International Journal of General Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Machine Learning-Assisted Ensemble Analysis for the Prediction of Acute Pancreatitis with Acute Kidney Injury

Yi Yang[1], Wei Xiao[2], Xingtai Liu[1], Yan Zhang[1], Xin Jin[1], Xiao Li[1]

[1]Department of Clinical Laboratory, The Third Clinical Medical College of the Three Gorges University, Gezhouba Central Hospital of Sinopharm, Yichang, People's Republic of China; [2]Department of Gastroenterology, Xianning central Hospital, The First Affiliated Hospital of Hubby University of Science and Technology, Xianning, People's Republic of China

Correspondence: Xiao Li, Department of Clinical Laboratory, The Third Clinical Medical College of the Three Gorges University, Gezhouba Central Hospital of Sinopharm, Yichang, 443002, People's Republic of China, Tel +86 717-672020, Email yi8847307962629@126.com

Purpose: Acute kidney injury (AKI) is a frequent complication of severe acute pancreatitis (AP) and carries a very poor prognosis. The present study aimed to construct a model capable of accurately identifying those patients at high risk of harboring occult acute kidney injury (AKI) characteristics.

Patients and Methods: We retrospectively recruited a total of 424 consecutive patients at the Gezhouba central hospital of Sinopharm and Xianning central hospital between January 1, 2016, and October 30, 2021. ML-assisted models were developed from candidate clinical features using two-step estimation methods. The receiver operating characteristic curve (ROC), decision curve analysis (DCA), and clinical impact curve (CIC) were performed to evaluate the robustness and clinical practicability of each model.

Results: Finally, a total of 30 candidate variables were included, and the AKI prediction model was established by an ML-based algorithm. The areas under the ROC curve (AUCs) of the random forest classifier (RFC) model, support vector machine (SVM), eXtreme gradient boosting (XGBoost), artificial neural network (ANN), and decision tree (DT) ranged from 0.725 (95% CI 0.223–1.227) to 0.902 (95% CI 0.400–1.403). Among them, RFC obtained the optimal prediction efficiency via adding inflammatory factors, which are serum creatinine (Scr), C-reactive protein (CRP), platelet-to-lymphocyte ratio (PLR), neutrophil-to-lymphocyte ratio (NLR), neutrophil-to-albumin ratio (NAR), and CysC, respectively.

Conclusion: We successfully developed ML-based prediction models for AKI, particularly the RFC, which can improve the prediction of AKI in patients with AP. The practicality of prediction and early detection may be greatly beneficial to risk stratification and management decisions.

Keywords: acute pancreatitis, acute kidney injury, serum cytokines, cystatin-C, machine learning algorithms, prediction

Introduction

The incidence of acute pancreatitis (AP) is a universal gastrointestinal cause for hospital admission worldwide, with cases distributed across all ages and both sexes.[1,2] However, acute kidney injury (AKI) is a frequent complication of severe AP and develops late in the course of the disease, usually after the failure of other organs.[3] Despite the availability of advanced practice guidelines for the management of AP, however, the morbidity and mortality remain stubbornly high.[4] According to the Global Burden of Disease Study (GBD) report, there were 2814972.3 incident cases and 115053.2 deaths due to acute pancreatitis that occurred in 2019 globally.[5] Of note, the high mortality of severe AP is mainly related to other organ failure and secondary infection, which is the most important determinant of outcome in AP.[6,7] Herein, the symptoms and signs of organ failure (eg, respiratory, cardiovascular, and kidney) in patients with AP should be evaluated clinically for the appropriate classification.

Previous studies reported that mortality (especially the mortality of AKI requiring dialysis) can rise sharply to 75%.[3,8] Contrast-enhanced computed tomography is the most useful imaging technique, especially after 72 hours, to assess the extent of the disease.[9] Consequently, the direct medical cost of injury could cause considerable economic expenditure,

Received: 6 February 2022
Accepted: 23 March 2022
Published: 17 May 2022

even the high mortality and disability rate of patients. Additionally, a diagnostic test used for the detection of AKI should be minimally invasive, widely available, cheap, easy to conduct, and reproducible.[10] However, there is still a lack of comprehensive data for the early prediction of AKI risk in all patients with severe AP. It is also increasingly of interest to measure the change in key candidate factors (eg, systemic serum cytokines, imaging indicators, and genetics markers) and to identify crucial tools for AKI prevention. In addition, many studies have evaluated the cytokine profile in different grades of severity of AP, existing evidence has shown that cytokine surge is higher in patients with AP developing AKI.[11,12] Given this situation, we speculate that cytokines may be a potential predictor of AP complicated with AKI.

Nowadays, with the continuous improvement of the availability of electronic health data, the application and exploration of more robust and advanced computing methods such as machine learning in the field of disease prediction become more practical. Machine learning (ML) is an emerging field and gradually infiltrates into medical research. Worthy of note is that ML analysis relies on different depth iterative algorithms to integrate candidate variables, so it can obtain high-precision prediction efficiency.[13–15]

With this in mind, we aimed to develop an AKI risk prediction model for patients with AP that utilizes clinical medical data in this study. The capability of enabling expeditious and accurate risk stratification platforms may facilitate more timely interventions that are conducive to high-risk AKI management via early identification, which can be instrumental in intensive care.

Patients and Methods

Patients Selection

The clinical data of AP patients hospitalized in Gezhouba central hospital of Sinopharm and Xianning central hospital between January 1, 2016, and October 30, 2021, were analyzed retrospectively in this study. The inclusion criteria were as follows: (i) The patient was older than 18 years old; (ii) Patients who met the diagnostic criteria of AP and were hospitalized for AP; (iii) Patients with complete case data, traceable imaging, pathology, and laboratory examination-related indicators. The exclusion criteria were as follows: (i) Patients with incomplete clinical data, chronic pancreatitis, or other inflammatory diseases; (ii) Patients with severe cardiopulmonary function or severe immune deficiency, as well as tumor diseases. This retrospective study was following the declaration of Helsinki and was ethically reviewed and approved by the Institutional Ethics Committee of Gezhouba central hospital of Sinopharm (Reference: 2020006). Since the patient information contained in this study was anonymous, written informed consent was not obtained from all participants. The detailed research flow chart was displayed in Figure 1.

Diagnostic Criteria of AP and AKI

The diagnostic criteria of AP adopted the Kidney Disease Improving Global Outcomes (KDIGO) guideline,[16] as follows: (i) Abdominal pain consistent with AP (persistent, severe, acute onset of upper abdominal pain, usually radiating to the back); (ii) Serum lipase or amylase levels are at least three times the upper limit of normal; (iii) The imaging examination was consistent with the characteristic imaging findings of AP. The diagnostic criteria of AKI were based on the guidelines of the Kidney Disease Improving Global Outcomes criteria,[3] as follows: (i) Serum creatinine (Scr) increased by more than 26.5 μmol/L (0.3mg/dL); (ii) The urine volume lasted for more than 6 hours and was less than 0.5mg/kg/h; (iii) Serum creatinine (Scr) increased 1.5 times higher than the baseline level.

