scientific reports

Check for updates

OPEN

Machine learning for risk prediction of acute kidney injury in patients with diabetes mellitus combined with heart failure during hospitalization

Guojing Li[1,2], Zhiqiang Zhao[1,2], Zongliang Yu[1], Junyi Liao[1] & Mengyao Zhang[1]✉

This study aimed to develop a machine learning (ML) model for predicting the risk of acute kidney injury (AKI) in diabetic patients with heart failure (HF) during hospitalization. Using data from 1,457 patients in the MIMIC-IV database, the study identified twenty independent risk factors for AKI through LASSO regression and logistic regression. Six ML algorithms were evaluated, including LightGBM, random forest, and neural networks. The LightGBM model demonstrated superior performance with the highest prediction accuracy, with AUC values of 0.973 and 0.804 in the training and validation sets, respectively. The Shapley additive explanations algorithm was used to visualize the model and identify the most relevant features for AKI risk. Clinical impact curves further confirmed the strong discriminatory ability and generalizability of the LightGBM model. This study highlights the potential of ML models, particularly LightGBM, to effectively predict AKI risk in diabetic patients with HF, enabling early identification of high-risk patients and timely interventions to improve prognosis.

Keywords Acute kidney injury, Diabetes mellitus, Heart failure, Machine learning, Prediction model, MIMIC-IV database

Abbreviations

AKI	Acute kidney injury
ICUs	Intensive care units
HF	Heart failure
DM	Diabetes mellitus
LR	Logistic regression
ML	Machine learning
BIDMC	Beth Israel Deaconess Medical Center
ICD-9	International Classification of Diseases 9th edition
RRT	Renal replacement therapy
CRRT	Continuous RRT
HR	Heart rate
RR	Respiratory rate
BP	Blood pressure
PT	Prothrombin time
INR	International normalized ratio
PTT	Partial thromboplastin time
NT-PROBNP	Triglycerides, N-terminal pro-brain natriuretic peptide
GCS	Glasgow coma scale
APSIII	Acute physiology score III
SIRS	Systemic inflammatory response syndrome

[1]Department of Cardiology, Kunshan First People's Hospital, Affiliated Kunshan Hospital of Jiangsu University, No.566 Tongfeng East Road, Kunshan Development Zone, Kunshan, Suzhou City 215300, Jiangsu Province, China. [2]Guojing Li and Zhiqiang Zhao have contributed equally to this work and share first authorship. ✉email: ksshyy2006@sina.com

SOFA	Sequential organ failure assessment
ACEIs	Angiotensin-converting enzyme inhibitors
ARBs	Angiotensin II receptor blockers
ECMO	Extracorporeal membrane oxygenation
eGFR	Estimated glomerular filtration rate
AIC	Akaike information criteria
LightGBM	Light gradient boosting
RF	Random forest
GBM	Gradient boosting
NN	Neural network
KNN	K nearest neighbors
ROC	Receiver operating characteristic
DCA	Decision curve analysis
SHAP	Shapley additive explanations
CIC	Clinical impact curve

Acute kidney injury (AKI) is a complex heterogeneous syndrome. It presents as a rapid deterioration in kidney function and is frequently observed in intensive care units (ICUs)[1,2]. Currently, AKI has become a significant global public health issue. The incidence rate is as high as 60-70%[3,4]. The incidence of heart failure (HF) continues to increase with population growth and aging, affecting more than 40 million people worldwide[5]. It is important to note that AKI is increasingly prevalent among hospitalized patients with cardiovascular disease. It occurs in up to 47% of patients with HF[6,7]. In recent years, there has been increasing evidence of a strong association between diabetes mellitus (DM) and AKI. DM is an independent risk factor for AKI, with approximately 10–20% of DM patients experiencing this condition[8,9]. Mina et al.[10] demonstrated that patients with DM, especially older people and those with other comorbidities like heart failure, have a notably higher risk of developing AKI. DM not only increases HF risk, but also leads to a poor prognosis, especially in ICU patients[11,12]. DM and HF often occur simultaneously, significantly increasing the risk of AKI in patients. The mechanism may be related to various factors such as altered renal hemodynamics, inflammatory response, oxidative stress, and drug toxicity (Fig. 1)[13,14].

AKI not only leads to increased medical costs and prolonged hospital stays but also results in higher risks of mortality and adverse renal events. The mortality for hospitalized patients with AKI is 10.8%[15,16]. Therefore,

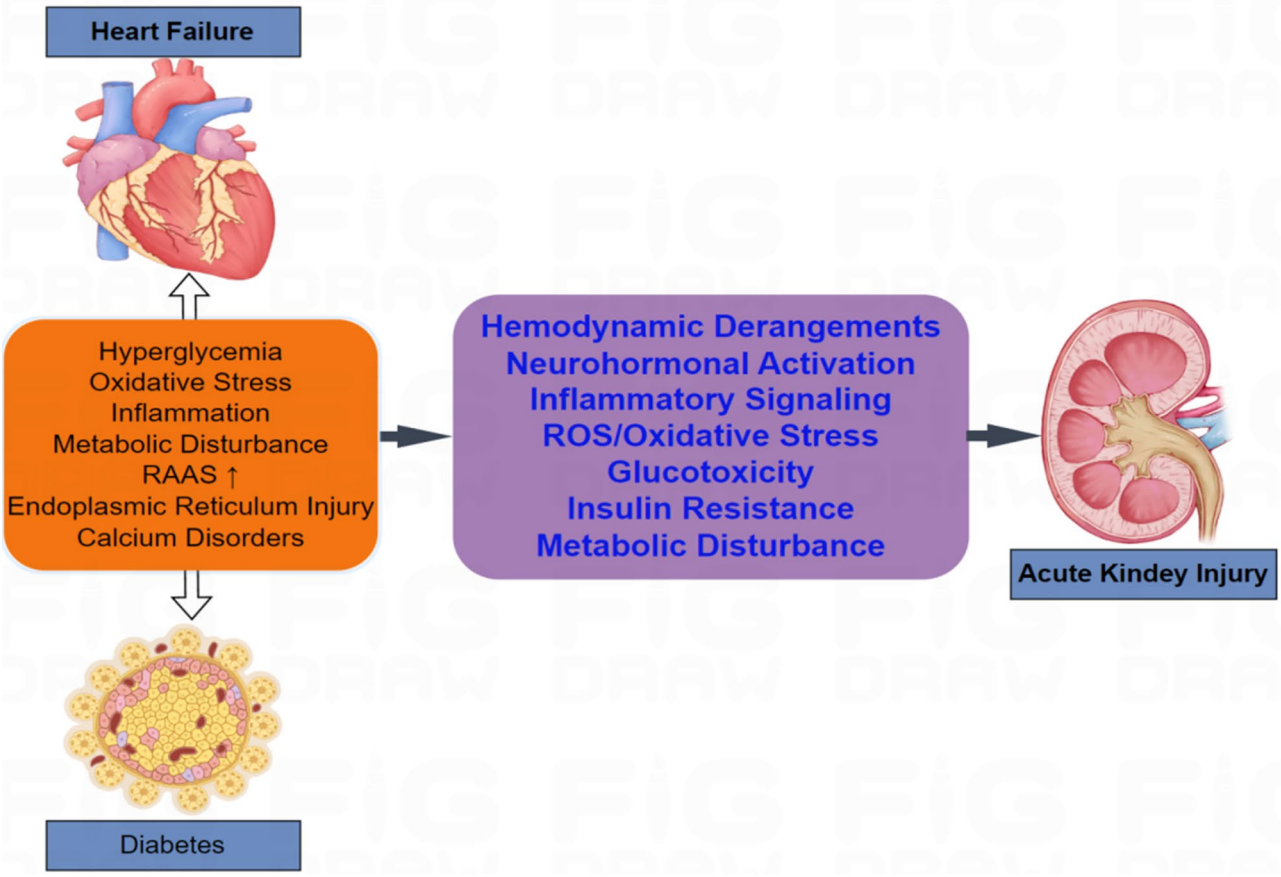

Fig. 1. Diabetes mellitus combined with heart failure leading to acute kidney injury.

early identification and diagnosis of AKI can provide valuable references for clinicians. This enables clinicians to reverse early AKI through timely intervention, thereby improving patient prognosis and reducing AKI-related mortality[17].

Traditional AKI risk assessment methods, such as clinical scoring systems like the risk injury failure loss of function end stage renal disease criteria and the acute kidney injury network criteria, are primarily employed to assess the severity of AKI. However, these methods are not very effective in predicting the risk of developing AKI[18]. In addition, traditional predictive models, such as logistic regression (LR) and Cox proportional risk models, may fail to capture complex nonlinear relationships due to their linear assumptions, limiting predictive performance[19]. Compared with traditional prediction models, machine learning (ML), as a powerful data analysis tool, can better handle complex nonlinear relationships and shows great potential in disease prediction[20]. ML algorithms can learn complex patterns from a large number of patient data, identify potential risk factors, and be used to build prediction models, which have shown good prediction performance in predicting cardiovascular diseases and other diseases[20–22].

Currently, there is no existing risk prediction model for AKI in DM patients with HF. As of August 2024, a search utilizing the keywords "acute kidney injury", "diabetes", "heart failure", "prediction", and "machine learning" in the PubMed, Google Scholar, Embase, and Web of Science databases found no published AKI risk prediction models for this specific population. Therefore, this study aims to develop six ML models for predicting AKI risk during hospitalization in DM patients with HF. Ultimately, the goal is to identify the model that demonstrates the best predictive performance. This model will help clinicians better identify high-risk patients and take timely interventions to improve patient prognosis.

Methods

Data sources

The data were sourced from the large critical care database of MIMIC-IV 2.2. The database used in this study received approval from both the Beth Israel Deaconess Medical Center (BIDMC) and the Massachusetts Institute of Technology Institutional Review Board. It encompasses detailed clinical data on patients admitted to the BIDMC ICU between 2008 and 2019, including demographic information, laboratory indicators, medication records, and vital signs. Data privacy was achieved by removing patient identification numbers for anonymization and offset over time. Therefore, patient consent was not required for this study[23]. After signing a data use agreement, completing the National Institutes of Health web-based training course and the protecting human study exam, Guojing Li (certification No.: 12410114) was certified and entitled to data access and use privileges.

Study population

Patients meeting the following criteria were included: (i) aged 19 to 89 years old; (ii) admitted to ICU for the first time and stayed in the hospital for more than 24 h; (iii) DM patients with HF diagnosed according to the codes in the International Classification of Diseases 9th edition (ICD-9) (249, 250, and 428) and ICD-10 (E08, E09, E10, E11, E13, and I50); and (iv) DM patients with HF that were assessed for the development of AKI within 30 days of hospitalization according to the criteria of the Kidney Disease: Improving Global Outcomes guideline (Table 1)[2]. Patients were excluded due to: (i) no ICU hospitalization record; (ii) follow-up time less than 24 h or greater than 30 days; (iii) previous history of AKI; and (iv) renal replacement therapy (RRT) or continuous RRT (CRRT) on admission. In the end, we included 1,457 patients. The study population screening flow chart is shown in Fig. 2.

Data extraction and missing value processing

This study used structured query language to extract complete electronic medical record data from the MIMIC-IV database. The data pertained to the first ICU admission of DM patients with HF who met the inclusion criteria. In the study, the following basic characteristics of the patients were extracted: (i) demographic characteristics, including age, gender, race, marriage, type of admission, length of hospitalization, and whether or not patients died in-hospital; (ii) vital signs, including temperature, heart rate (HR), respiratory rate (RR), blood pressure (BP), and oxygen saturation (SpO2); (iii) laboratory parameters, including serum creatinine, lactate, PH, anion gap, oxygen partial pressure, carbon dioxide partial pressure, bicarbonate, base excess, potassium, sodium, chloride, calcium, phosphate, white blood cell count, red blood cell count, platelet count, red blood cell distribution width, mean red blood cell volume, and hematocrit, hemoglobin, glucose, blood urea nitrogen, prothrombin time (PT), international normalized ratio (INR), partial thromboplastin time (PTT), triglycerides,

Stage	Serum creatinine	Urine output
AKI definition	0.3 mg/dL or higher increase within 48 h; and 1.5 times or more increase within 7 days	< 0.5 mL/(kg·h) for < 6 h
AKI staging		
Stage 1	1.5 ~ 1.9 times baseline or ≥ 26.5 μmol/L (0.3 mg/dl) increase	< 0.5 ml/(kg·h) for 6 ~ 12 h
Stage 2	2.0 ~ 2.9 times baseline increase	< 0.5 ml/(kg·h) for ≥ 12 h
Stage 3	3 times baseline or absolute value ≥ 353.6 μmol/L (4.0 mg/dl) increase, or initiation of RRT, or in patients < 18 years old and eGFR < 35 ml/(min·1.73m^2)	< 0.3 ml/(kg·h) for ≥ 24 h or anuria ≥ 12 h

Table 1. Definition and staging standard of acute kidney injury in the kidney disease: improving global outcomes.

