the 10-year study period. Detailed patient demographic data are presented in Table 1. The median patient age was 79.0 (IQR 70.0-88.0) years, and 16,359 (55.1%) patients were male. Additionally, 14,993 (50.5%) patients had hypertension, 7908 (26.6%) patients had diabetes mellitus, and 1768 (6%) patients had cancer. Within the entire CKD cohort, 649 (2.2%) patients progressed to ESRD, including 519 of 23,741 (2.2%) in the training set and 130 of 5936 (2.2%) in the testing set. All-cause mortality occurred in 3631 of 29,677 patients (12.2%), including 2905 of 23,741 (12.2%) in the training set and 726 of 5936 (12.2%) in the testing set.

**Table 1.** Baseline demographic and clinical characteristics of the study cohort.

Characteristics	Full cohort (N=29,677)	Training set (n=23,741)	Testing set (n=5936)
Demographic data			
Age (y), median (IQR)	79.0 (70.0-88.0)	79.0 (70.0-88.0)	79.0 (70.0-88.0)
Male gender, n (%)	16,359 (55.1)	13,092 (55.2)	3267 (55.0)
Alcohol, n (%)	2990 (10.1)	2370 (10.0)	620 (10.5)
Betel nut, n (%)	608 (2.1)	487 (2.1)	121 (2.1)
Smoking, n (%)	3960 (13.4)	3158 (13.4)	802 (13.6)
BW <sup>a</sup> (kg), median (IQR)	62.5 (54.0-71.8)	62.5 (54.0-71.7)	62.6 (54.0-72.2)
BMI (kg/m <sup>2</sup> ), median (IQR)	24.5 (21.9-27.4)	24.5 (21.9-27.4)	24.7 (21.9-27.6)
SBP <sup>b</sup> (mm Hg), median (IQR)	133.9 (122.0-146.9)	134.0 (122.0-146.8)	133.8 (122-147)
DBP <sup>c</sup> (mm Hg), median (IQR)	72.1 (65-80)	72.0 (65-80)	72.2 (65-80)
Comorbidities, n (%)			
Hypertension	14,993 (50.5)	11,968 (50.4)	3025 (51.0)
Diabetes mellitus	7908 (26.6)	6335 (26.7)	1573 (26.5)
Cancer	1768 (6.0)	1422 (6.0)	346 (5.8)
Laboratory data, median (IQR)			
Creatinine (mg/dL)	1.2 (1.0-1.5)	1.2 (1.0-1.5)	1.2 (1.0-1.5)
BUN <sup>d</sup> (mg/dL)	22.0 (18.0-29.0)	22.0 (18.0-29.0)	22.0 (17.0-29.0)
Sodium (mmol/L)	140 (138.0-142.0)	140 (138.0-142.0)	140 (138.0-142.0)
Potassium (mmol/L)	4.2 (3.9-4.5)	4.2 (3.9-4.5)	4.2 (3.9-4.5)
Chloride (mmol/L)	105.0 (101.0-107.0)	105.0 (101.0-107.0)	104.0 (101.0-107.0)
Uric acid (mg/dL)	6.4 (5.2-7.7)	6.4 (5.2-7.7)	6.5 (5.3-7.7)
Calcium (mg/dL)	9.1 (8.7-9.5)	9.1 (8.7-9.5)	9.1 (8.7-9.5)
Phosphate (mg/dL)	3.5 (3.1-4.0)	3.6 (3.1-4.0)	3.5 (3.1-4.0)
Albumin (g/dL)	3.9 (3.5-4.3)	3.9 (3.5-4.3)	3.9 (3.5-4.2)
Cholesterol (mg/dL)	167.0 (142.0-195.0)	167.0 (141.0-194.0)	168.0 (143.0-195.0)
Bicarbonate (mmol/L)	23.4 (20.2-26.7)	23.4 (20.2-26.6)	23.6 (20.3-27.0)
HbA <sub>1c</sub> <sup>e</sup> (%)	6.3 (5.8-7.3)	6.4 (5.8-7.3)	6.3 (5.7-7.2)
HDL-C <sup>f</sup> (mg/dL)	43.0 (36.0-53.0)	44.0 (36.0-53.0)	43.0 (36.0-52.0)
LDL-C <sup>g</sup> (mg/dL)	95.0 (74.8-119.0)	95.0 (74.0-118.0)	96.0 (76.0-120.0)
Triglyceride (mg/dL)	114.0 (83.0-163.0)	114.0 (83.0-162.0)	117.0 (85.0-165.0)
Total protein (g/dL)	6.5 (5.9-7.2)	6.5 (5.9-7.2)	6.5 (5.8-7.1)

Characteristics	Full cohort (N=29,677)	Training set (n=23,741)	Testing set (n=5936)
CK <sup>h</sup> (U/L)	61.0 (37.0-114.0)	61.0 (37.0-112.0)	63.0 (37.0-120.0)
Glucose (mg/dL)	128.0 (106.0-168.0)	128.0 (106.0-168.0)	128.0 (106.0-168.0)
ALT <sup>i</sup> (U/L)	19.0 (14.0-30.0)	19.0 (14.0-30.0)	19.0 (14.0-31.0)
ALKP <sup>j</sup> (U/L)	76.0 (60.0-100.0)	76.0 (60.0-100.0)	75.0 (60.0-100.0)
GGT <sup>k</sup> (U/L)	40.0 (21.0-92.0)	40.0 (21.0-92.0)	39.0 (21.0-91.0)
Total bilirubin (mg/dL)	0.5 (0.3-0.8)	0.5 (0.3-0.8)	0.5 (0.3-0.8)
INR <sup>l</sup>	1.0 (1.0-1.1)	1.0 (1.0-1.1)	1.0 (1.0-1.1)
Free T4 <sup>m</sup> (ng/dL)	1.1 (0.9-1.2)	1.1 (0.9-1.2)	1.1 (0.9-1.2)
TSH <sup>n</sup> (μIU/mL)	1.5 (0.8-2.6)	1.5 (0.8-2.6)	1.5 (0.8-2.5)
WBC <sup>o</sup> (per μL)	7100 (5600-9400)	7100 (5600-9400)	7160 (5600-9400)
Hemoglobin (g/dL)	12.3 (10.8-13.7)	12.3 (10.8-13.7)	12.4 (10.8-13.7)
PLT <sup>p</sup> (per μL)	208,000 (164,000-259,000)	208,000 (164,000-259,000)	208,000 (164,000-258,000)
UPCR <sup>q</sup> (g/g)	0.5 (0.1-2.2)	0.5 (0.1-2.2)	0.5 (0.2-2.3)
UACR <sup>r</sup> (g/g)	0.1 (0.0-0.4)	0.1 (0.0-0.4)	0.0 (0.0-0.3)
Cardiac function, median (IQR)			
LVEF <sup>s</sup> (%)	58.0 (53.0-63.0)	58.0 (53.0-63.0)	58.0 (53.0-62.0)
Concomitant medications, n (%)			
Metformin	5038 (17.0)	4056 (17.1)	982 (16.5)
SGLT2 <sup>t</sup> inhibitors	801 (2.7)	647 (2.7)	154 (2.6)
GLP1 RAs <sup>u</sup>	114 (0.4)	91 (0.4)	23 (0.4)
Sulfonylureas	2493 (8.4)	1986 (8.4)	507 (8.5)
Alpha-glucosidase inhibitors	889 (3.0)	724 (3.0)	165 (2.8)
Thiazolidinediones	492 (1.7)	395 (1.7)	97 (1.6)
DPP4 <sup>v</sup> inhibitors	1196 (4.0)	971 (4.1)	225 (3.8)
Insulin	3968 (13.4)	3194 (13.5)	774 (13)
Statins	7557 (25.5)	6023 (25.4)	1534 (25.8)
Aspirin	5538 (18.7)	4406 (18.6)	1132 (19.1)
Clopidogrel	3033 (10.2)	2414 (10.2)	619 (10.4)
DOACs <sup>w</sup>	1739 (5.9)	1396 (5.9)	343 (5.8)
Warfarin	674 (2.3)	530 (2.2)	144 (2.4)
Alpha blockers	2188 (7.4)	1731 (7.3)	457 (7.7)
Beta blockers	8378 (28.2)	6640 (28.0)	1738 (29.3)
CCB <sup>x</sup>	10,652 (35.9)	8497 (35.8)	2155 (36.3)
ACE <sup>y</sup> inhibitors	856 (2.9)	664 (2.8)	192 (3.2)
ARB <sup>z</sup>	9699 (32.7)	7762 (32.7)	1937 (32.6)
Thiazides	902 (3.0)	724 (3.0)	178 (3.0)
Loop diuretics	5417 (18.3)	4312 (18.2)	1105 (18.6)
MRAs <sup>aa</sup>	2976 (10.0)	2382 (10.0)	594 (10.0)
Nitrates	3406 (11.5)	2680 (11.3)	726 (12.2)
NSAIDs <sup>ab</sup>	2490 (8.4)	1978 (8.3)	512 (8.6)
COX-2 <sup>ac</sup> inhibitors	3569 (12.0)	2850 (12.0)	719 (12.1)
Sacubitril or valsartan	297 (1.0)	231 (1.0)	66 (1.1)

<sup>a</sup>BW: body weight.