Blood Specimen Collection

The relevant laboratory indexes of the first peripheral venous blood sample within 24 hours after admission were recorded and the clinical data of all selected patients were collected for statistical analysis.

Data Collection and Quality Assessment

According to the principle of inclusion of whole candidate variables, we sorted all the variables that can be collected. The baseline demographics include age, gender, body mass index (BMI), chronic diseases history, pathology, and pancreatic texture. The routine laboratory measurements were also collected, including neutrophil count (109/L), lymphocyte count

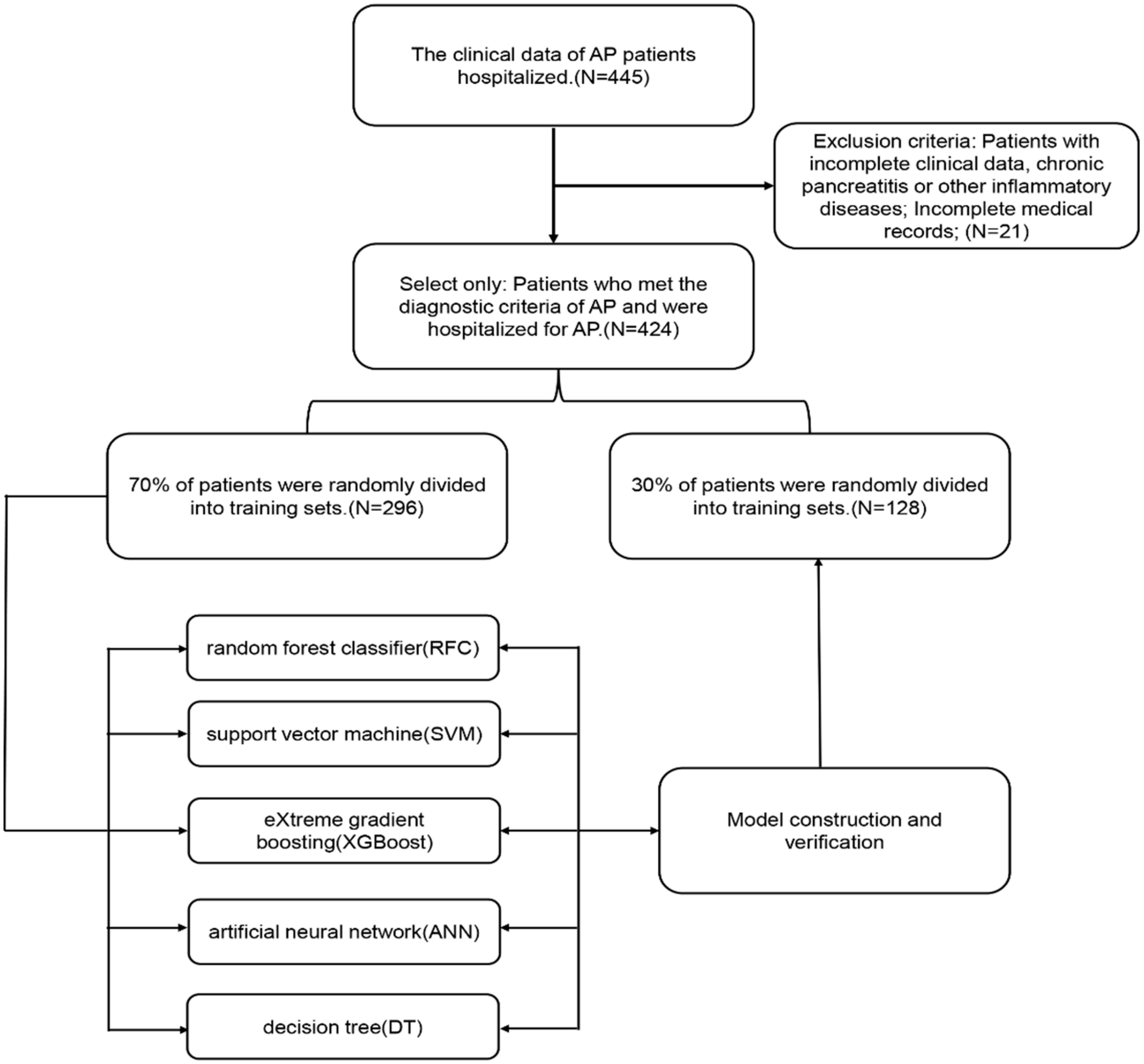

Figure 1 The flow chart of patient selection and data process.

(109/L), platelet count (109/L), monocyte count (109/L), hemoglobin, albumin, and globulin. Among them, the platelet-to-lymphocyte ratio (PLR), neutrophil-to-lymphocyte ratio (NLR), neutrophil-to-albumin ratio (NAR) were obtained by the ratio of lymphocyte count, platelet count, neutrophil count, and albumin, respectively. According to the fact that the missing value was greater than 10% or more of the overall variable, the variable was directly discarded and not included in the final model variable screening.[17] Eventually, a total of 30 variables met the inclusion criteria and were used to build the ML-based model.

Development and Validation of ML-Based Models

We randomly divide the data set into two parts using the caret software package (70% for model training and 30% for model testing). As for model construction, a total of five mL-based algorithms were implemented to establish the prediction model. According to the principle of " OOB error",[18] we have gradually screened the model variables, as follows:

$$\text{Gini (D)} = 1 - \sum_{i=1}^{m} P_i^2$$

The characteristic variable is marked as X and the target variable is marked as Y. The X and Y were evenly divided into two parts, namely X1, Y1, and X2, Y2. Gini index measures the purity of data partition or training tuple set D. Briefly, by sorting the intersection of variable sets, the optimal subset modeling is obtained. The model was evaluated by inspection, discrimination, and calibration. The receiver operating characteristic (ROC) curve was used to evaluate the recognition ability of the prediction model in the training data set and the test data set; The discrimination ability of each model was quantified by the area under the ROC curve (AUC), decision curve analysis (DCA), and clinical impact curve (CIC).

Statistical Analysis

For descriptive analysis, continuous variables were presented as median with interquartile range (IQR). Categorical variables were presented as numbers (%). The Chi-square test or the Mann–Whitney *U*-test as appropriate was used to compare the differences of baseline clinical information between the AKI and non-AKI cohorts. Additionally, a linear regression model was used as a reference model and nomogram visualization. The stepwise regression based on the Akaike information criterion minimum was used to select variables for inclusion in the nomogram, the predictive performance of the nomogram was measured by concordance index (C-index) and calibration with 1000 bootstrap samples to decrease the overfit bias. All analysis was performed using the *R* Project for Statistical Computing (version 4.0.4, http://www.r-project.org/). In all analyses, $P<0.05$ was considered to indicate statistical significance.

