Interpretable Machine Learning for Mortality Prediction in S-AKI Patients Undergoing Hemodialysis

Kunyuan Xu [1, *], Yun Huang [2]

[1] Department of Endocrinology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China.

[2] School of Clinical Medicine, Jiangxi University of Traditional Chinese Medicine, Jiangxi, China.

* Corresponding Author Email: xky3371@163.com

Abstract. This study developed a machine learning model to predict in-hospital mortality risk among ICU patients with sepsis-associated acute kidney injury (S-AKI) undergoing hemodialysis. A retrospective analysis of 1,467 patients from the MIMIC-IV database and 226 from the eICU-CRD database was conducted, with models evaluated internally and externally. The RF model achieved excellent performance, with AUROCs of 0.798 (95% CI: 0.754 – 0.843) and 0.790 (95% CI: 0.723 – 0.857) in internal and external validations, respectively. Decision curve analysis indicated a net benefit of ~0.2 at a 10% mortality threshold, demonstrating good clinical applicability. SHAP analysis identified prothrombin time, APS III, systolic blood pressure, mean arterial pressure, and respiration rate as key predictors, with increased mortality risk associated with prothrombin time >10s, APS III >80, Nibp_systolic <110 mmHg, and Nibp_mean <70 mmHg. This model offers potential for supporting prognosis management and individualized treatment in clinical practice.

Keywords: Sepsis-associated acute kidney injury; hemodialysis; mortality; random forest; Shapley additive explanations.

1. Introduction

Sepsis-associated acute kidney injury (S-AKI), affecting up to 70% of ICU patients [1, 2], has a high in-hospital mortality risk, reaching 50% even after renal replacement therapy (RRT) [3]. While hemodialysis is commonly used, its timing, dosage, and outcomes remain debated, emphasizing the need for accurate mortality prediction [4-6]. To date, no effective model exists to predict S-AKI mortality post-hemodialysis. This study developed an interpretable machine learning model to predict in-hospital mortality and identify key prognostic factors, providing insights to improve clinical management.

2. Methods

This study developed a machine learning model using de-identified data from MIMIC-IV[7] 2.2(n=1,467) for training and eICU-CRD 2.0 [8] (n=226) for external validation. Access to the MIMIC-IV database was granted under approval from MIT's IRB (approval number 62263389). Inclusion criteria: Age ≥18, ICU stay ≥48 hours, diagnosis of S-AKI (per Sepsis-3 and KDIGO guidelines), and complete hemodialysis data. Exclusion criteria: Age <18, missing >20% of critical variables, ESRD with long-term dialysis before ICU, and ICU stays <24 hours. Six algorithms (XGBoost, LightGBM, CART, SVM, RF, and LR) were tested on a 70% training, 30% testing split. Performance metrics included AUC, calibration curves, and DCA. SHAP and partial dependence plots visualized feature importance. The data analysis was conducted using Python (version 3.9.7) and R (version 4.2.2). A significance criterion of $P < 0.05$ is established.

3. Results

3.1. Baseline Characteristics

In-hospital mortality rates were 32.9% in the training cohort (483/1467) and 35.3% in the validation cohort (59/167). Non-survivors in the MIMIC-IV cohort were older and had higher APACHE III and SAPS II scores, heart and respiratory rates, but lower systolic blood pressure, diastolic blood pressure, MAP, and temperature compared to survivors (all P < 0.001). They also showed higher glucose levels and prolonged PT and PTT (P < 0.001). The eICU-CRD cohort exhibited similar patterns in APACHE III, SAPS II, heart rate, blood pressure, MAP, respiratory rate, and PT (P < 0.05). Differences were observed in age, temperature, SpO2, glucose, hemoglobin, and PTT between the two cohorts (P > 0.05) (Tables 1).

Table 1. The comparison of baseline demographics and clinical characteristics between survivors and non-survivors in the MIMIC-IV database and eICU-CRD database.