<sup>b</sup>SBP: systolic blood pressure.

<sup>c</sup>DBP: diastolic blood pressure.

<sup>d</sup>BUN: blood urea nitrogen.

<sup>e</sup>HbA<sub>1c</sub>: glycated hemoglobin.

<sup>f</sup>HDL-C: high-density lipoprotein cholesterol.

<sup>g</sup>LDL-C: low-density lipoprotein cholesterol.

<sup>h</sup>CK: creatine kinase.

<sup>i</sup>ALT: alanine aminotransferase.

<sup>j</sup>ALKP: alkaline phosphatase.

<sup>k</sup>GGT: gamma-glutamyl transferase.

<sup>l</sup>INR: international normalized ratio.

<sup>m</sup>Free T4: free thyroxine.

<sup>n</sup>TSH: thyroid-stimulating hormone.

<sup>o</sup>WBC: white blood cell count.

<sup>p</sup>PLT: platelet count.

<sup>q</sup>UPCR: urine protein-to-creatinine ratio.

<sup>r</sup>UACR: urine albumin-to-creatinine ratio.

<sup>s</sup>LVEF: left ventricular ejection fraction.

<sup>t</sup>SGLT2: sodium-glucose cotransporter 2.

<sup>u</sup>GLP1 RA: glucagon-like peptide-1 receptor agonist.

<sup>v</sup>DPP4: dipeptidyl peptidase-4.

<sup>w</sup>DOAC: direct oral anticoagulant.

<sup>x</sup>CCB: calcium channel blocker.

<sup>y</sup>ACE: angiotensin-converting enzyme.

<sup>z</sup>ARB: angiotensin II receptor blocker.

<sup>aa</sup>MRA: mineralocorticoid receptor antagonist.

<sup>ab</sup>NSAID: nonsteroidal anti-inflammatory drug.

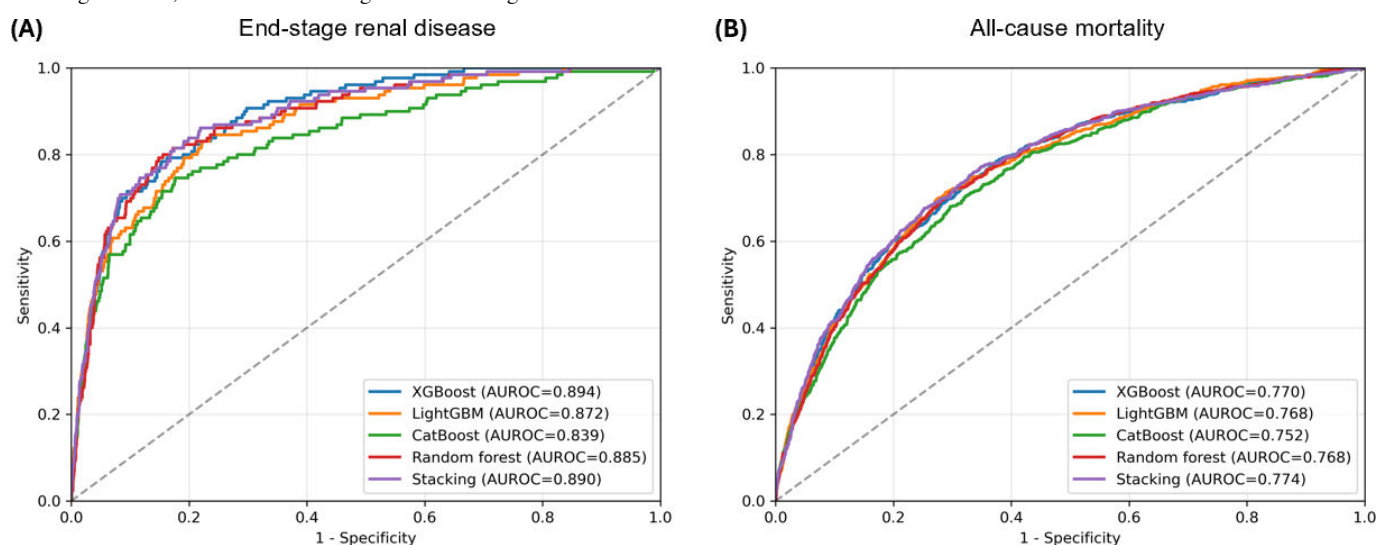
<sup>ac</sup>COX-2: cyclooxygenase-2.

## Model Performance

Five machine learning algorithms, including XGBoost, LightGBM, CatBoost, random forest, and a stacking classifier, were trained on 80% ( $n=23,741$ ) of the cohort and evaluated on the remaining 20% ( $n=5936$ ). [Figure 1A](#) displays ROC curves illustrating the predictive performance

of each algorithm for ESRD prediction in the testing set. Overall, all models showed good discrimination, with AUROC values ranging from 0.839 to 0.894. XGBoost achieved the highest AUROC of 0.894, followed by the stacking classifier (0.890), random forest (0.885), LightGBM (0.872), and CatBoost (0.839).

**Figure 1.** Receiver operating characteristic curves of the machine learning models for prediction of (A) end-stage renal disease and (B) all-cause mortality in the testing set. AUROC: area under the receiver operating characteristic curve; CatBoost: categorical boosting; LightGBM: light gradient boosting machine; XGBoost: extreme gradient boosting.

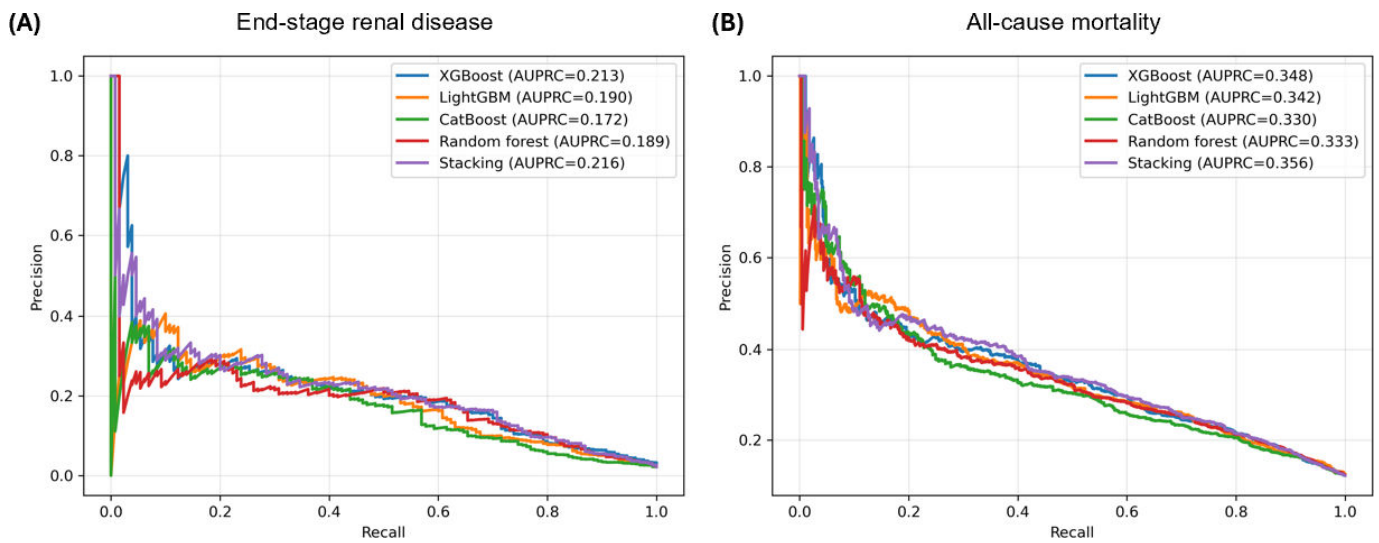


Predictive performance for all-cause mortality was lower than that for ESRD, with AUROC values ranging from 0.752 to 0.774 ([Figure 1B](#)). The stacking classifier achieved the highest AUROC of 0.774, followed by XGBoost (0.770), LightGBM (0.768), random forest (0.768), and CatBoost (0.752).

