



## OPEN ACCESS

## EDITED BY

Herwig Unger,  
University of Hagen, Germany

## REVIEWED BY

Anirach Mingkhwan,  
King Mongkut's University of Technology  
North Bangkok, Thailand  
Peter Kropf,  
University of Neuchatel, Switzerland

## \*CORRESPONDENCE

Xianyao Wan  
✉ 13322210199@163.com  
Jia Wang  
✉ emily.jia.1987@163.com  
Congfeng Zhang  
✉ 18241514131@163.com

†These authors have contributed  
equally to this work

RECEIVED 05 May 2025

ACCEPTED 23 July 2025

PUBLISHED 08 August 2025

## CITATION

Wei Y, Xu W, Yang S, Zhang C, Wang J  
and Wan X (2025) Significant adverse  
prognostic events in patients with  
urosepsis: a machine learning based model  
development and validation study.  
*Front. Cell. Infect. Microbiol.* 15:1623109.  
doi: 10.3389/fcimb.2025.1623109

## COPYRIGHT

© 2025 Wei, Xu, Yang, Zhang, Wang and Wan.  
This is an open-access article distributed under  
the terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Significant adverse prognostic events in patients with urosepsis: a machine learning based model development and validation study

Yiqu Wei<sup>1,2†</sup>, Wanqing Xu<sup>3†</sup>, Shuo Yang<sup>4†</sup>, Congfeng Zhang<sup>2\*</sup>,  
Jia Wang<sup>1\*</sup> and Xianyao Wan<sup>1\*</sup>

<sup>1</sup>Department of Critical Care Medicine, The First Affiliated Hospital of Dalian Medical University, Dalian, China, <sup>2</sup>Department of Critical Care Medicine, Dandong Central Hospital, Dandong, China, <sup>3</sup>Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, <sup>4</sup>Department of Anesthesiology, The Second Affiliated Hospital of Dalian Medical University, Dalian, China

**Background:** Urosepsis is a subset of sepsis with a high mortality rate. Currently, the ranking of urosepsis in sepsis etiology is on the rise. Our goal is to use machine learning (ML) methods to construct and validate an interpretable prognosis prediction model for patients with urosepsis.

**Method:** Data were collected from the Intensive Care Medical Information Mart IV database version 3.1 and divided into a training cohort and a validation cohort in a 7:3 ratio. Random Forest (RF), Lasso, Boruta, and eXtreme Gradient Boosting (XGBoost) were used to identify the most influential variables in the model development dataset, and the optimal variables were selected based on achieving the  $\lambda_{1se}$  value. Model development includes seven machine learning methods and ten cross validations. Accuracy and Decision Curve Analysis (DCA) were used to evaluate the performance of the model in order to select the optimal model. Internal validation of the model included area under the ROC curve (AUC), sensitivity, specificity, Matthews correlation coefficient, and F1-score. Finally, SHapley Additive exPlans (SHAP) was used to explain ML models.

**Result:** A total of 1389 patients with urosepsis were included. Optimal predictors were selected through statistical regularization, yielding a parsimonious set of 9 variables for model development. The performance of XGBoost model is the best and the accuracy of XGBoost was 0.818, with an AUC of 0.904 (95% CI: 0.886–0.923). The internal validation accuracy was 0.797, AUC was 0.869 (95% CI: 0.834–0.904), sensitivity was 0.797, specificity was 0.752, Matthews correlation coefficient was 0.597, and F1-score was 0.791. This indicates that the predictive model performs well in internal validation. SHAP-based summary graphs and diagrams were used to globally explain the XGBoost model.

**Conclusion:** ML demonstrates strong prognostic capability in urosepsis, with the SHAP method providing clinically intuitive explanations of model predictions. This enables clinicians to identify critical prognostic factors and personalize

treatments. While our model achieved high predictive accuracy, its retrospective derivation from a single-center database necessitates external validation in diverse populations, which should be addressed through future prospective multicenter studies to establish clinical generalizability.

#### KEYWORDS

machine learning, urosepsis, prognostic model, MIMIC-IV database, SHAP

## 1 Introduction

Urosepsis is caused by infection of the genitourinary system, which accounts for about 9% to 31% of sepsis cases, and is one of the worst prognosis diseases for patients with urinary tract infections (Porat et al., 2025). For certain specific populations, the case fatality rate of urosepsis is approximately 25%-60%, highlighting the significant clinical importance of improving the early diagnosis and management of urosepsis (Li et al., 2024a). Considering the high incidence rate and mortality of urosepsis, it is necessary to establish a reliable and effective prognosis model.

Several risk prediction models for urosepsis patients have been widely studied and established. Villanueva-Congote (Villanueva-Congote et al., 2024) et al. have shown that the Neutrophil-to-Lymphocyte Ratio (NLR) and Platelet-to-Lymphocyte Ratio (PLR) may be valuable prognostic indicators for predicting the risk of urosepsis. They can help clinicians with early risk stratification, timely intervention, and resource allocation. Croghan (Croghan et al., 2023) et al. conducted a prospective multi-institutional study, recording the baseline and continuous ureteroscopic intrarenal pressure (IRP) of patients during ureteroscopy (URS) surgery, and found that there seemed to be a relationship between elevated IRP and postoperative urosepsis. Canat (Canat et al., 2020) et al. found that the elevated level of procalcitonin (PCT) on the second day after prostate biopsy was a statistically significant independent predictor of urosepsis. It can be used as an early biomarker to predict the occurrence of urosepsis after prostate biopsy. However, the above predictive models are based on traditional COX regression that suffers from limitations including its reliance on the proportional hazards assumption, inability to estimate baseline hazard functions directly, sensitivity to multicollinearity, and inadequate handling of high-dimensional data and nonlinear relationships, necessitating more advanced methods for complex datasets.

In recent years, various machine learning (ML) algorithms, a method of data analysis that develops algorithms to predict outcomes by learning from data, have been studied for the early detection of urosepsis. It is superior to traditional statistical methods and does not require assumptions about input variables and their relationship with outputs. The advantage of fully data-

driven learning without relying on rule-based programming is that ML constitutes a reasonable approach. Researchers have successfully applied a variety of ML paradigms to improve the generalization ability of models in complex clinical scenarios, such as Gradient Boosting Decision Trees (GBDT), Random Forests (RF) and Deep Neural Networks (DNN). These methods further enhance the interpretability of the model through the pathophysiological associations revealed by feature importance analysis (Su et al., 2023). The occurrence of urosepsis after PCNL surgery is one of the main reasons for the increased mortality. Li (Li et al., 2024b) et al. collected important preoperative and intraoperative clinical data of patients and established a model combined with ML methods that for predicting the occurrence of urosepsis after PCNL. The results showed that the model had a good predictive effect on the occurrence of urosepsis after PCNL (AUC = 0.89). In addition, they also found that the change of platelet counts before and after surgery was an important predictive factor (Li et al., 2024b). However, although ML algorithms have performed well in previous studies, due to the “black box” nature of ML algorithms, it is difficult to interpret which characteristics of patients are responsible for a given prediction. In addition, the main result of the above study is to detect the occurrence of urosepsis, rather than adverse clinical outcomes. Moreover, the sample size included in the study is too small, resulting in low clinical credibility.

Therefore, we use large-scale data based on MIMIC-IV database, to develop a prognosis prediction model for critically ill patients with urosepsis to improve the reliability of research conclusions. Our feature selection ensemble—incorporating Random Forest, Lasso, Boruta, and XGBoost—mitigates algorithm selection bias by spanning diverse learning paradigms, thereby generating robust variable rankings essential for clinical modeling. In addition, to explain the results of the ML model, we combine advanced ML algorithms based on SHapley Additive exPlanations (SHAP), a popular ML technique for a deeper understanding of the complex relationship between features and predictions. In addition to optimizing the predictive performance of mortality risk in critically ill patients with urosepsis, this study also provides intuitive explanations that will help clinicians fully understand how the developed model makes specific predictions and increase opportunities for early intervention.