Variables	Mimic database		*P* value	eICU-CRD database		*P* value
	Survivors (n=984)	Non-survivors (n=483)		Survivors (n=167)	Nonsurvivors (n=59)	
Age (years)	63.0 (53.0, 73.0)	67.0 (57.0, 76.0)	<0.001	64.0 (56.0, 72.0)	69.0 (56.5, 76.0)	0.028
Gender, n(%)			0.951			0.046
Male	596 (60.6)	291 (60.2)		61 (36.5)	31 (52.5)	
Female	388 (39.4)	192 (39.8)		106 (63.5)	28 (47.5)	
AKI stage, n(%)			0.932			0.299
Stage 1	243 (24.7)	120 (24.8)		46 (27.5)	13 (22.0)	
Stage 2	259 (26.3)	131 (27.1)		8 (4.8)	6 (10.2)	
Stage 3	482 (49.0)	232 (48.0)		113 (67.7)	40 (67.8)	
Weight (kg)	83.5 (70.7, 100.0)	82.1 (68.9, 97.2)	0.122	82.0 (68.0, 99.0)	83.0 (67.0, 93.0)	0.657
Length of Stay (days)	10.1 (4.9, 18.2)	11.2 (5.9, 18.0)	0.067	4.3 (2.20, 8.52)	8.9 (4.13, 14.9)	0.001
APS III	75.0 (66.0, 92.0)	94.0 (74.0, 112.0)	<0.001	71.0 (58.5, 86.5)	87.0 (70.5, 112.0)	<0.001
SAPS II	47.0 (39.0, 57.0)	53.0 (43.8, 62.0)	<0.001	59.0 (46.5, 70.0)	74.0 (56.5, 95.5)	<0.001
Heart rate (bpm)	86.0 (76.0, 96.0)	91.0 (81.0-100)	<0.001	81.0 (72.0, 88.0)	85.0 (76.0, 112.0)	0.006
Respiratory rate (breaths/min)	20.0 (17.0, 22.0)	22.0 (18.0-25.0)		18.0 (17.0, 21.5)	22.0 (20.0, 26.0)	0.001
Nibp_systolic (mmHg)	119.0 (108.0, 133.0)	106.0 (96.0, 119.0)	<0.001	113.0 (103.0, 132.0)	107.0 (96.0, 117.0)	0.002
Nibp diastolic (mmHg)	61.0 (53.0, 69.0)	55.0 (48.0, 63.5)	<0.001	59.0 (54.0, 66.0)	56.0 (47.0, 66.0)	0.017
Nibp_mean (mmHg)	75.0 (68.0, 84.0)	68.0 (61.0, 75.8)	<0.001	75.0 (69.0, 84.0)	69.0 (63.5, 77.0)	0.001
SpO2 (%)	97.5 (96.0, 99.0)	98.0 (96.0, 99.0)	0.798	96.0 (95.0, 100.0)	95.0 (94.0, 97.0)	0.005
Temperature (°C)	36.9 (36.7, 37.1)	36.7 (36.4, 37.1)	<0.001	36.7 (36.5, 36.9)	36.8 (36.5, 37.3)	0.061
Hemoglobin (g/dL)	8.70 (8.00, 9.7)	8.7 (7.90, 9.6)	0.289	10.3 (9.30, 11.4)	10.5 (9.55, 11.6)	0.494
Glucose (mg/dL)	124.0 (105.0, 151.0)	133.0 (113.0, 165.0)	<0.001	168 (134, 229)	179 (150, 276)	0.061
Prothrombin (s)	14.1 (12.7, 16.6)	17.5 (14.3, 23.4)	<0.001	16.4 (14.0, 22.8)	19.2 (15.8, 25.5)	0.006
PTT (s)	33.7 (28.7, 49.4)	44.4 (33.2, 59.8)	<0.001	36.0 (31.6, 46.0)	39.0 (33.0, 45.0)	0.473

Abbreviation: APS III: Acute Physiology Score III; SAPS II: Simplified Acute Physiology Score II; AKI: acute kidney injury; Nibp: noninvasive blood pressure; SpO2: O2 saturation; PTT: partial thromboplastin time.