Given the low incidence of ESRD in the study cohort, precision-recall analysis was additionally performed. The AUPRC ranged from 0.172 to 0.216 ([Figure 2A](#)). The stacking classifier achieved the highest AUPRC of 0.216, followed by XGBoost (0.213), LightGBM (0.190), random

forest (0.189), and CatBoost (0.172). For all-cause mortality, the AUPRC ranged from 0.330 to 0.356 ([Figure 2B](#)). The stacking classifier achieved the highest AUPRC of 0.356, followed by XGBoost (0.348), LightGBM (0.342), random forest (0.333), and CatBoost (0.330). Additional performance metrics, including accuracy, precision, recall,  $F_1$  score, specificity, and log loss, are summarized in Tables S1 and S2 in [Multimedia Appendix 1](#). Confusion matrices for ESRD and all-cause mortality prediction using the XGBoost model are shown in [Figure S1](#) in [Multimedia Appendix 1](#).

**Figure 2.** Precision-recall curves of the machine learning models for prediction of (A) end-stage renal disease and (B) all-cause mortality in the testing set. AUPRC: area under the precision-recall curve; CatBoost: categorical boosting; LightGBM: light gradient boosting machine; XGBoost: extreme gradient boosting.

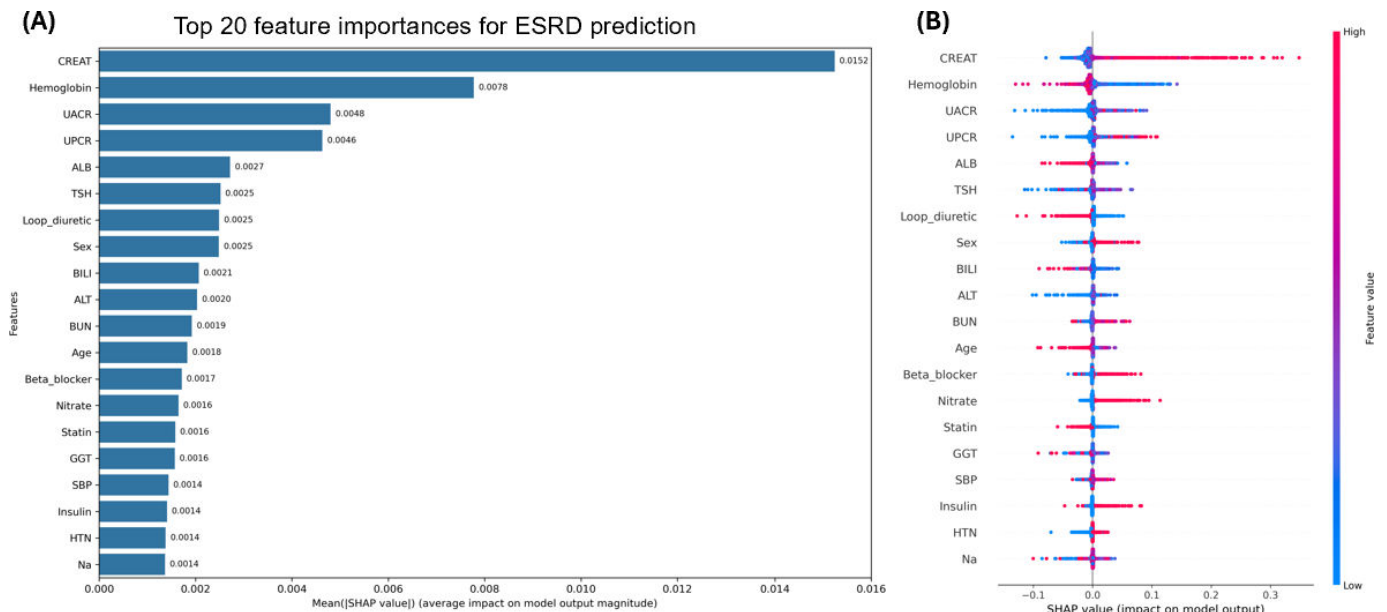


### Ranks of Feature Importance and SHAP Values in the XGBoost Model

Feature attribution analyses using SHAP identified key predictors for ESRD and all-cause mortality. Figure 3A presents feature importance rankings from the XGBoost model for ESRD prediction, arranged in descending order based on impact. The top 5 most influential features were

serum creatinine, hemoglobin, UACR, UPCR, and albumin. Figure 3B illustrates the SHAP summary plot, where red points represent higher feature values and blue points indicate lower values. Positive SHAP values suggest increased ESRD risk, while negative values indicate lower risk. Higher values of serum creatinine, UACR, and UPCR, as well as lower hemoglobin and albumin levels, were associated with a higher predicted risk of ESRD.

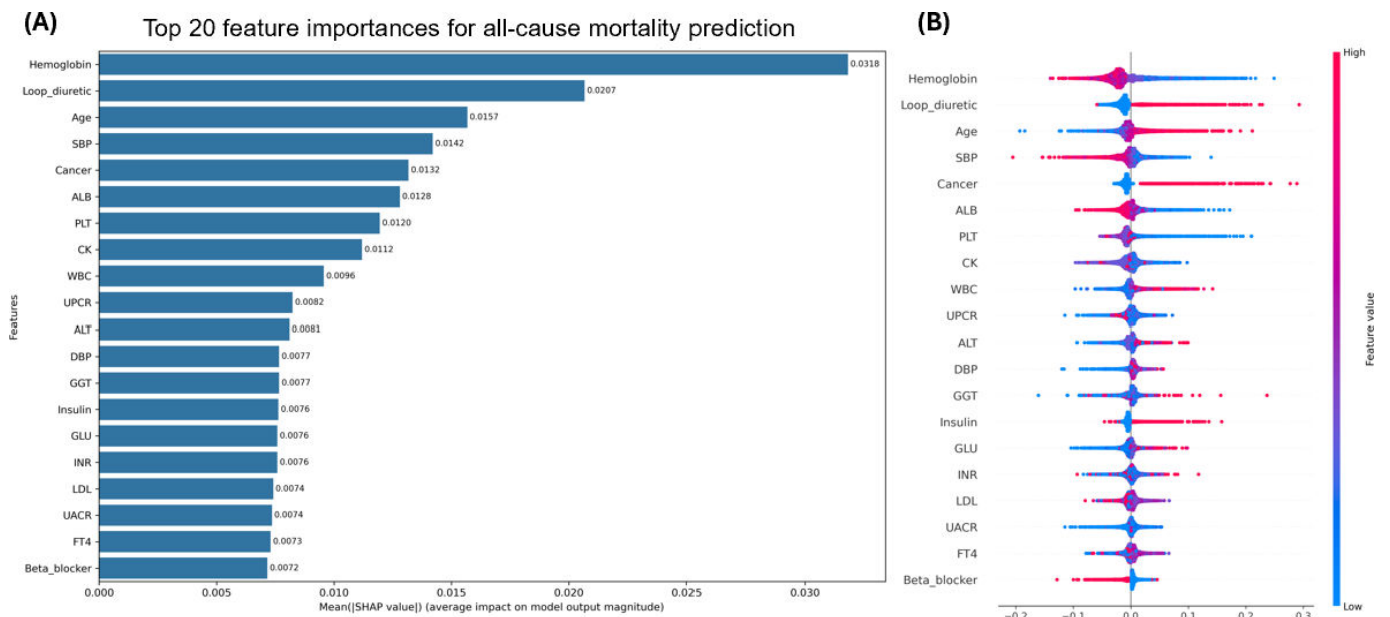
**Figure 3.** Shapley additive explanations (SHAP) analysis for the prediction of end-stage renal disease (ESRD). (A) Mean absolute SHAP values showing the top 20 most influential features in the extreme gradient boosting (XGBoost) model. (B) SHAP summary plot showing the direction and magnitude of feature effects on the model output. Red indicates higher feature values, and blue indicates lower feature values. ALB: albumin; ALT: alanine aminotransferase; BILI: total bilirubin; BUN: blood urea nitrogen; CREAT: creatinine; CI: chloride; GGT: gamma-glutamyl transferase; HTN: hypertension; Na: sodium; SBP: systolic blood pressure; TSH: thyroid-stimulating hormone; UACR: urine albumin-to-creatinine ratio; UPCR: urine protein-to-creatinine ratio.