Results

Baseline Clinicopathological Characteristics of the Study Cohort

The detailed clinical characteristics and pathological baseline data of 424 patients were displayed in Table 1. There were 67 (15.80%) patients who had developed AKI according to the KDIGO guideline. The number of patients with stages I, II, and III were 38, 22, and 7, respectively (56.72%, 32.84%, and 10.45%). For internal validation, the whole patients were randomly split into a training set (N = 296, 70%) and validation set (N = 128, 30%) via the caret package. 49 (16.55%) and 18 (14.06%) patients developed AKI in the training and validation cohort, respectively.

Feature Variable Selection in ML-Based Algorithm

Feature selection is the area of machine learning that focuses on this problem.[13] Herein, the candidate covariates of each algorithm are filtered by the iterative analysis. A total of 30 variables were executed via correlation analysis. As shown in Figure 2A, the correlation matrix revealed that AKI presented a significant correlation with inflammatory factors and some clinical variables, including PLR, NAR, NLR, and CRP. Additionally, according to the contribution of each meaningful candidate variable to the predictive model, as shown in Figure 2B, CRP, PLR, NAR, NLR, Scr, and CysC contributed to the ML-based model. Consistent with the results of correlation analysis, the five top-ranked predictors were CRP, PLR, NAR, NLR, and CysC.

Construction of ML-Based Risk Stratification Platform

For training data, each patient has a result (positive or negative training), and the final judgment result was output. As shown in the formula: Gini (D)=1- $\sum_{i=1}^{m} P_i^2$. The RFC algorithm represents a computational method for effectively navigating in the free parameter space to obtain a robust model (Figure 3A). The variable Gini index in the RFC model was depicted in Supplementary Table S1. Consistent with the predicted results, the top 6 candidate variables were CRP, PLR, NAR, NLR, Scr, and CysC, respectively. Additionally, data mining through the DT model is very useful, as shown by impurity analysis: Gini (p)= $\sum_{K=1}^{K} Pk(1-Pk)$. As depicted in Figure 3B, with the addition of inflammatory factor indicators, relevant PCT, BMI, and remnant texture acted as an irreplaceable weight at the branch of DT. Meanwhile, the ANN model also shows more robust prediction efficiency than other models but is inferior to the RFC

Table 1 Baseline Demographic and Clinical Characteristics of Patients

Variables	Training Set			P-value	Testing Set			P-value
	Overall (N=296)	Yes (N=49)	No (N=247)		Overall (N=128)	Yes (N=18)	No (N=110)	
Gender (%)								
Male	188 (63.5)	33 (67.3)	155 (62.8)	0.654	80 (62.5)	8 (44.4)	72 (65.5)	0.149
Female	108 (36.5)	16 (32.7)	92 (37.2)		48 (37.5)	10 (55.6)	38 (34.5)	
Age (median [IQR])	49.00 [41.00, 56.00]	60.00 [54.00, 66.00]	46.00 [40.00, 53.00]	<0.001	48.00 [39.75, 54.25]	60.00 [51.00, 62.00]	45.00 [39.00, 51.75]	<0.001
BMI (median [IQR]), kg/m^2	24.10 [21.25, 26.72]	24.30 [20.20, 26.70]	24.10 [21.40, 26.70]	0.391	24.40 [22.00, 27.30]	27.05 [25.32, 27.75]	24.05 [21.80, 26.87]	0.001
Diabetes (%)								
Yes	158 (53.4)	22 (44.9)	136 (55.1)		66 (51.6)	9 (50.0)	57 (51.8)	1
No	138 (46.6)	27 (55.1)	111 (44.9)	0.252	62 (48.4)	9 (50.0)	53 (48.2)	
Hypertension (%)								
Yes	143 (48.3)	23 (46.9)	120 (48.6)		54 (42.2)	9 (50.0)	45 (40.9)	0.641
No	153 (51.7)	26 (53.1)	127 (51.4)	0.957	74 (57.8)	9 (50.0)	65 (59.1)	
Predisposing factors (%)								
High lipogenic	92 (31.1)	15 (30.6)	77 (31.2)	0.818	47 (36.7)	6 (33.3)	41 (37.3)	0.646
Biliary origin	98 (33.1)	18 (36.7)	80 (32.4)		38 (29.7)	7 (38.9)	31 (28.2)	
Other factors[&]	106 (35.8)	16 (32.7)	90 (36.4)		43 (33.6)	5 (27.8)	38 (34.5)	
Treatment (%)								
CRRT	156 (52.7)	26 (53.1)	130 (52.6)	1	64 (50.0)	10 (55.6)	54 (49.1)	0.799
Others	140 (47.3)	23 (46.9)	117 (47.4)		64 (50.0)	8 (44.4)	56 (50.9)	
Hospitalization (median [IQR]), days	19.00 [15.00, 23.25]	22.00 [17.00, 24.00]	19.00 [15.00, 23.00]	0.161	20.00 [15.00, 24.00]	19.50 [15.25, 23.75]	20.00 [15.25, 24.00]	0.633
Scr (median [IQR])	78.00 [66.00, 88.00]	177.00 [131.00, 212.00]	74.00 [64.00, 84.00]	<0.001	76.00 [66.00, 86.25]	175.00 [137.75, 216.50]	74.00 [65.25, 81.75]	<0.001
BUN (median [IQR])	5.20 [4.01, 6.38]	11.71 [8.43, 14.58]	4.76 [3.92, 5.84]	<0.001	5.10 [3.93, 6.16]	10.80 [9.19, 14.03]	4.69 [3.82, 5.82]	<0.001
UA (median [IQR])	361.50 [296.00, 418.00]	449.00 [361.00, 548.00]	346.00 [287.50, 405.50]	<0.001	358.50 [301.50, 421.00]	476.50 [389.00, 543.00]	340.50 [288.00, 412.00]	<0.001
eGFR (median [IQR])	85.50 [75.00, 96.00]	53.00 [46.00, 59.00]	89.00 [80.00, 98.00]	<0.001	86.00 [76.75, 94.25]	53.50 [45.25, 62.75]	88.00 [81.00, 95.75]	<0.001
CRP (median [IQR])	39.00 [26.00, 52.00]	99.00 [53.00, 119.00]	35.00 [24.50, 47.00]	<0.001	41.00 [24.75, 52.00]	78.00 [44.75, 123.25]	39.00 [24.00, 49.75]	<0.001
CysC (median [IQR])	0.92 [0.81, 1.01]	1.69 [1.38, 1.95]	0.88 [0.80, 0.96]	<0.001	0.90 [0.81, 1.02]	1.48 [1.30, 1.81]	0.86 [0.80, 0.97]	<0.001
Ca (median [IQR])	2.28 [2.19, 2.35]	2.10 [1.92, 2.30]	2.28 [2.22, 2.35]	<0.001	2.26 [2.19, 2.35]	2.12 [1.90, 2.31]	2.27 [2.22, 2.36]	0.001
Cl (median [IQR])	105.00 [85.00, 129.00]	103.00 [99.00, 107.00]	108.00 [80.50, 134.00]	0.427	105.50 [87.75, 131.25]	102.00 [100.25, 107.75]	106.50 [83.25, 136.00]	0.699
Alb (median [IQR])	43.80 [41.27, 46.10]	38.50 [36.20, 41.30]	44.40 [42.15, 46.55]	<0.001	43.95 [41.88, 46.40]	40.55 [38.55, 43.42]	44.50 [42.10, 46.70]	<0.001
WBC (median [IQR])	10.21 [8.04, 11.90]	11.68 [9.25, 13.56]	10.01 [7.79, 11.70]	<0.001	10.45 [8.18, 12.52]	11.37 [9.15, 12.92]	10.38 [8.00, 12.17]	0.156
NEUT (median [IQR])	7.96 [5.93, 9.84]	11.07 [9.18, 12.92]	7.40 [5.66, 9.55]	<0.001	7.31 [5.69, 9.52]	11.21 [7.41, 12.41]	7.18 [5.49, 8.90]	<0.001
LYM (median [IQR])	1.39 [1.03, 1.97]	1.05 [0.79, 1.27]	1.56 [1.08, 2.06]	<0.001	1.44 [0.96, 2.04]	1.04 [0.73, 1.21]	1.60 [1.08, 2.08]	<0.001
HCT (median [IQR])	41.85 [39.00, 44.40]	35.30 [33.40, 39.10]	42.90 [39.70, 44.90]	<0.001	41.00 [38.70, 44.12]	39.80 [36.05, 40.75]	41.45 [39.02, 44.30]	0.002
PLT (median [IQR])	180.00 [145.75, 209.00]	184.00 [154.00, 225.00]	177.00 [145.00, 207.00]	0.134	181.00 [145.00, 209.00]	194.50 [138.25, 216.00]	178.50 [145.00, 206.75]	0.437
NAR (median [IQR])	0.08 [0.07, 0.09]	0.14 [0.12, 0.16]	0.07 [0.06, 0.08]	<0.001	0.08 [0.07, 0.09]	0.16 [0.14, 0.17]	0.08 [0.07, 0.08]	<0.001
NLR (median [IQR])	5.59 [3.93, 7.06]	10.59 [6.46, 14.76]	5.20 [3.71, 6.61]	<0.001	5.24 [3.53, 7.20]	12.53 [9.63, 14.41]	4.80 [3.26, 6.45]	<0.001
PLR (median [IQR])	123.30 [96.65, 161.05]	173.60 [133.40, 229.40]	117.70 [92.20, 147.20]	<0.001	120.85 [91.77, 155.07]	166.65 [127.20, 196.43]	113.70 [86.88, 145.18]	<0.001

