The Heart Surgery Forum

Heart Surg. Forum **2025**; 28(12): 49196
https://doi.org/10.31083/HSF49196

Article

Interpretable Machine Learning Analysis of Preoperative NT-proBNP and Creatinine for Predicting Acute Kidney Injury After Cardiac Valve Surgery

Huan Fu[1], Ying Tian[1], Shuigen Song[2], Fengping Huang[2], Dingde Long[1], Yang Dong[1], Tianyuan Li[1,*]

[1]Department of Anesthesiology, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 330006 Nanchang, Jiangxi, China
[2]Department of Anesthesiology, People's Hospital of Jinggangshan, 343604 Jinggangshan, Jiangxi, China
*Correspondence: berlinli456@163.com (Tianyuan Li)
Academic Editor: Ferdinand Aurel Vogt
Submitted: 26 August 2025 Revised: 31 October 2025 Accepted: 12 November 2025 Published: 17 December 2025

Abstract

Background: Acute kidney injury (AKI) is a common and serious complication of cardiac valve surgery, and is associated with high mortality and healthcare costs. Existing prediction models for AKI are often incapable of capturing complex biomarker interactions. This study aimed to build an interpretable machine learning (ML) model that incorpates preoperative N-terminal pro-B-type natriuretic peptide (NT-proBNP) and serum creatinine (SCr) levels to predict of AKI risk in valve surgery patients. **Methods**: Consecutive adults who underwent isolated valve surgery with cardiopulmonary bypass (CPB) in the first affiliated hospital of Nanchang University from October 2016 to October 2021 were included in this retrospective cohort study. Patients who had preoperative dialysis or were having an emergency surgery were excluded as well as those with missing NT-proBNP/SCr data. The main outcome was any stage AKI within 7 days after surgery (Kidney Disease: Improving Global Outcomes, KDIGO criteria). Utilizing preoperative variables, five ML models Logistic regression, support vector machine (SVM), Random Forest (RF), extreme gradient boosting (XGBoost), and K-nearest neighbors (KNN) were developed after handling class imbalance synthetic minority oversampling technique (SMOTE). Key predictors were identified through feature selection techniques. Evaluation of model performance was done at area under the curve (AUC), sensitivity, specificity and decision curve analysis (DCA). SHapley Additive exPlanations (SHAP) values provided interpretability. **Results**: Among 333 patients eligible for inclusion, 106 experienced AKI (31.8%). Seven predictors were consistently selected: age, NT-proBNP, SCr, CPB duration, aortic cross-clamp (ACC) duration, hemoglobin and albumin. Overall, the RandomForest model outperformed the other models, with AUC of 0.872, accuracy of 0.835, sensitivity of 0.718, specificity of 0.923 and F1-score of 0.789 in the testing cohort (n = 91). DCA demonstrated excellent calibration and the highest net benefit with this model. SHAP analysis identified NT-proBNP, SCr, and duration of ACC as the three leading risk factors with clear, personalized risk evaluation. **Conclusions**: This novel, interpretable ML model leverages preoperative NT-proBNP and SCr to accurately predict AKI after cardiac valve surgery. It demonstrated promising predictive performance in internal validation, with the potential to surpass traditional models and have future potential for clinical application. Prospective trials are needed to assess whether model-guided interventions can truly reduce AKI incidence.

Keywords: NT-proBNP; creatinine; machine learning; AKI

1. Introduction

Acute kidney injury (AKI) remains a devastating complication following cardiac valve surgery, occurring in 37.6% of isolated valve patients [1–3]. AKI is associated with a 3–8 fold increase in mortality, prolonged intensive care and hospital stays, increased risk of chronic kidney disease (CKD), and significantly increased healthcare costs [4–8]. Despite advances in perioperative management, effective therapeutic interventions for established AKI remain limited. Consequently, the early identification of high-risk patients is paramount for implementing targeted preventative strategies.

Current AKI risk prediction models such as the Cleveland Clinic score, Simplified Renal Index score, and Mehta score rely predominantly on preoperative clinical variables (e.g., age, baseline renal function, diabetes, heart failure) and intraoperative factors (e.g., cardiopulmonary bypass (CPB) duration, transfusion requirements) [9–11]. These models, are often built using traditional logistic regression, and while useful, they have significant limitations. First, they do not always adequately capture complex, non-linear interactions between risk factors [12,13]. Second, solitary biomarkers lack sensitivity for predicting AKI. Many lack robust integration of potent preoperative biomarkers that may reflect underlying pathophysiological processes directly relevant to AKI susceptibility [14]. Although still underexploited, two biomarkers stand out as particularly promising yet underexploited in predictive modeling for valve surgery patients: N-terminal pro-B-type natriuretic peptide (NT-proBNP) and serum creatinine (SCr).

NT-proBNP is a well-established marker of ventricular wall stress and myocardial dysfunction. It is strongly associated with postoperative cardiovascular complications and mortality [15–18]. Rrecent large-scale studies have demonstrated independent predictive value for NT-proBNP in cardiac surgery-associated AKI (CSA-AKI) across all severities, but particularly in severe AKI and dialysis [19–25]. A study of over 35,000 patients by Wang *et al.* [1] found that preoperative NT-proBNP significantly enhanced the prediction of AKI beyond conventional risk factors, with remarkable net reclassification improvements (NRI: 0.24–0.47). The association was especially pronounced in valve surgery patients.

Beyond its diagnostic role, SCr exhibits dynamic interplay with NT-proBNP: NT-proBNP is renally cleared, rendering its levels creatinine-dependent. Conversely, AKI exacerbates NT-proBNP accumulation through fluid overload and cardiorenal crosstalk. Notably, preoperative NT-proBNP-SCr synergy significantly improved the prediction of AKI prediction in non-cardiac surgery cases (AUC: 0.74, NRI: 17%) [26]. However, the relevance of preoperative NT-proBNP to valve surgery—where CPB intensifies hemolysis-mediated nephrotoxicity and inflammation—remains unknown.

Machine learning (ML) offers a paradigm shift for the prediction of AKI. Unlike logistic regression, ML algorithms (e.g., Random Forest (RF), extreme gradient boosting (XGBoost)) excel at identifying complex, non-linear relationships and interactions within high-dimensional data without rigid a priori assumptions [27,28]. Recent studies have validated the superiority of ML models over regression in CSA-AKI prediction (AUCs: 0.83–0.86) [1], yet their potential to leverage NT-proBNP-SCr synergy specifically in valve surgery has yet to be harnessed.

Therefore, this study aimed to develop and validate the first ML model integrating preoperative NT-proBNP, SCr, and valve surgery-specific covariates (e.g., CPB time, transfusion volume) to accurately identify patients at risk of developing any-stage AKI following cardiac valve surgery. We hypothesize this model will leverage on the complementary pathophysiological insights of these biomarkers, as well as the capacity of ML to uncover complex, non-linear predictive patterns.

2. Materials & Methods

2.1 Study Population

This retrospective cohort study included consecutive adult patients (≥18 years) undergoing isolated valve surgery (aortic, mitral, tricuspid repair/replacement) with CPB between October 2016 and October 2021 at the first affiliated hospital of Nanchang University. Exclusion criteria were: (1) preoperative dialysis or CKD Stage 4–5 (eGFR <30 mL/min/1.73 m^2); (2) emergent/salvage surgery; (3) concomitant non-valve cardiac procedures (e.g., coronary artery bypass grafting (CABG), aortic root replacement); (4) missing preoperative NT-proBNP or creatinine values; (5) re-do sternotomy within 30 days. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the first affiliated hospital of Nanchang University (Approval No: 2021-8-003), with waiver of informed consent due to the retrospective nature.

2.2 Data Collection

Data were collected on demographic characteristics, comorbidities, laboratory biomarkers, intraoperative information, and postoperative information. Plasma NT-proBNP concentration was routinely measured by our clinical laboratory upon hospital admission. SCr concentrations were determined upon hospital admission and daily whilst patients remained in critical care.