For all-cause mortality, Figure 4A displays feature importance rankings from the XGBoost model. Hemoglobin and loop diuretic use were the most impactful features, followed by age, systolic blood pressure, and cancer. Figure 4B shows

the SHAP summary plot for all-cause mortality. Lower hemoglobin levels, loop diuretic use, older age, lower systolic blood pressure, and the presence of cancer were generally associated with a higher predicted mortality risk.

**Figure 4.** Shapley additive explanations (SHAP) analysis for the prediction of all-cause mortality. (A) Mean absolute SHAP values showing the top 20 most influential features in the extreme gradient boosting (XGBoost) model. (B) SHAP summary plot showing the direction and magnitude of feature effects on the model output. Red indicates higher feature values, and blue indicates lower feature values. ALB: albumin; ALT: alanine aminotransferase; CK: creatine kinase; DBP: diastolic blood pressure; FT4: free thyroxine; GGT: gamma-glutamyl transferase; GLU: glucose; INR: international normalized ratio; LDL: low-density lipoprotein cholesterol; PLT: platelet count; SBP: systolic blood pressure; UACR: urine albumin-to-creatinine ratio; UPCR: urine protein-to-creatinine ratio; WBC: white blood cell count.

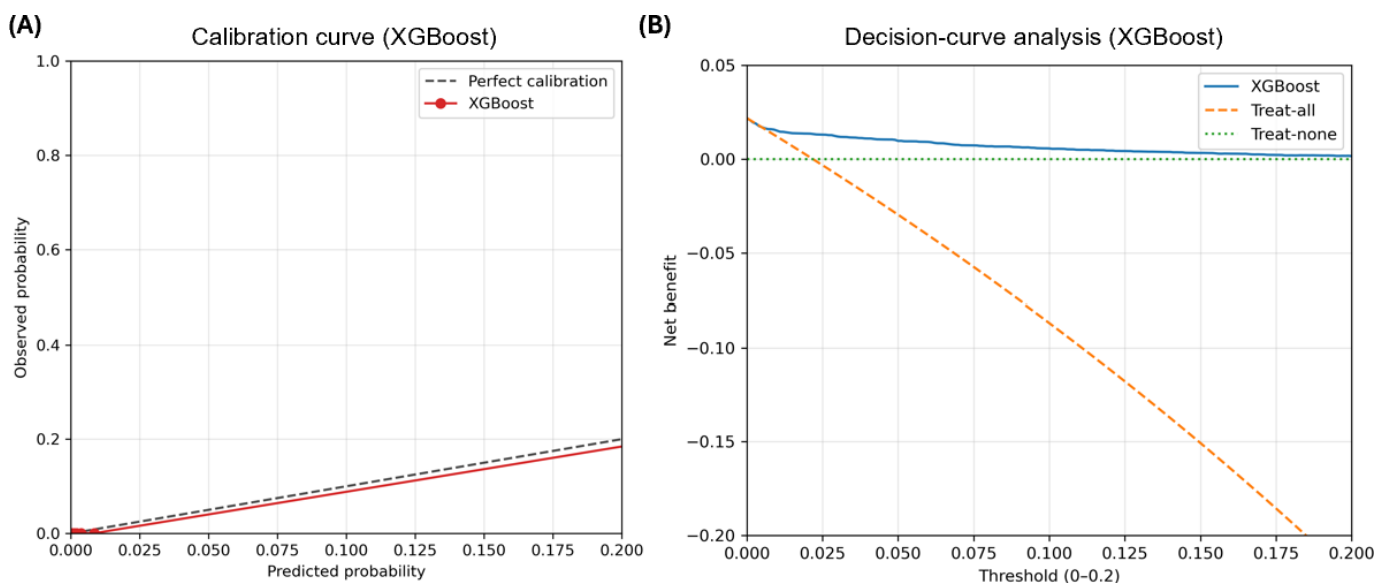


**Calibration and Decision-Curve Analysis**

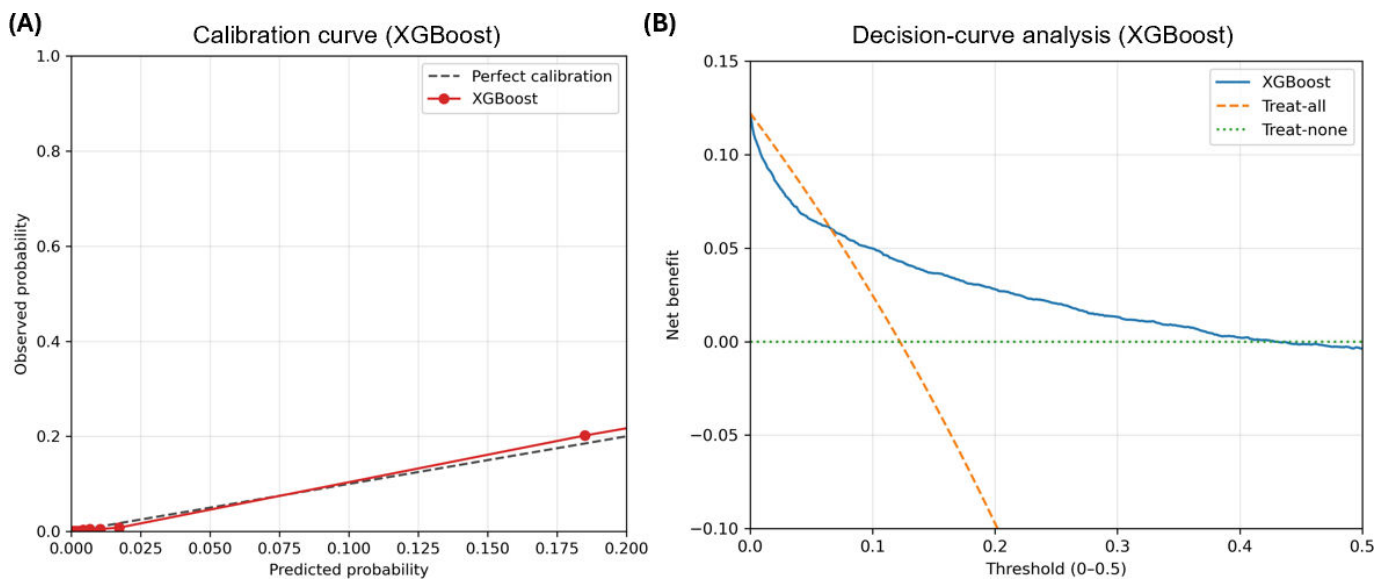
To further assess clinical usefulness, calibration and decision-curve analyses of the XGBoost model were performed in the testing set. The calibration curve showed overall agreement between predicted and observed ESRD risks (Figure 5A). In the decision-curve analysis, the XGBoost model showed a higher net benefit than the treat-all and treat-none strategies across threshold probabilities ranging from 0 to

0.20 (Figure 5B). For all-cause mortality, the calibration curve also showed overall agreement between predicted and observed risks (Figure 6A). In the decision-curve analysis, the XGBoost model showed a positive net benefit across a broad range of threshold probabilities, although the treat-all strategy provided slightly higher net benefit at very low thresholds (Figure 6B).

**Figure 5.** Calibration and decision-curve analyses of the extreme gradient boosting (XGBoost) model for end-stage renal disease prediction in the testing set. (A) Calibration curve comparing predicted and observed event probabilities. (B) Decision-curve analysis comparing the net benefit of the XGBoost model with the treat-all and treat-none strategies.