Note: [&]Other factors: Alcoholic, autoimmune, idiopathic, traumatic, etc.

Abbreviations: IQR, inter-quartile range; CRRT, continuous renal replacement therapy; Scr, serum creatinine; BUN, blood urea nitrogen; UA, uric acid; eGFR, estimated glomerular filtration rate; CRP, C-reactive protein; CysC, cystatin C; Alb, albumin; WBC, white blood cell; NEUT, neutrophil count; LYM, lymphocyte count; HCT, hematocrit; PLT, platelet count; NAR, neutrophil-to-albumin ratio; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio.

https://doi.org/10.2147/IJGM.S361330 DovePress

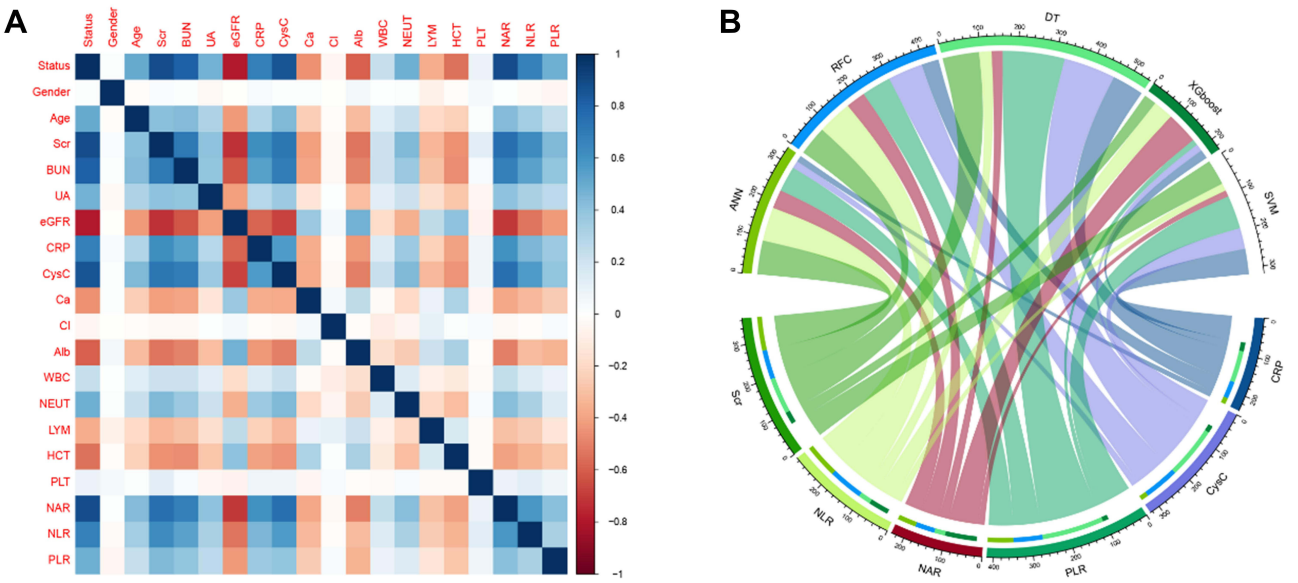

Figure 2 Variable screening and weight allocation. (**A**) Correlation matrix analysis of candidate features. (**B**) The weight distribution of the candidate variables of each ML-based model.

Abbreviations: Scr, serum creatinine; BUN, blood urea nitrogen; UA, uric acid; eGFR, estimated glomerular filtration rate; CRP, C-reactive protein; CysC, cystatin C; Alb, albumin; WBC. white blood cell; NEUT, neutrophil count; LYM, lymphocyte count; HCT, hematocrit; PLT, platelet count; NAR, neutrophil-to-albumin ratio; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio.