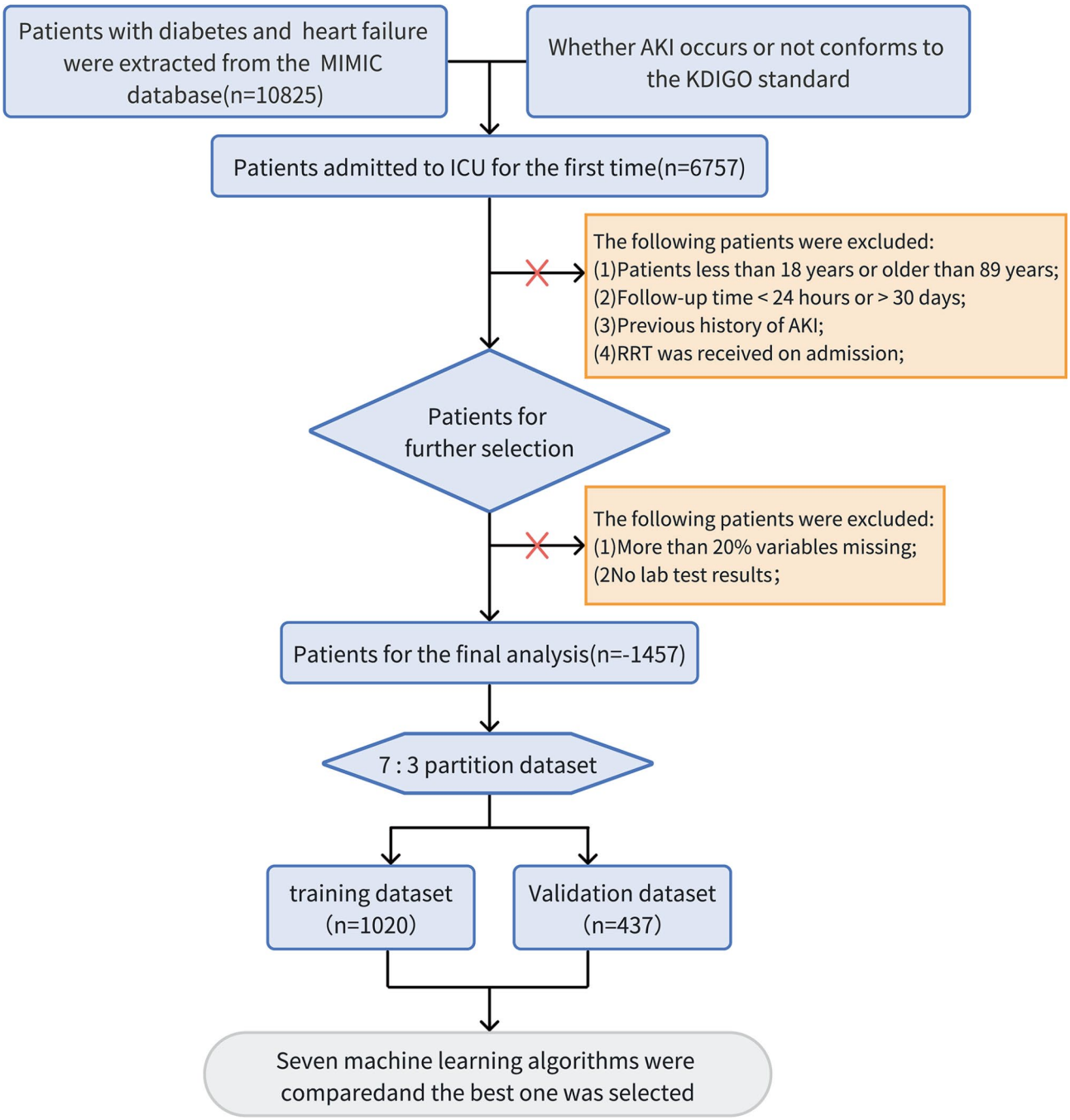

Fig. 2. Patient selection process flow chart.

N-terminal pro-brain natriuretic peptide (NT-PROBNP), aspartate aminotransferase, glycated hemoglobin A 1c, and urinary output; (iv) comorbidities, including anemia, hyperlipidemia, hypertension, respiratory failure, pneumonia, hypotension, sepsis (Sepsis 3), cerebrovascular disease, atrial fibrillation, atherosclerosis, myocardial infarction, obesity, electrolyte disorders, sleep disorders, pulmonary heart disease, urinary tract obstruction, and urinary tract infection; (v) personal history, including alcoholism and smoking; (vi) surgical history, including percutaneous coronary intervention, revascularization, coronary artery bypass grafting, intra-aortic balloon pump, and urologic surgery; (vii) scores related to the severity of diseases, including the Glasgow coma scale (GCS) score, the acute physiology score III (APSIII), the Oxford acute severity of illness score, the simplified acute physiology score (SAPSII), the systemic inflammatory response syndrome (SIRS) criteria, and the sequential organ failure assessment (SOFA) score; and (viii) drugs and interventions, including antiarrhythmics, diuretics, angiotensin-converting enzyme inhibitors (ACEIs)/angiotensin II receptor blockers (ARBs), aldosterone receptor antagonists, beta-blockers, statins, nonsteroidal anti-inflammatory drugs, insulin, positive inotropic agents, anticoagulants, albumin, calcium gluconate, antibiotics, calcium channel blockers,

mechanical ventilation, RRT, and extracorporeal membrane oxygenation (ECMO). The estimated glomerular filtration rate (eGFR) was determined using the equation: 175 * serum creatinine – 1.154 * age – 0.203 (* 0.742 if female)[24].

For laboratory indicators that were measured multiple times within 24 h of admission, the first measurement was used in this study. To mitigate reverse-causation bias, data collected after the occurrence of an outcome event were deemed invalid. Additionally, to minimize the bias stemming from missing data, this study employed multiple imputation (MI) based on random forest (RF) for variables with less than 20% missing values. The mice package in R software was leveraged for imputation, generating five imputed datasets[25,26]. Variables with missing values ≥ 20%, including lactate, pH, INR, SOFA, and GCS scores, were kept in the model owing to their clinical relevance, and missing data were handled during model training. Variables with missing data exceeding 20% and considered to have lower significance for AKI prediction were excluded from the model. Sepsis was diagnosed through Sepsis 3.0, and other comorbidities were diagnosed through ICD-9 and ICD-10[27]. The primary study endpoint was the occurrence of AKI within 30 days of ICU admission in DM patients with HF.

Statistical methods

This study used a multistep approach to develop and validate ML models for predicting AKI risk in DM patients with HF. The data were initially tested for normality by the Shapiro-Wilk method. Descriptive statistics and hypothesis testing were also used according to the type of data. Categorical variables were expressed as percentages (%). Differences were tested using Pearson's chi-square test or Fisher's exact probability method. Continuous variables were described using mean ± standard deviation or interquartile range [IQR], respectively, depending on normal distribution. In addition, Student's t-test or Wilcoxon rank sum test/Mann-Whitney U-test was employed to compare differences between groups, respectively[28].

To screen for key risk factors of AKI, LASSO regression, and stepwise LR models were used for feature selection. The LASSO regression model was employed to identify the optimal λ value based on the principle of minimum mean square error and one-standard error. Variables with non-zero coefficients were then screened. Subsequently, the screened variables were included in a stepwise LR model. Moreover, the ORs and 95% CIs of the stepwise LR model were calculated to further screen for significant risk factors for AKI using the Akaike information criteria (AIC) as a criterion ($P < 0.05$)[29].

Based on the screened features, six ML algorithms were utilized in this study to develop AKI prediction models, namely LR, light gradient boosting (LightGBM), RF, gradient boosting (GBM), neural network (NN), and K nearest neighbors (KNN). To prevent overfitting and optimize model performance, 5-fold cross-validation, and 100-bootstrap resampling were used to select parameters for models. Model performance was ultimately comprehensively assessed by receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA), including metrics such as the area under the ROC curve (AUC), sensitivity, specificity, calibration, and net clinical benefit. A nomogram was plotted to visualize the predicted results of the models. To further interpret the predicted results of models and identify key risk factors, the best-performing model was selected in this study. Moreover, the contribution of each feature to AKI prediction was analyzed using the Shapley additive explanations (SHAP) method. SHAP values can visually demonstrate the importance of each feature. The association between certain features and predicted results was analyzed in depth through partial dependence plots. Additionally, a clinical impact curve (CIC) based on the optimal model was plotted in this study to assess the clinical practicality and applicability of the model.

All statistical analyses and modeling procedures were done using R software (version 4.0.5). A two-sided P value of less than 0.05 was deemed to be statistically significant.

Results

Baseline characteristics

This study classified 1,457 DM patients with HF into the AKI group ($n = 1212$) and non-AKI group ($n = 245$) based on KDIGO standards. Table 2 presents the baseline characteristics of both groups, including median (interquartile range) or frequency (percentage) for each variable. Firstly, a marked difference in racial distribution was found ($P = 0.004$). While white patients accounted for 63.8% of the total population, they represented 65.1% of the AKI group, showing an increase compared to the non-AKI group (57.1%). Meanwhile, the proportion of Asian and Black patients in the AKI group was lower, indicating that race could be a factor influencing the incidence of AKI. Notably, the hospital mortality in the AKI group was noticeably higher than in the non-AKI group (14.5% vs. 6.1%, $P = 0.001$), suggesting a link between AKI and adverse outcomes. In terms of comorbidities, the incidence of anemia was notably higher in AKI patients compared to non-AKI patients (66.3% vs. 55.5%, $P = 0.002$). Additionally, the incidence of atrial fibrillation, arrhythmias, and obesity was also evidently higher in the AKI group (atrial fibrillation: 57.3% vs. 43.3%, $P < 0.001$; arrhythmias: 70.9% vs. 57.1%, $P < 0.001$; obesity: 31.9% vs. 18.8%, $P < 0.001$). The incidence of chronic obstructive pulmonary disease was noticeably higher in the AKI group compared to the non-AKI group (18.4% vs. 24.9%, $P = 0.024$), indicating that such chronic conditions may elevate the risk of AKI. Moreover, urine output was markedly lower in the AKI group than in the non-AKI group (70.00 ml vs. 130.00 ml, $P < 0.001$), emphasizing the severity of AKI, as this parameter is a critical indicator of kidney function. Finally, the SOFA scores and APACHE II scores, among other critical indicators, in the AKI group were notably higher than those in the non-AKI group ($P < 0.001$). This finding suggested a close connection between the severity of illness and the occurrence of AKI.

Furthermore, to construct and validate the prediction model, this study randomly divided 1,457 patients into a training set ($n = 1020$) and a validation set ($n = 437$) in a ratio of 7:3 (Table 3). P-values of both sets for features such as demographics, laboratory indices, vital signs, comorbidities, medications, and interventions were greater than 0.05. This demonstrated that no significant statistical differences in features were observed between the two groups of patients. Moreover, the two groups were comparable. Overall, AKI occurred within 30 days of

Characteristics	Total($n = 1457$)	Non-AKI($n = 245$)	AKI($n = 1212$)	*P* value
Race (%)				0.004
Asian	39 (2.7)	11 (4.5)	28 (2.3)	
Black	109 (7.5)	29 (11.8)	80 (6.6)	
Other	380 (26.1)	65 (26.5)	315 (26.0)	
White	929 (63.8)	140 (57.1)	789 (65.1)	
Marital status (%)				0.923
Divorced	106 (7.3)	19 (7.8)	87 (7.2)	
Married	644 (44.2)	102 (41.6)	542 (44.7)	
Other	177 (12.1)	30 (12.2)	147 (12.1)	
Single	299 (20.5)	52 (21.2)	247 (20.4)	
Widowed	231 (15.9)	42 (17.1)	189 (15.6)	
Admission type (%)				0.096
Elective	66 (4.5)	7 (2.9)	59 (4.9)	
Emergency	567 (38.9)	101 (41.2)	466 (38.4)	
Observation	272 (18.7)	55 (22.4)	217 (17.9)	
Surgical	103 (7.1)	20 (8.2)	83 (6.8)	
Urgent	449 (30.8)	62 (25.3)	387 (31.9)	
Hospital expire flag = Yes (%)	191 (13.1)	15 (6.1)	176 (14.5)	0.001
Gender = Male (%)	868 (59.6)	139 (56.7)	729 (60.1)	0.357
Age (median [IQR])	72.83 [64.88, 80.43]	72.93 [63.44, 80.79]	72.82 [65.04, 80.40]	0.999
Anemias = Yes (%)	940 (64.5)	136 (55.5)	804 (66.3)	0.002
Anxiety states = Yes (%)	230 (15.8)	45 (18.4)	185 (15.3)	0.263
Asthma = Yes (%)	172 (11.8)	25 (10.2)	147 (12.1)	0.457
Atrial fibrillation = Yes (%)	801 (55.0)	106 (43.3)	695 (57.3)	< 0.001
Cardiac dysrhythmias = Yes (%)	999 (68.6)	140 (57.1)	859 (70.9)	< 0.001
Cardiomyopathy = Yes (%)	242 (16.6)	40 (16.3)	202 (16.7)	0.971
Cerebrovascular disease = Yes (%)	271 (18.6)	37 (15.1)	234 (19.3)	0.146
Chronic kidney disease = Yes (%)	632 (43.4)	109 (44.5)	523 (43.2)	0.753
COPD = Yes (%)	284 (19.5)	61 (24.9)	223 (18.4)	0.024
Depressive disorder = Yes (%)	309 (21.2)	55 (22.4)	254 (21.0)	0.663
Esophageal diseases = Yes (%)	492 (33.8)	86 (35.1)	406 (33.5)	0.682
Gastrointestinal hemorrhage = Yes (%)	116 (8.0)	13 (5.3)	103 (8.5)	0.12
Hemorrhagic conditions = Yes (%)	332 (22.8)	45 (18.4)	287 (23.7)	0.085
Hyperlipidemia = Yes (%)	1020 (70.0)	176 (71.8)	844 (69.6)	0.543
Hypernatremia = Yes (%)	229 (15.7)	32 (13.1)	197 (16.3)	0.248
Hyperpotassemia = Yes (%)	258 (17.7)	44 (18.0)	214 (17.7)	0.983
Hypertension = Yes (%)	1286 (88.3)	216 (88.2)	1070 (88.3)	1
Hyponatremia = Yes (%)	269 (18.5)	36 (14.7)	233 (19.2)	0.115
Hypotension = Yes (%)	498 (34.2)	58 (23.7)	440 (36.3)	0.015
Ischemic heart disease = Yes (%)	1100 (75.5)	182 (74.3)	918 (75.7)	0.688
Liver disease = Yes (%)	215 (14.8)	27 (11.0)	188 (15.5)	0.087
Malignant neoplasm = Yes (%)	182 (12.5)	36 (14.7)	146 (12.0)	0.3
Metabolic and immunity disorders = Yes (%)	1097 (75.3)	186 (75.9)	911 (75.2)	0.866
Mineral metabolism disorders = Yes (%)	157 (10.8)	31 (12.7)	126 (10.4)	0.354
Myocardial infarction = Yes (%)	600 (41.2)	88 (35.9)	512 (42.2)	0.078
Obesity = Yes (%)	433 (29.7)	46 (18.8)	387 (31.9)	< 0.001
Peripheral nervous system diseases = Yes (%)	185 (12.7)	23 (9.4)	162 (13.4)	0.109
Peripheral vascular disease = Yes (%)	498 (34.2)	66 (26.9)	432 (35.6)	0.011
Pneumonia = Yes (%)	383 (26.3)	42 (17.1)	341 (28.1)	< 0.001
Respiratory failure = Yes (%)	651 (44.7)	85 (34.7)	566 (46.7)	0.001
Shock = Yes (%)	484 (33.2)	56 (22.9)	428 (35.3)	< 0.001
Sleep disorders = Yes (%)	371 (25.5)	49 (20.0)	322 (26.6)	0.038
Smoke = Yes (%)	732 (50.2)	134 (54.7)	598 (49.3)	0.145
Systemic inflammatory response syndrome = Yes (%)	352 (24.2)	47 (19.2)	305 (25.2)	0.056
Urinary obstruction = Yes (%)	236 (16.2)	54 (22.0)	182 (15.0)	0.009
Urinary tract infection = Yes (%)	356 (24.4)	59 (24.1)	297 (24.5)	0.953
Continued				