2.3 Data Preprocessing

The original dataset exhibited class imbalance, with fewer cases of postoperative AKI (AKI; n = 106) than cases without AKI (non-AKI; n = 227), resulting in biased towards the latter group. Once the final set of predictor variables was identified through feature selection on the original dataset (as detailed in the 'Feature screening' section), the class imbalance was addressed in the training set by employing the Synthetic Minority Oversampling Technique (SMOTE). SMOTE augments the minority class by synthetically generating new instances through interpolation between existing minority class samples, thereby balancing the dataset. The application of SMOTE expanded the original dataset in this study from 333 cases to 454 cases, with the postoperative AKI and non-AKI groups each comprising 227 cases. Median imputation was used for the few cases missing continuous variables, while mode imputation was used for categorical variables.

2.4 Assessment of AKI

The primary focus of the study was the occurrence of postoperative AKI at any stage. This was determined using the "Kidney Disease: Improving Global Outcomes" (KDIGO) guidelines, which rely on perioperative SCr measurements [29]. AKI was identified when the postoperative SCr level was more than 1.5-fold higher than the baseline value or when the SCr level increased by 0.3 mg/dL within 48 hours after surgery.

2.5 Estimation of Sample Size

Literature reviews indicate that the incidence of AKI following acute cardiac valve surgery is 37.6% [1]. For a sample size of at least 100 AKI events, this requires, the inclusion of at least 266 patients [30]. An autonomous algorithm, separate from the research team, was employed to segregate the sample into two groups. A training dataset comprising 80% of the participants, which was used to develop the model, while a testing dataset comprising the re-

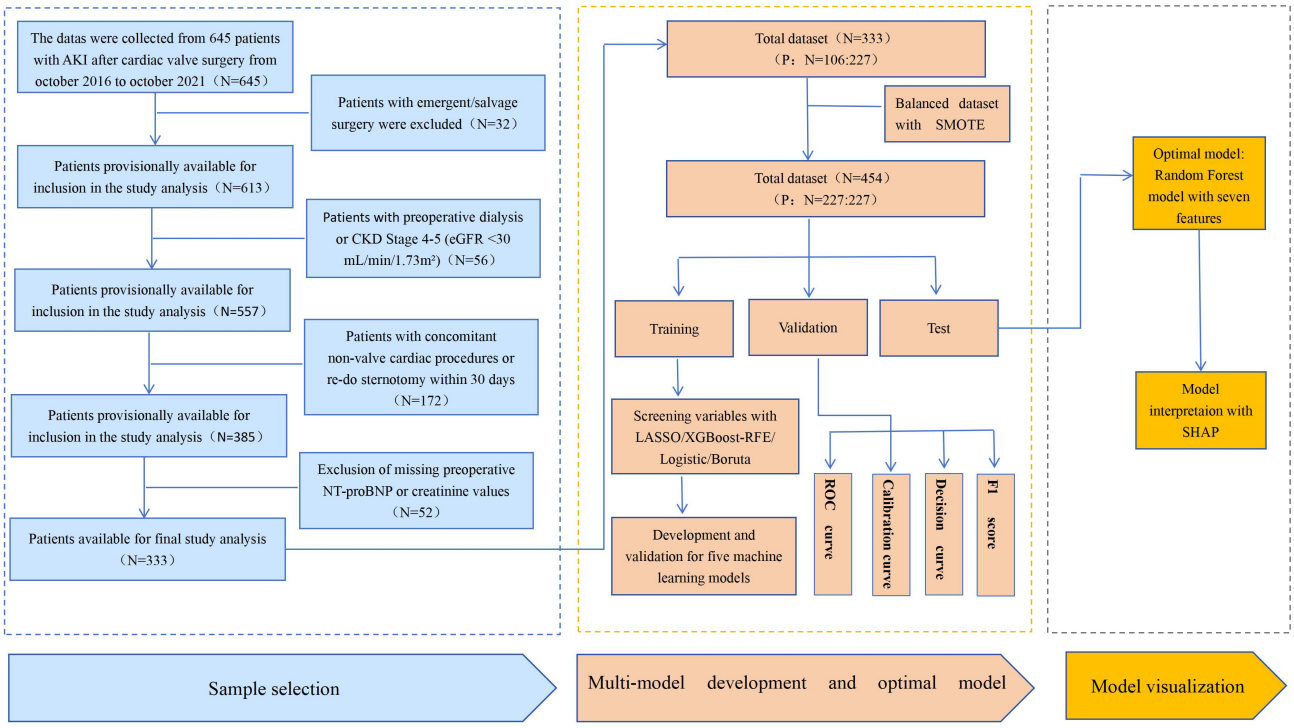

Fig. 1. Flow chart for model development and validation. AKI, Acute kidney injury; SMOTE, Synthetic Minority Oversampling Technique; LASSO, least absolute shrinkage and selection operator; SHAP, SHapley Additive exPlanations; XGBoost-RFE, extreme gradient boosting recursive feature elimination; CKD, chronic kidney disease; NT-proBNP, N-terminal pro-B-type natriuretic peptide; eGFR, Estimated Glomerular Filtration Rate; ROC, Receiver Operating Characteristic Curve.

maining 20% of the participants was used to evaluate the model's ability to predict outcomes. In order to satisfy these requirements, it was necessary to enrol a minimum of 333 patients and to randomly allocate the study population to training and testing datasets.

2.6 Feature Screening

This study focused on the prediction of AKI associated with cardiac surgery, characterized by AKI occurring within the first week after surgery. Drawing from prior research [9–11], the potential predictors for AKI examined in this analysis were: age, sex, body mass index (BMI), number of affected valves, New York Heart Association (NYHA) status, left ventricular ejection fraction (LVEF), duration of cardiopulmonary bypass (CPB) and aortic cross-clamping (ACC), smoking and alcohol use, diabetes, high blood pressure, arrhythmia, chronic obstructive pulmonary disease (COPD), history of stroke, white blood cells (WBC) count, red blood cells (RBC) count, concentration of hemoglobin (HGB), platelet count (PLT), glucose levels (Glu), aspartate aminotransferase (AST), alanine aminotransferase (ALT), SCr, blood urea nitrogen (BUN), albumin, need for blood transfusion, urine output, and NT-proBNP. To avoid overfitting in the predictive model, four techniques for selecting features—least absolute shrinkage and selection operator (LASSO), Boruta, logistic regression, and extreme gradient boosting recursive feature elimination (XGBoost-RFE) [31,32]—were employed to screen the variables and identify key predictors in the original, non-resampled dataset (n = 333).

2.7 Development of Machine Learning Models

To predict postoperative AKI, we developed and evaluated the performance of five distinct machine learning models: logistic regression (LR), support vector machine (SVM), RandomForest (RF), extreme gradient boosting (XGBoost), and K-nearest neighbors (KNN). The XGBoost model was implemented using the XGBoost package (https://xgboost.readthedocs.io). The remaining four models were built using the scikit-learn package (https://scikit-learn.org). Given the critical impact of hyperparameter tuning on model performance, the full dataset of 454 patients was randomly partitioned into an 80% training dataset and a 20% testing dataset. Each model was trained and evaluated using five-fold cross-validation. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), sensitivity (recall), specificity, positive predictive value (PPV, precision), negative predictive value (NPV), and F1-score.

2.8 Machine Learning Explainable Tool

SHapley Additive exPlanations (SHAP) was utilized to interpreting the prediction model, thus providing an integrated method to accurately assess how each feature con-

Table 1. Unadjusted comparison of AKI vs. non-AKI groups (original cohort).