## 2 Methods

### 2.1 Source of data

An open and free intensive care database called Medical Information Mart for Intensive Care IV (MIMIC-IV) version 3.1 (Johnson et al., 2023; Wang et al., 2024b), which contains the latest version of comprehensive clinical data of patients admitted to Beth Israel Deaconess Medical Center in Massachusetts from 2008 to 2022. MIMIC-IV contains data on 65000 patients admitted to the ICU and 200000 patients admitted to the emergency room. The clinical data in the database include demographic characteristics, vital signs, imaging examinations, laboratory test results, data dictionaries and documents containing the codes of the ninth and tenth editions of the international classification of diseases (ICD-9 and ICD-10, respectively), as well as hourly physiological data records beside the monitors verified by ICU nurses. The health information obtained from the MIMIC-IV database could not be identified, so the informed consent of patients was not required (Goldberger et al., 2000; Oweira et al., 2018). This study was approved to extract data from the database for research purposes (certification number: 58407754). The database has been approved by the Massachusetts Institute of Technology (MIT) institutional review board (IRB).

### 2.2 Study design and population

The study focused on patients with urosepsis who were subsequently hospitalized and admitted to the ICU for the first time. According to the definition of Sepsis-3.0, sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection (Ohbe et al., 2024). Organ dysfunction represents at least two points identified as acute and infection-related changes in the Sequential Organ Failure Assessment (SOFA). Urosepsis is sepsis originating from the urogenital tract. The diagnosis of urosepsis requires both suspicion of sepsis as well as evidence of a urinary tract infections (UTI). In this study, we included common types of urinary tract infections, including pyelonephritis, cystitis, and urinary tract infections. The codes in ICD-9 and ICD-10 include 590, 595, 599.0 and N30. Patients who met the following criteria in the database were selected for this study: [1] first admission to ICU; [2] ICU stay >24 hours; [3] Age >18 years old; [4] It meets the diagnostic criteria of sepsis 3.0; [5] There is conclusive evidence of urinary tract infection, such as positive urine culture. The diagnosis code contains the diagnosis related to urinary tract infection and has a higher priority than other infections (Griebing, 2019). In the MIMIC-IV database, ICD-9 (99591, 99592, and 78552) and ICD-10 (R65.20, R65.21) codes were used to identify patients with sepsis. Following these criteria, we screened 1389 patients for the study (Figure 1). The final patient cohort was allocated to training cohort and validation cohort according to the ratio of 7:3 through stratified random partitioning for model establishment.

### 2.3 Data extraction

We first obtained the raw data using the Structured Query language of Navicat Premium software (version 16.3.8), including sociodemographic characteristics, vital signs, laboratory parameters, complications, and microbial information (Supplementary Figure S1) (Yang et al., 2020). We extracted the following demographic data: age, sex, race, weight, height, admission route, length of stay in the ICU, and hospital expiration flag (records of in-hospital deaths in the database) at the time of first admission to the ICU. Next, the vital signs of patients in the first 24 hours of ICU stay were collected, including mean arterial pressure (MAP), heart rate (HR), body temperature (T), respiratory rate (RR), saturation of peripheral oxygen (SpO<sub>2</sub>), urine volume, and then the laboratory parameters in the first 24 hours were collected, including blood routine examination, liver and kidney function, blood glucose, and arterial blood gas (ABG). In addition, advanced life support records such as mechanical ventilation and renal replacement therapy were also recorded. We removed more than 20% of the variables with missing observations, such as height and serum albumin level, to promote and ensure the accuracy of the study. Then, we used the mice and VIM package to process the missing data. Missing data were completely random. With the help of the RF algorithm, we performed 5 imputations of 50 iterations for the original missing data and completed the sensitivity analysis. When combining the characteristics of vital signs and relevant laboratory parameters, the maximum, minimum, and average values were used and considered as independent characteristics to be included in the study.

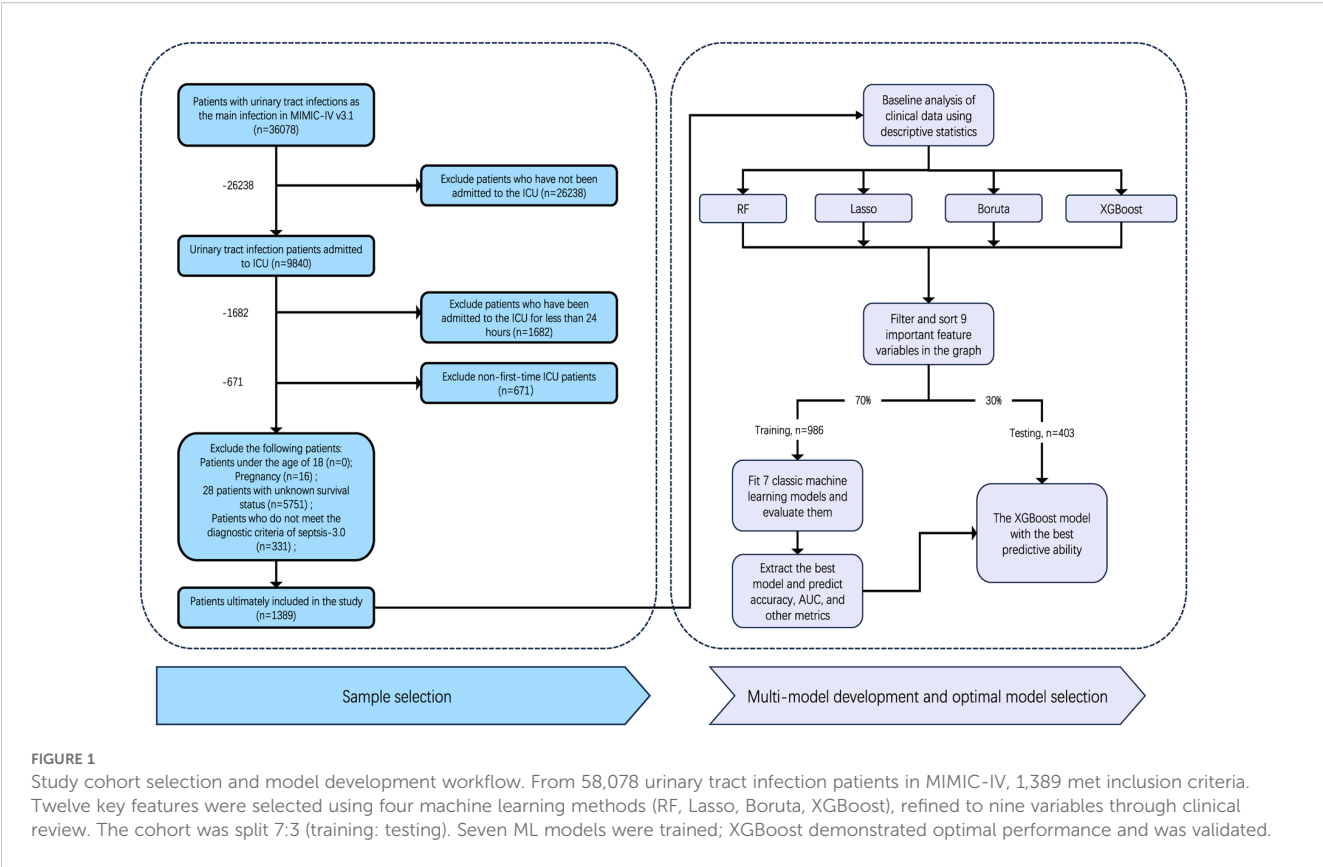
### 2.4 Clinical outcomes

The clinical outcome of the current study was 28 days All-Cause Mortality (ACM). Crucially, the time of death were specified as occurrences of death within a defined period following admission to the ICU, rather than merely identifying whether the patient was deceased at a specific time point.