**Figure 6.** Calibration and decision-curve analyses of the extreme gradient boosting (XGBoost) model for all-cause mortality prediction in the testing set. (A) Calibration curve comparing predicted and observed event probabilities. (B) Decision-curve analysis comparing the net benefit of the XGBoost model with the treat-all and treat-none strategies.

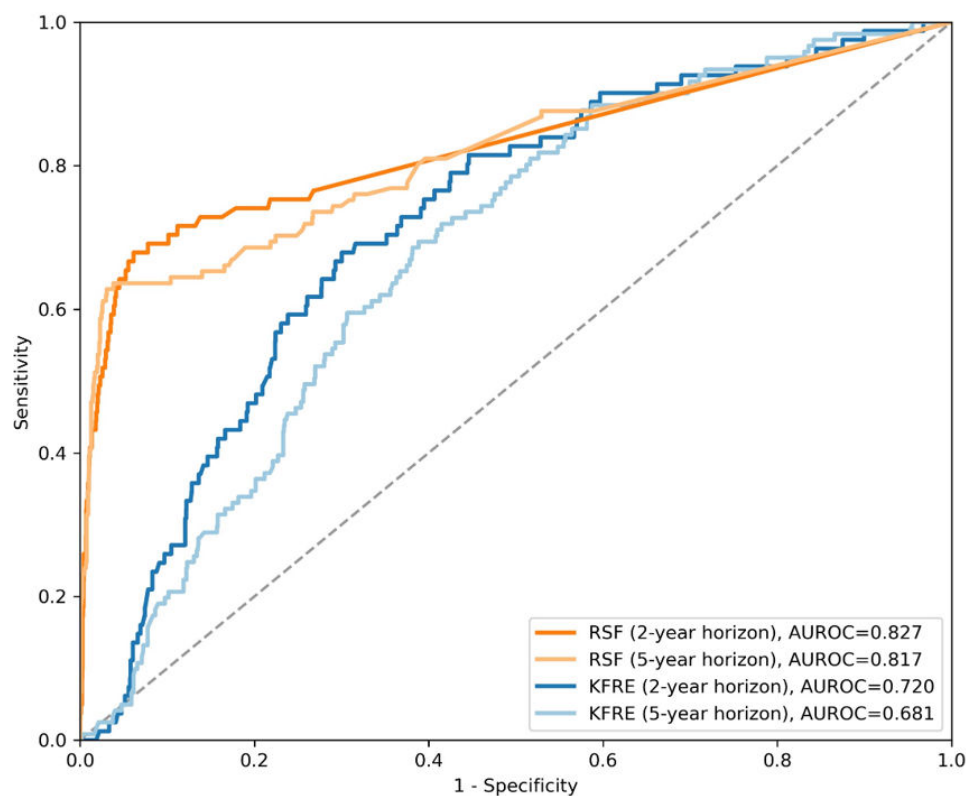


### Survival Analysis Using RSF and Comparison With KFRE

To better account for the time-to-event nature of ESRD, we additionally performed survival analysis using RSF. The

predictive performance of RSF was then compared with that of the KFRE at the 2- and 5-year horizons (Figure 7). RSF demonstrated better discrimination than KFRE at both time points, with AUROCs of 0.827 versus 0.720 at 2 years and 0.817 versus 0.681 at 5 years.

**Figure 7.** Receiver operating characteristic curves comparing the random survival forest (RSF) and the kidney failure risk equation (KFRE) for the prediction of end-stage renal disease at the 2- and 5-year horizons. AUROC: area under the receiver operating characteristic curve.



### Sensitivity Analyses

In the age-stratified analysis, model discrimination remained similar between patients aged  $\geq 60$  years and  $< 60$  years,

with AUROCs of 0.896 (95% CI 0.867-0.921) and 0.866 (95% CI 0.777-0.939), respectively (Figure S2 in Multimedia Appendix 1). In analyses without imputation, AUROCs for ESRD ranged from 0.888 to 0.903 and AUROCs for all-cause

mortality ranged from 0.797 to 0.808 across models (Figure S3 in [Multimedia Appendix 1](#)), showing results comparable to the primary analysis.

## Discussion

In this study, we developed and validated machine learning models to predict the risk of ESRD and all-cause mortality in a large cohort of patients with CKD. Our findings showed that the evaluated machine learning algorithms, including XGBoost, LightGBM, CatBoost, random forest, and a stacking classifier, achieved good discrimination for ESRD prediction, with a maximum AUROC of 0.894. Predictive performance for all-cause mortality was more modest, with a maximum AUROC of 0.774. In addition, calibration and decision-curve analyses supported the overall reliability and clinical utility of the model. A previous study developed predictive models for CKD progression using traditional statistical methods and clinical variables from Canadian cohorts collected between 2001 and 2008 [7]. These models, based on Cox proportional hazards regression, showed reasonable predictive performance by incorporating key demographic and laboratory parameters, such as eGFR and albuminuria. However, their reliance on conventional statistical techniques and a limited number of variables may restrict their ability to fully capture complex relationships among CKD progression risk factors. Furthermore, evolving clinical practices, therapeutic strategies, and diagnostic methods over time might limit their applicability to contemporary patient populations.

Using advanced machine learning techniques allows the integration of a broader range of clinical, demographic, and laboratory features into predictive models, enabling more precise risk assessment for CKD progression [16,17]. Unlike traditional statistical methods, which are often constrained by linear assumptions and limited predictors, machine learning algorithms can capture complex, nonlinear interactions among variables. Additionally, these models manage high-dimensional data, autonomously identify relevant features, and continuously adapt to evolving clinical practices and patient populations. The incorporation of explainability tools, such as SHAP, further enhances interpretability by clarifying individual risk factor contributions [18]. A recent study by Tangri et al [19] validated the Klinrisk machine learning model using data from the CANVAS (Canagliflozin Cardiovascular Assessment Study) and CREDENCE (Canagliflozin and Renal Events in Diabetes With Established Nephropathy Clinical Evaluation) clinical trials, demonstrating superior performance compared to KDIGO (Kidney Disease: Improving Global Outcomes) risk classification, with an AUROC of 0.88 at 3 years. In addition, our survival-based comparative analysis showed that RSF outperformed the KFRE at both the 2- and 5-year horizons, suggesting that machine learning approaches incorporating time-to-event information may provide improved discrimination over conventional risk equations.

Another study used machine learning models to predict ESRD risk in newly diagnosed patients with type

2 diabetes mellitus (T2DM), illustrating the effectiveness of these approaches in risk stratification [10]. Among the evaluated algorithms, XGBoost achieved the highest predictive performance. This study also used SHAP to interpret predictions, identifying baseline serum creatinine, mean serum creatinine within 1 year before T2DM diagnosis, high-sensitivity C-reactive protein, and UPCR as key predictors. However, its limitation to patients with diabetes restricts generalizability to broader CKD populations. Given CKD's heterogeneity, with significantly varying etiologies and progression rates between patients with diabetes and those without diabetes, models trained exclusively on T2DM populations might not fully capture CKD's complexity. Our study extends the application of machine learning in CKD risk prediction by analyzing CKD cohort with varied etiologies, offering a comprehensive ESRD risk assessment beyond diabetes-related kidney disease. Previous machine learning studies predicting CKD progression often had limited sample sizes and narrow feature sets [20-22]. Our study, in contrast, uses a significantly larger cohort and broader clinical, laboratory, and demographic features, likely enhancing observed model performance. This comprehensive approach improves the applicability of machine learning-based models in clinical settings, aiding early identification and intervention for patients at high risk of ESRD.

We identified baseline serum creatinine, UACR, and UPCR as top predictors of ESRD, consistent with established CKD progression knowledge. These biomarkers have long been recognized as key indicators of kidney function decline, and their prominence reinforces their critical role in risk assessment. Hemoglobin also emerged as a strong predictor, suggesting anemia's potentially significant role in CKD progression. Anemia is common in CKD, primarily due to reduced erythropoietin production, iron deficiency, and chronic inflammation [23]. It has been linked to increased cardiovascular risk and impaired quality of life, particularly among patients with CKD [23-25]. A study using data from the SOLVD trial, enrolling patients with heart failure with significantly reduced left ventricular ejection fractions of 35% or less, showed anemia significantly associated with accelerated eGFR decline, especially pronounced in patients with baseline CKD [26]. The complex interaction among anemia, cardiovascular disease, and renal function decline underscores the importance of comprehensive clinical management in CKD [27].