Characteristics	Total(n = 1457)	Non-AKI(n = 245)	AKI(n = 1212)	P value
Valvular heart disease = Yes (%)	418 (28.7)	66 (26.9)	352 (29.0)	0.557
Volume depletion = Yes (%)	212 (14.6)	48 (19.6)	164 (13.5)	0.019
Time (median [IQR])	1.65 [0.71, 4.79]	6.69 [4.72, 9.08]	1.34 [0.61, 3.18]	< 0.001
AG (median [IQR])	15.00 [13.00, 18.00]	15.00 [12.00, 17.00]	15.00 [13.00, 18.00]	0.593
BE (median [IQR])	0.00 [-2.00, 2.00]	0.00 [-3.00, 2.00]	0.00 [-2.00, 2.00]	0.155
BUN (median [IQR])	25.00 [18.00, 39.00]	26.00 [17.00, 41.00]	25.00 [18.00, 39.00]	0.669
Calcium (median [IQR])	8.60 [8.10, 9.00]	8.50 [8.10, 9.10]	8.60 [8.10, 9.00]	0.411
Chloride (median [IQR])	102.00 [99.00, 106.00]	103.00 [99.00, 107.00]	102.00 [99.00, 106.00]	0.15
Creatinine (median [IQR])	1.20 [0.90, 2.10]	1.10 [0.90, 1.70]	1.60 [0.90, 2.60]	0.032
Glucose (median [IQR])	156.00 [121.00, 216.00]	152.00 [111.00, 205.00]	157.00 [123.00, 216.00]	0.02
Hb (median [IQR])	11.10 [9.60, 12.60]	10.90 [9.70, 12.70]	11.10 [9.60, 12.60]	0.941
HCO (median [IQR])	24.00 [21.00, 27.00]	24.00 [20.00, 26.00]	24.00 [21.00, 27.00]	0.032
HCT (median [IQR])	33.90 [29.60, 38.30]	33.60 [30.30, 38.30]	34.00 [29.50, 38.32]	0.927
INR (median [IQR])	1.20 [1.10, 1.50]	1.20 [1.10, 1.50]	1.20 [1.10, 1.50]	0.006
Lactate (median [IQR])	1.60 [1.10, 2.20]	1.50 [1.20, 2.20]	1.60 [1.10, 2.20]	0.82
MCV (median [IQR])	90.00 [87.00, 94.00]	91.00 [87.00, 94.00]	90.00 [87.00, 94.25]	0.227
PCO2 (median [IQR])	42.00 [36.00, 48.00]	42.00 [37.00, 47.00]	42.00 [36.00, 48.00]	0.969
PH (median [IQR])	7.38 [7.32, 7.43]	7.37 [7.32, 7.42]	7.38 [7.32, 7.43]	0.099
Phosphate (median [IQR])	3.60 [3.10, 4.40]	3.70 [3.10, 4.40]	3.60 [3.10, 4.40]	0.977
PLT (median [IQR])	203.00 [156.00, 253.00]	199.00 [154.00, 256.00]	203.12 [156.00, 253.00]	0.849
PO2 (median [IQR])	137.00 [66.00, 321.00]	113.00 [58.00, 309.00]	141.00 [69.00, 326.25]	0.028
Potassium (median [IQR])	4.30 [3.90, 4.70]	4.30 [3.90, 4.80]	4.30 [3.90, 4.70]	0.429
PT (median [IQR])	13.60 [12.20, 16.00]	13.10 [11.80, 15.90]	13.80 [12.30, 16.00]	0.003
PTT (median [IQR])	31.30 [27.60, 40.60]	31.10 [27.90, 37.00]	31.40 [27.60, 41.23]	0.277
RBC (median [IQR])	3.77 [3.27, 4.29]	3.75 [3.22, 4.30]	3.78 [3.28, 4.28]	0.514
RDW (median [IQR])	14.50 [13.50, 15.80]	14.50 [13.40, 15.50]	14.50 [13.50, 15.90]	0.187
Sodium (median [IQR])	139.00 [136.00, 141.00]	139.00 [136.00, 141.00]	139.00 [136.00, 141.00]	0.489
WBC (median [IQR])	10.30 [7.60, 14.30]	10.50 [7.70, 14.20]	10.30 [7.60, 14.40]	0.929
DBP (median [IQR])	62.00 [53.00, 74.00]	62.00 [52.00, 76.00]	62.00 [53.00, 73.00]	0.757
HR (median [IQR])	84.00 [75.00, 98.00]	83.00 [75.00, 96.00]	85.00 [75.00, 98.00]	0.172
MBP (median [IQR])	79.00 [69.00, 91.00]	79.00 [69.00, 91.00]	79.00 [69.00, 91.00]	0.843
RR (median [IQR])	17.00 [14.00, 22.00]	17.00 [14.00, 22.00]	17.00 [14.00, 22.00]	0.836
SBP (median [IQR])	120.00 [104.00, 137.00]	121.00 [104.00, 137.00]	119.00 [104.00, 137.00]	0.457
SPO2 (median [IQR])	99.00 [95.00, 100.00]	98.00 [95.00, 100.00]	99.00 [96.00, 100.00]	0.229
ECG = Yes (%)	725 (49.8)	127 (51.8)	598 (49.3)	0.52
MV = Yes (%)	979 (67.2)	112 (45.7)	867 (71.5)	< 0.001
Cardiovascular pacer data = Yes (%)	508 (34.9)	81 (33.1)	427 (35.2)	0.564
ECMO = Yes (%)	3 (0.2)	1 (0.4)	2 (0.2)	1
IABP = Yes (%)	76 (5.2)	6 (2.4)	70 (5.8)	0.048
Adrenaline receptor agonist = Yes (%)	896 (61.5)	124 (50.6)	772 (63.7)	< 0.001
Antiarrhythmic drugs = Yes (%)	911 (62.5)	201 (82.0)	710 (58.6)	< 0.001
Antibiotics = Yes (%)	924 (63.4)	164 (66.9)	760 (62.7)	0.237
CCB = Yes (%)	210 (14.4)	38 (15.5)	172 (14.2)	0.663
Diuretic = Yes (%)	742 (50.9)	189 (77.1)	553 (45.6)	< 0.001
β-blocker = Yes (%)	794 (54.5)	187 (76.3)	607 (50.1)	< 0.001
Positive inotropic = Yes (%)	154 (10.6)	25 (10.2)	129 (10.6)	0.928
Vasoconstrictor drug = Yes (%)	577 (39.6)	75 (30.6)	502 (41.4)	0.002
NASIDS = Yes (%)	1043 (71.6)	195 (79.6)	848 (70.0)	0.003
Insulin = Yes (%)	1321 (90.7)	223 (91.0)	1098 (90.6)	0.929
Anticoagulants = Yes (%)	1089 (74.7)	213 (86.9)	876 (72.3)	< 0.001
NE = Yes (%)	393 (27.0)	48 (19.6)	345 (28.5)	0.006
Epinephrine = Yes (%)	203 (13.9)	26 (10.6)	177 (14.6)	0.122
Milrinone = Yes (%)	59 (4.0)	8 (3.3)	51 (4.2)	0.614
Vasopressin = Yes (%)	99 (6.8)	10 (4.1)	89 (7.3)	0.087
Phenylephrine = Yes (%)	511 (35.1)	68 (27.8)	443 (36.6)	0.011
Digoxin = Yes (%)	65 (4.5)	13 (5.3)	52 (4.3)	0.594
Continued				

Characteristics	Total($n=1457$)	Non-AKI($n=245$)	AKI($n=1212$)	*P* value
ACEI/ARB = Yes (%)	320 (22.0)	110 (44.9)	210 (17.3)	< 0.001
Albumin = Yes (%)	298 (20.5)	52 (21.2)	246 (20.3)	0.809
Antianemia drug = Yes (%)	77 (5.3)	23 (9.4)	54 (4.5)	0.003
Antiplatelet drugs = Yes (%)	798 (54.8)	182 (74.3)	616 (50.8)	< 0.001
Aspirin = Yes (%)	770 (52.8)	180 (73.5)	590 (48.7)	< 0.001
Immunosuppress = Yes (%)	10 (0.7)	5 (2.0)	5 (0.4)	0.017
MRA = Yes (%)	52 (3.6)	20 (8.2)	32 (2.6)	< 0.001
Statins = Yes (%)	808 (55.5)	178 (72.7)	630 (52.0)	< 0.001
Blood products or colloids = Yes (%)	623 (42.8)	89 (36.3)	534 (44.1)	0.031
Calcium gluconate = Yes (%)	432 (29.6)	81 (33.1)	301 (24.8)	0.018
Dobutamine = Yes (%)	34 (2.3)	5 (2.0)	29 (2.4)	0.92
Dopamine = Yes (%)	56 (3.8)	8 (3.3)	48 (4.0)	0.738
Vancomycin = Yes (%)	586 (40.2)	102 (41.6)	484 (39.9)	0.672
β-lactam antibiotics = Yes (%)	715 (49.1)	144 (58.8)	571 (47.1)	0.012
SIRS (median [IQR])	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	0.002
SAPSII (median [IQR])	39.00 [32.00, 48.00]	35.00 [29.00, 42.00]	40.00 [33.00, 49.00]	< 0.001
OASIS (median [IQR])	33.00 [28.00, 39.00]	29.00 [25.00, 34.00]	34.00 [29.00, 40.00]	< 0.001
APSIII (median [IQR])	44.00 [34.00, 59.00]	39.00 [31.00, 50.00]	45.00 [35.00, 61.00]	< 0.001
SOFA (median [IQR])	5.00 [3.00, 8.00]	5.00 [3.00, 7.00]	6.00 [4.00, 8.00]	< 0.001
GCS (median [IQR])	15.00 [15.00, 15.00]	15.00 [15.00, 15.00]	15.00 [15.00, 15.00]	0.66
Sepsis3 = Yes (%)	1011 (69.4)	133 (54.3)	878 (72.4)	< 0.001
Urine output (median [IQR])	80.00 [40.00, 150.00]	130.00 [75.00, 250.00]	70.00 [40.00, 140.00]	< 0.001
RRT = Yes (%)	53 (3.6)	5 (2.0)	48 (3.9)	0.049
Ventilation = Yes (%)	1319 (90.5)	208 (84.9)	1111 (91.7)	0.001
CAGB = Yes (%)	409 (28.1)	62 (25.3)	347 (28.6)	0.328
Cardiovascular surgery = Yes (%)	1121 (76.9)	145 (59.2)	976 (80.5)	< 0.001
Dialysis = Yes (%)	53 (3.6)	1 (0.4)	52 (4.3)	0.006
PCI = Yes (%)	45 (3.1)	4 (1.6)	41 (3.4)	0.214
Urinary system surgery = Yes (%)	19 (1.3)	5 (2.0)	14 (1.2)	0.42
Vascular reconstruction = Yes (%)	281 (19.3)	34 (13.9)	247 (20.4)	0.024
EGFR (median [IQR])	58.14 [38.78, 84.79]	58.05 [39.45, 80.77]	58.15 [38.25, 85.09]	0.936

Table 2. Baseline characteristics of patients in AKI group and non-AKI group.

hospitalization in 1,212 patients (83.2%), and 191 patients (13.1%) died during hospitalization. Demographically, 868 patients (59.6%) were male, accounting for the majority of the total cohort. Moreover, 929 patients (63.8%) were white. 567 patients (38.9%) were admitted as emergency cases. The patients had a median age of 72.8 years old [IQR: 64.9–80.4]. The median follow-up time was 1.65 days [IQR: 0.71–4.79], and the median eGFR value was 58.14 mL/min [38.78, 84.79] ($P=0.098$). DM patients with HF were more likely to suffer from anemia (64.5%), atrial fibrillation (55.0%), hyperlipidemia (70.0%), hypertension (88.3%), and sepsis (69.4%).

Feature selection

The results of cross-validation and coefficient curves for variable screening by LASSO are shown in Fig. 3. Figure 3a is a trend diagram of model error changing with penalty coefficient λ in the cross-validation process of LASSO regression. Figure 3b illustrates the variation curve of variable coefficients in LASSO regression. Specifically, the x-axis at the bottom of Fig. 3b represents the log(λ) values, while the x-axis at the top indicates the number of selected variables. Each line represents the coefficient trend of a variable, while the y-axis displays the coefficient for that variable. With an increase in the log(λ) value, the coefficient of the variable gradually tends toward 0, reflecting a diminishing importance of the feature. Once the optimal λ value is selected, the number of selected variables decreases, accomplishing the objective of feature selection. In order to obtain a high-performance and streamlined model, this study used 10-fold cross-validation on the LASSO regression model, selecting the number of variables ($n=40$) and parameter λ ($\lambda=0.0198$) corresponding to one standard deviation from the minimum mean square error (vertical dotted line on the right side of Fig. 3a). There were 40 risk factors with non-zero coefficients that were associated with AKI within 30 days of hospitalization in DM patients with HF.