Characteristic	Overall (N = 333)	AKI (N = 106)	Non-AKI (N = 227)	*p*-value
Demographics				
Age, years	58.0 [53.0–64.0]	60.0 [54.3–65.8]	56.0 [51.0–63.0]	<0.001
Female, n (%)	143 (42.9%)	52 (49.1%)	91 (40.1%)	0.155
Preoperative biomarkers				
NT-proBNP, pg/mL	1812 [1001–3640]	2633 [1546–4364]	1439 [764–3121]	<0.001
Serum creatinine, µmol/L	78.0 [60.0–116.0]	100.0 [79.3–152.8]	69.0 [59.0–98.0]	<0.001
Albumin, g/L	35.6 [31.5–39.8]	33.0 [30.3–36.9]	37.5 [32.7–40.8]	<0.001
Haemoglobin, g/L	119.2 ± 21.6	113.1 ± 18.9	122.0 ± 22.2	<0.001
Operative factors				
CPB duration, min	110 [105–119]	115 [109–125]	110 [103–116]	<0.001
ACC duration, min	67 [63–73]	72 [65–78]	65 [62–70]	<0.001

Data are presented as median [IQR], mean ± SD, or n (%). Only key predictors and selected significant variables are shown. ACC, aortic cross-clamp; CPB, cardiopulmonary bypass; NT-proBNP, N-terminal pro-B-type natriuretic peptide. A full comparison of all variables is available in **Supplementary Table 1**. *p*-values are unadjusted for multiple comparisons and should be interpreted as descriptive.

tributes to and impacts the ultimate predictions [33]. SHAP values indicate the extent to which each predictor affects the target variable, highlighting both positive and negative contributions. Additionally, the specific SHAP values for each data point allow for individual interpretation within the dataset.

2.9 Statistical Analysis

Data with a normal distribution are shown as mean ± standard deviation (SD), whereas data with a non-normal distribution are depicted via the median and interquartile range (IQR). Categorical variables were shown as proportions (%). Continuous variables were compared using independent or paired *t*-tests, ANOVA, or the Kruskal-Wallis test depending on their distribution. Categorical data was compared with the χ^2 test or Fisher's exact test. To identify the most influential predictors, feature selection was performed using the LASSO, Boruta, logistic regression, and XGboost-RFE methods. This led to the development of five predictive models. Model performance was assessed through discrimination and calibration metrics. Clinical decision curve analysis (DCA) was employed to assess the model's clinical applicability of each model [34]. The metrics used to evaluate model discrimination were the Area Under the Receiver Operating Characteristic curve (AUROC) and the Brier score were the metrics used to evaluate model discrimination. After determining the optimal model, the SHAP technique was applied to elucidate feature importance and interactions. All statistical analyses were performed using R software version 3.3.2 (The R Foundation for Statistical Computing, Vienna, Austria) (http://www.R-project.org), with a two-tailed p-value < 0.05 indicating statistical significance.

3. Results

3.1 Patient Characteristics

The study sample comprised a sample of 333 individuals who underwent cardiac surgery, of whom 106 (31.8%) experienced AKI following the procedure. Participants were categorized into two distinct groups: AKI and non-AKI, according to the development of AKI within the first postoperative week. The flowchart illustrating patient selection is shown in Fig. 1.

Key baseline characteristics are summarized in Table 1. Patients who developed AKI were significantly older and had higher preoperative levels of NT-proBNP and serum creatinine, lower levels of haemoglobin and albumin, and longer durations of cardiopulmonary bypass and aortic cross-clamping (all $p < 0.001$). No significant differences were found in gender distribution or other common comorbidities. A comprehensive comparison of all assessed variables is provided in **Supplementary Table 1**.

The SMOTE technique was employed to address data imbalance, leading to an expanded cohort of 454 patients (227 in each of the AKI and non-AKI groups). The above differences between the AKI and non-AKI groups were also present in the larger cohort (all $p < 0.001$). Furthermore, significant differences ($p < 0.05$) were also apparent in lifestyle factors like smoking, and drinking, health conditions such as hypertension, COPD, and stroke, as well as the necessity for blood transfusions ($p < 0.05$) as detailed in Table 2. A complete comparison of all variables is available in **Supplementary Table 2**.

The expanded dataset of 454 individuals was split randomly into 80% for training purposes and 20% for testing. No differences were observed between the training and testing datasets in regard to patients' characteristics (**Supplementary Table 3**), with the exception of RBC

Table 2. Key characteristics of the study cohort after SMOTE oversampling.

Variable	Overall (N = 454)	AKI (N = 227)	Non-AKI (N = 227)	*p*-value
Demographics				
Age, years	59.0 [53.0–64.0]	60.0 [55.0–65.0]	56.0 [51.0–63.0]	<0.001
Preoperative biomarkers				
NT-proBNP, pg/mL	1893 [1049–3744]	2663 [1552–4414]	1439 [764–3121]	<0.001
Serum creatinine, μmol/L	89.0 [63.0–122.0]	100.0 [80.5–139.0]	69.0 [59.0–98.0]	<0.001
Haemoglobin, g/L	117.0 [104.0–128.0]	116.0 [101.0–123.5]	119.0 [108.0–139.0]	<0.001
Albumin, g/L	34.2 [31.1–38.9]	32.9 [30.5–35.7]	37.5 [32.7–40.8]	<0.001
Operative factors				
CPB duration, min	112 [106–120]	115 [109–122]	110 [103–116]	<0.001
ACC duration, min	67 [63–74]	71 [65–77]	65 [62–70]	<0.001
Selected clinical features				
Hypertension, n (%)	112 (24.7%)	40 (17.6%)	72 (31.7%)	0.001
COPD, n (%)	31 (6.8%)	6 (2.6%)	25 (11.0%)	0.001
Stroke, n (%)				0.028
No	416 (91.630%)	215 (94.714%)	201 (88.546%)	
Yes	38 (8.370%)	12 (5.286%)	26 (11.454%)	
Blood transfusion, n (%)				0.007
No	415 (91.410%)	216 (95.154%)	199 (87.665%)	
Yes	39 (8.590%)	11 (4.846%)	28 (12.335%)	

Data are presented as median [IQR] or n (%). *p*-values from Mann-Whitney U or Chi-squared test. ACC, aortic cross-clamp; COPD, chronic obstructive pulmonary disease; CPB, cardiopulmonary bypass; NT-proBNP, N-terminal pro-B-type natriuretic peptide. A full comparison of all variables is available in **Supplementary Table 2**.

count ($p = 0.02$) and history of stroke ($p = 0.03$). However, the magnitude of this imbalance was minor.

3.2 Screening of Predictors

This research leveraged the Boruta algorithm, an enhancement of the Random Forest methodology, to effectively identify the true set of features by assessing the significance of each. The Boruta algorithm revealed 18 crucial elements in the original, non-resampled dataset (n = 333), which include age, BMI, the number of valves affected, LVEF, NT-proBNP levels, SCr, BUN, CPB duration, ACC duration, COPD, WBC count, RBC count, hemoglobin levels, glucose levels, AST, ALT, albumin levels, and the requirement for blood transfusion. LASSO regression identified the following factors in the original, non-resampled dataset such as age, NT-proBNP level, SCr, CPB duration, ACC duration, hemoglobin, AST, ALT, albumin level, and urine output. The logistic regression method detected 13 significant characteristics in the original, non-resampled dataset, encompassing age, NT-proBNP level, SCr, BUN, CPB duration, ACC duration, glucose level, hemoglobin, valve involvement, COPD, albumin levels, NYHA classification, and blood transfusion necessity. The XGboost-RFE algorithm identified 21 vital attributes were singled out in the original, non-resampled dataset, including age, BMI, valve involvement count, NYHA classification, LVEF, NT-proBNP levels, SCr, BUN, CPB duration, ACC duration, hypertension, PLT, COPD, WBC count, RBC count, hemoglobin, glucose level, AST, albumin level, urine output, and need for blood transfusion. Comparison of results from the LASSO and logistic regressions, XGboost-RFE, and Boruta algorithm revealed a shared subset of feature variables that were common to all four approaches. The final predictor variables were determined by selecting those consistently identified across all four methods: age, NT-proBNP, SCr, CPB duration, ACC duration, HGB, and albumin (Fig. 2).