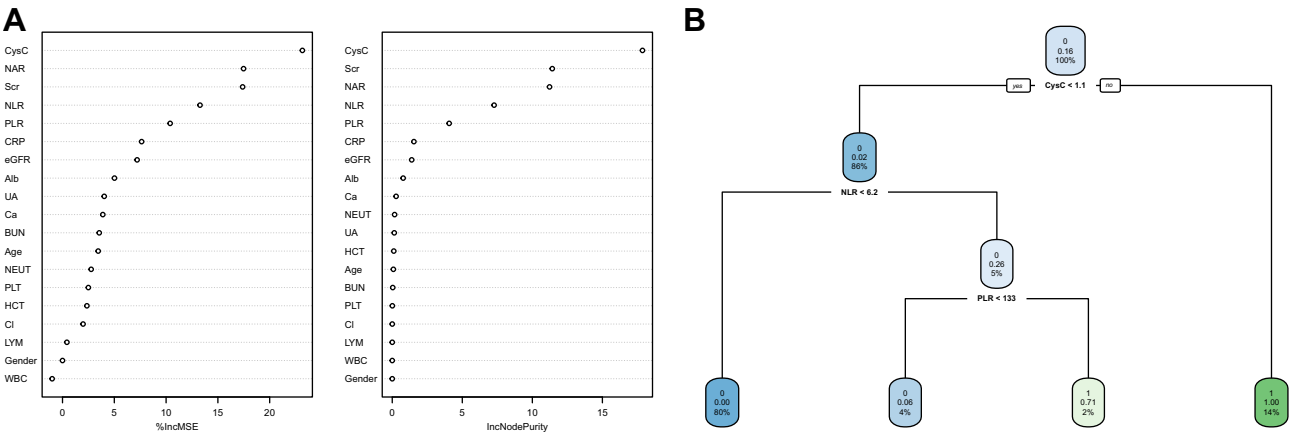

Figure 3 Predictive model visualization based on ML-based algorithm. (**A**) RFC model. (**B**) DT model.

Notes: The candidate factors associated with AKI were ordered via RFC algorithm (**A**) and (**B**) prediction node and weight was allocated via DT algorithm.

(Figure 4). We also constructed nomograms, which depend on the parameters obtained by logistic regression (LR). Compared with the conventional predictive model, inflammatory factors also accounted for an important proportion.

Comparison of Prediction Efficiency of ML-Based Models

To explore whether ML-Based models can elevate the prediction performance, we further used five supervised learning models for AKI assessment, attempted. As expected, the RFC model can better distinguish whether patients with a high risk of AKI or not. As shown in Figure 5, the DCA also exhibited that the RFC model was equipped with a robust prediction performance in the training and validation cohorts, respectively. Additionally, the AUCs of RFC models reached a plateau when 6 variables were introduced, followed by ANN, DT, SVM, and XGBoost. The detailed predictive performance of ML-based models was summarized in Table 2. Undoubtedly, the prediction efficiency of RFC was superior to the generalized

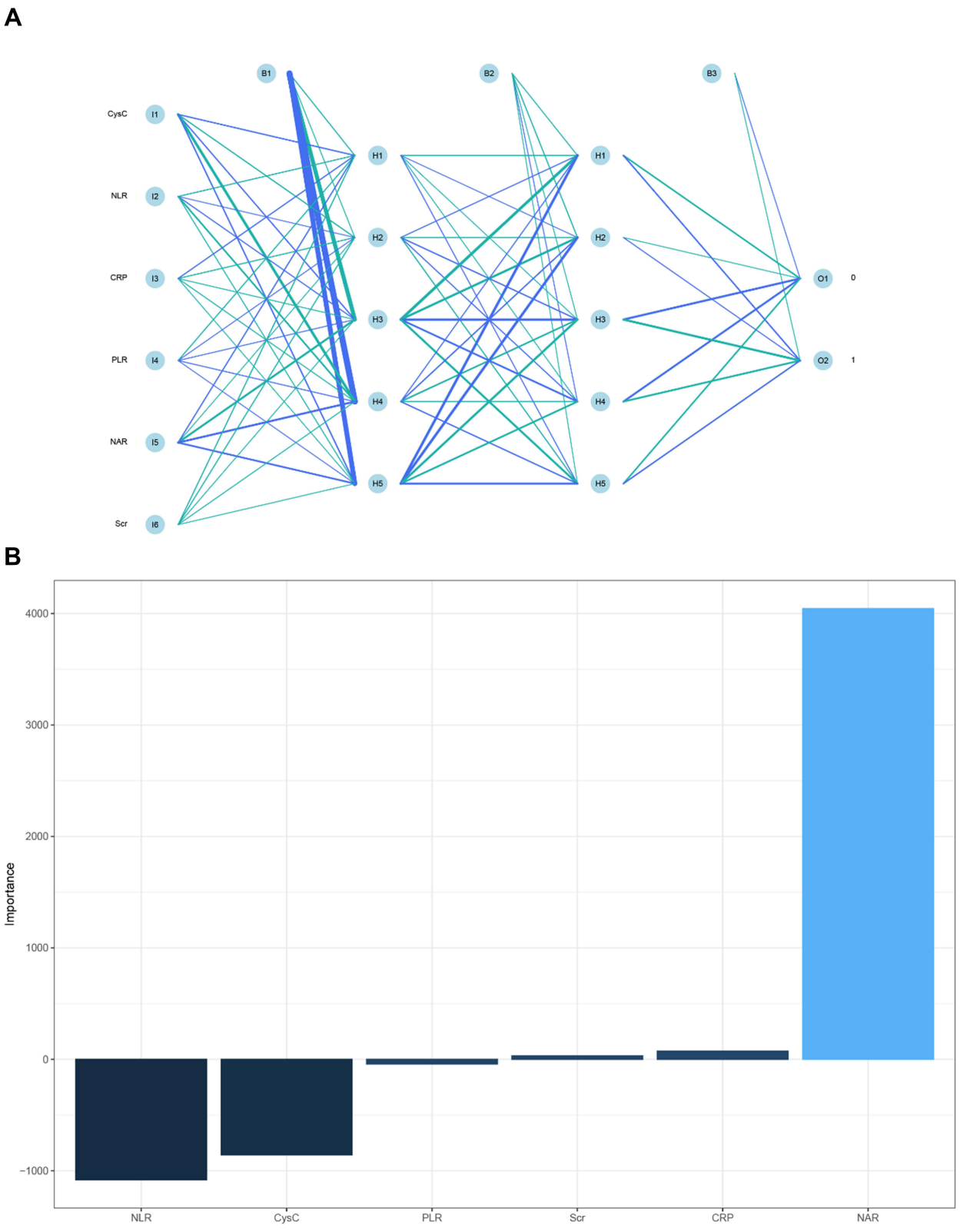

Figure 4 Predictive model visualization based on ANN algorithm. (**A**) ANN model. (**B**) Variable importance using connection weight.
Notes: The candidate factors associated with AKI were ordered via ANN algorithm (**A**) and (**B**) prediction node and weight was allocated via ANN algorithm.