### 2.5 Statistical analyses

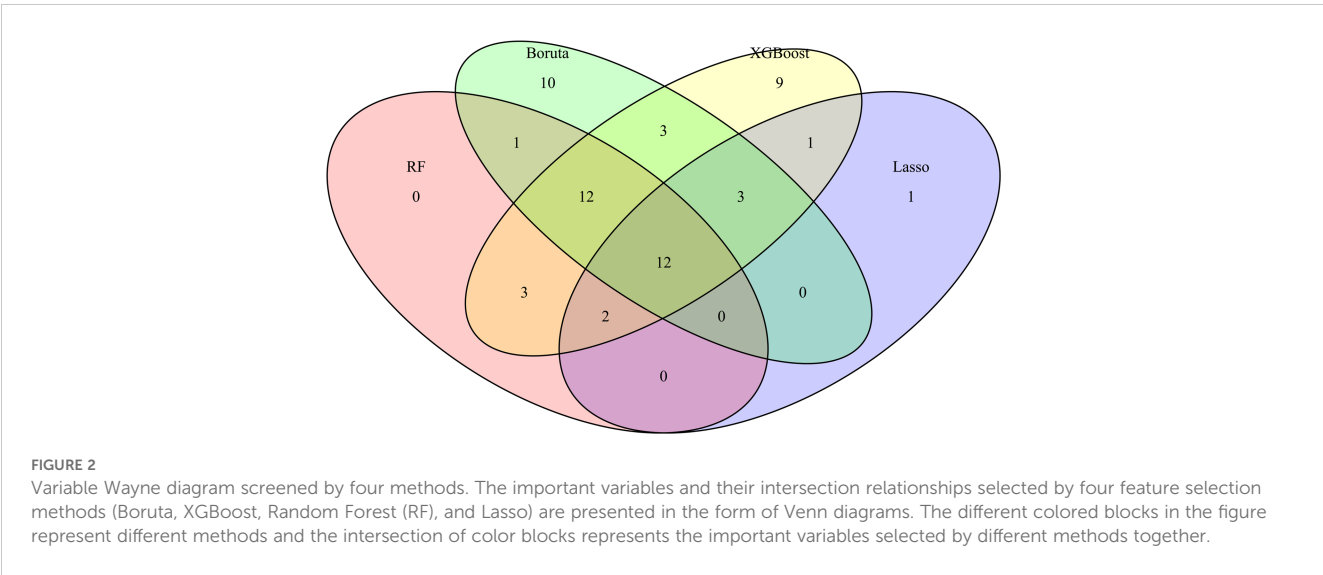
Shapiro Wilks test was used for the normality test. Continuous variables with normal distribution were expressed as mean (SD, standard deviation) and compared with independent samples by T-test. Non normally distributed variables were expressed as median (interquartile range) and compared with the Kruskal Wallis test. Categorical variables were described as percentages and compared using the Chi-Square test.

In addition, candidate data variables were additionally screened according to the principle of variable reduction to determine whether they were included in the model (Figure 2) (Venkatesh et al., 2023). RF is a classification algorithm composed of multiple decision trees. It constructs a machine learning model by randomly



sampling training data and searching for the optimal segmentation solution. Each decision tree in the RF was constructed using feature measures aligned to dataset attributes (Farhadian et al., 2020; Shi et al., 2023), effectively evaluating the importance of each feature (Nachouki et al., 2023). Lasso can select variables through a series of parameters and reduce the complexity of the model, thus avoiding overfitting. Lasso's complexity is controlled by  $\lambda$ , which eventually leads to a model with fewer variables. Compared with the traditional feature selection algorithm, Boruta is a packaging-

based method to select features. Its goal is to identify the feature set with the greatest correlation with the dependent variable, rather than focusing solely on creating an optimized compact subset for a particular model (Zhou et al., 2023). By iteratively eliminating low correlation features, it effectively reduces signal noise and produces consistent classification performance (Sun et al., 2022b). At the same time, XGBoost stands out as an influential ensemble learning technology rooted in the classification tree framework; It combines low-accuracy classifiers into high-accuracy classifiers through







in Figure 3. Finally, there are 9 variables used as predictive indicators, including urea nitrogen minimum, age, urine output (24-hour average), urine output (6-hour average), alkaline phosphatase maximum, SpO<sub>2</sub>, alkaline phosphatase minimum, MAP, and OASIS.

### 3.3 Model comparison

Following the identification of key predictive factors through rigorous feature selection, we developed and evaluated seven ML models (Bayes, Bayesian Network; DT, Decision Tree; LR, Logistic Regression; MLP, Multilayer Perceptron; RF, Random Forest; SVM, Support Vector Machine; and XGBoost, eXtreme Gradient Boosting) to assess their predictive performance, which demonstrated robust discrimination ability, with the AUC (95% CI) of the training cohort being 0.819 (0.792, 0.845), 0.841 (0.814, 0.865), 0.851 (0.828, 0.875), 0.87 (0.847, 0.892), 0.906 (0.887, 0.924), 0.897 (0.877, 0.916), and 0.904 (0.886, 0.923), respectively (Figure 4). The RF algorithm model showed that the training cohort has the highest AUC. The XGBoost model was second only to RF in performance and significantly better than the other five models. The AUC (95% CI) of the validation cohort were 0.817 (0.775, 0.859), 0.779 (0.735, 0.822), 0.856 (0.819, 0.893), 0.781 (0.737, 0.825), 0.835 (0.797, 0.873), 0.847 (0.809, 0.885), 0.869 (0.834, 0.904) (Figure 5). In the validation cohort, the XGBoost model has the highest AUC, followed by LR, and the RF model has the fourth highest AUC. The results of the accuracy, precision, recall, and F1-score of the seven models were shown in Figure 6 and Supplementary Figure S4. The performance of XGBoost classification model is better than other models. According to the

DCA results of the seven prediction models (Figure 7), the net benefit of RF is greater than that of other models. In this study, the ROC curve and DCA curve of the training cohort and the validation cohort were evaluated, and the classification ability, calibration degree, and clinical application value of each model were compared. The XGBoost model was the best, although its net income was slightly less than that of RF.

The XGBoost model demonstrated excellent calibration, with Brier scores of 0.142 (training) and 0.178 (validation)—well below the 0.25 random prediction threshold—indicating high probabilistic precision. The strong calibration performance – evidenced by Brier score 0.143, calibration slope 1.18, and visual alignment in Figure 8 – confirms that predicted probabilities reliably reflect actual mortality risk.

Both RF and XGBoost demonstrated excellent discrimination in the training cohort (RF AUC=0.906; XGBoost AUC=0.904). However, RF showed poorer generalization in validation (AUC=0.835 vs. XGBoost's 0.869), indicating potential overfitting from RF's deep-tree architecture capturing noise in training data. Complementing this, calibration metrics revealed a similar pattern: while RF achieved a Brier score of 0.1816 in validation, XGBoost attained slightly better calibration (Brier=0.1783), confirming its probabilistic reliability. This enhanced stability stems from XGBoost's regularization mechanisms (L1/L2 penalties, column subsampling), which mitigate overfitting whereas RF lacks comparable constraints. Thus, XGBoost was selected as the optimal model, as its superior validation performance (higher AUC + better Brier score) reflects greater clinical utility for probability-based decisions, despite marginally lower training performance. This addition reinforces our selection rationale through both discrimination and calibration perspectives.