3.3 Model Performance

Upon identifying these 7 variables, machine learning algorithms were deployed to predict the onset of AKI following surgery. The efficacy of these predictive models was examined using AUC, Brier Scores, and DCA as indicators. The Random Forest model showed a notably enhanced and lower Brier score relative to the other models. Fig. 3 presents the calibration charts for the five models, DCA revealed the Random Forest model was the most effective diagnostic tool for AKI (Fig. 4).

The Random Forest and XGboost models achieved outstanding predictive accuracy on the training dataset with both having an AUC of 1.00. For the LR model, the AUC value was 0.837 (95% CI: 0.793–0.881), for SVM it was 0.910 (95% CI: 0.876–0.944), and for KNN it was 0.928 (95% CI : 0.901–0.955) (refer to Fig. 5A and **Supplementary Table 4**).

The F1 scores were reported as 0.767 for LR, 0.999 for RF, 1.000 for XGboost, 0.844 for SVM, and 0.855 for KNN (**Supplementary Table 4**).

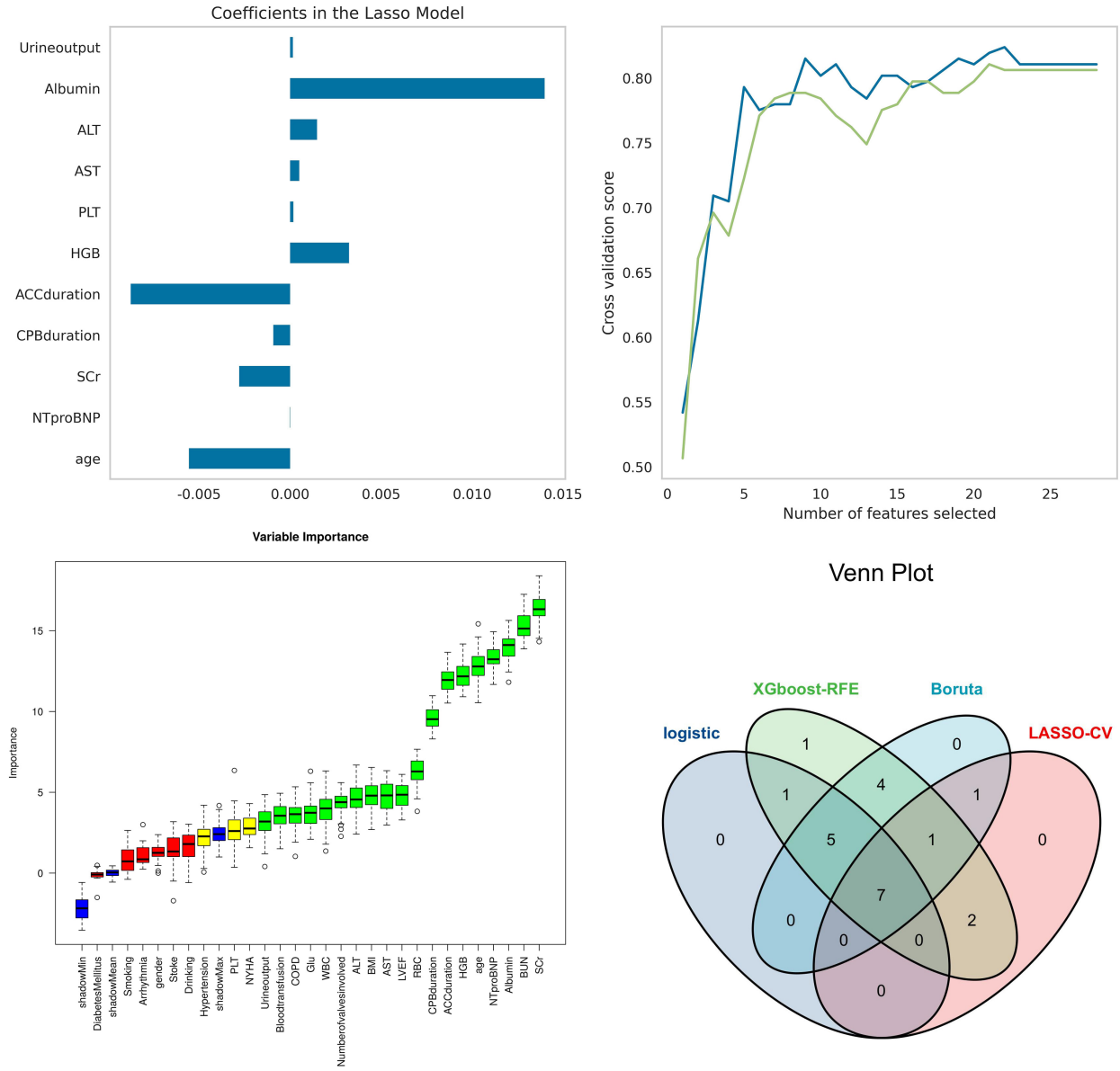

Fig. 2. Predictor screening results. LASSO-CV, Least absolute shrinkage and selection operator with Cross-Validation; XGBoost-RFE, extreme gradient boosting-Recursive Feature Elimination.

Following 5-fold cross-validation, the resulting AUCs for were 0.899 (95% CI: 0.825–0.973) for XGboost, 0.813 (95% CI: 0.715–0.911) for LR, 0.903 (95% CI: 0.833–0.973) for RF, 0.838 (95% CI: 0.746–0.930) for SVM, and 0.831 (95% CI: 0.735–0.927) for KNN (refer to Fig. 5B and **Supplementary Table 5**). **Supplementary Table 6** details the key hyperparameters and package versions used for all five machine learning models.

Corresponding F1 scores for these models were 0.753, 0.735, 0.822, 0.762, and 0.785 respectively for XGboost, LR, RF, SVM, and KNN. A forest plot depicting the AUC scores for the multiple models was generated from the AUROC values for each (Fig. 6).

For each model, the metrics of accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were calculated and juxtaposed (**Supplementary Table 4**). While XGboost demonstrated excellent performance on the training dataset, the Random Forest model was ultimately preferred due to potential overfitting concerns. Moreover, it showed promising predictive performance in independent validation data (AUC = 0.903 vs 0.899 for XGboost, **Supplementary Table 5**), and tighter confidence intervals (95% CI: 0.833–0.973 vs 0.825–0.973, respectively). Data from 91 patients were gathered in the testing phase to validate the efficacy of the Random Forest model. This revealed AUC = 0.872 (95% CI: 0.796–0.948), accuracy = 0.835, sensitivity = 0.718, specificity = 0.923, PPV = 0.875, NPV = 0.814, and F1 score = 0.789.

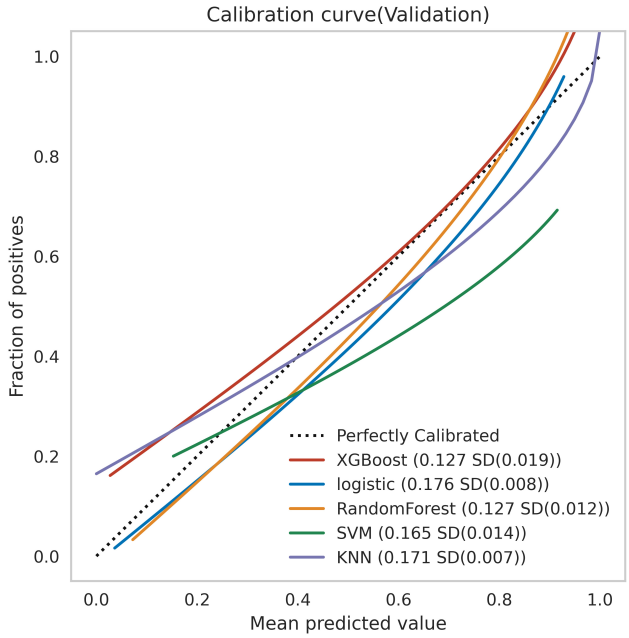

Fig. 3. Calibration plots of five models. The Random Forest achieved lower (better) Brier scores compared with the other models.