The predictive performance for all-cause mortality was lower than that for ESRD, likely reflecting the greater heterogeneity of mortality as an outcome. Unlike ESRD, which is more directly linked to renal function and proteinuria-related measures, all-cause mortality is influenced by a broader range of renal and nonrenal factors. Routinely collected structured EHR variables may not fully capture important dimensions of mortality risk, such as frailty, functional status, nutritional reserve, cognitive impairment, and overall disease severity. Incorporating these variables may improve mortality prediction in future studies. Top predictors identified by SHAP analysis were hemoglobin,

loop diuretic use, age, systolic blood pressure, cancer, and albumin. In patients with CKD, lower hemoglobin levels may reflect reduced erythropoietin production, chronic inflammation, malnutrition, and a greater comorbidity burden, all of which can contribute to vulnerability and adverse outcomes [28]. Loop diuretic use may indicate underlying volume overload or more advanced heart or kidney disease. Lower systolic blood pressure also emerged as an important predictor of mortality. This finding is consistent with prior studies in older adults with CKD and may be especially relevant in our cohort, given its advanced age distribution [29]. In this setting, lower systolic blood pressure may reflect frailty, impaired cardiac function, malnutrition, or overall physiologic vulnerability. Serum albumin is a well-established marker of both nutritional status and systemic inflammation [30,31]. Hypoalbuminemia has been consistently linked to increased short- and long-term mortality across diverse patient populations, including those with and without chronic kidney disease [32-34]. Proteinuria, commonly used to assess renal progression, also reflects systemic endothelial dysfunction and microvascular damage, contributing to increased cardiovascular morbidity and mortality. Studies have linked proteinuria with increased risk of coronary artery disease, stroke, heart failure, and arrhythmias [35-39]. Even low-grade albuminuria has been associated with adverse cardiac remodeling and unfavorable cardiovascular outcomes. A recent large-scale cohort study further demonstrated that proteinuria significantly increases the risk of sudden cardiac arrest in young adults, particularly in those with CKD [40]. Despite its well-established prognostic value in cardiovascular disease, proteinuria remains underutilized in routine cardiovascular risk stratification [41]. Our findings reinforce the clinical importance of proteinuria and support its broader integration into multidisciplinary risk evaluation strategies, especially within CKD populations.

From a clinical perspective, the practical value of these machine learning models lies in their potential integration into routine electronic health record workflows. Such risk estimates may help clinicians identify patients at high risk for ESRD or death who may benefit from closer follow-up, earlier nephrology referral, intensified management of modifiable risk factors, and more timely preparation

for kidney replacement therapy. In addition, the decision-curve analysis suggests that the model may support risk-based clinical decision-making across a range of threshold probabilities. However, the optimal thresholds for intervention and the clinical impact of implementing such a model should be established in future prospective studies.

The strengths of our study include a large real-world CKD dataset with comprehensive clinical and laboratory information. Multiple machine learning algorithms, enhanced by advanced techniques such as SMOTE and SHAP, strengthen our findings' robustness and applicability. Consistency in model performance across different frameworks further supports reliability. However, several limitations should be acknowledged. First, this was a retrospective, single-center study, which may introduce selection bias and limit generalizability. In particular, the cohort had a relatively advanced age distribution, likely reflecting the older patient population served by our institution. Nevertheless, age-stratified subgroup analysis showed similar model discrimination in patients aged  $\geq 60$  and  $< 60$  years, although external validation in younger populations remains necessary. Second, although we included a broad set of routinely available clinical variables, some potentially relevant factors, such as genetic markers and socioeconomic influences, were not included because they were not consistently available in structured form in our EHRs. Future studies incorporating multi-omics data and social determinants may enhance predictive accuracy. Third, preprocessing methods may have introduced bias. KNN imputation may smooth extreme laboratory values, and SMOTE may generate synthetic minority samples that do not fully reflect the true clinical distribution, particularly for laboratory-based predictors. However, sensitivity analyses without imputation showed comparable model discrimination, suggesting that the overall results were broadly similar without imputation. Finally, the relative homogeneity of this hospital-based cohort may limit broader generalizability, and external validation in independent and more diverse CKD populations remains necessary.

In conclusion, machine learning models may provide useful tools for risk stratification of ESRD and all-cause mortality in patients with CKD. However, further validation is required before broader clinical application.

## Acknowledgments

This study is based in part on data from the Big Data Center, Taipei Veterans General Hospital. The interpretations and conclusions contained herein do not represent the position of Taipei Veterans General Hospital. We also appreciate the computational resources provided by TVGH Cloud 1. We disclose that a large language model was used only for preliminary English-language editing of the manuscript. The manuscript was subsequently reviewed and edited through a formal professional English-editing service. The authors carefully reviewed and verified the entire manuscript and remain fully responsible for the accuracy, originality, integrity, references, and citations of all content.

## Funding

This study was funded by research grants from the National Science and Technology Council of Taiwan (NSTC 114-2314-B-075-005, NSTC110-2320-B-075-004-MY3, and NSTC113-2320-B-075-010), Taipei Veterans General Hospital (V113B-021, V114C-149, V114E-006-2, and V115C-048), and the Yin Shu-Tien Foundation Taipei Veterans General Hospital-National Yang Ming Chiao Tung University Excellent Physician Scientists Cultivation Program, No 113-V-B-104. The funders were not involved in the study design, data collection and analysis, interpretation of the results, manuscript preparation, or the decision to publish.

---

## Data Availability

The datasets generated and analyzed during this study are not publicly available due to institutional data governance regulations. However, data may be made available from the corresponding author upon reasonable request and with the permission from the Big Data Center, Taipei Veterans General Hospital.

---

## Authors' Contributions

Conceptualization: THC  
Data curation: THC, KHL, YCC  
Formal analysis: THC, WCT  
Funding acquisition: THC  
Investigation: THC  
Methodology: THC, KHL, WCT  
Project administration: THC, KHL, YH, YCC, DCT  
Resources: THC, KHL, YCC  
Supervision: YCC, DCT  
Validation: THC, KHL  
Visualization: THC, KHL  
Writing – original draft: THC, KHL  
Writing – review and editing: YCC, DCT

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Model performance metrics, confusion matrices, age-stratified receiver operating characteristic curves, and sensitivity analyses without imputation for the prediction of end-stage renal disease and all-cause mortality.

[\[DOCX File \(Microsoft Word File\), 524 KB-Multimedia Appendix 1\]](#)