3.2. Model Comparison

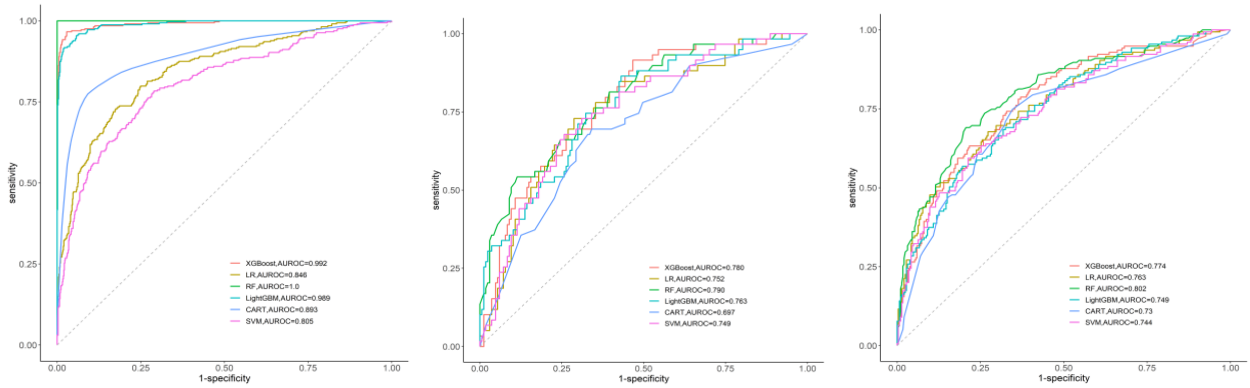
Table 2 summarizes the performance of six machine learning models in predicting mortality for hemodialysis patients with sepsis-associated AKI. RF showed the best performance in the training cohort (AUROC, sensitivity, specificity, PPV, NPV = 1.00), indicating overfitting. In internal validation, RF achieved an AUROC of 0.798 (95% CI: 0.754–0.843) but had lower sensitivity (0.690 vs 0.781) and NPV (0.825 vs 0.843) compared to XGBoost, and slightly lower specificity than LightGBM (0.794 vs 0.804). In external validation, RF had the highest AUROC (0.790, 95% CI: 0.723–0.857) and specificity (0.886) but lower sensitivity (0.542 vs 0.915) and NPV (0.846 vs 0.947) compared to XGBoost.

Table 2. Performance of each model for prediction.