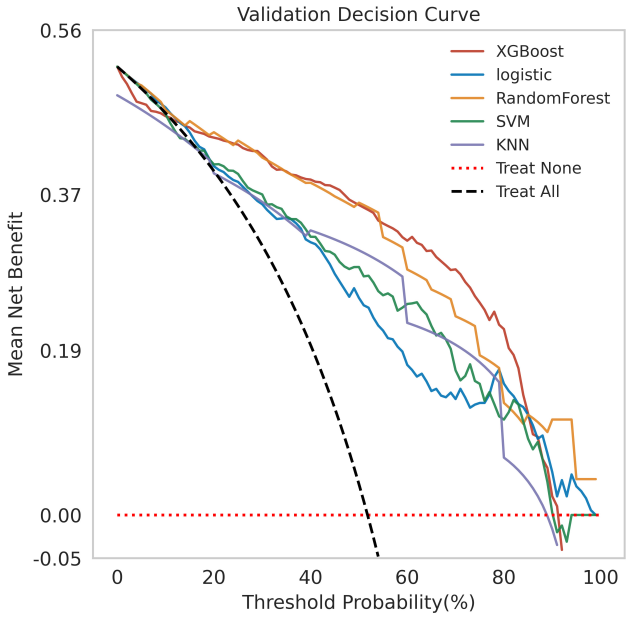

Fig. 4. Decision curve analysis for five machine learning models. The Random Forest model can serve as the best diagnostic tool for AKI.

3.4 Analysis of SHAP-Based Model Interpretability

Next, we examined the significance of different factors that could affect the susceptibility to CAS-AKI. Fig. 7A provides a visual depiction of this hierarchy, where each marker represents a sample, and the color gradient from blue to red signifies the level of the sample eigenvalues. The vertical axis presents the ranking of feature significance, showing the correlation and distribution of each eigenvalue with the SHAP value. Fig. 7B demonstrates the influence of the top nine features on prediction outcomes. It highlights the hierarchical importance of features in the RandomForest model, with the vertical axis listing features in order of significance and the horizontal axis illustrating average SHAP values. This analysis identified NT-proBNP, SCr, and ACC duration as the three most important features, underscoring their substantial impact on the presence of AKI. To better understanding of the model's decision-making at an individual level, an in-depth interpretability analysis was performed on two representative samples (Fig. 7C). This visualization of SHAP values allowed us to ascertain the effect of each feature on the model's predictions for these specific cases.

4. Discussion

This study establishes the first interpretable ML model integrating preoperative NT-proBNP and SCr to predict AKI after cardiac valve surgery. Our analysis of 333 patients (AKI incidence of 31.8%) identified seven robust predictors: age, NT-proBNP, SCr, CPB duration, ACC duration, HGB, and albumin. The Random Forest model showed promising predictive performance (AUC: 0.872) on internal validation, with the potential to surpass traditional models and have a clinical net benefit. Crucially, our model resolves two critical limitations of existing AKI risk tools. First, it leverages ML to capture non-linear interactions between biomarkers and clinical factors, surpassing traditional regression approaches. Second, it validates the synergistic predictive power of NT-proBNP and SCr specifically in patients undergoing valve surgery, a population uniquely vulnerable to CPB-induced hemodynamic stress and nephrotoxicity. Furthermore, the SHAP framework further provides clinician-friendly interpretability, identifying NT-proBNP, SCr, and ACC duration as the dominant drivers of risk for AKI.

The prominence of NT-proBNP and SCr in our model aligns with the cardiorenal pathophysiology inherent to valve surgery. The top ranking of NT-proBNP's top ranking underscores its dual role: it reflects preoperative ventricular dysfunction (a key AKI risk factor) and is renally cleared. This creates a mechanistic loop in which impaired renal function elevates NT-proBNP, which in turn predicts AKI susceptibility to AKI [35,36]. This mechanism is corroborated across other surgical settings. In lung cancer surgery, NT-proBNP/SCr synergy improves the prediction of AKI (AUC: 0.74) via hemodynamic instability and inflammation [1]. In cardiac surgery, the association of NT-proBNP's association with severe AKI (Stage 3 AUC: 0.83) underscores its role in ischemia-reperfusion injury [2]. The synergy with SCr likely arises from CPB-exacerbated hemolysis and inflammation, which simultaneously impair renal tubular function (increasing SCr) and

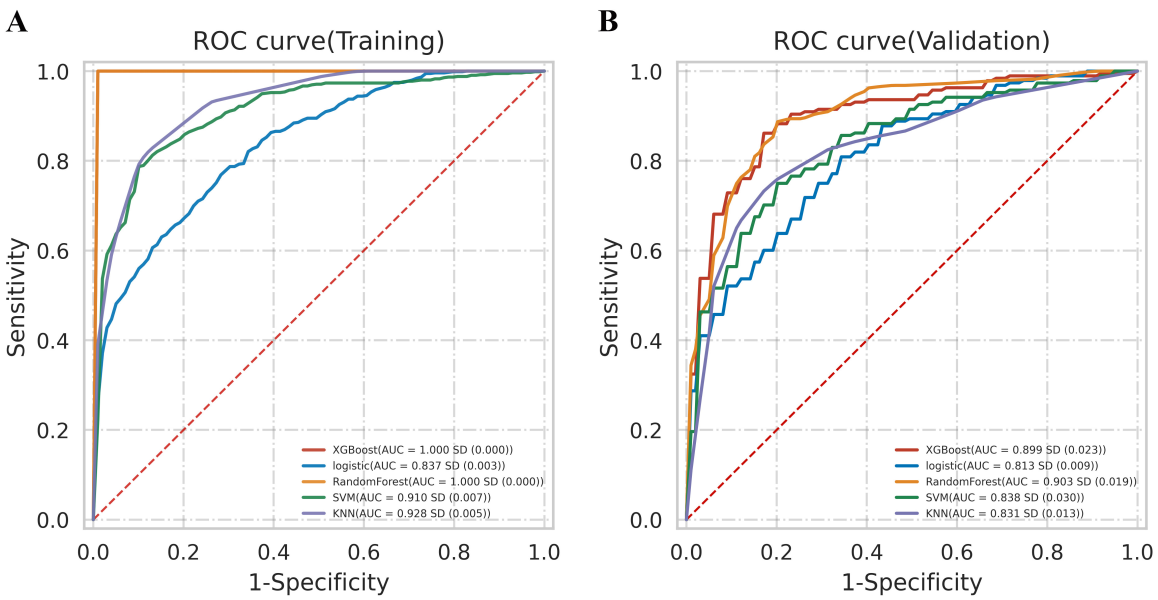

Fig. 5. Receiver-operating characteristic curves for five machine learning models. The Random Forest model achieved a larger (better) AUROC compared with the other models. AUROC, Area Under the Receiver Operating Characteristic curve. (A) ROC curves for different models on the training dataset. (B) ROC curves for different models on the validation dataset.