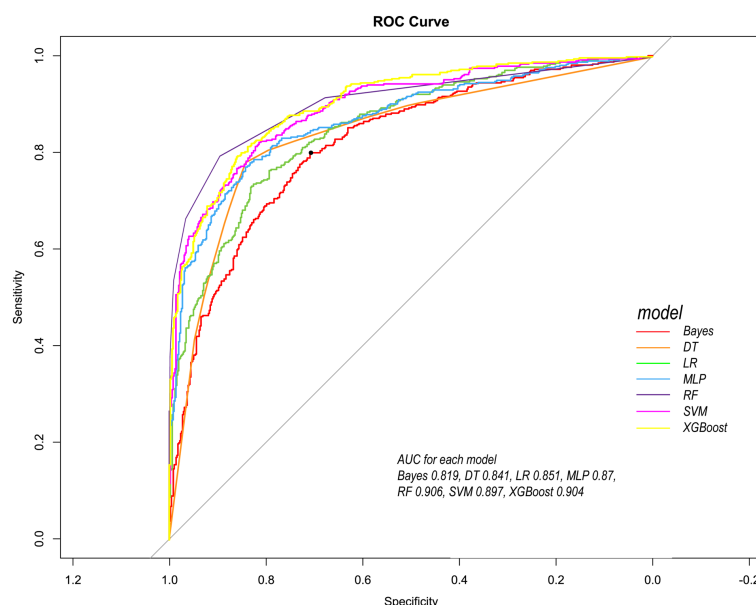
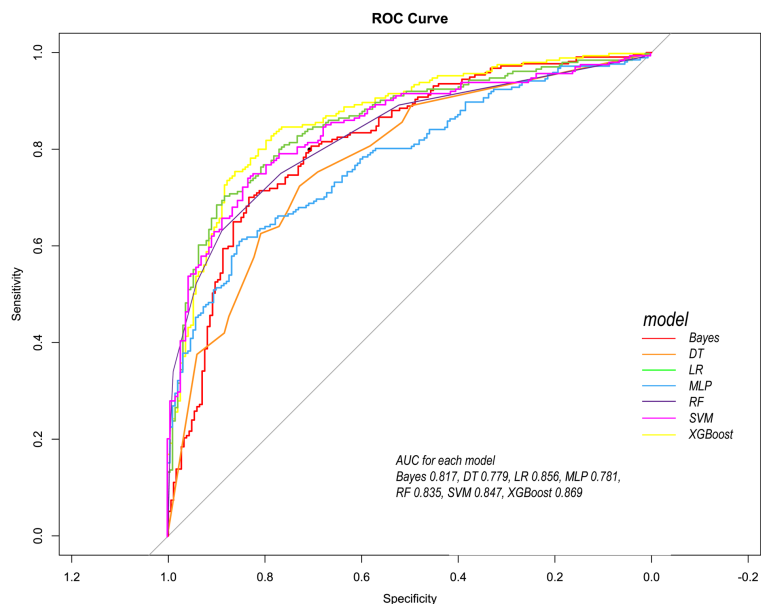


FIGURE 4

Comparison of ROC curves of seven models in the training cohort. Red line = Bayes model, orange line = DT model, green line = LR model, blue line = MLP model, dark purple line = RF model, bright purple line = SVM model, yellow line = XGBoost model.



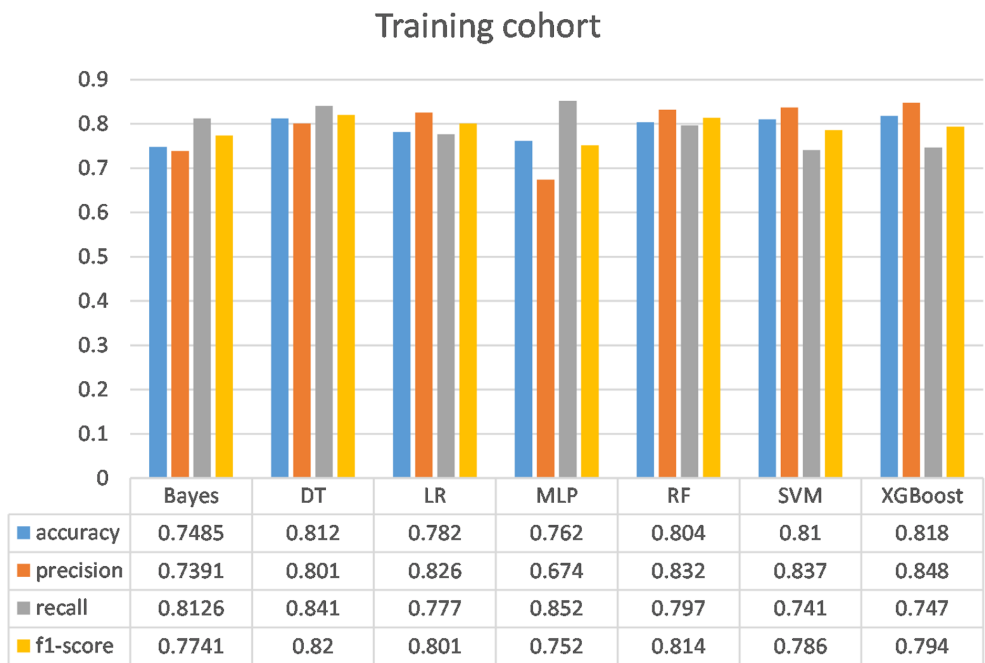
**FIGURE 5**  
Comparison of ROC curves of seven models in the internal validation cohort. Red line = Bayes model, orange line = DT model, green line =LR model, blue line =MLP model, dark purple line =RF model, bright purple line = SVM model, yellow line = XGBoost model.

### 3.4 Interpretability analysis

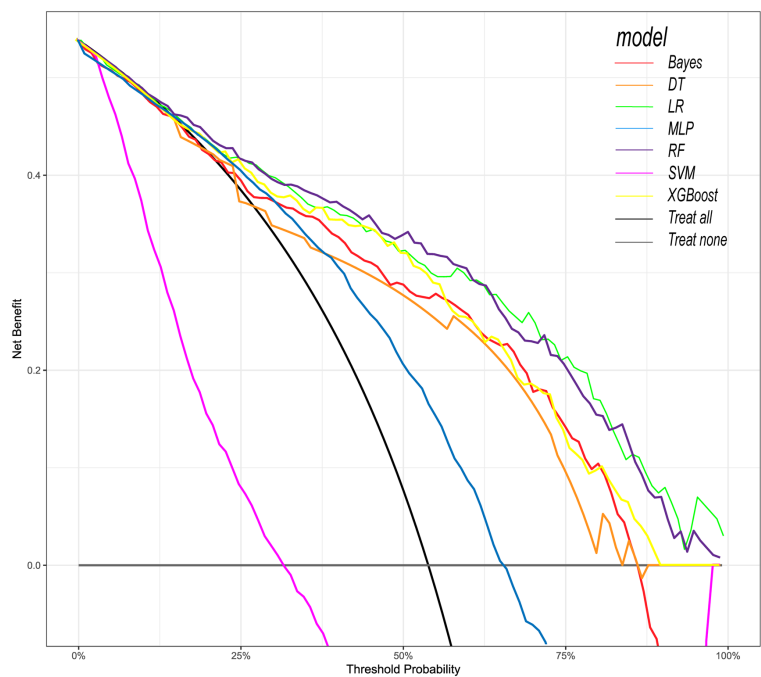
First, the global interpretability of the baseline model was studied. The XGBoost model was considered the baseline model because it was found to be the best-performing model. Feature importance estimates were based on the overall sample of the

training cohort. The global importance of each feature we estimated in SHAP was used to understand the general impact of various features in all samples (Figure 9).