Models	AUROC[b] (95%CI)	Cutoff	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV[d] (95%CI)
Training cohort						
XGBoost	0.992(0.987-0.997)	0.412	0.966(0.947-0.986)	0.971(0.959-0.984)	0.941(0.915-0.966)	0.984(0.975-0.993)
LightGBM	0.989(0.984-0.994)	0.338	0.954(0.932-0.977)	0.943(0.925-0.960)	0.887(0.854-0.92)	0.978(0.967-0.989)
CART	0.893(0.870-0.916)	0.351	0.774(0.729-0.820)	0.908(0.887-0.930)	0.799(0.755-0.843)	0.895(0.873-0.918)
SVM	0.805(0.776-0.835)	0.276	0.784(0.739-0.828)	0.711(0.677-0.744)	0.560(0.514-0.605)	0.875(0.848-0.902)
RF	1.000(1.000-1.000)	0.467	1.000(1.000-1.000)	1.000(1.000-1.000)	1.000(1.000-1.000)	1.000(1.000-1.000)
LR	0.846(0.820-0.872)	0.356	0.729(0.681-0.777)	0.825(0.797-0.853)	0.662(0.613-0.711)	0.866(0.840-0.892)
Internal validation cohort						
XGBoost	0.774(0.728-0.820)	0.362	0.781(0.715-0.846)	0.636(0.581-0.692)	0.538(0.473-0.603)	0.843(0.794-0.891)
LightGBM	0.749(0.702-0.796)	0.422	0.568(0.490-0.646)	0.804(0.758-0.850)	0.611(0.531-0.691)	0.774(0.727-0.822)
CART	0.730(0.680-0.780)	0.139	0.748(0.680-0.817)	0.654(0.599-0.709)	0.540(0.473-0.606)	0.827(0.778-0.877)
SVM	0.744(0.695-0.792)	0.361	0.632(0.556-0.708)	0.748(0.698-0.799)	0.576(0.502-0.651)	0.790(0.741-0.838)
RF	0.798(0.754-0.843)	0.402	0.690(0.618-0.763)	0.794(0.747-0.841)	0.645(0.572-0.717)	0.825(0.781-0.870)
LR	0.763(0.716-0.810)	0.319	0.677(0.604-0.751)	0.727(0.676-0.779)	0.574(0.502-0.645)	0.806(0.758-0.854)
External validation cohort						
XGBoost	0.780(0.714-0.846)	0.371	0.915(0.844-0.986)	0.533(0.457-0.609)	0.409(0.325-0.493)	0.947(0.901-0.992)
LightGBM	0.763(0.693-0.833)	0.158	0.864(0.777-0.952)	0.569(0.494-0.644)	0.415(0.328-0.502)	0.922(0.871-0.974)
CART	0.697(0.620-0.775)	0.458	0.695(0.577-0.812)	0.659(0.587-0.731)	0.418(0.321-0.516)	0.859(0.799-0.920)
SVM	0.749(0.678-0.820)	0.480	0.678(0.559-0.797)	0.743(0.676-0.809)	0.482(0.374-0.589)	0.867(0.811-0.923)
RF	0.790(0.723-0.857)	0.607	0.542(0.415-0.669)	0.886(0.838-0.934)	0.627(0.495-0.760)	0.846(0.792-0.899)
LR	0.752(0.680-0.824)	0.425	0.729(0.615-0.842)	0.713(0.644-0.781)	0.473(0.370-0.575)	0.881(0.827-0.936)

Abbreviation: AUROC: area under the receiver operating characteristic curve; PPV: positive predictive value; NPV: negative predictive value; XGBoost: extreme gradient boosting; LightGBM: light gradient boosting machine; CART: classification and regression tree; SVM: support vector machine; RF: random forest; LR: logistic regression.

The ROC curve (Figures 1A–C) shows RF performed optimally in both training and validation.



A B C

Figure 1. (A) ROC curves in training set; (B) ROC curves in external validation set; (C) ROC curves in internal validation set. Abbreviation: ROC, Receiver operating characteristic.

For decision curve analysis (DCA), three top-performing algorithms (RF, LightGBM, and XGBoost) were evaluated. At a 10% mortality prediction threshold, all models achieved a net benefit of ~0.2, indicating good clinical applicability. As thresholds increased, LightGBM showed slightly higher net benefits, but RF and XGBoost outperformed LightGBM in the 5–25% range, reducing ineffective interventions (Figure 2A). Calibration curves (Figure 2B) showed LightGBM and RF had minimal deviation from the ideal line in the 0.1–0.3 threshold range. Supplementary Figures 2C–D for the internal validation cohort confirmed RF maintained high net benefits and strong calibration.