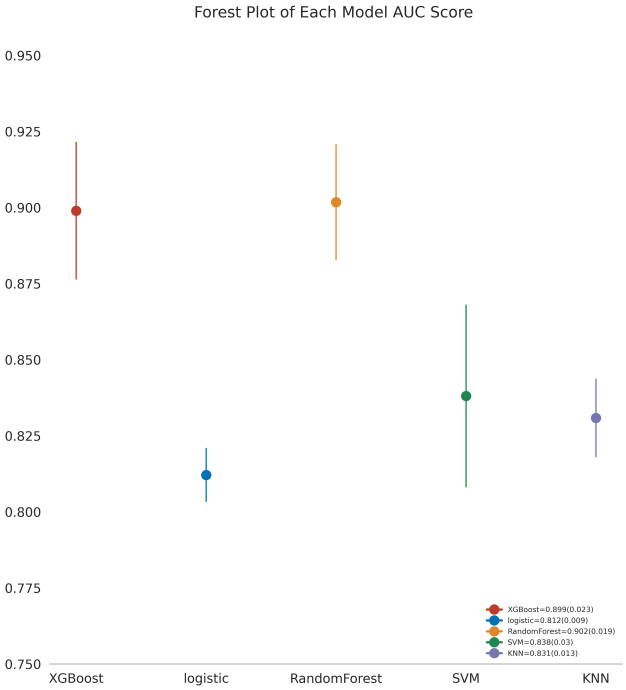

Fig. 6. Forest plot of the AUC score of the five models. The Random Forest model achieved a smaller (better) standard deviation (SD) compared with the other models. AUC, area under the curve.

amplify cardiac wall stress, thereby increasing NT-proBNP [37,38]. Furthermore, the duration of ACC—a modifiable intraoperative factor—emerged as the third key predictor, implicating ischemia-reperfusion injury as a central AKI pathway [13]. Our model also identified haemoglobin (HGB) and albumin as key predictors, reflecting two underappreciated pathways in valve surgery-associated AKI. Anaemia reduces renal oxygen delivery during CPB, exacerbating ischemic injury in the context of hemodilution and microemboli [39,40]. This is especially critical in valve surgery where longer aortic cross-clamp (ACC) times impair renal autoregulation [13]. The intraoperative nadir hemoglobin level is a recognized independent risk factor for CSA-AKI, as anemia compromises tubular oxygen supply during CPB-induced hemodynamic instability [41]. The antioxidant effects of albumin could safeguard the endothelial glycocalyx, preserving the function and structural integrity of the glomerular filtration barrier [42]. Research indicates that low albumin levels before and soon after surgery are independent predictors of kidney dysfunction and decreased long-term survival [43]. Crucially, hypoalbuminemia disrupts glycocalyx integrity, amplifying CPB-induced inflammation and oxidative stress, both of which key drivers of AKI [44].

Our findings should be contextualized within the growing body of literature applying ML to predict CSA-AKI. Several previous studies have demonstrated the superiority of ML models over traditional regression, with reported AUCs often ranging between 0.83 and 0.86 [1,27]. For instance, Tseng *et al.* [27] developed an XGBoost model (AUC: 0.843) that effectively leveraged complex intraoperative variables like urine output and transfusion volume; however, their model was derived from a cohort dominated by coronary artery bypass grafting (CABG) and did not incorporate specific preoperative biomarkers like NT-proBNP, potentially limiting its valve-specific applicability. Conversely, Wang *et al.* [1] established the powerful inde-

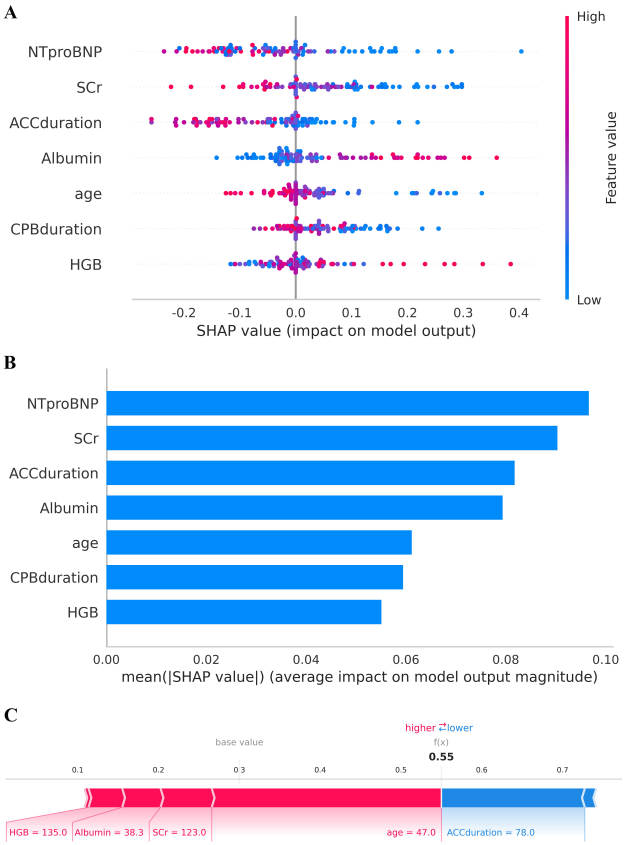

Fig. 7. SHAP analysis of the Random Forest model. A visual representation of each feature of the Random Forest model, showing the relationship between the importance of each feature. The color represents the value of the variable, with red representing the larger value and blue representing the smaller value. (A) SHAP feature contribution distribution plot. (B) Feature importance ranking based on SHAP values. (C) Individual sample prediction explanation plot.

pendent predictive value of preoperative NT-proBNP in a large, mixed cardiac surgery cohort using a linear model, but their approach could not fully capture the non-linear synergy between NT-proBNP and SCr that our ML framework elucidates. While Penny-Dimri *et al.* [13] provided a comprehensive risk profiling of CSA-AKI using ML, their study also focused on a broad cardiac surgery population. Our model thus distinguishes itself by specifically targeting the valve surgery population, a group uniquely exposed to prolonged CPB and ACC times, and by formally integrating and interpreting the interaction between two key cardiorenal biomarkers within an interpretable ML framework. This specific focus allows our model to identify valve-surgery-specific risk drivers, such as ACC duration, with greater clarity. While the referenced study by Vogt *et al.* [45] utilizes dynamic creatinine changes and logistic regression for general cardiac surgery AKI detection, our model specifically for valve surgery patients leverages the synergistic power of preoperative NT-proBNP and creatinine within an interpretable machine learning framework to enable earlier and more personalized risk stratification.

Our work represents a significantly advances beyond the conventional AKI prediction models:

Traditional scores (e.g., Cleveland Clinic, Mehta) rely on clinical variables but ignore biomarkers, achieving modest AUCs (0.76–0.83) [9–11]. By integrating NT-proBNP—a proven marker of ventricular stress and renal clearance—our model improved the AUC to 0.872. This aligns with Wang *et al.* [1], and specifically validates the power of this model in valve surgery cohorts where CPB intensifies hemolysis-mediated injury. While the large cohort size (n = 35,337) in the study by Wang *et al.* [1] established the independent association of NT-proBNP with AKI, their linear model could not capture complex NT-proBNP-SCr interactions. Tseng *et al.* [27] prioritized intraoperative variables (e.g., urine output, transfusion), but their model (AUC: 0.843) lacked valve-specific covariates and biomarker integration. Our SHAP analysis confirmed ACC duration (valve surgery-specific) as the third-most influential predictor, highlighting how procedural complexity modulates AKI risk—a nuance absent in CABG-dominated cohorts.

This study presents a interpretable ML framework with future potential for clinical translation. By quantifying individualized AKI risk preoperatively using widely available biomarkers (NT-proBNP, SCr), the model could, following successful external validation, enable the early triage of high-risk patients for nephroprotective strategies. Its interpretability provides a mechanistic transparency that allows clinicians to understand the drivers of risk—such as cardiac strain, renal vulnerability, or anticipated surgical complexity—which is essential for building trust in ML tools and guiding personalized interventions. Thus, our work establishes a roadmap for developing clinically actionable models that harmonize biomarker biology with clinical covariates.