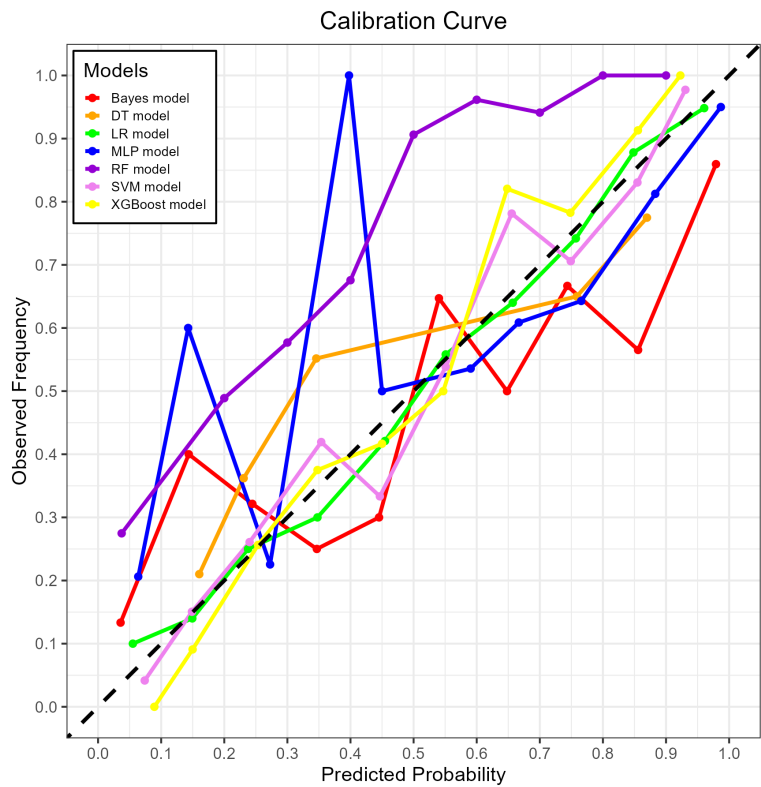
The SHAP summary graph illustrates the entire distribution of the impact of each feature on the model output. Color enables us to understand how changes in eigenvalues affect changes in results.



**FIGURE 6**  
Comparison of the performance of the seven models in the training cohort. Bayes, Bayesian Network; DT, Decision tree; LR, Logistic regression model; MLP, Multilayer perceptron; RF, Random Forest model; SVM, Support vector machine; XGBoost, eXtreme Gradient Boosting.



**FIGURE 7**  
Decision curve analysis (DCA) of seven prediction models. The net benefit curve of the prognostic model was shown. The x-axis represents the threshold probability of intensive care outcome, and the y-axis represents the net benefit. Red line = Bayes model, orange line = DT model, green line = LR model, blue line = MLP model, dark purple line = RF model, bright purple line = SVM model, yellow line = XGBoost model, black line = Treat all, gray line = Treat none.



**FIGURE 8**  
The calibration curve plot of the seven models. Red line = Bayes model, orange line = DT model, green line = LR model, blue line = MLP model, dark purple line = RF model, bright purple line = SVM model, yellow line = XGBoost model.



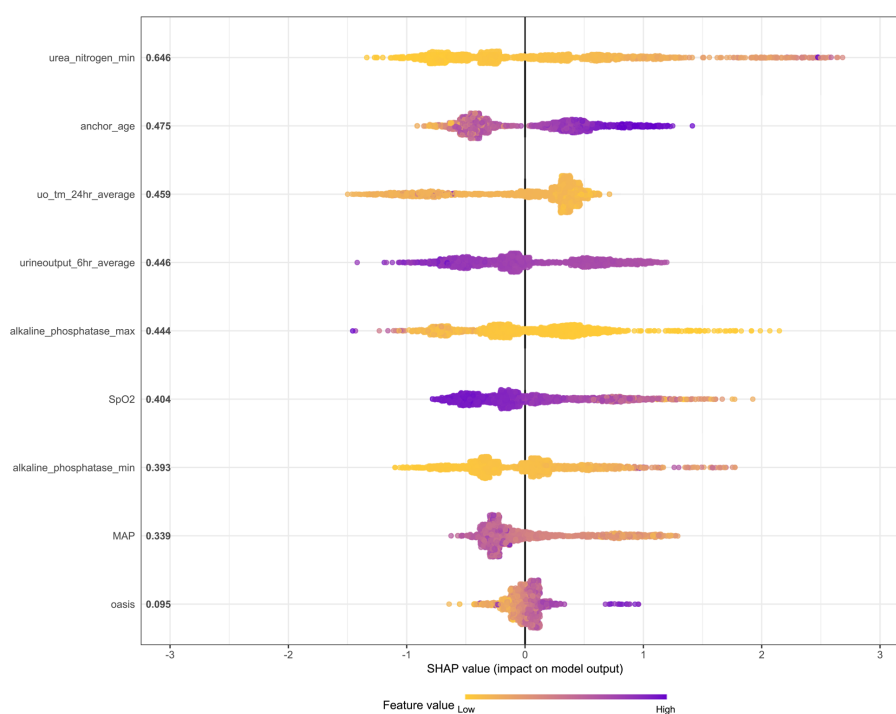


FIGURE 9

The SHAP method is used to analyze the important features of the XGBoost model. Create a point for each feature attribute value of each patient's model, thereby assigning a point to each patient on the line for each feature. Dots are colored according to the eigenvalues of the corresponding patients and accumulate vertically to depict the density. Purple indicates high eigenvalues (death in this case), while yellow indicates low eigenvalues. The farther the point is from the baseline SHAP value, the greater the impact on the output.

Purple represents high eigenvalues, while yellow represents low eigenvalues. The farther a point is from the baseline SHAP value zero, the greater its impact on the output. This allows a better understanding of the relationship between features and the SHAP value (as well as the predicted output). It can be seen from the figure that urea nitrogen min plays a crucial role compared with other risk factors (such as MAP and OASIS).

In addition, local interpretation analyzed the results of specific predictions for individual patients. Figure 10A presents data from a

urosepsis patient who died during ICU hospitalization. Our prediction model assigned this patient a mortality probability of 96%. The figure demonstrates that urea nitrogen min, urine output 6hr average, OASIS, anchor age, MAP, and SpO<sub>2</sub> contributed to biasing the prediction towards mortality, whereas alkaline phosphatase max, alkaline phosphatase min, and uo tm 24hour average reduced the predicted risk of death. Figure 10B illustrates data from a surviving urosepsis patient during ICU hospitalization, specifically highlighting features favoring mortality and their actual

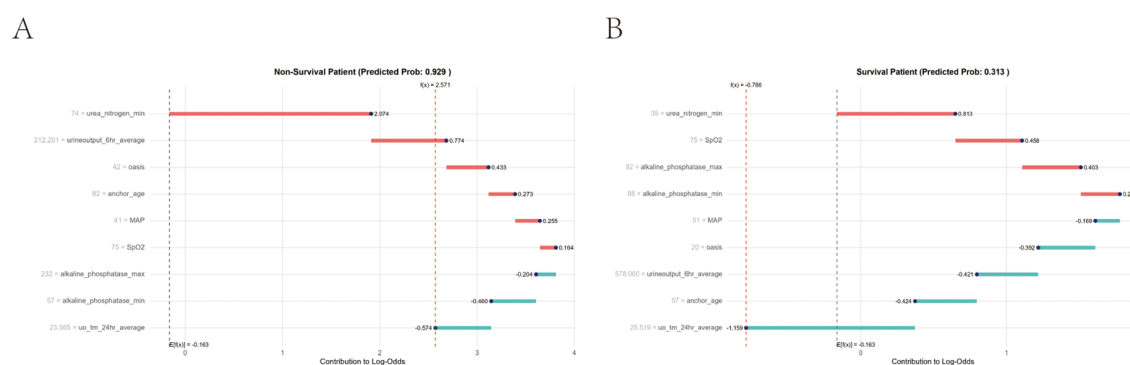


FIGURE 10

Local model explanation by the Shapley Additive Explanations (SHAP) method. (A) Non-survival patient. (B) Survival patient. Each patient is represented by the x-axis, while the feature contribution is represented by the y-axis: an increased red part for each individual patient represents a greater probability toward the decision of "Non-survival".

measurements. In this case, the model predicted a 31.3% probability of mortality. The x-axis denotes individual patients, and the y-axis represents feature contributions. For each patient, the extent of the red area indicates the magnitude of contribution towards a 'non-survival' prediction.