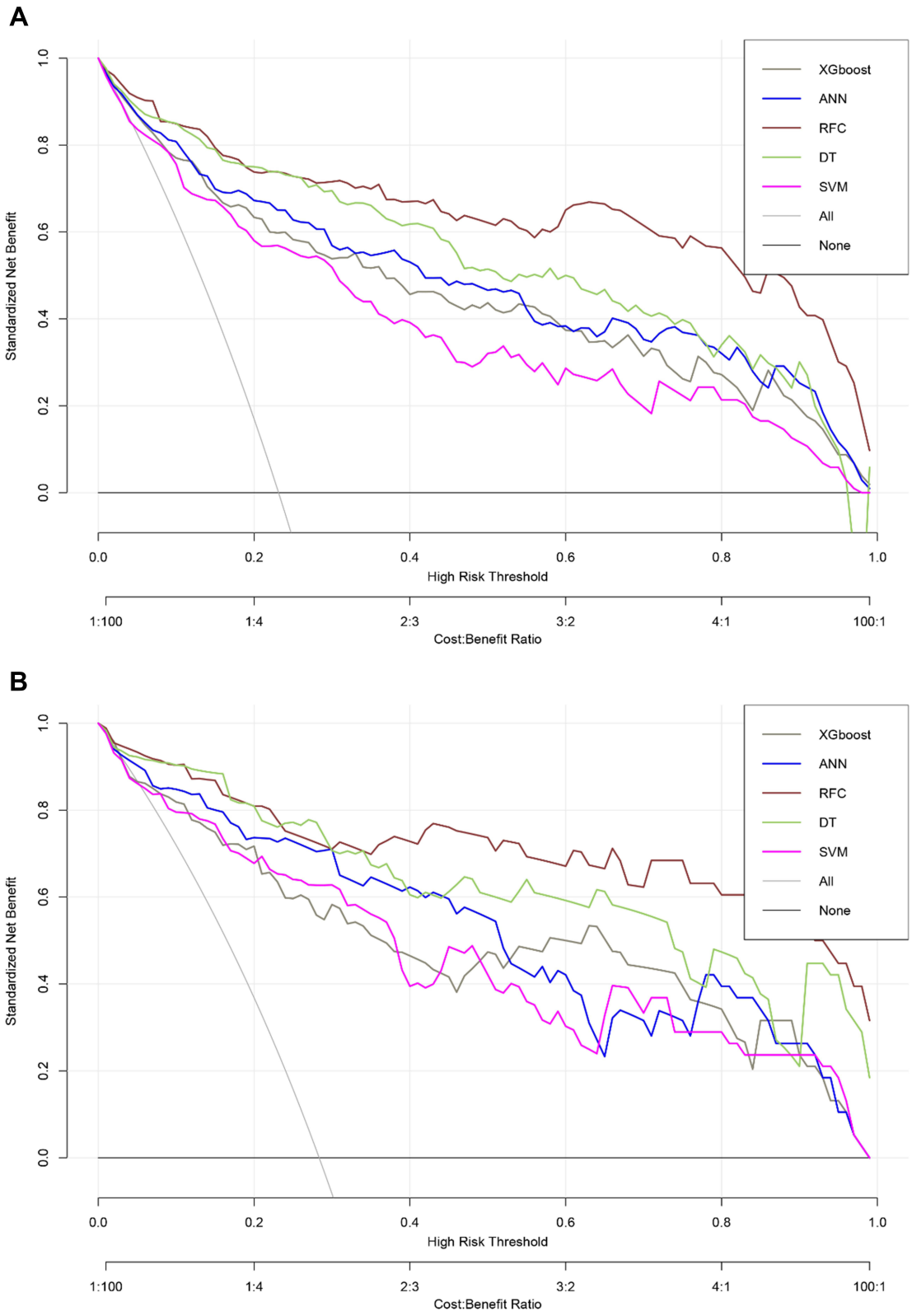

Figure 5 Prediction performance of candidate models based on ML-based algorithm. (**A**) DCA for five ML-based models in the training set. (**B**) DCA for five ML-based models in the testing set.

Abbreviations: RFC, random forest classifier; SVM, support vector machine; DT, decision tree; ANN, artificial neural network; XGboost, eXtreme gradient boosting.

linear model (Supplementary Table S2). Collectively, using the iterative algorithm analysis of supervised learning, both RFC and DT (machine learning-assisted decision-support) models were appropriately used to guide AKI prediction.

Internal and External Verification of Optimal RFC Prediction Model

To further validate the predictive performance of the RFC model, we also used CIC to evaluate the accuracy, as illustrated in Supplementary Figure S1, the CIC demonstrated that the stratification of AKI could be achieved in the training cohorts. These results were also consistent with the results of validation cohorts, indicating RFC had the best performance across the metrics of discrimination, calibration, and overall performance, especially the candidate systemic inflammation markers were highly relevant to AKI.

Discussion

AKI is still widely regarded as the largest contributor to the main mortality of patients with AP because AKI can develop in the early or late course of the disease, so it has become an important determinant of prognosis.[19] Given this situation, it is particularly important to accurately predict which patients are more likely to develop AKI, to actively carry out symptomatic prevention and treatment. To date, only a few studies are dealing with AP and AKI, mainly focusing on epidemiology, pathogenesis, causes, and management of AKI.[3,8,20] In the retrospective study, we established a risk stratification platform, an integrated model derived from five ML-based algorithms, which can accurately predict AKI of AP patients in advance by using the clinical information in electronic health records at admission. Importantly, the AUC displayed by the platform in the training and validation queue ranges from 0.725 to 0.902. The impact of risk stratification may help to promote a more responsive health system for high-risk AKI patients through early identification, subsequent immediate intervention, and intensive care and monitoring, which is expected to help save lives.