Notwithstanding these strengths, several limitations of this study should be acknowledged. First, this research is based on data from a single center and was conducted retrospectively. Therefore, it is crucial to carry out external validation across various groups or in a multicenter setting to ensure the findings can be broadly applied. And although SMOTE alleviates class imbalance, synthetic samples may introduce bias. External validation is required to verify the model's performance on real-world imbalanced data. Second, while key intraoperative factors (CPB/ACC duration) were included, real-time hemodynamic data (e.g., blood pressure fluctuations) were unavailable and could refine predictions. Third, NT-proBNP assays lack standardization across centers, and hence future work should establish risk thresholds adaptable to local assays. Finally, prospective trials must assess whether model-guided interventions (e.g., preemptive hemodynamic optimization in high-risk patients) actually reduce the incidence of AKI. We recom-

mend embedding this tool within electronic health records to automate risk alerts, as well as linking it to dynamic postoperative biomarker trends for real-time AKI monitoring.

5. Conclusions

This novel, interpretable ML model leverages preoperative NT-proBNP and SCr to accurately predict AKI after cardiac valve surgery. It demonstrates promising predictive performance (AUC: 0.872) on internal validation, with the potential to surpass traditional models, and have future potential for clinical translation. Prospective trials are essential to assess whether model-guided interventions can truly reduce AKI incidence.

Availability of Data and Materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

TYL: conception and design of the study, data acquisition and interpretation, manuscript writing and revision. HF and YT: study design, data interpretation, manuscript revision. YD and DDL: data collection, data analysis, writing the first draft; SGS and FPH: data collection, data analysis, manuscript writing and revision. All authors approved the final version for publication. All authors agree to be accountable for all aspects of the work and ensure that any questions regarding the accuracy or integrity of any part of the work are properly addressed and resolved.

Ethics Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University (Approval No: 2021-8-003) with a waiver of informed consent due to its retrospective nature.

Acknowledgment

We would like to thank the Extreme Analytics platform for statistical support (https://www.xsmartanalysis.com).

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/HSF49196.

References

[1] Wang C, Gao Y, Tian Y, Wang Y, Zhao W, Sessler DI, *et al*. Prediction of acute kidney injury after cardiac surgery from preoperative N-terminal pro-B-type natriuretic peptide. British Journal of Anaesthesia. 2021; 127: 862–870. https://doi.org/10.1016/j.bja.2021.08.015.

[2] Dasta JF, Kane-Gill SL, Durtschi AJ, Pathak DS, Kellum JA. Costs and outcomes of acute kidney injury (AKI) following cardiac surgery. Nephrology, Dialysis, Transplantation. 2008; 23: 1970–1974. https://doi.org/10.1093/ndt/gfm908.

[3] Schunk SJ, Zarbock A, Meersch M, Küllmar M, Kellum JA, Schmit D, *et al*. Association between urinary dickkopf-3, acute kidney injury, and subsequent loss of kidney function in patients undergoing cardiac surgery: an observational cohort study. Lancet. 2019; 394: 488–496. https://doi.org/10.1016/S0140-6736(19)30769-X.

[4] Rydén L, Sartipy U, Evans M, Holzmann MJ. Acute kidney injury after coronary artery bypass grafting and long-term risk of end-stage renal disease. Circulation. 2014; 130: 2005–2011. https://doi.org/10.1161/CIRCULATIONAHA.114.010622.

[5] Alshaikh HN, Katz NM, Gani F, Nagarajan N, Canner JK, Kacker S, *et al*. Financial Impact of Acute Kidney Injury After Cardiac Operations in the United States. The Annals of Thoracic Surgery. 2018; 105: 469–475. https://doi.org/10.1016/j.athoracsur.2017.10.053.

[6] Kork F, Balzer F, Spies CD, Wernecke KD, Ginde AA, Jankowski J, *et al*. Minor Postoperative Increases of Creatinine Are Associated with Higher Mortality and Longer Hospital Length of Stay in Surgical Patients. Anesthesiology. 2015; 123: 1301–1311. https://doi.org/10.1097/ALN.0000000000000891.

[7] Griffin BR, Teixeira JP, Ambruso S, Bronsert M, Pal JD, Cleveland JC, *et al*. Stage 1 acute kidney injury is independently associated with infection following cardiac surgery. The Journal of Thoracic and Cardiovascular Surgery. 2021; 161: 1346–1355.e3. https://doi.org/10.1016/j.jtcvs.2019.11.004.

[8] Turan A, Cohen B, Adegboye J, Makarova N, Liu L, Mascha EJ, *et al*. Mild Acute Kidney Injury after Noncardiac Surgery Is Associated with Long-term Renal Dysfunction: A Retrospective Cohort Study. Anesthesiology. 2020; 132: 1053–1061. https://doi.org/10.1097/ALN.0000000000003109.

[9] Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP. A clinical score to predict acute renal failure after cardiac surgery. Journal of the American Society of Nephrology. 2005; 16: 162–168. https://doi.org/10.1681/ASN.2004040331.

[10] Wijeysundera DN, Karkouti K, Dupuis JY, Rao V, Chan CT, Granton JT, *et al*. Derivation and validation of a simplified predictive index for renal replacement therapy after cardiac surgery. JAMA. 2007; 297: 1801–1809. https://doi.org/10.1001/jama.297.16.1801.

[11] Mehta RH, Grab JD, O'Brien SM, Bridges CR, Gammie JS, Haan CK, *et al*. Bedside tool for predicting the risk of postoperative dialysis in patients undergoing cardiac surgery. Circulation. 2006; 114: 2208–2216; quiz 2208. https://doi.org/10.1161/CIRCULATIONAHA.106.635573.

[12] Filiberto AC, Ozrazgat-Baslanti T, Loftus TJ, Peng YC, Datta S, Efron P, *et al*. Optimizing predictive strategies for acute kidney injury after major vascular surgery. Surgery. 2021; 170: 298–303. https://doi.org/10.1016/j.surg.2021.01.030.

[13] Penny-Dimri JC, Bergmeir C, Reid CM, Williams-Spence J, Cochrane AD, Smith JA. Machine Learning Algorithms for Predicting and Risk Profiling of Cardiac Surgery-Associated Acute Kidney Injury. Seminars in Thoracic and Cardiovascular Surgery. 2021; 33: 735–745. https://doi.org/10.1053/j.semtcvs.2020.09.028.

[14] Fan R, Qin W, Zhang H, Guan L, Wang W, Li J, *et al*. Ma-

chine learning in the prediction of cardiac surgery associated acute kidney injury with early postoperative biomarkers. Frontiers in Surgery. 2023; 10: 1048431. https://doi.org/10.3389/fsurg.2023.1048431.

[15] Tuñón J, Blanco-Colio L, Cristóbal C, Tarín N, Higueras J, Huelmos A, *et al*. Usefulness of a combination of monocyte chemoattractant protein-1, galectin-3, and N-terminal probrain natriuretic peptide to predict cardiovascular events in patients with coronary artery disease. The American Journal of Cardiology. 2014; 113: 434–440. https://doi.org/10.1016/j.amjcard.2013.10.012.

[16] Daniels LB, Maisel AS. Natriuretic peptides. Journal of the American College of Cardiology. 2007; 50: 2357–2368. https://doi.org/10.1016/j.jacc.2007.09.021.

[17] Liu HH, Cao YX, Jin JL, Guo YL, Zhu CG, Wu NQ, *et al*. Prognostic value of NT-proBNP in patients with chronic coronary syndrome and normal left ventricular systolic function according to glucose status: a prospective cohort study. Cardiovascular Diabetology. 2021; 20: 84. https://doi.org/10.1186/s12933-021-01271-0.

[18] Ndrepepa G, Braun S, Niemöller K, Mehilli J, von Beckerath N, von Beckerath O, *et al*. Prognostic value of N-terminal probrain natriuretic peptide in patients with chronic stable angina. Circulation. 2005; 112: 2102–2107. https://doi.org/10.1161/CIRCULATIONAHA.105.550715.