40 variables were analyzed by the stepwise LR. Variables with $P\geq0.05$ were excluded through bidirectional stepwise selection based on AIC. The final prediction models included 20 features identified as independent predictors of AKI in DM patients with HF, including admission type, anemia, atrial fibrillation, hypotension, obesity, pneumonia, abnormal creatinine values, antiarrhythmics, diuretics, anticoagulants, ACEI/ARB, aspirin, statins, calcium gluconate, β lactam antibiotics, RRT, ECMO, APSIII score, sepsis, and a history of cardiovascular

Characteristics	Total($n = 1457$)	Train($n = 1020$)	Test($n = 437$)	p
Aki stage = Yes (%)	1212 (83.2)	845 (82.8)	367 (84.0)	0.648
Time (median [IQR])	1.65 [0.71, 4.79]	1.65 [0.74, 4.89]	1.67 [0.67, 4.39]	0.462
Hospital expire flag = Yes (%)	191 (13.1)	136 (13.3)	55 (12.6)	0.762
Demographic variables, n (%)				
Ethnicity (%)				0.869
Asian	39 (2.7)	25 (2.5)	14 (3.2)	
Black	109 (7.5)	77 (7.5)	32 (7.3)	
Other	380 (26.1)	268 (26.3)	112 (25.6)	
White	929 (63.8)	650 (63.7)	279 (63.8)	
Marital status (%)			0.96	
Divorced		106 (7.3)	72 (7.1)	34 (7.8)
Married	644 (44.2)	448 (43.9)	196 (44.9)	
Other	177 (12.1)	124 (12.2)	53 (12.1)	
Single	299 (20.5)	214 (21.0)	85 (19.5)	
Widowed	231 (15.9)	162 (15.9)	69 (15.8)	
Admission type (%)			0.73	
Elective		66 (4.5)	45 (4.4)	21 (4.8)
Emergency	567 (38.9)	394 (38.6)	173 (39.6)	
Observation	272 (18.7)	185 (18.1)	87 (19.9)	
Surgical	103 (7.1)	77 (7.5)	26 (5.9)	
Urgent	449 (30.8)	319 (31.3)	130 (29.7)	
Gender = Male (%)	868 (59.6)	622 (61.0)	246 (56.3)	0.107
Age (median [IQR])	72.8 [64.9, 80.4]	72.7 [65.0, 80.0]	73.4 [64.5, 81.1]	0.732
Comorbidity, n (%)				
Acidosis = Yes (%)	441 (30.3)	310 (30.4)	131 (30.0)	0.924
Alkalosis = Yes (%)	159 (10.9)	115 (11.3)	44 (10.1)	0.559
Anemias = Yes (%)	940 (64.5)	657 (64.4)	283 (64.8)	0.946
Anxiety states = Yes (%)	230 (15.8)	167 (16.4)	63 (14.4)	0.39
Asthma = Yes (%)	172 (11.8)	128 (12.5)	44 (10.1)	0.209
Atherosclerosis = Yes (%)	125 (8.6)	94 (9.2)	31 (7.1)	0.221
Atrial fibrillation = Yes (%)	801 (55.0)	554 (54.3)	247 (56.5)	0.472
Cardiac dysrhythmias = Yes (%)	999 (68.6)	693 (67.9)	306 (70.0)	0.47
Cardiomyopathy = Yes (%)	242 (16.6)	168 (16.5)	74 (16.9)	0.888
Cerebrovascular disease = Yes (%)	271 (18.6)	190 (18.6)	81 (18.5)	1
Chronic kidney disease = Yes (%)	632 (43.4)	446 (43.7)	186 (42.6)	0.724
Chronic lower respiratory diseases = Yes (%)	247 (17.0)	179 (17.5)	68 (15.6)	0.395
Chronic rheumatic heart disease = Yes (%)	208 (14.3)	141 (13.8)	67 (15.3)	0.501
COPD = Yes (%)	284 (19.5)	202 (19.8)	82 (18.8)	0.699
Dementias = Yes (%)	66 (4.5)	49 (4.8)	17 (3.9)	0.528
Depressive disorder = Yes (%)	309 (21.2)	220 (21.6)	89 (20.4)	0.657
Drink alcohol = Yes (%)	95 (6.5)	62 (6.1)	33 (7.6)	0.353
Drugs poisoning = Yes (%)	229 (15.7)	157 (15.4)	72 (16.5)	0.658
Esophageal diseases = Yes (%)	492 (33.8)	345 (33.8)	147 (33.6)	0.994
Gastroduodenal diseases = Yes (%)	147 (10.1)	104 (10.2)	43 (9.8)	0.911
Gastrointestinal hemorrhage = Yes (%)	116 (8.0)	84 (8.2)	32 (7.3)	0.628
Glomerular disease = Yes (%)	53 (3.6)	37 (3.6)	16 (3.7)	1
Gout = Yes (%)	100 (6.9)	71 (7.0)	29 (6.6)	0.911
Hemorrhagic conditions = Yes (%)	332 (22.8)	227 (22.3)	105 (24.0)	0.502
Hyperlipidemia = Yes (%)	1020 (70.0)	714 (70.0)	306 (70.0)	1
Hypernatremia = Yes (%)	229 (15.7)	172 (16.9)	57 (13.0)	0.079
Hyperpotassemia = Yes (%)	258 (17.7)	192 (18.8)	66 (15.1)	0.103
Hypertension = Yes (%)	1286 (88.3)	897 (87.9)	389 (89.0)	0.62
Hyponatremia = Yes (%)	269 (18.5)	190 (18.6)	79 (18.1)	0.862
Hypopotassemia = Yes (%)	121 (8.3)	78 (7.6)	43 (9.8)	0.198
Hypotension = Yes (%)	398 (27.3)	275 (27.0)	123 (28.1)	0.688
Hypothyroidism = Yes (%)	269 (18.5)	175 (17.2)	94 (21.5)	0.059
Continued				

Characteristics	Total($n=1457$)	Train($n=1020$)	Test($n=437$)	p
Ischemic heart disease = Yes (%)	1100 (75.5)	773 (75.8)	327 (74.8)	0.747
Liver disease = Yes (%)	215 (14.8)	146 (14.3)	69 (15.8)	0.518
Malignant neoplasm = Yes (%)	182 (12.5)	127 (12.5)	55 (12.6)	1
Metabolic and immunity disorders = Yes (%)	1097 (75.3)	773 (75.8)	324 (74.1)	0.549
Mineral metabolism disorders = Yes (%)	157 (10.8)	114 (11.2)	43 (9.8)	0.508
Myocardial infarction = Yes (%)	600 (41.2)	417 (40.9)	183 (41.9)	0.768
Nutritional deficiencies = Yes (%)	184 (12.6)	126 (12.4)	58 (13.3)	0.691
Obesity = Yes (%)	433 (29.7)	318 (31.2)	115 (26.3)	0.072
Pain = Yes (%)	146 (10.0)	105 (10.3)	41 (9.4)	0.663
Pancreas and gallbladder diseases = Yes (%)	132 (9.1)	90 (8.8)	42 (9.6)	0.704
Peripheral nervous system diseases = Yes (%)	185 (12.7)	130 (12.7)	55 (12.6)	1
Peripheral vascular disease = Yes (%)	498 (34.2)	353 (34.6)	145 (33.2)	0.641
Pneumonia = Yes (%)	383 (26.3)	267 (26.2)	116 (26.5)	0.935
Pulmonary heart disease = Yes (%)	319 (21.9)	228 (22.4)	91 (20.8)	0.564
Respiratory failure = Yes (%)	651 (44.7)	456 (44.7)	195 (44.6)	1
Shock = Yes (%)	484 (33.2)	337 (33.0)	147 (33.6)	0.871
Sleep disorders = Yes (%)	371 (25.5)	265 (26.0)	106 (24.3)	0.531
Smoking history = Yes (%)	732 (50.2)	519 (50.9)	213 (48.7)	0.489
Systemic inflammatory response Syndrome = Yes (%)	352 (24.2)	242 (23.7)	110 (25.2)	0.6
Urinary obstruction = Yes (%)	236 (16.2)	171 (16.8)	65 (14.9)	0.412
Urinary tract infection = Yes (%)	356 (24.4)	252 (24.7)	104 (23.8)	0.762
Valvular heart disease = Yes (%)	418 (28.7)	285 (27.9)	133 (30.4)	0.368
Volume depletion = Yes (%)	212 (14.6)	149 (14.6)	63 (14.4)	0.989
Sepsis3 = Yes (%)	1011 (69.4)	702 (68.8)	309 (70.7)	0.513
Laboratory parameters, (median [IQR])				
Creatinine (mg/dL)	1.20 [0.90, 1.60]	1.20 [0.90, 1.70]	1.20 [0.90, 1.60]	0.556
Estimated GFR, EGFR (mL/min)	58.14 [38.78, 84.79]	59.19 [39.67, 84.92]	56.31 [35.72, 83.43]	0.098
Anion Gap (mEq/L)	15.00 [13.00, 18.00]	15.00 [13.00, 18.00]	15.00 [13.00, 17.00]	0.846
Base Excess (mEq/L)	0.00 [-2.00, 2.00]	0.00 [-3.00, 2.00]	0.00 [-2.00, 2.00]	0.559
Blood urea nitrogen (mg/dL)	25.00 [18.00, 39.00]	25.00 [18.00, 40.00]	25.00 [18.00, 39.00]	0.473
Potassium (mEq/L)	4.30 [3.90, 4.70]	4.30 [3.90, 4.70]	4.20 [3.90, 4.70]	0.344
Sodium (mEq/L)	139.00 [136.00, 141.00]	139.00 [136.00, 141.00]	139.00 [136.00, 141.00]	0.165
Phosphate (mg/dL)	3.60 [3.10, 4.40]	3.61 [3.10, 4.50]	3.60 [3.10, 4.40]	0.639
Calcium (mg/dL)	8.60 [8.10, 9.00]	8.50 [8.10, 9.00]	8.60 [8.10, 9.00]	0.098
Chloride (mEq/L)	102.00 [99.00, 106.00]	102.00 [99.00, 106.00]	102.00 [99.00, 106.00]	0.755
Bicarbonate (mEq/L)	24.00 [21.00, 27.00]	24.00 [21.00, 27.00]	24.00 [21.00, 26.00]	0.998
Lactate (mmol/L)	1.60 [1.10, 2.20]	1.60 [1.10, 2.20]	1.50 [1.10, 2.20]	0.772
PH (units)	7.38 [7.32, 7.43]	7.38 [7.32, 7.43]	7.38 [7.33, 7.43]	0.126
Glucose (mg/dL)	156.00 [121.00, 216.00]	157.00 [121.45, 216.00]	153.00 [121.00, 211.00]	0.285
Hemoglobin (g/dL)	11.10 [9.60, 12.60]	11.05 [9.60, 12.60]	11.20 [9.60, 12.60]	0.595
Hematocrit, HCT (%)	33.90 [29.60, 38.30]	33.85 [29.80, 38.23]	34.10 [29.30, 38.60]	0.877
Mean corpuscular volume, MCV (fL)	90.00 [87.00, 94.00]	91.00 [87.00, 95.00]	89.38 [86.00, 94.00]	0.027
Platelet (K/uL)	203.00 [156.00, 253.00]	202.50 [159.00, 252.25]	203.00 [151.00, 256.00]	0.706
Red Blood Cells, RBC (m/uL)	3.77 [3.27, 4.29]	3.77 [3.26, 4.24]	3.78 [3.28, 4.37]	0.278
Red blood cell distribution width, RDW (%)	14.50 [13.50, 15.80]	14.50 [13.50, 15.90]	14.50 [13.50, 15.60]	0.738
White Blood Cells, WBC (K/uL)	10.30 [7.60, 14.30]	10.40 [7.63, 14.43]	10.20 [7.50, 13.70]	0.273
INR	1.20 [1.10, 1.50]	1.20 [1.10, 1.50]	1.20 [1.10, 1.50]	0.51
PT (sec)	13.60 [12.20, 16.00]	13.60 [12.20, 15.90]	13.90 [12.30, 16.40]	0.437
PTT (sec)	31.30 [27.60, 40.60]	31.40 [27.30, 41.32]	31.10 [28.30, 38.70]	0.571
PO2 (mmHg)	137.00 [66.00, 321.00]	140.00 [68.00, 312.25]	133.00 [65.00, 335.00]	0.835
PCO2 (mmHg)	42.00 [36.00, 48.00]	42.00 [37.00, 48.00]	41.00 [36.00, 47.00]	0.337
Vital signs, (median [IQR])				
Systolic blood pressure, SBP (mmHg)	120.00 [104.00, 137.00]	121.00 [104.00, 137.00]	116.00 [102.00, 137.00]	0.118
Diastolic blood pressure, DBP (mmHg)	62.00 [53.00, 74.00]	62.00 [53.00, 74.00]	62.00 [54.00, 74.00]	0.696
Mean blood pressure, MBP (mmHg)	79.00 [69.00, 91.00]	80.00 [69.00, 91.00]	79.00 [69.00, 92.00]	0.777
Heart rate, HR (bpm)	84.00 [75.00, 98.00]	84.00 [75.00, 98.00]	84.00 [75.00, 96.00]	0.736
Continued				