Cytokines may be involved in the pathogenesis of AKI, such as IL-1β, IL-8, and IL-6, which act on endothelial cells leading to kidney ischemia, thrombosis, and release of oxygen free radicals.[21] Meanwhile, inflammatory mediators may increase mucosal permeability and lead to endotoxin translocation, that is, endotoxin promotes the development of AKI by increasing the level of endothelin, which leads to vasoconstriction, decreased renal blood flow, and tubular necrosis.[21–23] Consistent with the conclusions of the above literature, we found that inflammatory factors played a crucial role in monitoring AKI. Increasing experimental and clinical studies revealed that inflammatory response plays an irreplaceable role in the pathophysiology of AKI.[21,24–26] Of note, the systemic inflammatory response in the process of AKI may be caused by local inflammation of renal tissue.[21,27] Benefiting from this enlightenment, we monitored peripheral blood-related inflammatory markers in patients with AP, including CRP, PLR, NAR, and NLR. It's not surprising that systemic inflammatory markers such as neutrophils, lymphocytes, platelets, a combined ratio of albumin, C-reactive protein (CRP), and biomarkers may contribute to predicting AKI in patients with AP. Although the pathophysiology of systemic inflammatory response driven by local pancreatic injury has not been fully understood, studies have shown that both the innate immune system (including neutrophils, monocytes, and macrophages) and the adaptive immune system (mainly composed of lymphocytes) play an important role in disease progression.[28,29]

Table 2 The ROC Curve Analyses for Predicting AKI in Each ML-Based Model

Model	Training Set			Testing Set		
	AUC Mean	AUC 95% CI	Variables[&]	AUC Mean	AUC 95% CI	Variables[&]
RFC	0.902	0.400–1.403	6	0.913	0.364–1.462	6
SVM	0.725	0.223–1.227	8	0.758	0.209–1.307	8
DT	0.887	0.385–1.389	9	0.891	0.342–1.440	9
ANN	0.872	0.370–1.374	8	0.868	0.339–1.397	8
XGboost	0.791	0.289–1.293	10	0.801	0.272–1.330	10

Note: [&]Variables included in the model.

Abbreviations: RFC, random forest classifier; SVM, support vector machine; DT, decision tree; ANN, artificial neural network; XGboost, eXtreme gradient boosting; AUC, area under curve; 95% CI, 95% confidence interval.

Collectively, monitoring the level of inflammatory factors can precisely predict the risk of AKI. Additionally, our study also revealed that Scr and CysC, were all significantly associated with high-risk of AKI, these findings elucidated that added value of inflammatory factors can contribute to the prediction of AKI in patients with AP.

Low molecular-weight proteins have been used to estimate the value of eGFR.[30] Among them, cystatin-C has been identified as a superior GFR marker to creatinine in chronic renal insufficiency with small variability.[31–33] For instance, El-Gammacy et al[34] reported that serum cystatin-C on day 3 of life can predict AKI earlier than serum Cr and eGFR. Ahlström et al[35] reported that serum cystatin-C was as good as plasma creatinine in detecting ARF in intensive care patients. Consistent with previous studies, our study showed that serum cystatin-C in patients with AP complicated with AKI was significantly higher than that in patients without AKI, suggesting that the increase of baseline serum cystatin-C may be related to AKI in patients with AP. Compared with the Kidney Disease Improving Global Outcomes (KDIGO) guideline, the added value of cystatin-C and systemic inflammation markers can easily be applied to clinical practice for predicting the occurrence of AKI in patients with AP.

Nowadays, given the increasing applicability and effectiveness of supervised machine learning algorithms in predictive disease modeling, the breadth of research seems to progress. The well-known supervised learning classifiers, including support vector machine, random forest, convolutional neural network, and decision tree, have been gradually applied to clinical practice.[36] In this study, we successfully screened the rank order of risk factors predicting AKI, and with the help of machine learning classification, it showed that the machine learning-assisted decision-support model has more advantages than the traditional linear regression model. RFC is an integrated classifier composed of many DTS, which is equivalent to the set of many branch trees.[37,38] This study relies on the training results of the RFC model (more than 500 trees) on different feature subsets and then uses out-of-a-bag (OOB) with classification accuracy to evaluate its performance. We used bootstrap resampling technology to select feature sets through random sampling and random selection. In short, the average reduction of the Gini impurity index was used to evaluate the importance of variables. Each variable can get the corresponding weight according to the Gini index, and then the calculated sum was used as the risk score to obtain more robust prediction efficiency.

We acknowledged that this study has some limitations. First, our data only come from two medical centers, so there may be some deviation in the distribution of the actual data, which may lead to the trained model can not handle all the data well. In the future, we will consider randomly selecting more data from different medical centers to verify the model in our study and reduce the differences between different data centers. Second, some clinical and molecular traits were inadequate, it is still necessary to screen and explore cutting-edge molecular markers, such as immunodiagnostic biomarkers and genetic analysis. Third, the divergent candidate factors in muti-ensemble analyses showed that even if the ML-based algorithm is used, which could be attributed to improving predictive performance but also to be interpreted with caution.

Conclusion

Taken together, the machine learning-assisted decision-support model developed in this study was shown to be a potentially useful tool in determining the high-risk and predicting the possibility of AKI in patients with AP. As such, it may be useful for clinicians to use in combination with other biomarkers to determine which patients need effective intervention and treatment, as well as to alleviate the economic burden of hospitalization.

Acknowledgments

The author thanks all the medical staff of the tertiary medical center for their contributions, including data entry, sorting, and information verification. At the same time, the author also thanks all the patients who participated in the study for their support and cooperation.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis, and interpretation, or in all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors declare no conflicts of interest for this work.