[19] de Geus HRH, Betjes MG, Bakker J. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. Clinical Kidney Journal. 2012; 5: 102–108. https://doi.org/10.1093/ckj/sfs008.

[20] Yamashita T, Seino Y, Ogawa A, Ogata KI, Fukushima M, Tanaka K, *et al*. N-terminal pro-BNP is a novel biomarker for integrated cardio-renal burden and early risk stratification in patients admitted for cardiac emergency. Journal of Cardiology. 2010; 55: 377–383. https://doi.org/10.1016/j.jjcc.2010.01.008.

[21] Patel UD, Garg AX, Krumholz HM, Shlipak MG, Coca SG, Sint K, *et al*. Preoperative serum brain natriuretic peptide and risk of acute kidney injury after cardiac surgery. Circulation. 2012; 125: 1347–1355. https://doi.org/10.1161/CIRCULATIONAHA.111.029686.

[22] Moltrasio M, Cabiati A, Milazzo V, Rubino M, De Metrio M, Discacciati A, *et al*. B-type natriuretic peptide and risk of acute kidney injury in patients hospitalized with acute coronary syndromes*. Critical Care Medicine. 2014; 42: 619–624. https://doi.org/10.1097/CCM.0000000000000025.

[23] de Cal M, Haapio M, Cruz DN, Lentini P, House AA, Bobek I, *et al*. B-type natriuretic Peptide in the critically ill with acute kidney injury. International Journal of Nephrology. 2011; 2011: 951629. https://doi.org/10.4061/2011/951629.

[24] Breidthardt T, Christ-Crain M, Stolz D, Bingisser R, Drexler B, Klima T, *et al*. A combined cardiorenal assessment for the prediction of acute kidney injury in lower respiratory tract infections. The American Journal of Medicine. 2012; 125: 168–175. https://doi.org/10.1016/j.amjmed.2011.07.010.

[25] Nowak A, Breidthardt T, Dejung S, Christ-Crain M, Bingisser R, Drexler B, *et al*. Natriuretic peptides for early prediction of acute kidney injury in community-acquired pneumonia. Clinica Chimica Acta; International Journal of Clinical Chemistry. 2013; 419: 67–72. https://doi.org/10.1016/j.cca.2013.01.014.

[26] Cardinale D, Cosentino N, Moltrasio M, Sandri MT, Petrella F, Colombo A, *et al*. Acute kidney injury after lung cancer surgery: Incidence and clinical relevance, predictors, and role of N-terminal pro B-type natriuretic peptide. Lung Cancer. 2018; 123: 155–159. https://doi.org/10.1016/j.lungcan.2018.07.009.

[27] Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, *et al*. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. Critical Care. 2020; 24: 478. https://doi.org/10.1186/s13054-020-03179-9.

[28] Luo XQ, Kang YX, Duan SB, Yan P, Song GB, Zhang NY, *et al*. Machine Learning-Based Prediction of Acute Kidney Injury Following Pediatric Cardiac Surgery: Model Development and Validation Study. Journal of Medical Internet Research. 2023; 25: e41142. https://doi.org/10.2196/41142.

[29] Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron. Clinical Practice. 2012; 120: c179–c184. https://doi.org/10.1159/000339789.

[30] Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. Journal of Clinical Epidemiology. 2005; 58: 475–483. https://doi.org/10.1016/j.jclinepi.2004.06.017.

[31] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1996; 58: 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[32] Muthukrishnan R, Rohini R. LASSO: A feature selection technique in predictive modeling for machine learning. In 2016 IEEE international conference on advances in computer applications (ICACA) (pp. 18–20). IEEE. 2016.

[33] Jiang C, Xiu Y, Qiao K, Yu X, Zhang S, Huang Y. Prediction of lymph node metastasis in patients with breast invasive micropapillary carcinoma based on machine learning and SHapley Additive exPlanations framework. Frontiers in Oncology. 2022; 12: 981059. https://doi.org/10.3389/fonc.2022.981059.

[34] Zhang Z, Rousson V, Lee WC, Ferdynus C, Chen M, Qian X, *et al*. Decision curve analysis: a technical note. Annals of Translational Medicine. 2018; 6: 308. https://doi.org/10.21037/atm.2018.07.02.

[35] Takase H, Dohi Y. Kidney function crucially affects B-type natriuretic peptide (BNP), N-terminal proBNP and their relationship. European Journal of Clinical Investigation. 2014; 44: 303–308. https://doi.org/10.1111/eci.12234.

[36] Harrison TG, Shukalek CB, Hemmelgarn BR, Zarnke KB, Ronksley PE, Iragorri N, *et al*. Association of NT-proBNP and BNP With Future Clinical Outcomes in Patients With ESKD: A Systematic Review and Meta-analysis. American Journal of Kidney Diseases. 2020; 76: 233–247. https://doi.org/10.1053/j.ajkd.2019.12.017.

[37] Paparella D, Yau TM, Young E. Cardiopulmonary bypass induced inflammation: pathophysiology and treatment. An update. European Journal of Cardio-Thoracic Surgery. 2002; 21: 232–244. https://doi.org/10.1016/s1010-7940(01)01099-5.

[38] Rosner MH, Okusa MD. Acute kidney injury associated with cardiac surgery. Clinical Journal of the American Society of Nephrology. 2006; 1: 19–32. https://doi.org/10.2215/CJN.00240605.

[39] Schetz M, Bove T, Morelli A, Mankad S, Ronco C, Kellum JA. Prevention of cardiac surgery-associated acute kidney injury. The International Journal of Artificial Organs. 2008; 31: 179–189. https://doi.org/10.1177/039139880803100211.

[40] Ng RRG, Chew STH, Liu W, Shen L, Ti LK. Identification of modifiable risk factors for acute kidney injury after coronary artery bypass graft surgery in an Asian population. The Journal of Thoracic and Cardiovascular Surgery. 2014; 147: 1356–1361. https://doi.org/10.1016/j.jtcvs.2013.09.040.

[41] Zhang H, Yan S, Bian L, Wang J, Wang T, Liu G, *et al*. Intraoperative 20% albumin infusion and acute kidney injury in on-pump cardiac surgery: a focus on preoperative albumin levels. Renal Failure. 2025; 47: 2522327. https://doi.org/10.1080/0886022X.2025.2522327.

[42] Moret E, Jacob MW, Ranucci M, Schramko AA. Albumin-Beyond Fluid Replacement in Cardiopulmonary Bypass Surgery: Why, How, and When? Seminars in Cardio-

thoracic and Vascular Anesthesia. 2014; 18: 252–259. https://doi.org/10.1177/1089253214535667.

[43] de la Cruz KI, Bakaeen FG, Wang XL, Huh J, LeMaire SA, Coselli JS, *et al*. Hypoalbuminemia and long-term survival after coronary artery bypass: a propensity score analysis. The Annals of Thoracic Surgery. 2011; 91: 671–675. https://doi.org/10.1016/j.athoracsur.2010.09.004.

[44] Miralles Bagán J, Parrilla Quiles L, Paniagua Iglesias P, Betbesé Roig AJ, Sabaté Tenas S, Pérez García S, *et al*. The Potential Role of Albumin in Reducing Cardiac Surgery-Associated Acute Kidney Injury: A Randomized Controlled Trial. Journal of Cardiothoracic and Vascular Anesthesia. 2025; 39: 453–460. https://doi.org/10.1053/j.jvca.2024.10.012.

[45] Vogt F, Zibert J, Bahovec A, Pollari F, Sirch J, Fittkau M, *et al*. Improved creatinine-based early detection of acute kidney injury after cardiac surgery. Interactive Cardiovascular and Thoracic Surgery. 2021; 33: 19–26. https://doi.org/10.1093/icvts/ivab034.