Beyond interpretability, SHAP results translate into actionable clinical protocols. For patients with critically elevated urea nitrogen levels, intensivists should prioritize dynamic renal function monitoring and initiate early nephroprotective interventions to mitigate acute kidney injury—a dominant predictor of mortality. Similarly, sustained depression of mean arterial pressure (MAP) below clinically significant thresholds warrants immediate hemodynamic optimization, including fluid resuscitation and vasopressor escalation when indicated. These data-driven alerts, derived from SHAP's quantification of feature contributions, can be integrated into ICU monitoring systems to proactively guide bedside decisions, converting model insights into tangible clinical workflows.

## 4 Discussion

In recent years, ML algorithms have become increasingly popular in the medical field, helping clinicians diagnose diseases faster and more accurately, while achieving personalized treatment plans. In this study, we first used ML methods to construct a predictive model for major adverse prognostic events in patients with urosepsis. It has been proven that ML methods can explain the key characteristics of patients with urosepsis and establish high-precision death prediction models. Compared with traditional risk models, this ML based approach considers the complex interactions between variables and can dynamically adjust based on individual patient characteristics.

Our research analyzed demographic data, vital signs, 80 laboratory indicators, advanced life support data, and comorbidities of patients with urosepsis within 24 hours after admission. Specifically, we use RF, Lasso, Boruta, and XGBoost to model the variables of influencing factors. Then, we rank the variables with non-zero coefficients according to their impact on the outcome variables and identify common variables by taking the intersection of variables selected by all four methods. Finally, 12 common variables were obtained. Combined with systematic review, meta-analysis, and expert clinical opinions, nine factors affecting the outcome were identified as predictors, including urea nitrogen minimum, age, urine output (24-hour average), urine output (6-hour average), alkaline phosphatase maximum, SpO<sub>2</sub>, alkaline phosphatase minimum, MAP, and OASIS.

Urine output and urea nitrogen are closely related to the main adverse prognostic events of patients (Sun et al., 2022a). Urine output plays an important role in predicting the mortality rate of urosepsis patients. This result has been confirmed in many related studies (Zhang et al., 2021; Kasugai et al., 2023; Yamamoto et al., 2025). Surprisingly, urea nitrogen is the most important predictor of mortality in patients with urosepsis. Urea nitrogen is a metabolic waste produced by the liver, which enters the kidneys through the

bloodstream and is then filtered out by the kidneys (Li et al., 2021). Urea nitrogen levels are important indicators of kidney function, water balance, and protein metabolism. Compared to infection and inflammation markers, urea nitrogen is a low-cost, easily accessible indicator that can reflect kidney damage in patients. Previous research results have shown that urosepsis significantly reduces renal blood flow and renal function, thereby further increasing urea nitrogen levels (Tóth-Heyn et al., 2000; Langenberg et al., 2006; Seely et al., 2011). Therefore, urea nitrogen levels can reflect the organ dysfunction status of patients with urosepsis, and severe organ dysfunction is one of the main causes of death in patients (Wang et al., 2024a). However, previous death prediction models for patients with urosepsis did not use this key factor (Mauk, 2018; Gou et al., 2023).

In addition, our research also suggests that alkaline phosphatase (AP) is an important predictor of prognosis in patients with urosepsis. AP is an endogenous detoxifying enzyme that exists throughout the body in four different isoenzymes: germ cell AP, intestinal AP, placental AP, and a non-specific form primarily derived from the kidneys, liver, and bones (Millán, 2006). AP can dephosphorylate endotoxins (Poelstra et al., 1997; Bentala et al., 2002; Verweij et al., 2004). Through its dephosphorylation ability, AP can not only detoxify endotoxins, but also various compounds. The AP measured in serum is currently used as a diagnostic tool for liver disease (Siddique and Kowdley, 2012), bone disease (Ross et al., 2000), and testicular cancer (Neumann et al., 2011). AP may also play an important role in the treatment of critical illnesses. Its dephosphorylation ability can counteract the adverse cascade reactions of molecules such as PAMP (Koyama et al., 2002) and DAMPs (Picher et al., 2003) during urosepsis. Our study found that AP can also serve as a predictor of outcomes in patients with urosepsis. The maximum and minimum values of AP in patients with urosepsis within 24 hours of admission are important predictors of patient prognosis. This discovery may help clinicians evaluate the patient's condition within 24 hours of admission and determine the patient's prognosis early.

We used the above nine features to construct ML models to predict the prognosis of patients. Among the ML models, the XGBoost model performs the best. XGBoost is a popular ML algorithm in recent years, characterized by fast computing speed, strong generalization ability, and high predictive performance (Hou et al., 2020; Yue et al., 2022). In the ROC curve, XGBoost performs second only to the RF model in the training cohort, but shows the best discriminative ability in the internal validation cohort. When the threshold probability is between 12.5% and 70%, The clinical intervention guided by the XGBoost model provided greater net benefits in the training cohort. DCA indicates that the RF model has the greatest benefit within a reasonable threshold probability, but considering its poor performance in the validation cohort ROC curve, this means that the RF model may not be optimal. Based on the performance of the model in the training and validation cohorts, the XGBoost model has higher clinical application value and good clinical practicality compared to other models. Finally, we used SHAP values to reveal the 'black box' of ML (Molnar, 2019). The SHAP summary diagram illustrates the overall distribution of the

impact of each feature on the model output. SHAP is a flexible method that can be used to interpret individual predictions and global interpretations. SHAP tries to provide an intuitive visualization of how different characteristics affect the predicted outcome. One advantage of SHAP for global interpretation is that SHAP reveals not only the importance of features but also their relationship to output. In addition, the prediction of SHAP is reasonably distributed among eigenvalues. These factors are essential to ensure trust in the technology. Our interpretability framework adheres to established principles for explainable clinical AI (Molnar, 2020), transforming complex model predictions into clinically intuitive decision support.

Our findings demonstrate that ML models substantially outperform conventional prognostic tools—both in accuracy and clinical interpretability. Specifically, while traditional methods like logistic regression (LR)—representing linear modeling approaches—achieved competent validation performance (AUC: 0.856), XGBoost surpassed it by a clinically significant margin ( $\Delta$ AUC: +0.013; Figure 5). This gap arises from ML's ability to capture complex, non-linear interactions that traditional models intrinsically miss. Furthermore, ML uniquely identified dynamic predictors like alkaline phosphatase extrema, whereas conventional methods favor static, guideline-driven variables. Thus, ML transcends incremental accuracy gains; it uncovers pathophysiology-driven decision pathways, converting rigid scores into adaptive, patient-specific prognostication.