---

## References

1. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Suppl* (2011). Apr 2022;12(1):7-11. [doi: [10.1016/j.kisu.2021.11.003](https://doi.org/10.1016/j.kisu.2021.11.003)] [Medline: [35529086](https://pubmed.ncbi.nlm.nih.gov/35529086/)]
2. Chen W, Howard K, Gorham G, et al. Costs and healthcare use of patients with chronic kidney disease in the Northern Territory, Australia. *BMC Health Serv Res*. Jul 9, 2024;24(1):791. [doi: [10.1186/s12913-024-11258-8](https://doi.org/10.1186/s12913-024-11258-8)] [Medline: [38982437](https://pubmed.ncbi.nlm.nih.gov/38982437/)]
3. Lee CC, Hsu CC, Lin MH, Sung JM, Kuo TH. Healthcare utilization and expenditure among individuals with end-stage kidney disease in Taiwan. *J Formos Med Assoc*. Feb 2022;121:S47-S55. [doi: [10.1016/j.jfma.2021.12.010](https://doi.org/10.1016/j.jfma.2021.12.010)] [Medline: [34980549](https://pubmed.ncbi.nlm.nih.gov/34980549/)]
4. Ang MD, Cheng CY, Tsai WC, et al. National prevalence, regional distribution, and risk factors for CKD in Taiwan. *Clin J Am Soc Nephrol*. Apr 1, 2026;21(4):703-715. [doi: [10.2215/CJN.0000000892](https://doi.org/10.2215/CJN.0000000892)] [Medline: [41105475](https://pubmed.ncbi.nlm.nih.gov/41105475/)]
5. Levey AS, Coresh J. Chronic kidney disease. *Lancet*. Jan 14, 2012;379(9811):165-180. [doi: [10.1016/S0140-6736\(11\)60178-5](https://doi.org/10.1016/S0140-6736(11)60178-5)] [Medline: [21840587](https://pubmed.ncbi.nlm.nih.gov/21840587/)]
6. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int*. Apr 2024;105(4S):S117-S314. [doi: [10.1016/j.kint.2023.10.018](https://doi.org/10.1016/j.kint.2023.10.018)] [Medline: [38490803](https://pubmed.ncbi.nlm.nih.gov/38490803/)]
7. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. Apr 20, 2011;305(15):1553-1559. [doi: [10.1001/jama.2011.451](https://doi.org/10.1001/jama.2011.451)] [Medline: [21482743](https://pubmed.ncbi.nlm.nih.gov/21482743/)]
8. Rahman A, Debnath T, Kundu D, et al. Machine learning and deep learning-based approach in smart healthcare: recent advances, applications, challenges and opportunities. *AIMS Public Health*. 2024;11(1):58-109. [doi: [10.3934/publichealth.2024004](https://doi.org/10.3934/publichealth.2024004)] [Medline: [38617415](https://pubmed.ncbi.nlm.nih.gov/38617415/)]
9. Allen A, Iqbal Z, Green-Saxena A, et al. Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. Jan 2022;10(1):e002560. [doi: [10.1136/bmjdr-2021-002560](https://doi.org/10.1136/bmjdr-2021-002560)] [Medline: [35046014](https://pubmed.ncbi.nlm.nih.gov/35046014/)]
10. Ou SM, Tsai MT, Lee KH, et al. Prediction of the risk of developing end-stage renal diseases in newly diagnosed type 2 diabetes mellitus using artificial intelligence algorithms. *BioData Min*. Mar 10, 2023;16(1):8. [doi: [10.1186/s13040-023-00324-2](https://doi.org/10.1186/s13040-023-00324-2)] [Medline: [36899426](https://pubmed.ncbi.nlm.nih.gov/36899426/)]
11. Moss L, Corsar D, Shaw M, Piper I, Hawthorne C. Demystifying the black box: the importance of interpretability of predictive models in neurocritical care. *Neurocrit Care*. Aug 2022;37(Suppl 2):185-191. [doi: [10.1007/s12028-022-01504-4](https://doi.org/10.1007/s12028-022-01504-4)] [Medline: [35523917](https://pubmed.ncbi.nlm.nih.gov/35523917/)]

12. Navarro CLA, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*. Oct 20, 2021;375:n2281. [doi: [10.1136/bmj.n2281](https://doi.org/10.1136/bmj.n2281)] [Medline: [34670780](https://pubmed.ncbi.nlm.nih.gov/34670780/)]
13. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. Jun 2019;110:12-22. [doi: [10.1016/j.jclinepi.2019.02.004](https://doi.org/10.1016/j.jclinepi.2019.02.004)] [Medline: [30763612](https://pubmed.ncbi.nlm.nih.gov/30763612/)]
14. Gallitto G, Englert R, Kincses B, et al. External validation of machine learning models-registered models and adaptive sample splitting. *GigaScience*. Jan 6, 2025;14:giaf036. [doi: [10.1093/gigascience/giaf036](https://doi.org/10.1093/gigascience/giaf036)] [Medline: [40366867](https://pubmed.ncbi.nlm.nih.gov/40366867/)]
15. Inker LA, Eneanya ND, Coresh J, et al. New creatinine-and cystatin C-based equations to estimate GFR without race. *N Engl J Med*. Nov 4, 2021;385(19):1737-1749. [doi: [10.1056/NEJMoa2102953](https://doi.org/10.1056/NEJMoa2102953)] [Medline: [34554658](https://pubmed.ncbi.nlm.nih.gov/34554658/)]
16. Aoki J, Kaya C, Khalid O, et al. CKD progression prediction in a diverse US population: a machine-learning model. *Kidney Med*. Sep 2023;5(9):100692. [doi: [10.1016/j.xkme.2023.100692](https://doi.org/10.1016/j.xkme.2023.100692)] [Medline: [37637863](https://pubmed.ncbi.nlm.nih.gov/37637863/)]
17. Delrue C, De Bruyne S, Speeckaert MM. Application of machine learning in chronic kidney disease: current status and future prospects. *Biomedicines*. Mar 3, 2024;12(3):568. [doi: [10.3390/biomedicines12030568](https://doi.org/10.3390/biomedicines12030568)] [Medline: [38540181](https://pubmed.ncbi.nlm.nih.gov/38540181/)]
18. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using Shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed*. Feb 2022;214:106584. [doi: [10.1016/j.cmpb.2021.106584](https://doi.org/10.1016/j.cmpb.2021.106584)] [Medline: [34942412](https://pubmed.ncbi.nlm.nih.gov/34942412/)]
19. Tangri N, Ferguson TW, Bamforth RJ, et al. Machine learning for prediction of chronic kidney disease progression: validation of the Klinrisk model in the CANVAS Program and CREDENCE trial. *Diabetes Obes Metab*. Aug 2024;26(8):3371-3380. [doi: [10.1111/dom.15678](https://doi.org/10.1111/dom.15678)] [Medline: [38807510](https://pubmed.ncbi.nlm.nih.gov/38807510/)]
20. Bai Q, Su C, Tang W, Li Y. Machine learning to predict end stage kidney disease in chronic kidney disease. *Sci Rep*. May 19, 2022;12(1):8377. [doi: [10.1038/s41598-022-12316-z](https://doi.org/10.1038/s41598-022-12316-z)] [Medline: [35589908](https://pubmed.ncbi.nlm.nih.gov/35589908/)]
21. Takkavatakarn K, Oh W, Cheng E, Nadkarni GN, Chan L. Machine learning models to predict end-stage kidney disease in chronic kidney disease stage 4. *BMC Nephrol*. Dec 19, 2023;24(1):376. [doi: [10.1186/s12882-023-03424-7](https://doi.org/10.1186/s12882-023-03424-7)] [Medline: [38114923](https://pubmed.ncbi.nlm.nih.gov/38114923/)]
22. Saito H, Yoshimura H, Tanaka K, et al. Predicting CKD progression using time-series clustering and light gradient boosting machines. *Sci Rep*. Jan 19, 2024;14(1):1723. [doi: [10.1038/s41598-024-52251-9](https://doi.org/10.1038/s41598-024-52251-9)] [Medline: [38242985](https://pubmed.ncbi.nlm.nih.gov/38242985/)]
23. Pereira AA, Sarnak MJ. Anemia as a risk factor for cardiovascular disease: management of comorbidities in kidney disease in the 21st century: anemia and bone disease. *Kidney Int*. Nov 2003;64(87):S32-9. [doi: [10.1046/j.1523-1755.64.s87.6.x](https://doi.org/10.1046/j.1523-1755.64.s87.6.x)] [Medline: [14531771](https://pubmed.ncbi.nlm.nih.gov/14531771/)]
24. Vlagopoulos PT, Tighiouart H, Weiner DE, et al. Anemia as a risk factor for cardiovascular disease and all-cause mortality in diabetes: the impact of chronic kidney disease. *J Am Soc Nephrol*. Nov 2005;16(11):3403-3410. [doi: [10.1681/ASN.2005030226](https://doi.org/10.1681/ASN.2005030226)] [Medline: [16162813](https://pubmed.ncbi.nlm.nih.gov/16162813/)]
25. van Haalen H, Jackson J, Spinowitz B, Milligan G, Moon R. Impact of chronic kidney disease and anemia on health-related quality of life and work productivity: analysis of multinational real-world data. *BMC Nephrol*. Mar 7, 2020;21(1):88. [doi: [10.1186/s12882-020-01746-4](https://doi.org/10.1186/s12882-020-01746-4)] [Medline: [32143582](https://pubmed.ncbi.nlm.nih.gov/32143582/)]
26. Bansal N, Tighiouart H, Weiner D, et al. Anemia as a risk factor for kidney function decline in individuals with heart failure. *Am J Cardiol*. Apr 15, 2007;99(8):1137-1142. [doi: [10.1016/j.amjcard.2006.11.055](https://doi.org/10.1016/j.amjcard.2006.11.055)] [Medline: [17437743](https://pubmed.ncbi.nlm.nih.gov/17437743/)]
27. Gomez-Aviles P, Gomez-Cruz JR, Almeda-Valdes P, Guerra EC, Martinez-Dominguez P, Jimenez-Hernandez M. Anemia in patients with chronic kidney disease and its association with cardiac abnormalities. *Eur Heart J*. Oct 28, 2024;45(Supplement\_1):ehae666.3271. [doi: [10.1093/eurheartj/ehae666.3271](https://doi.org/10.1093/eurheartj/ehae666.3271)]
28. Palaka E, Grandy S, van Haalen H, McEwan P, Darlington O. The impact of CKD anaemia on patients: incidence, risk factors, and clinical outcomes—a systematic literature review. *Int J Nephrol*. 2020;2020:7692376. [doi: [10.1155/2020/7692376](https://doi.org/10.1155/2020/7692376)] [Medline: [32665863](https://pubmed.ncbi.nlm.nih.gov/32665863/)]
29. Weiss JW, Johnson ES, Petrik A, Smith DH, Yang X, Thorp ML. Systolic blood pressure and mortality among older community-dwelling adults with CKD. *Am J Kidney Dis*. Dec 2010;56(6):1062-1071. [doi: [10.1053/j.ajkd.2010.07.018](https://doi.org/10.1053/j.ajkd.2010.07.018)] [Medline: [20961677](https://pubmed.ncbi.nlm.nih.gov/20961677/)]
30. Cederholm T, Barazzoni R, Austin P, et al. ESPEN guidelines on definitions and terminology of clinical nutrition. *Clin Nutr*. Feb 2017;36(1):49-64. [doi: [10.1016/j.clnu.2016.09.004](https://doi.org/10.1016/j.clnu.2016.09.004)] [Medline: [27642056](https://pubmed.ncbi.nlm.nih.gov/27642056/)]
31. Eckart A, Struja T, Kutz A, et al. Relationship of nutritional status, inflammation, and serum albumin levels during acute illness: a prospective study. *Am J Med*. Jun 2020;133(6):713-722.e7. [doi: [10.1016/j.amjmed.2019.10.031](https://doi.org/10.1016/j.amjmed.2019.10.031)] [Medline: [31751531](https://pubmed.ncbi.nlm.nih.gov/31751531/)]
32. Akirov A, Masri-Iraqi H, Atamna A, Shimon I. Low albumin levels are associated with mortality risk in hospitalized patients. *Am J Med*. Dec 2017;130(12):1465.e11. [doi: [10.1016/j.amjmed.2017.07.020](https://doi.org/10.1016/j.amjmed.2017.07.020)] [Medline: [28803138](https://pubmed.ncbi.nlm.nih.gov/28803138/)]