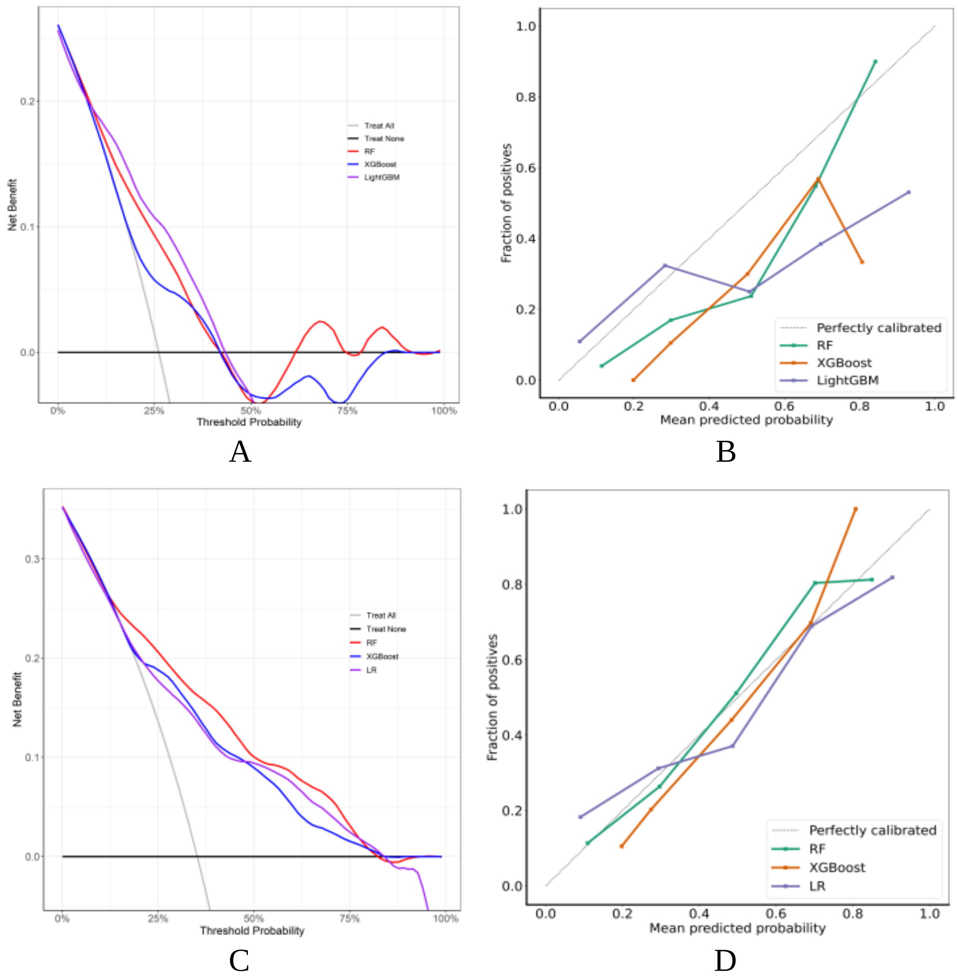
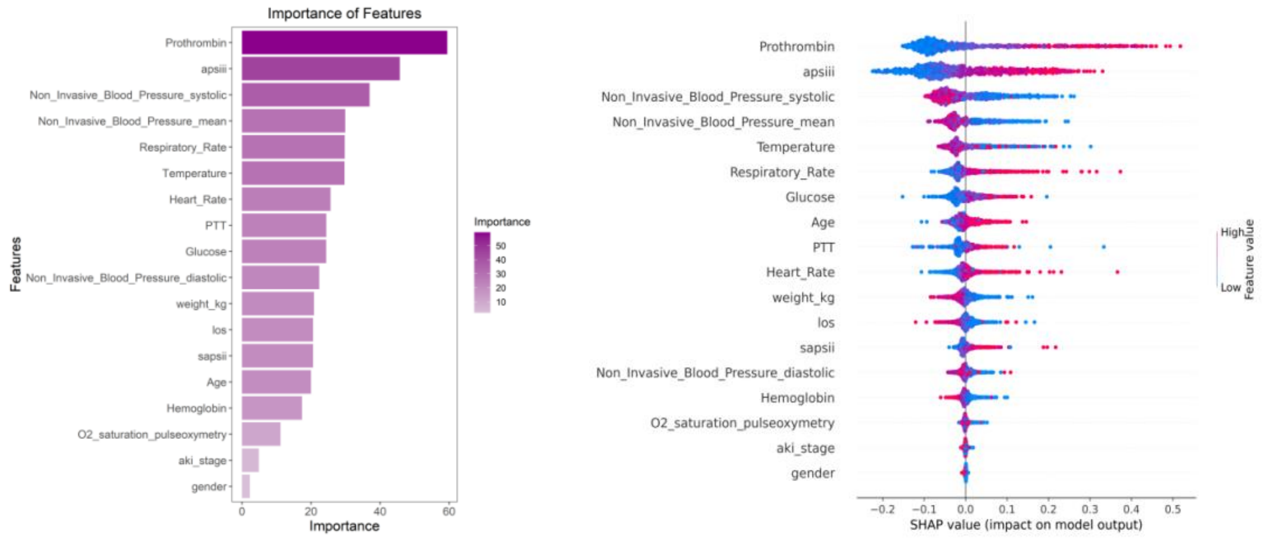


Figure 2. (A) Calibration curve of external validation set;
(B) DAC of external validation set; (C) Calibration curve of internal validation set;
(D) DCA of internal validation set. Abbreviation: DAC, Decision curve analysis curves.