However, our research is not without limitations. Firstly, our training cohort and internal validation cohort are both from the MIMIC-IV database, with the majority of patients coming from Western countries; Secondly, we did not conduct a more comprehensive study of the database, which may have led us to overlook some key variables, resulting in potential errors; Thirdly, the retrospective and observational nature of this study may lead to selection bias, which may result in the inclusion of patients who do not fully represent all patients in that category. Therefore, conducting a prospective evaluation is necessary to assess the performance of the model in real-world situations. To further evaluate generalizability, future validation will utilize the eICU Collaborative Research Database—capturing heterogeneous ICU practices across U.S. healthcare systems—to assess model performance using identical endpoints (28-day ACM) and metrics (AUC, sensitivity, specificity, F1-score, Brier score, and DCA-derived net benefit). This will be followed by prospective multi-center testing with local hospital data, maintaining consistent predictor variables, outcome definitions, and performance thresholds to ensure cross-population comparability and clinical utility quantification.

## 5 Conclusions

In conclusion, the ML method is a reliable tool for predicting the prognosis of patients with urosepsis. Combining global and local interpretability methods to interpret the intrinsic information from the XGBoost model may prove clinically useful and help

clinicians customize precise management to maximize the survival of patients with urosepsis.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: This study utilized publicly available datasets, which can be accessed at <https://mimic.mit.edu/>.

## Ethics statement

The studies involving humans were approved by The Institutional Review Boards at Massachusetts Institute of Technology and Beth Israel Deaconess Medical Center. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because Patient data were fully anonymized, eliminating the requirement for individual informed consent in this investigation.

## Author contributions

YW: Data curation, Writing – original draft, Conceptualization. WX: Formal Analysis, Data curation, Writing – original draft. SY: Writing – original draft. CZ: Writing – review & editing, Conceptualization. JW: Conceptualization, Writing – review & editing. XW: Writing – review & editing, Conceptualization.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

We acknowledge the contributions of the MIMIC-IV program registry for their work in creating and maintaining the MIMIC IV database.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2025.1623109/full#supplementary-material>

## References

- Bentala, H., Verweij, W. R., Huizinga-Van der Vlag, A., van Loenen-Weemaes, A. M., Meijer, D. K., and Poelstra, K. (2002). Removal of phosphate from lipid A as a strategy to detoxify lipopolysaccharide. *Shock* 18, 561–566. doi: 10.1097/00024382-200212000-00013
- Canat, H. L., Can, O., Atalay, H. A., Akkas, F., and Ötünçtemur, A. (2020). Procalcitonin as an early indicator of urosepsis following prostate biopsy. *A. ging Male* 23, 431–436. doi: 10.1080/13685538.2018.1512964
- Croghan, S. M., Cunnane, E. M., O'Meara, S., Muheilan, M., Cunnane, C. V., Patterson, K., et al. (2023). *In vivo* ureteroscopic intrarenal pressures and clinical outcomes: a multi-institutional analysis of 120 consecutive patients. *BJU Int.* 132, 531–540. doi: 10.1111/bju.16169
- Farhadian, M., Torkaman, S., and Mojarad, F. (2020). Random forest algorithm to identify factors associated with sports-related dental injuries in 6 to 13-year-old athlete children in Hamadan, Iran-2018 -a cross-sectional study. *BMC Sports Sci. Med. Rehabil.* 12, 69. doi: 10.1186/s13102-020-00217-5
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101, E215–E220. doi: 10.1161/01.cir.101.23.e215
- Gou, J. J., Zhang, C., Han, H. S., and Wu, H. W. (2023). Risk factors of concurrent urinary sepsis in patients with diabetes mellitus comorbid with upper urinary tract calculi. *World J. Diabetes* 14, 1403–1411. doi: 10.4239/wjd.v14.i9.1403
- Griebbling, T. L. (2019). Re: the clinical impact of bacteremia on outcomes in elderly patients with pyelonephritis or urinary sepsis: A prospective multicenter study. *J. Urol* 201, 8. doi: 10.1097/01.ju.0000550105.63529.31
- Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., et al. (2020). Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGBoost. *J. Transl. Med.* 18, 462. doi: 10.1186/s12967-020-02620-5
- Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., et al. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* 10, 1. doi: 10.1038/s41597-022-01899-x
- Kang, J., Choi, Y. J., Kim, I. K., Lee, H. S., Kim, H., Baik, S. H., et al. (2021). LASSO-based machine learning algorithm for prediction of lymph node metastasis in T1 colorectal cancer. *Cancer Res. Treat* 53, 773–783. doi: 10.4143/crt.2020.974
- Kasugai, D., Nakashima, T., and Goto, T. (2023). Clinical implications of urine output-based sepsis-associated acute kidney injury. *Intensive Care Med.* 49, 1263–1265. doi: 10.1007/s00134-023-07190-w
- Koyama, I., Matsunaga, T., Harada, T., Hokari, S., and Komoda, T. (2002). Alkaline phosphatases reduce toxicity of lipopolysaccharides *in vivo* and *in vitro* through dephosphorylation. *Clin. Biochem.* 35, 455–461. doi: 10.1016/s0009-9120(02)00330-2
- Langenberg, C., Wan, L., Egi, M., May, C. N., and Bellomo, R. (2006). Renal blood flow in experimental septic acute renal failure. *Kidney Int.* 69, 1996–2002. doi: 10.1038/sj.ki.5000440
- Li, X., Li, T., Wang, J., Dong, G., Zhang, M., Xu, Z., et al. (2021). Higher blood urea nitrogen level is independently linked with the presence and severity of neonatal sepsis. *Ann. Med.* 53, 2192–2198. doi: 10.1080/07853890.2021.2004317
- Li, P., Tang, Y., Zeng, Q., Mo, C., Ali, N., Bai, B., et al. (2024b). Diagnostic performance of machine learning in systemic infection following percutaneous nephrolithotomy and identification of associated risk factors. *Heliyon* 10, e30956. doi: 10.1016/j.heliyon.2024.e30956
- Li, H., Zhou, J., Wang, Q., Zhu, Y., Zi, T., Qin, X., et al. (2024a). Study on the predictive value of renal resistive index combined with  $\beta$ 2-microglobulin in patients with urosepsis complicated with acute kidney injury. *J. Inflammation Res.* 17, 9583–9599. doi: 10.2147/jir.S492858
- Mauk, M. G. (2018). Calling in the test: Smartphone-based urinary sepsis diagnostics. *EBioMedicine* 37, 11–12. doi: 10.1016/j.ebiom.2018.10.047
- Millán, J. L. (2006). Alkaline Phosphatases: Structure, substrate specificity and functional relatedness to other members of a large superfamily of enzymes. *Purinergic Signal* 2, 335–341. doi: 10.1007/s11302-005-5435-6
- Molnar, C. (2019). *Interpret the Machine Learning: A Guide for Making Black Box Models Explainable*. Available online at: <https://christophm.github.io/interpretable-ml-book/>.
- Molnar, C. (2020). *Interpretable machine learning* (Morrisville, NC: Lulu. com) (Accessed July 02, 2025).
- Moore, A., and Bell, M. (2022). XGBoost, A novel explainable AI technique, in the prediction of myocardial infarction: A UK biobank cohort study. *Clin. Med. Insights Cardiol.* 16, 11795468221133611. doi: 10.1177/11795468221133611
- Nachouki, M., Mohamed, E. A., Mehdi, R., and Abou Naaj, M. (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. *Trends Neurosci. Educ.* 33, 100214. doi: 10.1016/j.tine.2023.100214
- Neumann, A., Keller, T., Jocham, D., and Doehn, C. (2011). Human placental alkaline phosphatase (hPLAP) is the most frequently elevated serum marker in testicular cancer. *Aktuelle Urol* 42, 311–315. doi: 10.1055/s-0031-1271545
- Ohbe, H., Satoh, K., Totoki, T., Tanikawa, A., Shirasaki, K., Kuribayashi, Y., et al. (2024). Definitions, epidemiology, and outcomes of persistent/chronic critical illness: a scoping review for translation to clinical practice. *Crit. Care* 28, 435. doi: 10.1186/s13054-024-05215-4
- Oweira, H., Schmidt, J., Mehrabi, A., Kulaksiz, H., Schneider, P., Schöb, O., et al. (2018). Comparison of three prognostic models for predicting cancer-specific survival among patients with gastrointestinal stromal tumors. *Future Oncol.* 14, 379–389. doi: 10.2217/fon-2017-0450
- Picher, M., Burch, L. H., Hirsh, A. J., Spychala, J., and Boucher, R. C. (2003). Ecto 5'-nucleotidase and nonspecific alkaline phosphatase. Two AMP-hydrolyzing ectoenzymes with distinct roles in human airways. *J. Biol. Chem.* 278, 13468–13479. doi: 10.1074/jbc.M300569200
- Poelstra, K., Bakker, W. W., Klok, P. A., Kamps, J. A., Hardonk, M. J., and Meijer, D. K. (1997). Dephosphorylation of endotoxin by alkaline phosphatase *in vivo*. *Am. J. Pathol.* 151, 1163–1169.
- Porat, A., Bhutta, B. S., and Kesler, S. (2025). "Urosepsis," in *StatPearls* (StatPearls Publishing, Treasure Island (FL)). Available online at: <https://www.ncbi.nlm.nih.gov/books/NBK482344/>.
- Ross, P. D., Kress, B. C., Parson, R. E., Wasnich, R. D., Armour, K. A., and Mizrahi, I. A. (2000). Serum Bone alkaline phosphatase calcaneus Bone density predict fractures: prospective study. *Osteoporos Int.* 11, 76–82. doi: 10.1007/s001980050009
- Seely, K. A., Holthoff, J. H., Burns, S. T., Wang, Z., Thakali, K. M., Gokden, N., et al. (2011). Hemodynamic changes in the kidney in a pediatric rat model of sepsis-induced acute kidney injury. *Am. J. Physiol. Renal Physiol.* 301, F209–F217. doi: 10.1152/ajprenal.00687.2010
- Shi, G., Liu, G., Gao, Q., Zhang, S., Wang, Q., Wu, L., et al. (2023). A random forest algorithm-based prediction model for moderate to severe acute postoperative pain after