33. Sanz-García C, Rodríguez-García M, Górriz-Teruel JL, et al. Differences in association between hypoalbuminaemia and mortality among younger versus older patients on haemodialysis. *Clin Kidney J.* Jan 2025;18(1):sfae339. [doi: [10.1093/ckj/sfae339](https://doi.org/10.1093/ckj/sfae339)] [Medline: [39781478](https://pubmed.ncbi.nlm.nih.gov/39781478/)]
34. de la Cruz KI, Bakaen FG, Wang XL, et al. Hypoalbuminemia and long-term survival after coronary artery bypass: a propensity score analysis. *Ann Thorac Surg.* Mar 2011;91(3):671-675. [doi: [10.1016/j.athoracsur.2010.09.004](https://doi.org/10.1016/j.athoracsur.2010.09.004)] [Medline: [21352977](https://pubmed.ncbi.nlm.nih.gov/21352977/)]
35. Li M, Cheng A, Sun J, Fan C, Meng R. The role of urinary albumin-to-creatinine ratio as a biomarker to predict stroke: a meta-analysis and systemic review. *Brain Circ.* 2021;7(3):139-146. [doi: [10.4103/bc.bc\\_64\\_20](https://doi.org/10.4103/bc.bc_64_20)] [Medline: [34667897](https://pubmed.ncbi.nlm.nih.gov/34667897/)]
36. Ozyol A, Yucel O, Ege MR, Zorlu A, Yilmaz MB. Microalbuminuria is associated with the severity of coronary artery disease independently of other cardiovascular risk factors. *Angiology.* Aug 2012;63(6):457-460. [doi: [10.1177/0003319711423528](https://doi.org/10.1177/0003319711423528)] [Medline: [22007029](https://pubmed.ncbi.nlm.nih.gov/22007029/)]
37. Rein P, Saely CH, Vonbank A, Fraunberger P, Drexel H. Is albuminuria a myocardial infarction risk equivalent for atherothrombotic events? *Atherosclerosis.* May 2015;240(1):21-25. [doi: [10.1016/j.atherosclerosis.2015.02.037](https://doi.org/10.1016/j.atherosclerosis.2015.02.037)] [Medline: [25746373](https://pubmed.ncbi.nlm.nih.gov/25746373/)]
38. Blecker S, Matsushita K, Köttgen A, et al. High-normal albuminuria and risk of heart failure in the community. *Am J Kidney Dis.* Jul 2011;58(1):47-55. [doi: [10.1053/j.ajkd.2011.02.391](https://doi.org/10.1053/j.ajkd.2011.02.391)] [Medline: [21549463](https://pubmed.ncbi.nlm.nih.gov/21549463/)]
39. Kim ED, Soliman EZ, Coresh J, Matsushita K, Chen LY. Two-week burden of arrhythmias across CKD severity in a large community-based cohort: the ARIC study. *J Am Soc Nephrol.* Mar 2021;32(3):629-638. [doi: [10.1681/ASN.2020030301](https://doi.org/10.1681/ASN.2020030301)] [Medline: [33510037](https://pubmed.ncbi.nlm.nih.gov/33510037/)]
40. Jeong JH, Kim YG, Han KD, et al. Proteinuria is associated with an increased risk of sudden cardiac arrest in the young population. *J Am Heart Assoc.* Mar 18, 2025;14(6):e036077. [doi: [10.1161/JAHA.124.036077](https://doi.org/10.1161/JAHA.124.036077)] [Medline: [40079300](https://pubmed.ncbi.nlm.nih.gov/40079300/)]
41. Barzilay JI, Farag YMK, Durthaler J. Albuminuria: an underappreciated risk factor for cardiovascular disease. *J Am Heart Assoc.* Jan 16, 2024;13(2):e030131. [doi: [10.1161/JAHA.123.030131](https://doi.org/10.1161/JAHA.123.030131)] [Medline: [38214258](https://pubmed.ncbi.nlm.nih.gov/38214258/)]

## Abbreviations

**AUPRC:** area under the precision-recall curve

**AUROC:** area under the receiver operating characteristic curve

**CANVAS:** Canagliflozin Cardiovascular Assessment Study

**CatBoost:** categorical boosting

**CKD:** chronic kidney disease

**CREDESCENCE:** Canagliflozin and Renal Events in Diabetes With Established Nephropathy Clinical Evaluation

**eGFR:** estimated glomerular filtration rate

**EHR:** electronic health record

**ESRD:** end-stage renal disease

**KDIGO:** Kidney Disease: Improving Global Outcomes

**KFRE:** kidney failure risk equation

**LightGBM:** light gradient boosting machine

**RSF:** random survival forest

**SHAP:** Shapley additive explanations

**SMOTE:** Synthetic Minority Over-Sampling Technique

**T2DM:** type 2 diabetes mellitus

**UACR:** urine albumin-to-creatinine ratio

**UPCR:** urine protein-to-creatinine ratio

**XGBoost:** extreme gradient boosting

*Edited by Arriel Benis; peer-reviewed by Chen-Cheng Yang, Heng Jin; submitted 23 Jul.2025; final revised version received 18.Mar.2026; accepted 16.May.2026; published 05.Jun.2026*

*Please cite as:*

*Chen TH, Lin KH, Ho Y, Tseng WC, Chu YC, Tarng DC*

*Predicting End-Stage Renal Disease and Mortality in Chronic Kidney Disease Using Machine Learning: Retrospective Cohort Study*

*JMIR Med Inform 2026;14:e81152*

*URL: <https://medinform.jmir.org/2026/1/e81152>*

*doi: [10.2196/81152](https://doi.org/10.2196/81152)*

© Tz-Heng Chen, Kuan-Hsun Lin, Yang Ho, Wei-Cheng Tseng, Yuan-Chia Chu, Der-Cherng Tarng. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 05.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.