References

1. Greenberg JA, Hsu J, Bawazeer M, et al. Clinical practice guideline: management of acute pancreatitis. *Can J Surg*. 2016;59(2):128–140. doi:10.1503/cjs.015015
2. Lankisch PG, Apte M, Banks PA. Acute pancreatitis. *Lancet*. 2015;386(9988):85–96. doi:10.1016/S0140-6736(14)60649-8
3. Nassar TI, Qunibi WY. AKI associated with acute pancreatitis. *CJASN*. 2019;14(7):1106–1115. doi:10.2215/CJN.13191118
4. Zerem E. Treatment of severe acute pancreatitis and its complications. *World J Gastroenterol*. 2014;20(38):13879–13892. doi:10.3748/wjg.v20.i38.13879
5. Li CL, Jiang M, Pan CQ, Li J, Xu LG. The global, regional, and national burden of acute pancreatitis in 204 countries and territories, 1990–2019. *BMC Gastroenterol*. 2021;21(1):332. doi:10.1186/s12876-021-01906-2
6. Garg PK, Singh VP. Organ failure due to systemic injury in acute pancreatitis. *Gastroenterology*. 2019;156(7):2008–2023. doi:10.1053/j.gastro.2018.12.041
7. James TW, Crockett SD. Management of acute pancreatitis in the first 72 hours. *Curr Opin Gastroenterol*. 2018;34(5):330–335. doi:10.1097/MOG.0000000000000456
8. Devani K, Charilaou P, Radadiya D, Brahmbhatt B, Young M, Reddy C. Acute pancreatitis: trends in outcomes and the role of acute kidney injury in mortality- A propensity-matched analysis. *Pancreatology*. 2018;18(8):870–877. doi:10.1016/j.pan.2018.10.002
9. Tenner S, Baillie J, DeWitt J, Vege SS. American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol*. 2013;108(9):1400–1415; 1416. doi:10.1038/ajg.2013.218
10. Wajda J, Dumnicka P, Maraj M, Ceranowicz P, Kuźniewski M, Kuśnierz-Cabala B. Potential prognostic markers of acute kidney injury in the early phase of acute pancreatitis. *Int J Mol Sci*. 2019;20(15):3714. doi:10.3390/ijms20153714
11. Dambrauskas Z, Giese N, Gulbinas A, et al. Different profiles of cytokine expression during mild and severe acute pancreatitis. *World J Gastroenterol*. 2010;16(15):1845–1853. doi:10.3748/wjg.v16.i15.1845
12. Prasada R, Muktesh G, Samanta J, et al. Natural history and profile of selective cytokines in patients of acute pancreatitis with acute kidney injury. *Cytokine*. 2020;133:155177. doi:10.1016/j.cyto.2020.155177
13. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593
14. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci*. 2019;50(4):477–487. doi:10.1016/j.jmir.2019.09.005
15. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin microb infect*. 2020;26(5):584–595. doi:10.1016/j.cmi.2019.09.009
16. Stevens PE, Levin A. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med*. 2013;158(11):825–830. doi:10.7326/0003-4819-158-11-201306040-00007
17. Miot HA. Anomalous values and missing data in clinical and experimental studies. *J vascu brasileiro*. 2019;18:e20190004. doi:10.1590/1677-5449.190004
18. Fan J, Lv J. A selective overview of variable selection in high dimensional feature space. *Stat Sin*. 2010;20(1):101–148.
19. Scurt FG, Bose K, Canbay A, Mertens PR, Chatzikyrkou C. Pankreatitisbedingte akute Nierenschädigung (AP-AKI): Definition, Pathophysiologie, Diagnostik und Therapie [Acute kidney injury following acute pancreatitis (AP-AKI): definition, Pathophysiology, Diagnosis and Therapy]. *Z Gastroenterol*. 2020;58(12):1241–1266. German. doi:10.1055/a-1255-3413
20. Li H, Qian Z, Liu Z, Liu X, Han X, Kang H. Risk factors and outcome of acute renal failure in patients with severe acute pancreatitis. *J Crit Care*. 2010;25(2):225–229. doi:10.1016/j.jcrc.2009.07.009
21. Zhang XP, Wang L, Zhou YF. The pathogenic mechanism of severe acute pancreatitis complicated with renal injury: a review of current knowledge. *Dig Dis Sci*. 2008;53(2):297–306. doi:10.1007/s10620-007-9866-5
22. Bose SM, Verma GR, Mazumdar A, Giridhar M, Ganguly NK. Significance of serum endotoxin and antiendotoxin antibody levels in predicting the severity of acute pancreatitis. *Surg Today*. 2002;32(7):602–607. doi:10.1007/s005950200109
23. Sandoval D, Gukovskaya A, Reavey P, et al. The role of neutrophils and platelet-activating factor in mediating experimental pancreatitis. *Gastroenterology*. 1996;111(4):1081–1091. doi:10.1016/S0016-5085(96)70077-X
24. Zhao K, Chen C, Shi Q, et al. Inhibition of glycogen synthase kinase-3β attenuates acute kidney injury in sodium taurocholate-induced severe acute pancreatitis in rats. *Mol Med Rep*. 2014;10(6):3185–3192. doi:10.3892/mmr.2014.2650
25. Nagao S, Taguchi K, Sakai H, et al. Carbon monoxide-bound hemoglobin vesicles ameliorate multiorgan injuries induced by severe acute pancreatitis in mice by their anti-inflammatory and antioxidant properties. *Int J Nanomed*. 2016;11:5611–5620. doi:10.2147/IJN.S118185
26. Kolodecik TR, Reed AM, Date K, et al. The serum protein renalase reduces injury in experimental pancreatitis. *J Biol Chem*. 2017;292(51):21047–21059. doi:10.1074/jbc.M117.789776
27. Yilmaz H, Cakmak M, Inan O, Darcin T, Akcay A. Can neutrophil-lymphocyte ratio be independent risk factor for predicting acute kidney injury in patients with severe sepsis? *Ren Fail*. 2015;37(2):225–229. doi:10.3109/0886022X.2014.982477
28. Zhang Y, Wu W, Dong L, Yang C, Fan P, Wu H. Neutrophil to lymphocyte ratio predicts persistent organ failure and in-hospital mortality in an Asian Chinese population of acute pancreatitis. *Medicine*. 2016;95(37):e4746. doi:10.1097/MD.0000000000004746
29. Bermejo-Martín JF, Tamayo E, Ruiz G, et al. Circulating neutrophil counts and mortality in septic shock. *Critical Care*. 2014;18(1):407. doi:10.1186/cc13728

30. Little DJ, Mascio HM, Altenburg RJ, et al. Implementing GFR estimation guidelines using cystatin c: a quality improvement project. *Am j Kidney Dis*. 2016;67(4):711–712. doi:10.1053/j.ajkd.2015.10.014
31. Star R, Hostetter T, Hortin GL. New markers for kidney disease. *Clin Chem*. 2002;48(9):1375–1376. doi:10.1093/clinchem/48.9.1375
32. Coll E, Botey A, Alvarez L, et al. Serum cystatin C as a new marker for noninvasive estimation of glomerular filtration rate and as a marker for early renal impairment. *Am j Kidney Dis*. 2000;36(1):29–34. doi:10.1053/ajkd.2000.8237
33. Dharnidharka VR, Kwon C, Stevens G. Serum cystatin C is superior to serum creatinine as a marker of kidney function: a meta-analysis. *Am J Kidney Dis*. 2002;40(2):221–226. doi:10.1053/ajkd.2002.34487
34. El-Gammacy TM, Shinkar DM, Mohamed NR, Al-Halag AR. Serum cystatin C as an early predictor of acute kidney injury in preterm neonates with respiratory distress syndrome. *Scand J Clin Lab Invest*. 2018;78(5):352–357. doi:10.1080/00365513.2018.1472803
35. Ahlström A, Tallgren M, Peltonen S, Pettilä V. Evolution and predictive power of serum cystatin C in acute renal failure. *Clin Nephrol*. 2004;62 (5):344–350. doi:10.5414/CNP62344
36. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019;19(1):281. doi:10.1186/s12911-019-1004-8
37. Baştanlar Y, Ozuysal M. Introduction to machine learning. *Methods Mol Biol*. 2014;1107:105–128.
38. Chen W, Li Y, Xue W, et al. Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Sci Total Environ*. 2020;701:134979. doi:10.1016/j.scitotenv.2019.134979