Characteristics	Total($n = 1457$)	Train($n = 1020$)	Test($n = 437$)	p
Respiratory rate, RR (insp/min)	17.00 [14.00, 22.00]	17.00 [14.00, 22.00]	17.00 [14.00, 22.00]	0.707
Pulse oxygen saturation, SPO2 (%)	99.00 [95.00, 100.00]	98.00 [95.00, 100.00]	99.00 [96.00, 100.00]	0.236
Urine output (median [IQR])(mL)	80.00 [40.00, 150.00]	80.00 [40.00, 150.00]	75.00 [40.00, 150.00]	0.821
Drug use, n (%)				
Adrenaline receptor agonist = Yes (%)	896 (61.5)	623 (61.1)	273 (62.5)	0.659
Antiarrhythmic drugs = Yes (%)	911 (62.5)	638 (62.5)	273 (62.5)	1
Antibiotics = Yes (%)	924 (63.4)	645 (63.2)	279 (63.8)	0.871
CCB = Yes (%)	210 (14.4)	142 (13.9)	68 (15.6)	0.462
Diuretic = Yes (%)	742 (50.9)	528 (51.8)	214 (49.0)	0.357
β blocker = Yes (%)	794 (54.5)	552 (54.1)	242 (55.4)	0.7
Positive inotropic = Yes (%)	154 (10.6)	112 (11.0)	42 (9.6)	0.493
Vasodilator drug = Yes (%)	757 (52.0)	531 (52.1)	226 (51.7)	0.95
Vasoconstrictor drug = Yes (%)	577 (39.6)	406 (39.8)	171 (39.1)	0.855
NASIDs = Yes (%)	1043 (71.6)	728 (71.4)	315 (72.1)	0.832
Insulin = Yes (%)	1321 (90.7)	926 (90.8)	395 (90.4)	0.889
Anticoagulants = Yes (%)	1089 (74.7)	771 (75.6)	318 (72.8)	0.285
Norepinephrine = Yes (%)	393 (27.0)	274 (26.9)	119 (27.2)	0.936
Epinephrine = Yes (%)	203 (13.9)	149 (14.6)	54 (12.4)	0.292
Milrinone = Yes (%)	59 (4.0)	42 (4.1)	17 (3.9)	0.955
Vasopressin = Yes (%)	99 (6.8)	71 (7.0)	28 (6.4)	0.786
Phenylephrine = Yes (%)	511 (35.1)	356 (34.9)	155 (35.5)	0.882
Digoxin = Yes (%)	65 (4.5)	47 (4.6)	18 (4.1)	0.783
ACEI/ARB = Yes (%)	320 (22.0)	232 (22.7)	88 (20.1)	0.302
Albumin = Yes (%)	298 (20.5)	204 (20.0)	94 (21.5)	0.559
Antianemia drug = Yes (%)	77 (5.3)	57 (5.6)	20 (4.6)	0.507
Antifungal drugs = Yes (%)	9 (0.6)	8 (0.8)	1 (0.2)	0.381
Antigout drug = Yes (%)	87 (6.0)	65 (6.4)	22 (5.0)	0.386
Antiplatelet drugs = Yes (%)	798 (54.8)	564 (55.3)	234 (53.5)	0.578
Antiviral drugs = Yes (%)	9 (0.6)	8 (0.8)	1 (0.2)	0.381
Aspirin = Yes (%)	770 (52.8)	543 (53.2)	227 (51.9)	0.693
Fibrinolytics = Yes (%)	25 (1.7)	17 (1.7)	8 (1.8)	0.999
Immunosuppress = Yes (%)	10 (0.7)	9 (0.9)	1 (0.2)	0.299
MRA = Yes (%)	52 (3.6)	38 (3.7)	14 (3.2)	0.735
Statins = Yes (%)	808 (55.5)	574 (56.3)	234 (53.5)	0.367
Blood products or colloids = Yes (%)	623 (42.8)	432 (42.4)	191 (43.7)	0.674
Calcium gluconate = Yes (%)	432 (29.6)	307 (30.1)	125 (28.6)	0.61
Dobutamine = Yes (%)	34 (2.3)	27 (2.6)	7 (1.6)	0.307
Dopamine = Yes (%)	56 (3.8)	39 (3.8)	17 (3.9)	1
Macrolides = Yes (%)	44 (3.0)	31 (3.0)	13 (3.0)	1
Metronidazole = Yes (%)	88 (6.0)	61 (6.0)	27 (6.2)	0.98
Quinolones = Yes (%)	91 (6.2)	62 (6.1)	29 (6.6)	0.776
Sodium bicarbonate = Yes (%)	50 (3.4)	37 (3.6)	13 (3.0)	0.638
Vancomycin = Yes (%)	586 (40.2)	422 (41.4)	164 (37.5)	0.189
β lactam antibiotics = Yes (%)	815 (55.9)	559 (54.8)	256 (58.6)	0.203
Disease score, (median [IQR])				
SIRS (median [IQR])	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	3.00 [2.00, 3.00]	0.502
SAPSII (median [IQR])	39.00 [32.00, 48.00]	39.00 [32.00, 48.00]	39.00 [32.00, 48.00]	0.998
OASIS (median [IQR])	33.00 [28.00, 39.00]	33.00 [28.00, 39.00]	33.00 [28.00, 39.00]	0.889
APSIII (median [IQR])	44.00 [34.00, 59.00]	44.00 [34.00, 60.00]	45.00 [34.00, 58.00]	0.957
SOFA (median [IQR])	5.00 [3.00, 8.00]	5.00 [3.00, 8.00]	6.00 [3.00, 8.00]	0.547
GCS (median [IQR])	15.00 [15.00, 15.00]	15.00 [15.00, 15.00]	15.00 [15.00, 15.00]	0.641
Surgery or intervention measures, n (%)				
CRRT = Yes (%)	5 (0.3)	4 (0.4)	1 (0.2)	1
RRT = Yes (%)	40 (2.7)	30 (2.9)	10 (2.3)	0.6
Mechanical Ventilation = Yes (%)	1319 (90.5)	925 (90.7)	394 (90.2)	0.829
CAGB = Yes (%)	409 (28.1)	288 (28.2)	121 (27.7)	0.881
Continued				

Characteristics	Total($n=1457$)	Train($n=1020$)	Test($n=437$)	p
Cardiovascular surgery = Yes (%)	1121 (76.9)	775 (76.0)	346 (79.2)	0.208
CRT = Yes (%)	60 (4.1)	40 (3.9)	20 (4.6)	0.665
Dialysis = Yes (%)	53 (3.6)	37 (3.6)	16 (3.7)	1
PCI = Yes (%)	45 (3.1)	36 (3.5)	9 (2.1)	0.187
Urinary system surgery = Yes (%)	19 (1.3)	15 (1.5)	4 (0.9)	0.546
Vascular reconstruction = Yes (%)	281 (19.3)	199 (19.5)	82 (18.8)	0.796
ECG = Yes (%)	725 (49.8)	501 (49.1)	224 (51.3)	0.489
Cardiovascular pacer data = Yes (%)	508 (34.9)	355 (34.8)	153 (35.0)	0.987
ECMO = Yes (%)	3 (0.2)	2 (0.2)	1 (0.2)	1
IABP = Yes (%)	76 (5.2)	50 (4.9)	26 (5.9)	0.487

Table 3. Baseline characteristics of patients in the training group and the validation group.

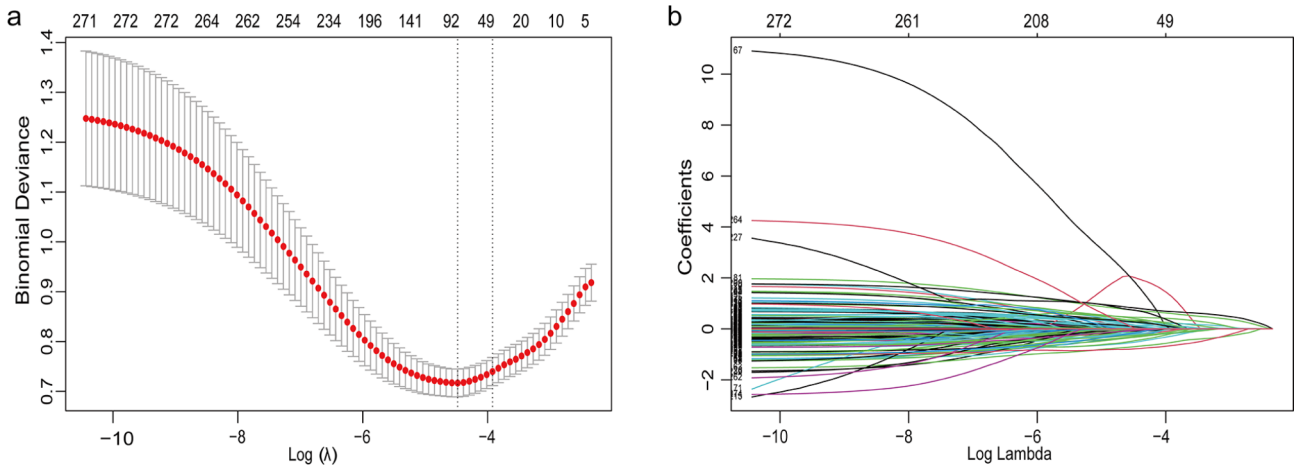

Fig. 3. Coefficient curves and cross-validation plots for acute kidney injury risk-related variable screening based on LASSO regression models.

surgery. Specific parameters can be found in Table 4. Regarding admission type, the results indicated that, compared to emergency admissions (baseline category), patients admitted for surgery had a markedly reduced risk of AKI (OR = 0.265, $p = 0.042$). No notable differences in AKI risk were noted among patients admitted electively, for observation, or via the emergency department. This indicated that admission type was related to AKI risk, and emergency admission may indicate more critical conditions, requiring further study. Comorbidities such as anemia, obesity, atrial fibrillation, and sepsis 3, with OR > 1, suggested that these were risk factors for the development of AKI. For instance, the risk of AKI in obese patients was 2.252 times higher than in non-obese patients (OR = 2.252, $P < 0.001$). Medications such as antiarrhythmics, diuretics, and statins, with an OR < 1, indicated that they were connected to a lower risk of AKI. As an example, patients taking antiarrhythmics had a notably lower risk of developing AKI compared to those not using the medication (OR = 0.424, $P < 0.001$). For each 1-point increased in the APSIII score, the risk of AKI raised by 2% (OR = 1.020, $P = 0.003$). These 20 selected predictors were employed to build subsequent machine learning models.

Construction of the AKI predictive risk nomogram and models

In order to improve the interpretability and clinical practicality of models, a nomogram (Fig. 4) based on 20 independent predictors screened by LR analysis was built. The nomogram was employed to predict AKI risk in DM patients with HF within 30 days during hospitalization. The nomogram linked each predictor to its respective risk score. The patient's total points were calculated by summing the individual points of all the predictors. The total points could be located on the nomogram to determine the corresponding AKI risk probability. Higher total points correlated with an increased risk of AKI. The total points on this nomogram mainly ranged from 115 to 400, corresponding to an AKI risk from 0 to 99.9%. Importantly, the nomogram was designed to provide an easy-to-use risk assessment tool by consolidating multiple predictors into a single risk score for clinical application.

Six ML models were constructed to predict AKI risk during hospitalization in DM patients with HF, including GBM, KNN, LightGBM, LR, NN, and RF models. Each model underwent five-fold cross-validation and 100-bootstrap resampling for parameter optimization. The final optimal parameters for each model were as follows: ntree = 405, mtry = 5, maxnodes = 93, and nodesize = 4 for the RF model; three-layer structure, 21 input features, 28 nodes in the hidden layer and 1 node in the output layer for the NN model; size = 2, decay = 2,

Characteristic	OR[1]	95% CI[1]	*p*-value
Admission type			
Emergency	—	—	
Elective	0.529	0.154, 1.529	0.300
Observation	0.572	0.164, 1.703	0.300
Surgical	0.265	0.068, 0.907	0.042
Urgent	0.951	0.277, 2.773	>0.900
Anemias			
No	—	—	
Yes	1.481	1.288, 2.218	0.048
Atrial fibrillation			
No	—	—	
Yes	1.709	1.156, 2.539	0.007
Hypotension			
No	—	—	
Yes	1.510	1.162, 2.416	0.034
Obesity			
No	—	—	
Yes	2.252	1.425, 3.644	<0.001
Pneumonia			
No	—	—	
Yes	1.835	1.104, 3.123	0.022
Creatinine			
Normal	—	—	
Abnormal	1.342	1.085, 2.046	0.020
Antiarrhythmic drugs			
No	—	—	
Yes	0.424	0.255, 0.687	<0.001
Diuretic			
No	—	—	
Yes	0.374	0.236, 0.583	<0.001
Anticoagulants			
No	—	—	
Yes	0.501	0.270, 0.890	0.022
ACEI/ARB			
No	—	—	
Yes	0.330	0.216, 0.502	<0.001
Aspirin			
No	—	—	
Yes	0.506	0.322, 0.786	0.003
Statins			
No	—	—	
Yes	0.638	0.395, 0.918	0.025
Calcium gluconate			
No	—	—	
Yes	0.622	0.406, 0.956	0.030
β lactam antibiotics			
No	—	—	
Yes	0.527	0.345, 0.796	0.003
RRT			
No	—	—	
Yes	0.267	0.081, 0.907	0.038
ECMO			
No	—	—	
Yes	0.014	0.000, 0.736	0.038
APSIII	1.020	1.007, 1.033	0.003
Sepsis3			
Continued			

Characteristic	OR[1]	95% CI[1]	*p*-value
No	—	—	
Yes	1.713	1.130, 2.598	0.011
Cardiovascular surgery			
No	—	—	
Yes	5.080	3.234, 8.082	< 0.001

Table 4. Multivariate logistic regression analysis of risk factors for AKI in patients with diabetes and heart failure. [1]OR = Odds Ratio, CI = Confidence Interval.

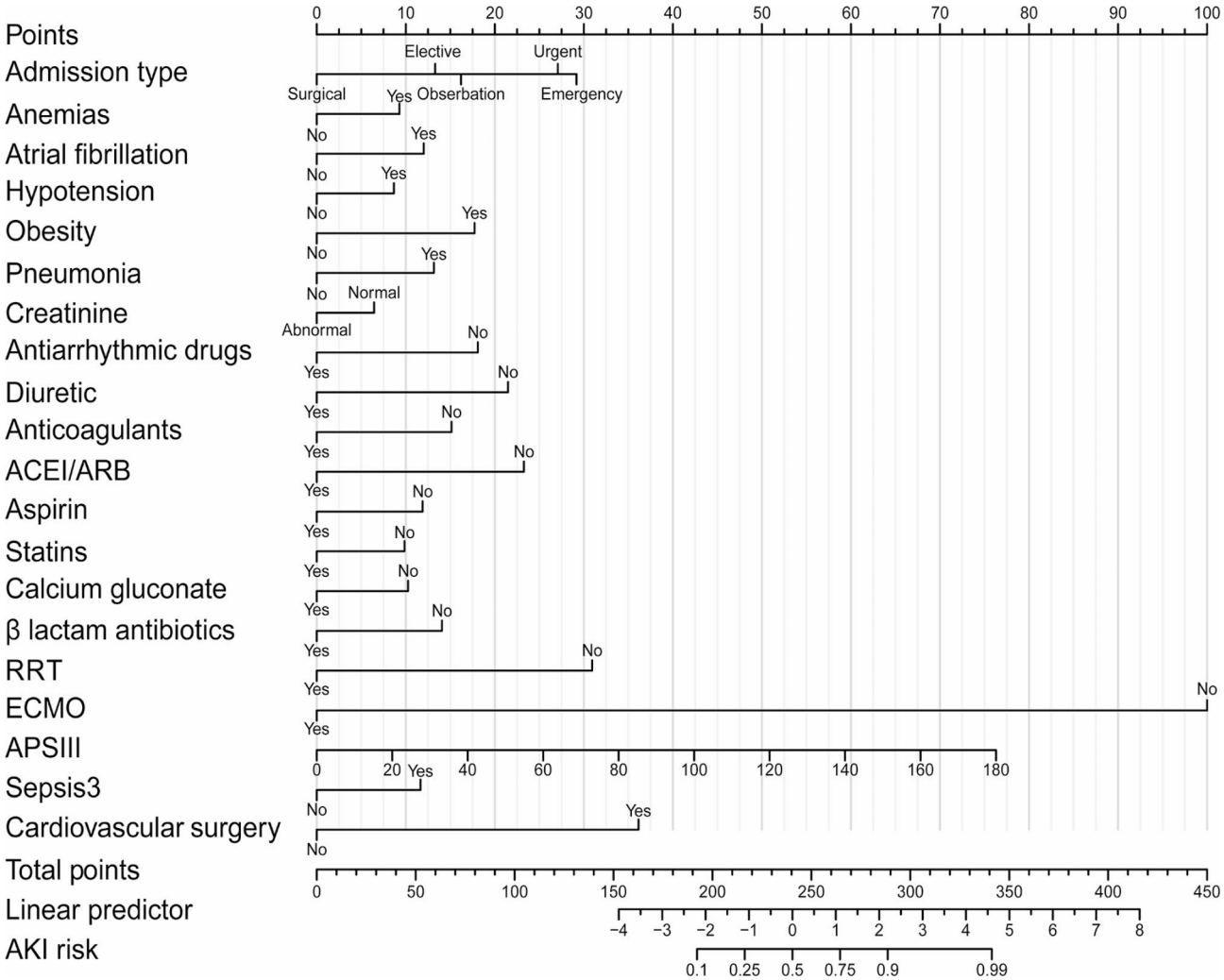

Fig. 4. Nomogram of predicted 30-day risk of acute kidney injury in patients with diabetes mellitus combined with heart failure. ACEI/ARB: angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker, RRT: renal replacement therapy, ECMO: extracorporeal membrane oxygenation, APSIII: acute physiology score III.

and maxit = 8778 for the LR model; n.trees = 200, interaction.depth = 1, and shrinkage = 0.3 for the GBM model; k = 13 for the KNN model; learning_rate = 0.0195, num_leaves = 37, n_estimators = 156, feature_fraction = 0.945, bagging_fraction = 0.745, and bagging_freq = 3 for the LightGBM model.