3.3. Model Interpretability

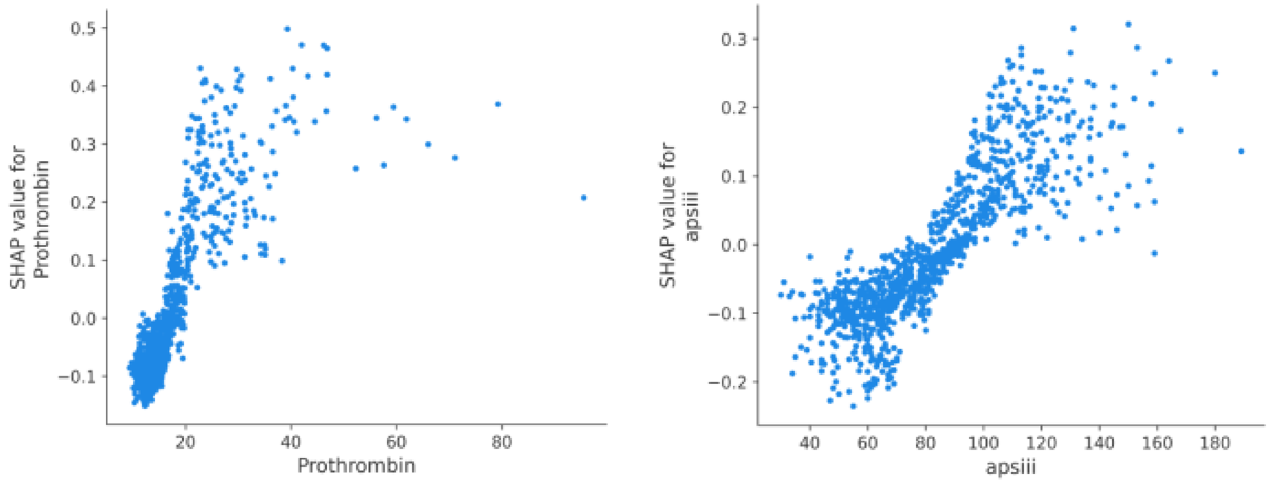
SHAP analysis of the training set identified the most critical predictors for the RF model: PT, APS III, Nibp_systolic, Nibp_mean, respiratory rate, temperature, heart rate, PTT, blood glucose, and Nibp_diastolic (Figure 4A). These variables significantly influenced mortality predictions. SHAP values revealed the relationships between predictors and mortality (Figure 4B). Higher PT, APS III, respiratory rate, blood glucose, age, PTT, and heart rate increased mortality risk, while higher non-invasive systolic blood pressure, Nibp_mean, and temperature reduced it, promoting survival predictions.