### SUPPLEMENTARY FIGURE 1

Raw data and Mice filled data density plot. Blue line = raw data, red line = interpolated data.

### SUPPLEMENTARY FIGURE 2

LASSO regression path showing the coefficients of variables across different values of the regularization parameter ( $\lambda$ ).

### SUPPLEMENTARY FIGURE 3

Cross-validation error plot for selecting the optimal  $\lambda$  in LASSO. The vertical dashed line represents the optimal  $\lambda$  where the minimal cross-validation error is achieved.

### SUPPLEMENTARY FIGURE 4

Comparison of the performance of the seven models in the validation cohort. Bayes, Bayesian Network; DT, Decision tree; LR, Logistic regression model; MLP, Multilayer perceptron; RF, Random Forest model; SVM, Support vector machine; XGBoost, eXtreme Gradient Boosting.

orthopedic surgery under general anesthesia. *BMC Anesthesiol* 23, 361. doi: 10.1186/s12871-023-02328-1

Siddique, A., and Kowdley, K. V. (2012). Approach to a patient with elevated serum alkaline phosphatase. *Clin. Liver Dis.* 16, 199–229. doi: 10.1016/j.cld.2012.03.012

Su, M., Guo, J., Chen, H., and Huang, J. (2023). Developing a machine learning prediction algorithm for early differentiation of urosepsis from urinary tract infection. *Clin. Chem. Lab. Med.* 61, 521–529. doi: 10.1515/cclm-2022-1006

Sun, J., Ge, X., Wang, Y., Niu, L., Tang, L., and Pan, S. (2022a). USF2 knockdown downregulates THBS1 to inhibit the TGF- $\beta$  signaling pathway and reduce pyroptosis in sepsis-induced acute kidney injury. *Pharmacol. Res.* 176, 105962. doi: 10.1016/j.phrs.2021.105962

Sun, Y., Zhang, Q., Yang, Q., Yao, M., Xu, F., and Chen, W. (2022b). Screening of gene expression markers for corona virus disease 2019 through boruta\_MCFS feature selection. *Front. Public Health* 10. doi: 10.3389/fpubh.2022.901602

Tóth-Heyn, P., Drukker, A., and Guignard, J. P. (2000). The stressed neonatal kidney: from pathophysiology to clinical management of neonatal vasomotor nephropathy. *Pediatr. Nephrol.* 14, 227–239. doi: 10.1007/s004670050048

Venkatesh, K. K., Jelovsek, J. E., Hoffman, M., Beckham, A. J., Bitar, G., Friedman, A. M., et al. (2023). Postpartum readmission for hypertension and pre-eclampsia: development and validation of a predictive model. *Bjog* 130, 1531–1540. doi: 10.1111/1471-0528.17572

Verweij, W. R., Bentala, H., Huizinga-van der Vlag, A., Miek van Loenen-Weemaes, A., Kooi, K., Meijer, D. K., et al. (2004). Protection against an *Escherichia coli*-induced sepsis by alkaline phosphatase in mice. *Shock* 22, 174–179. doi: 10.1097/01.shk.0000132485.05049.8a

Villanueva-Congote, J., Hinojosa-Gonzalez, D., Segall, M., and Eisner, B. H. (2024). The relationship between neutrophil/lymphocyte ratio, platelet/neutrophil ratio, and risk of urosepsis in patients who present with ureteral stones and suspected urinary tract infection. *World J. Urol* 42, 596. doi: 10.1007/s00345-024-05229-1

Wang, L., Li, D., He, W., Shi, G., Zhai, J., Cen, Z., et al. (2024a). Development and validation of a predictive model for post-percutaneous nephrolithotomy urinary sepsis: a multicenter retrospective study. *Minerva Urol Nephrol.* 76, 357–366. doi: 10.23736/s2724-6051.23.05396-x

Wang, J., Xu, Y., Liu, L., Wu, W., Shen, C., Huang, H., et al. (2023). Comparison of LASSO and random forest models for predicting the risk of premature coronary artery disease. *BMC Med. Inform Decis Mak* 23, 297. doi: 10.1186/s12911-023-02407-w

Wang, T., Yin, H., Shen, G., Cao, Y., Qin, X., Xu, Q., et al. (2024b). Effects of acetaminophen use on mortality of patients with acute respiratory distress syndrome: secondary data mining based on the MIMIC-IV database. *BMC Pulm Med.* 24, 568. doi: 10.1186/s12890-024-03379-x

Yamamoto, R., Yamakawa, K., Yoshizawa, J., Kaito, D., Umemura, Y., Homma, K., et al. (2025). Urine output and development of acute kidney injury in sepsis: A multicenter observational study. *J. Intensive Care Med.* 40, 191–199. doi: 10.1177/08850666241268390

Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., et al. (2020). Brief introduction of medical database and data mining technology in big data era. *J. Evid Based Med.* 13, 57–69. doi: 10.1111/jebm.12373

Yue, S., Li, S., Huang, X., Liu, J., Hou, X., Zhao, Y., et al. (2022). Machine learning for the prediction of acute kidney injury in patients with sepsis. *J. Transl. Med.* 20, 215. doi: 10.1186/s12967-022-03364-0

Zhang, L., Xu, F., Han, D., Huang, T., Li, S., Yin, H., et al. (2021). Influence of the trajectory of the urine output for 24 h on the occurrence of AKI in patients with sepsis in intensive care unit. *J. Transl. Med.* 19, 518. doi: 10.1186/s12967-021-03190-w

Zhou, H., Xin, Y., and Li, S. (2023). A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinf.* 24, 224. doi: 10.1186/s12859-023-05300-5