Performance evaluation and validation of AKI risk prediction models

In this study, AUC was leveraged to evaluate the discrimination of models. The calibration of models was assessed by calibration curves (Table 5; Fig. 5). The six ML models show good discrimination in both the training and validation sets. The AUCs were all above 0.75. Among them, the LightGBM model and RF model performed the best in the training set, with AUCs of 0.973 and 0.987, respectively. However, the RF model showed a decrease in AUC to 0.799 in the validation set, indicating a possible degree of overfitting. The LightGBM model demonstrated

Classification Models	AUC (95%CI)	Cutoff	Sensitivity	Specificity	FRP	TRP
Training set						
GBM	0.854(0.824–0.883)	0.576	0.799	0.777	0.223	0.799
KNN	0.863 (0.837–0.889)	0.579	0.728	0.851	0.149	0.728
Light GBM	0.9734(0.964–0.983)	0.838	0.918	0.920	0.080	0.918
LR	0.857 (0.829–0.885)	0.580	0.774	0.806	0.194	0.774
NN	0.853 (0.824–0.881)	0.574	0.802	0.771	0.229	0.802
RF	0.9867(0.981–0.993)	0.877	0.940	0.937	0.063	0.940
Validation set						
GBM	0.798 (0.744–0.852)	0.493	0.850	0.643	0.357	0.850
KNN	0.769 (0.709–0.830)	0.443	0.757	0.686	0.314	0.757
Light GBM	0.8034(0.748–0.859)	0.521	0.749	0.771	0.229	0.749
LR	0.797 (0.742–0.853)	0.494	0.837	0.657	0.343	0.837
NN	0.799(0.743–0.854)	0.488	0.774	0.714	0.286	0.774
RF	0.799(0.743–0.854)	0.490	0.733	0.757	0.243	0.733

Table 5. Assessment of the six prediction models.

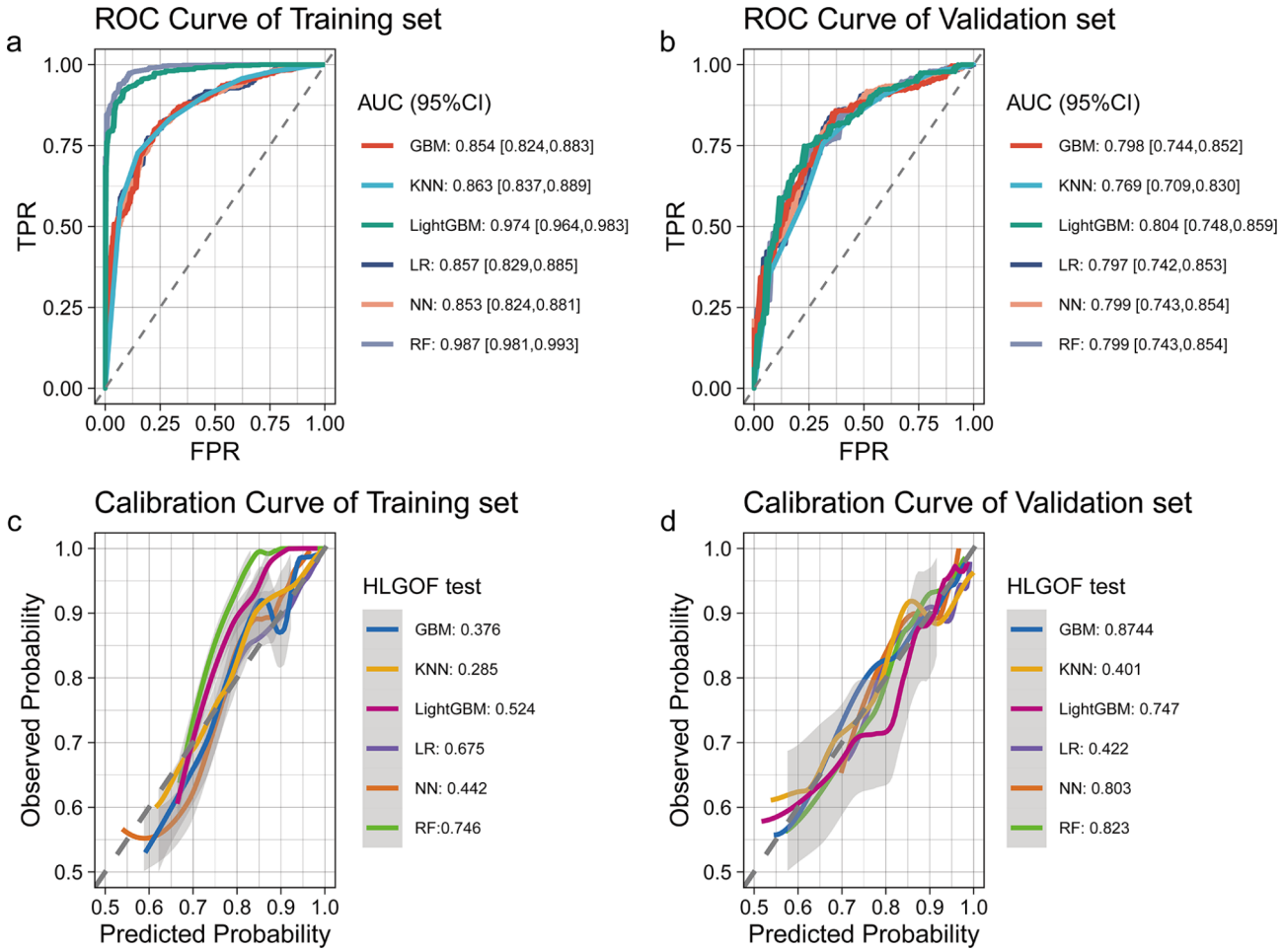

Fig. 5. ROC curve and calibration curve of the nomogram of the risk of acute kidney injury in patients with diabetes mellitus combined with heart failure in the training and validation sets. (**a**) ROC curve of the nomogram in the training set; (**b**) ROC curve of the nomogram in the validation set; (**c**) calibration curve of the nomogram in the training set; (**d**) calibration curve of the nomogram in the validation set. ROC: receiver operating characteristic curve, AUC: area under the ROC curve, CI: confidence interval, TPR: true positive rate, FPR: false positive rate, HLGOF: Hosmer-Lemeshow goodness-of-fit test.

the best performance in the validation set, with an AUC of 0.804, demonstrating stronger generalization ability. The calibration curves for each model were close to the reference line. The results of the Hosmer-Lemeshow test indicated P-values greater than 0.05, suggesting no notable difference between the predicted probabilities and the observed probabilities in the models. This finding indicated good calibration.

In this study, DCA was employed to evaluate the clinical performance of each model for predicting AKI risk in DM patients with HF (Fig. 6). The results showed that all models had better net benefits than the full intervention and no intervention strategies within a wide range of threshold probabilities, confirming their clinical predictive performance. The LightGBM model showed the highest net benefit in both the training and validation sets, indicating its optimal clinical performance.

In this study, SHAP values were used to quantify the significance of the top 15 features in the LightGBM model (Fig. 7). The results revealed that the use of ACEI/ARB drugs notably impacted AKI risk prediction, followed by cardiovascular surgery, diuretics, APSIII score, and aspirin. Attention should be paid to these important features in risk assessment and treatment planning in clinical practice.

The SHAP plot, based on the RF algorithm, illustrated the contribution of each feature to the prediction of the model. Unlike the nomogram, the SHAP plot was not intended for the direct calculation of risk probabilities. Rather, it served to identify the features that had the most significant influence on the prediction of the model. In the bar chart, positive values represented a positive contribution to AKI risk, negative values indicated a negative contribution, and zero values denoted no contribution.

For an in-depth analysis of the association between the features and AKI risk, the study plotted their associated partial dependence plots based on the four important features of ACEI/ARB use, diuretic dose, atrial fibrillation, and APSIII score (Fig. 8). The results demonstrated that ACEI/ARB use reduced the risk of AKI. Moreover, the protective effect may be dose-dependent. The impact of diuretics on AKI risk presented a bidirectional effect, with low doses potentially reducing risk and high doses potentially increasing risk. Atrial fibrillation was identified as an independent risk factor for AKI. However, APSIII score was positively correlated with AKI risk. The above results indicated that the LightGBM model can effectively identify key predictors and reveal their complex relationship with AKI risk, providing a reference for clinical risk assessment and personalized treatment.

To evaluate the clinical applicability of the LightGBM model, a clinical impact curve (CIC) was created in this study (Fig. 9). The findings revealed that, as expected, the number of high-risk patients predicted by the model gradually decreased as the high-risk threshold increased. The trends of the curves in both the training and validation sets were largely consistent, suggesting that the model exhibited good generalization ability without any noticeable overfitting. The curves for both the training and validation sets reached a better equilibrium at around a threshold value of 0.7. This indicated that choosing this threshold was effective in reducing false positives and avoiding excessive medical interventions while identifying a sufficient number of truly high-risk patients. This reflected a good cost-benefit ratio.

Discussion

The clinical data of 1,457 DM patients with HF from the MIMIC-IV database were retrospectively analyzed. ML models were successfully constructed to predict AKI risk during hospitalization. The study findings indicated that the LightGBM model exhibited the highest performance, with the highest AUC values, good calibration, and higher net clinical benefit. In addition, it exhibited good discrimination and generalization ability. This study represents the first attempt to construct models for predicting AKI risk in this population using common clinical indicators. This study offers an effective tool for the early identification of high-risk patients and timely intervention measures in clinical practice.

Analysis of study results and exploration of risk factors

This study included 1,457 DM patients with HF. It aimed to investigate the risk of developing AKI and related predictors. The results showed that up to 83.2% of patients developed AKI within 30 days of hospitalization. This suggests that AKI is very common in this population and requires early identification and intervention.

20 independent predictive factors associated with AKI risk were identified by the LASSO regression and stepwise logistic regression analysis. Independent risk factors for AKI development include emergency admission, abnormal creatinine values, anemia, atrial fibrillation, hypotension, obesity, pneumonia, APSIII, sepsis, and a history of cardiovascular surgery. Further analysis reveals that the above predictors affect the occurrence of AKI mainly through the following mechanisms: affecting renal perfusion (such as hypotension, anemia, and cardiovascular surgery)[30–33]; and inducing inflammatory responses or direct injury to the kidneys (such as sepsis, pneumonia, and certain drugs)[34,35]. Atrial fibrillation can lead to the formation of thrombosis in the atria. After thrombus dislodgement, renal artery embolism occurs, blocking renal blood flow and increasing the risk of AKI. Obesity damages kidney function through mechanisms such as chronic inflammation, insulin resistance, and oxidative stress[36,37]. The APSIII score reflects the critical condition of patients, often accompanied by shock, sepsis, and other conditions. Higher scores indicate a higher risk of developing AKI. Emergency admission often suggests that patients are in critical condition. There may be hemodynamic instability, infection, or shock, which may affect renal perfusion. On the contrary, the use of antiarrhythmic drugs, ACEI/ARB drugs, aspirin, statins, β lactam antibiotics, calcium gluconate, RRT, and ECMO is linked to a reduced AKI risk in DM patients with HF. Antiarrhythmic drugs can effectively reduce the risk of AKI caused by cardiogenic shock by controlling arrhythmia, improving heart function, and stabilizing blood pressure. ACEI/ARB drugs can delay the progression of chronic kidney disease by lowering blood pressure and reducing proteinuria. In addition, although aspirin and statins are primarily utilized for preventing cardiovascular events, their role in improving vascular endothelial function may indirectly reduce the risk of renal artery embolism caused by thrombosis. β lactam antibiotics can effectively control bacterial infections and reduce the associated risk of AKI. Calcium

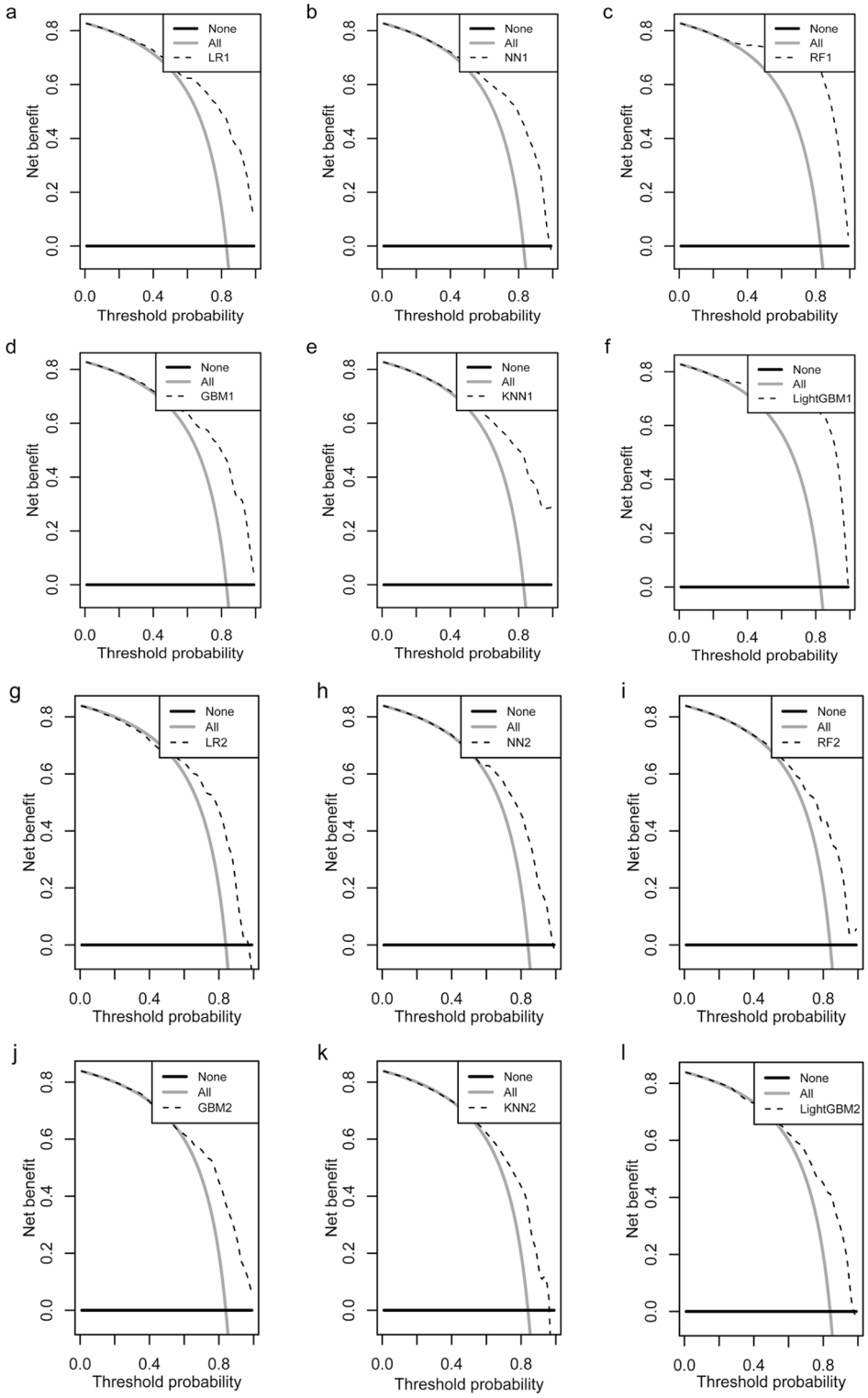

Fig. 6. Clinical decision curve of the nomogram of the risk of acute kidney injury in patients with diabetes mellitus combined with heart failure in the training and validation sets. (**a**–**f**) Clinical decision curves of the nomogram in the training set; (**g**–**l**) clinical decision curves of the nomogram in the validation set.