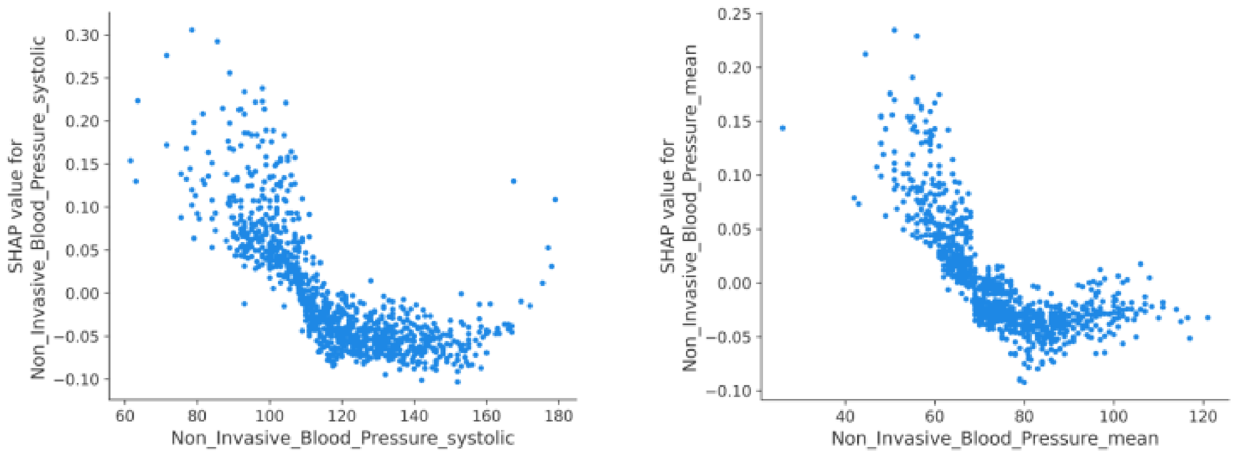
A B

Figure 3. Statistical plots of the SHAP analysis.
(A) Order plot of variable importance for SHAP analysis;
(B) statistical graph of variable contribution in SHAP analysis.
The color indicates high (red) or low (blue) levels of that feature value in the samples. Abbreviation: SHAP, SHapley Additive exPlanations.

SHAP dependence plots were created for the four most significant continuous variables (Figure 5A-5D). The risk of death significantly increased when PT > 10 s, APS III scores exceeded 80, systolic blood pressure fell below 110 mmHg, and Nibp_mean < 70 mmHg.



A B



C D

Figure 4. SHAP dependence plots of important features in RF model.
(A) Prothrombin time; (B) APS III; (C) Nibp_systolic; (D) Nibp_mean.

4. Summary

This study developed a machine learning model to predict in-hospital mortality in S-AKI patients receiving dialysis, with the RF model demonstrating strong performance and generalization. Key predictors included PT, APACHE III score and non-invasive blood pressure. Future research should focus on refining models with larger populations, incorporating additional variables, and validating in real-world clinical settings through prospective studies.

References

[1] Zarbock A, Koyner JL, Gomez H, Pickkers P, Forni L. Sepsis-associated acute kidney injury-treatment standard. Nephrol Dial Transplant. 2023; 39(1):26-35.

[2] Kounatidis D, Vallianou NG, Psallida S, Panagopoulos F, Margellou E, Tsilingiris D, et al. Sepsis-Associated Acute Kidney Injury: Where Are We Now? Medicina (Kaunas). 2024; 60(3).

[3] Bagshaw SM, Uchino S, Bellomo R, Morimatsu H, Morgera S, Schetz M, et al. Septic acute kidney injury in critically ill patients: clinical characteristics and outcomes. Clin J Am Soc Nephrol. 2007; 2(3):431-9.

[4] Wang Q, Liu F, Tao W, Qian K. Timing of renal replacement therapy in patients with sepsis-associated acute kidney injury: A systematic review and meta-analysis. Aust Crit Care. 2024; 37(2):369-79.

[5] Bagshaw SM, Berthiaume LR, Delaney A, Bellomo R. Continuous versus intermittent renal replacement therapy for critically ill patients with acute kidney injury: a meta-analysis. Crit Care Med. 2008; 36(2):610-7.

[6] Wang Y, Gallagher M, Li Q, Lo S, Cass A, Finfer S, et al. Renal replacement therapy intensity for acute kidney injury and recovery to dialysis independence: a systematic review and individual patient data meta-analysis. Nephrol Dial Transplant. 2018; 33(6):1017-24.

[7] Alistair Johnson LB, Tom Pollard, Steven Horng, Leo Anthony Celi, Roger Mark. MIMIC-IV (version 2.2). PhysioNet. 2023.

[8] Pollard TJ JA, Raffa JD, Celi LA, Mark RG, and Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research 2018 [Available from: https://eicu-crd.mit.edu/gettingstarted/access/.