gluconate may improve kidney function by regulating calcium and phosphorus metabolism. RRT can reduce the burden on the kidneys. For critically ill patients, ECMO treatment can replace cardiopulmonary function, improve systemic circulation, maintain blood pressure, and reduce the risk of AKI due to shock and other factors. It is worth noting that diuretics have a dual effect on AKI risk. Rational use of diuretics can reduce the heart load and improve renal perfusion, but excessive use may lead to dehydration and increase the risk of AKI.

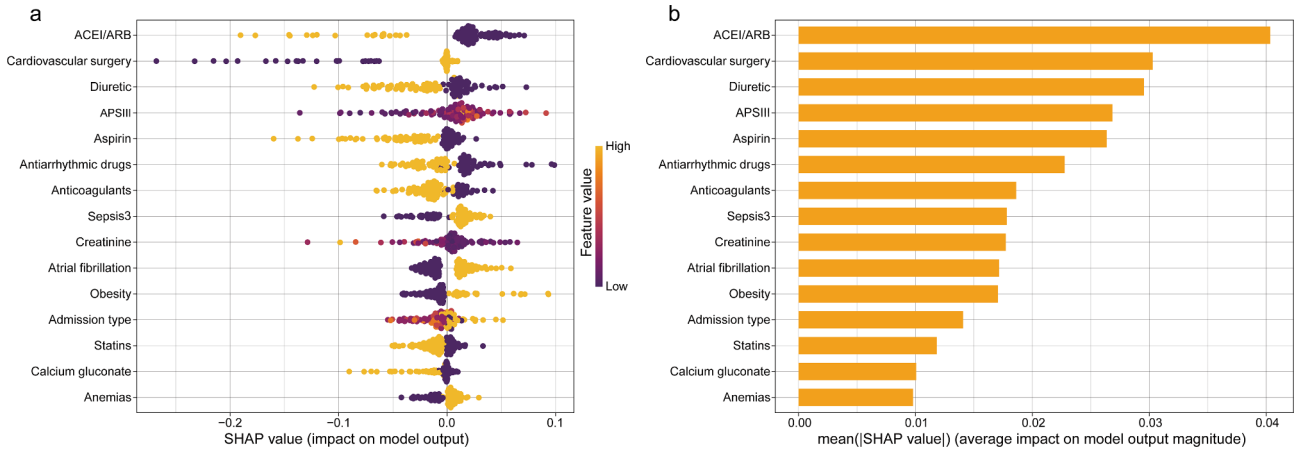

Fig. 7. Results of SHAP analysis. (**a**) Summary chart: the impact of the top 15 features in the LightGBM model calculated by SHAP values on acute kidney injury prediction; (**b**) bar chart: importance ranking of the features in the LightGBM model that were most relevant to acute kidney injury prediction. The color indicates the importance of the feature (yellow for high, purple for low) and each point represents a sample. SHAP: Shapley additive explanations, ACEI/ARB: angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker, APSIII: acute physiology score III.

In addition to the above 20 predictors, new biomarkers, genetic factors, and gut microbiota may also affect AKI risk in DM patients with HF, and provide a direction for the optimization of AKI risk prediction models in the future. For example, mid-stream urine biomarkers (such as NGAL, KIM-1, IL-18) and serum biomarkers (such as cystatin C, β 2-microglobulin) can reflect kidney injury earlier and more sensitively. It is expected to improve the prediction accuracy by incorporating these biomarkers into prediction models[38–41]. The polymorphism of genes such as TNF-α, IL-6, and eNOS may affect the susceptibility of patients to AKI. Conducting genetic testing and assessing genetic risk can help achieve individualized prevention of AKI[42]. Dysbiosis of the gut microbiota can lead to impaired intestinal barrier function, allowing bacteria and their metabolites to enter the blood, which causes a systemic inflammatory response, damages the kidneys, and increases the risk of AKI.

With the advancement of artificial intelligence, ML is becoming increasingly prevalent in the medical field. The advantage of ML lies in its ability to analyze large volumes of existing data and generate the most valuable insights. For example, ML can handle complex relationships, improving prediction accuracy, enabling personalized prediction, and aiding clinical decision-making in AKI prediction[43]. In this study, ML was used to build models for predicting AKI risk in DM patients with HF. The performances of models using six common algorithms (GBM, KNN, LightGBM, LR, NN, and RF) were compared. The findings indicated that the LightGBM model demonstrated superior performance, achieving the highest AUC values (0.973 and 0.804) in both the training and validation sets. It exhibited excellent calibration and clinical practicality.

For the convenience of clinical application, this study also constructed a visual nomogram based on the LR model to assist clinicians in rapidly assessing the risk of AKI in patients and formulating individualized prevention and treatment plans.

Comparison with previous studies

Previous studies have investigated the application of ML models in predicting AKI risk, but have focused primarily on the general population or disease-specific populations, such as patients with sepsis[44] and patients undergoing cardiac surgery[45,46]. However, there are relatively few AKI risk prediction studies for the DM population with HF. This study was the first to develop and validate risk prediction models for AKI for DM patients with HF based on the MIMIC-IV database. In addition, AKI risk predictors were screened for this population, providing a new reference for clinical practice.

Compared with previous studies, this study has certain innovations and strengths. (i) The study is highly targeted, focusing on the high-risk population of DM with HF, which is more conducive to early identification and intervention of AKI. (ii) The data source is reliable, using the large, publicly available, and high-quality MIMIC-IV database. (iii) The model construction and validation methods are scientific and reasonable. ML algorithms were used to construct models. Moreover, cross-validation and bootstrap resampling methods were used for parameter optimization to ensure the stability and generalization ability of models. (iv) The results were visualized. A visualized nomogram and SHAP analyses were developed for easy clinical application.

Theoretical significance and clinical application value of the study

This study confirmed the feasibility and effectiveness of ML in predicting AKI risk in DM patients with HF. Multiple predictive factors associated with the occurrence of AKI were revealed, providing a foundation for a more comprehensive understanding of the pathogenesis of AKI and the development of new prevention and treatment strategies. Compared to traditional methods for AKI risk assessments, the LightGBM prediction model and visual nomogram developed in this study can provide more accurate and objective risk assessment results.

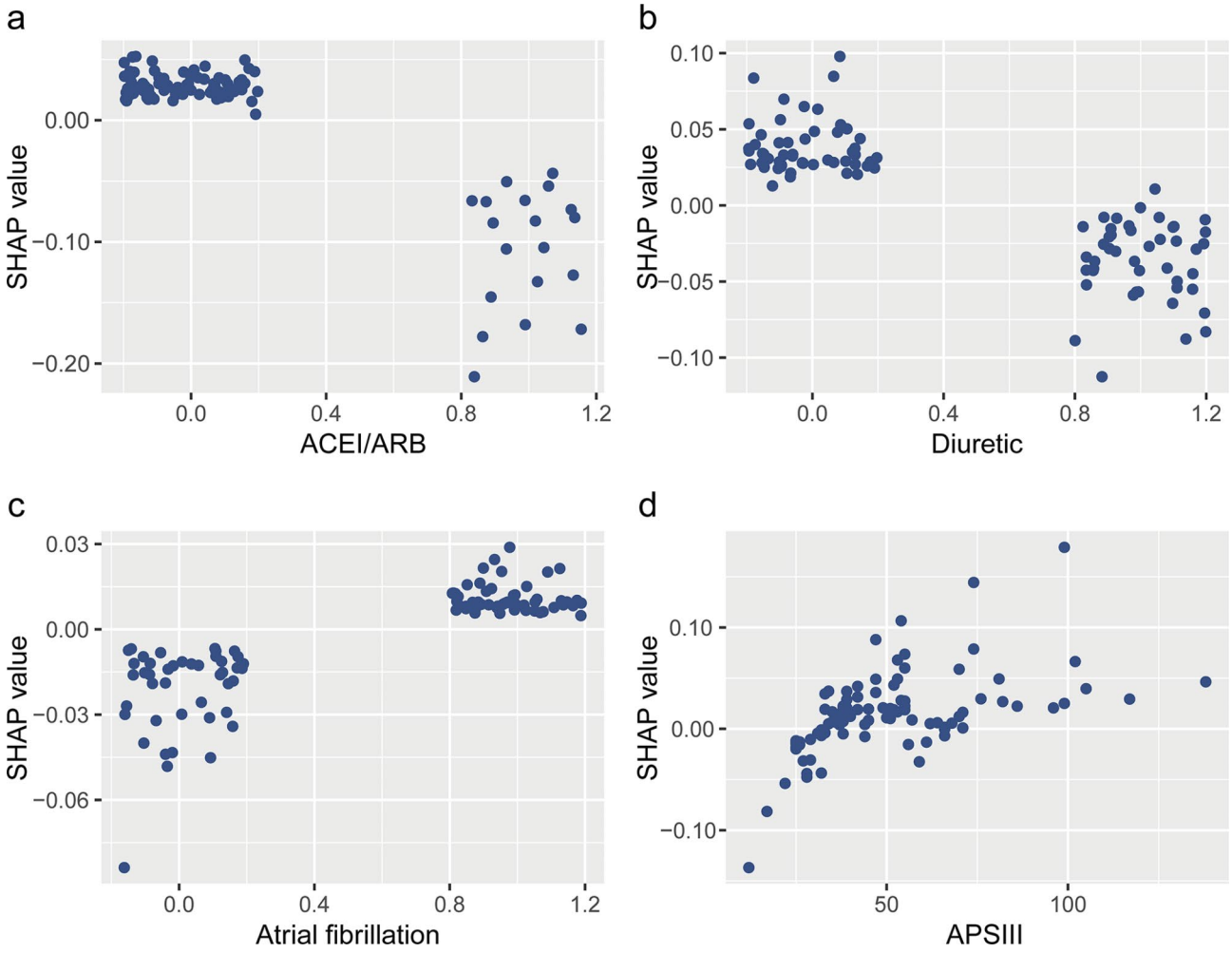

Fig. 8. Partial dependence plots of the LightGBM model used to predict the risk of acute kidney injury within 30 days of hospitalization in patients with diabetes mellitus combined with heart failure. SHAP: Shapley additive explanations, ACEI/ARB: angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker, APSIII: acute physiology score III.

These results can aid healthcare professionals in promptly identifying high-risk patients and implementing timely preventive measures (such as optimizing fluid management and avoiding nephrotoxic drugs), thus reducing the incidence of AKI and promoting rational allocation of healthcare resources.

Study limitations

Although the present study yielded some meaningful results, there are some limitations. (i) Due to the retrospective study design, the study may be subject to selection bias and information bias. It is difficult to completely exclude the effect of potential confounders. In addition, the MIMIC-IV database primarily includes critically ill patients. Therefore, the findings of the study might not generalize to other populations. (ii) The study data were single-center data, only including data from BIDMC in the United States. No data were collected in domestic hospitals for DM patients with HF. The generalization ability of the model needs to be further validated in other populations and different healthcare institutions to improve the extrapolation of the model. (iii) The model has not been externally validated. Its stability and reliability need to be externally validated using data from other databases to further assess its predictive performance and clinical application value. (iv) There were few features in the study. Due to limitations in the MIMIC-IV database, some special variables like cystatin C and renin were not included in the model, which are associated with AKI. (5) Some variables had high missing values. Although this study used MI to handle missing values, it may still have some impact on the results. Some alternative variables such as triglycerides, cholesterol, glycated hemoglobin, and NT-PROBNP were not included in this study due to too many missing values.

Conclusion

In conclusion, ML models built in this study, especially the LightGBM model, demonstrated effective predictive performance for assessing AKI risk in DM patients with HF during hospitalization. Multiple independent risk factors were revealed, providing a new reference for clinical risk assessment. In the future, larger sample

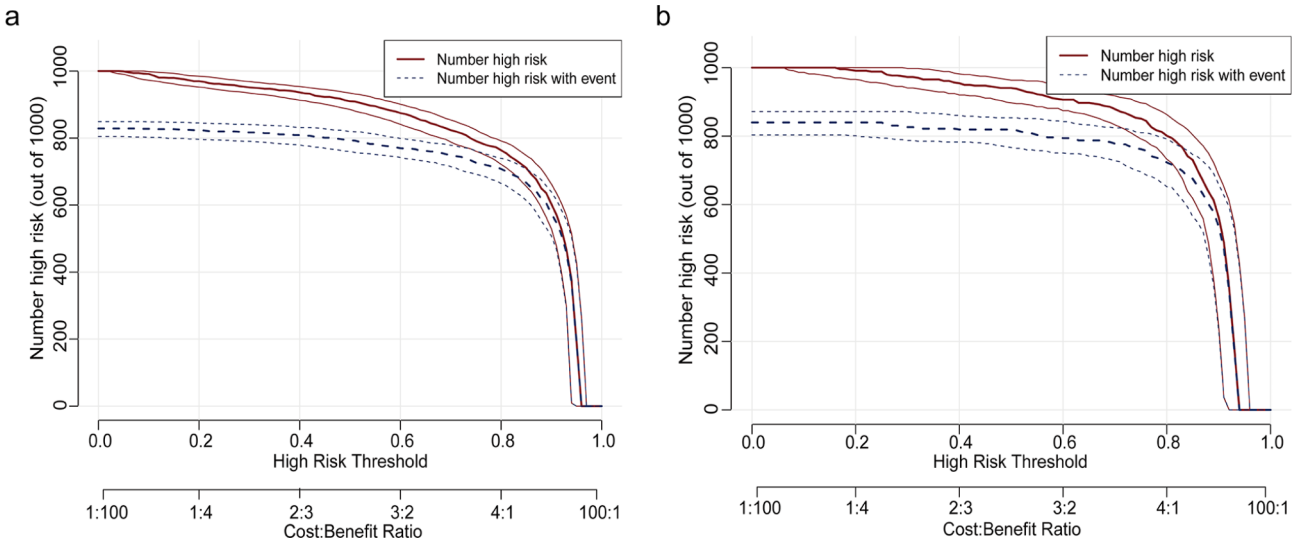

Fig. 9. Clinical impact curve for the LightGBM model. a represents the training set and b represents the validation set. The red curve represents the number of patients predicted to be at high risk by the model (number high risk); the blue curve represents the number of patients predicted to be at high risk with event (number high risk with event). The results of the clinical impact curves confirmed the clinical value of the LightGBM model.

sizes and multicenter prospective studies are required to further validate and optimize the model, in order to develop more accurate, reliable, and practical AKI risk prediction tools that provide stronger support for clinical decision-making.

Tables.

Data availability

All analysis code and synthetic datasets employed to demonstrate the model will be published on GitHub to ensure the reproducibility of the study (https://github.com/L-gj-cmd/AKI-Prediction-Model/blob/main/feature_selection_modeling_validation.R). To obtain the data from this study or for any inquiries, please contact the corresponding author at: [ksshyy2006@sina.com].

Received: 14 August 2024; Accepted: 17 January 2025
Published online: 28 March 2025

References

1. Gameiro, J. et al. Acute kidney disease and long-term outcomes in critically ill acute kidney injury patients with sepsis: A cohort analysis. *Clin. Kidney J.* **14**, 1379–1387. https://doi.org/10.1093/ckj/sfaa130 (2021).
2. Khwaja, A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron. Clinical practice* **120**, c179-184, (2012). https://doi.org/10.1159/000339789
3. Andonovic, M. et al. Short- and long-term outcomes of intensive care patients with acute kidney disease. *EClinicalMedicine* **44**, 101291. https://doi.org/10.1016/j.eclinm.2022.101291 (2022).
4. Hoste, E. A. J. et al. Global epidemiology and outcomes of acute kidney injury. *Nat. Rev. Nephrol.* **14**, 607–625. https://doi.org/10.1038/s41581-018-0052-0 (2018).
5. Yang, B., Zhu, Y., Lu, X., Shen, C. A. & Novel composite indicator of predicting mortality risk for heart failure patients with diabetes admitted to intensive care unit based on machine learning. *Front. Endocrinol.* **13**, 917838. https://doi.org/10.3389/fendo.2022.917838 (2022).
6. Jentzer, J. C. et al. Contemporary management of severe acute kidney injury and refractory cardiorenal syndrome: JACC council perspectives. *J. Am. Coll. Cardiol.* **76**, 1084–1101. https://doi.org/10.1016/j.jacc.2020.06.070 (2020).
7. Vandenberghe, W. et al. Acute kidney Injury in cardiorenal syndrome type 1 patients: a systematic review and meta-analysis. *Cardiorenal Med.* **6**, 116–128. https://doi.org/10.1159/000442300 (2016).
8. Gui, Y., Palanza, Z., Fu, H. & Zhou, D. Acute kidney injury in diabetes mellitus: Epidemiology, diagnostic, and therapeutic concepts. *FASEB Journal: Official Publication Federation Am. Soc. Experimental Biology.* **37**, e22884. https://doi.org/10.1096/fj.202201340RR (2023).
9. Medel-Herrero, A., Mitchell, D., Moyce, S., Udaltsova, I. & Schenker, M. The expanding burden of acute kidney injury in California: Impact of the epidemic of diabetes on kidney Injury Hospital admissions. *Nephrol. Nurs. Journal: J. Am. Nephrol. Nurses' Association.* **46**, 629–640 (2019).
10. Mina, G. S., Gill, P., Soliman, D., Reddy, P. & Dominic, P. Diabetes mellitus is associated with increased acute kidney injury and 1-year mortality after transcatheter aortic valve replacement: A meta-analysis. *Clin. Cardiol.* **40**, 726–731. https://doi.org/10.1002/clc.22723 (2017).
11. Gerber, Y. et al. A Contemporary Appraisal of the heart failure epidemic in Olmsted County, Minnesota, 2000 to 2010. *JAMA Intern. Med.* **175**, 996–1004. https://doi.org/10.1001/jamainternmed.2015.0924 (2015).
12. van den Berge, J. C. et al. Short- and long-term prognosis of patients with Acute Heart failure with and without diabetes: Changes over the last three decades. *Diabetes care* **41**, 143–149. https://doi.org/10.2337/dc17-0544 (2018).

13. Lawson, C. A. et al. 20-year trends in cause-specific heart failure outcomes by sex, socioeconomic status, and place of diagnosis: A population-based study. *Lancet Public. Health.* **4**, e406–e420. https://doi.org/10.1016/s2468-2667(19)30108-2 (2019).
14. Kristensen, S. L. et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: A systematic review and meta-analysis of cardiovascular outcome trials. *Lancet Diabets Endocrionol.* **7**, 776–785. https://doi.org/10.1016/s2213-8587(19)30249-9 (2019).
15. Cheng, Y. et al. Association between base excess and mortality among patients in ICU with acute kidney Injury. *Front. Med.* **8**, 779627. https://doi.org/10.3389/fmed.2021.779627 (2021).
16. Ronco, C., Bellomo, R. & Kellum, J. A. Acute kidney injury. *Lancet (London England)* **394**, 1949–1964. https://doi.org/10.1016/s0140-6736(19)32563-2 (2019).
17. Coelho, S., Cabral, G., Lopes, J. A. & Jacinto, A. Renal regeneration after acute kidney injury. *Nephrol. (Carlton Vic).* **23**, 805–814. https://doi.org/10.1111/nep.13256 (2018).
18. Campion, D. et al. A prospective, multicenter, three-cohort study evaluating contrast-induced acute kidney injury (CI-AKI) in patients with cirrhosis. *J. Hepatol.* **80**, 62–72. https://doi.org/10.1016/j.jhep.2023.10.010 (2024).
19. Yue, S. et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. *J. Translational Med.* **20** https://doi.org/10.1186/s12967-022-03364-0 (2022).
20. Cuocolo, R., Perillo, T., De Rosa, E., Ugga, L. & Petretta, M. Current applications of big data and machine learning in cardiology. *J. Geriatric Cardiology: JGC* **16**, 601–607. https://doi.org/10.11909/j.issn.1671-5411.2019.08.002 (2019).
21. Daghistani, T. A. et al. Predictors of in-hospital length of stay among cardiac patients: A machine learning approach. *Int. J. Cardiol.* **288**, 140–147. https://doi.org/10.1016/j.ijcard.2019.01.046 (2019).
22. Przewlocka-Kosmala, M., Marwick, T. H., Dabrowski, A. & Kosmala, W. Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: A classification based on machine learning. *J. Am. Soc. Echocardiography: Official Publication Am. Soc. Echocardiography*. **32**, 604–615e606. https://doi.org/10.1016/j.echo.2018.12.002 (2019).
23. Goldberger, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101**, E215–220. https://doi.org/10.1161/01.cir.101.23.e215 (2000).
24. Stevens, L. A., Coresh, J., Greene, T. & Levey, A. S. Assessing kidney function–measured and estimated glomerular filtration rate. *N. Engl. J. Med.* **354**, 2473–2483. https://doi.org/10.1056/NEJMra054415 (2006).
25. Li, Y. et al. dynr.mi: An R program for multiple imputation in dynamic modeling. *World Acad. Sci. Eng. Technol.* **13**, 302–311. https://doi.org/10.5281/zenodo.3298841 (2019).
26. Zhang, Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals Translational Med.* **4** https://doi.org/10.3978/j.issn.2305-5839.2015.12.63 (2016).
27. Singer, M. et al. The Third International Consensus definitions for Sepsis and septic shock (Sepsis-3). *Jama* **315**, 801–810. https://doi.org/10.1001/jama.2016.0287 (2016).
28. Zhang, Z., Gayle, A. A., Wang, J. & Zhang, H. Cardinal-Fernández, P. comparing baseline characteristics between groups: An introduction to the CBCgrps package. *Annals Translational Med.* **5**, 484. https://doi.org/10.21037/atm.2017.09.39 (2017).
29. Zhang, Z. Variable selection with stepwise and best subset approaches. *Annals Translational Med.* **4** https://doi.org/10.21037/atm.2016.03.35 (2016).
30. Habas, E. Anemia and hypoxia impact on chronic kidney disease onset and progression: Review and updates. *Cureus* **15**, e46737. https://doi.org/10.7759/cureus.46737 (2023).
31. Fage, N., Asfar, P., Radermacher, P. & Demiselle, J. Norepinephrine and vasopressin in hemorrhagic shock: A focus on renal hemodynamics. *Int. J. Mol. Sci.* **24** https://doi.org/10.3390/ijms24044103 (2023).
32. Villa, G. et al. Hemodynamic instability during acute kidney Injury and Acute Renal replacement therapy: Pathophysiology and clinical implications. *Blood Purif.* **50**, 729–739. https://doi.org/10.1159/000513942 (2021).
33. Ge, T. et al. A rare cause of abdominal pain managed unconventionally: acute renal infarction caused by atrial fibrillation: A case report. *J. Med. Case Rep.* **16** https://doi.org/10.1186/s13256-022-03608-z (2022).
34. Juan, C. X. et al. Exosome-mediated pyroptosis of mir-93-TXNIP-NLRP3 leads to functional difference between M1 and M2 macrophages in sepsis-induced acute kidney injury. *J. Cell. Mol. Med.* **25**, 4786–4799. https://doi.org/10.1111/jcmm.16449 (2021).
35. Bilgili, B., Haliloğlu, M. & Cinel, İ. Sepsis and acute kidney Injury. *Turkish J. Anaesthesiol. Reanimation*. **42**, 294–301. https://doi.org/10.5152/tjar.2014.83436 (2014).
36. Caus, M., Eritja, À. & Bozic, M. Role of microRNAs in obesity-related kidney disease. *Int. J. Mol. Sci.* **22** https://doi.org/10.3390/ijms222111416 (2021).
37. do Carmo, J. M. et al. Parental obesity predisposes offspring to kidney dysfunction and increased susceptibility to ischemia-reperfusion injury in a sex-dependent manner. *Am. J. Physiol. Renal. Physiol.* **326**, F727–f736. https://doi.org/10.1152/ajprenal.00294.2023 (2024).
38. Fassett, R. G. et al. Biomarkers in chronic kidney disease: A review. *Kidney Int.* **80**, 806–821. https://doi.org/10.1038/ki.2011.198 (2011).
39. Schrezenmeier, E. V., Barasch, J., Budde, K., Westhoff, T. & Schmidt-Ott, K. M. Biomarkers in acute kidney injury - pathophysiological basis and clinical performance. *Acta Physiol. (Oxford, England)* **219**, 554–572. https://doi.org/10.1111/apha.12764 (2017).
40. Zhou, F., Luo, Q., Wang, L. & Han, L. Diagnostic value of neutrophil gelatinase-associated lipocalin for early diagnosis of cardiac surgery-associated acute kidney injury: A meta-analysis. *Eur. J. cardio-thoracic Surgery: Official J. Eur. Association Cardio-thoracic Surg.* **49**, 746–755. https://doi.org/10.1093/ejcts/ezv199 (2016).
41. Wu, H. et al. IL-18 contributes to renal damage after ischemia-reperfusion. *J. Am. Soc. Nephrology: JASN.* **19**, 2331–2341. https://doi.org/10.1681/asn.2008020170 (2008).
42. Chen, G. et al. Ethnicity-stratified analysis of the Association between TNF-α Genetic Polymorphisms and acute kidney Injury: A systematic review and Meta-analysis. *Biomed. Res. Int.* **2020** (5262351). https://doi.org/10.1155/2020/5262351 (2020).
43. Wilson, F. P. Machine learning to predict acute kidney injury. *Am. J. Kidney Diseases: Official J. Natl. Kidney Foundation.* **75**, 965–967. https://doi.org/10.1053/j.ajkd.2019.08.010 (2020).
44. Fan, Z. et al. Construction and validation of prognostic models in critically ill patients with sepsis-associated acute kidney injury: Interpretable machine learning approach. *J. Translational Med.* **21**, 406. https://doi.org/10.1186/s12967-023-04205-4 (2023).
45. Lyu, Y. et al. High level of intraoperative lactate might predict acute kidney injury in aortic arch surgery via minimally invasive approach in patients with type A dissection. *Front. Cardiovasc. Med.* **10**, 1188393. https://doi.org/10.3389/fcvm.2023.1188393 (2023).
46. Jing, H. et al. Predicting the risk of acute kidney injury after cardiopulmonary bypass: Development and assessment of a new predictive nomogram. *BMC Anesthesiol.* **22**, 379. https://doi.org/10.1186/s12871-022-01925-w (2022).

Acknowledgements

Not applicable.

Author contributions

All authors contributed to the study conception and design. Writing - original draft preparation: Guojing Li; Writing - review and editing: Mengyao Zhang; Conceptualization: Guojing Li; Zhiqiang Zhao; Methodology:

Zongliang Yu; Junyi Liao; Formal analysis and investigation: Guojing Li; Zhiqiang Zhao; Funding acquisition: Mengyao Zhang; Zongliang Yu; Resources: Mengyao Zhang; Supervision: Zongliang Yu; Junyi Liao; Mengyao Zhang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Youth Science Foundation Project (81700737) and Jiangsu Province Suzhou Science and Technology Plan Project (SKY2023028).

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
The data used in this study originated from the MIMIC-IV database developed by the Beth Israel Deaconess Medical Center between 2008 and 2019, and therefore, patient consent and ethical approval were not required.

Consent for publication
Not applicable.

Additional information
Correspondence and requests for materials should be addressed to M.Z.