

Comprehensive Sepsis Risk Prediction in Leukemia Using a Random Forest Model and Restricted Cubic Spline Analysis

Yanqi Kou^{1,2,*}, Yuan Tian^{2,3,*}, Yanping Ha^{2,3,*}, Shijie Wang^{1,*}, Xiaobai Sun¹, Shuxin Lv¹, Botao Luo^{2,3}, Yuping Yang², Ling Qin¹

¹Department of Hematology, The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, Henan Province, People's Republic of China; ²Department of Gastroenterology, Affiliated Hospital of Guangdong Medical University, Guangdong Medical University, Zhanjiang City, Guangdong Province, People's Republic of China; ³Department of Pathology, Guangdong Medical University, Zhanjiang City, Guangdong Province, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yuping Yang, Department of Gastroenterology, Affiliated Hospital of Guangdong Medical University, Guangdong Medical University, No. 2 Wenming East Road, Xiashan, Zhanjiang, Guangdong, 524000, People's Republic of China, Email yangyupingchn@163.com; Ling Qin, Department of Hematology, The First Affiliated Hospital and College of Clinical Medicine, Henan University of Science and Technology, 24 Jinghua Road, Jianxi District, Luoyang, Henan Province, 471003, People's Republic of China, Email zqq777@126.com

Background: Sepsis is a severe complication in leukemia patients, contributing to high mortality rates. Identifying early predictors of sepsis is crucial for timely intervention. This study aimed to develop and validate a predictive model for sepsis risk in leukemia patients using machine learning techniques.

Methods: This retrospective study included 4310 leukemia patients admitted to the Affiliated Hospital of Guangdong Medical University from 2005 to 2024, using 70% for training and 30% for validation. Feature selection was performed using univariate logistic regression, LASSO, and the Boruta algorithm, followed by multivariate logistic regression analysis. Seven machine learning models were constructed and evaluated using receiver operating characteristic (ROC) curves and decision curve analysis (DCA). Shapley additive explanations (SHAP) were applied to interpret the results, and restricted cubic spline (RCS) regression explored the nonlinear relationships between variables and sepsis risk. Furthermore, we examined the interactions among predictors to better understand their potential interrelationships.

Results: The random forest (RF) model outperformed all others, achieving an AUC of 0.765 in the training cohort and 0.700 in the validation cohort. Key predictors of sepsis identified by SHAP analysis included C-reactive protein (CRP), procalcitonin (PCT), neutrophil count (Neut), lymphocyte count (Lymph), thrombin time (TT), red blood cell count (RBC), total bile acid (TBA), and systolic blood pressure (SBP). RCS analysis revealed significant non-linear associations between CPR, PCT, Neut, Lymph, TT, RBC and SBP with sepsis risk. Pairwise correlation analysis further revealed interactions among these variables.

Conclusion: The RF model exhibited robust predictive power for sepsis in leukemia patients, providing clinicians with a valuable tool for early risk assessment and the optimization of treatment strategies.

Keywords: leukemia, sepsis, prediction model, biomarkers, machine learning

Introduction

Sepsis is a life-threatening condition that significantly increases mortality in leukemia patients.¹ Due to the immunosuppressed state of leukemia patients, they are particularly vulnerable to infections, which can rapidly escalate to sepsis.²⁻⁴ Early identification of sepsis is critical for improving outcomes, yet accurately predicting sepsis risk remains challenging.⁵

Although several biomarkers and clinical indicators have been associated with sepsis, there is a need for a reliable, data-driven predictive model that can identify patients at high risk for sepsis.² This challenge arises due to the

multifactorial nature of sepsis and the intricate immune dysfunctions seen in leukemia.⁶ Leukemic blasts can significantly impair immune responses by inhibiting T cell ligands (such as PD-L1 and Gal-9) and diminishing the cytotoxic activity of NK cells, which increases vulnerability to infections.⁷ Furthermore, treatments like chemotherapy and hematopoietic stem cell transplantation further elevate the risk of severe infections.⁸ As a result, identifying more sensitive biomarkers is critical to improving sepsis prediction accuracy in leukemia patients.

In recent years, machine learning has offered new insights for the comprehensive analysis of biomarkers, holding promise for the development of more sensitive and specific predictive models to aid clinicians in identifying high-risk patients and optimizing treatment strategies.^{9,10} These methods are now widely used in clinical practice to facilitate early disease detection and timely prevention.⁹ However, their application in leukemia remains limited. For example, the Epic sepsis model, a proprietary tool for predicting sepsis, has been adopted by several hospitals across the United States.¹¹ Despite its widespread use, the model's sensitivity is relatively low, and its predictive performance has been suboptimal compared to current clinical practices.

This study aims to develop a predictive model using machine learning techniques to assess sepsis risk in leukemia patients. By incorporating clinical, demographic, and laboratory data, we seek to identify key predictors of sepsis and design a model that can be effectively applied in clinical settings to enhance the timely detection and management of this severe complication. To achieve this, we compared multiple machine learning models to evaluate their performance. Each model possesses unique strengths and limitations, making a comprehensive comparison essential to identify the most suitable approach for this specific clinical scenario. This comprehensive comparison was critical to identify the most clinically feasible and effective approach for sepsis prediction, prioritizing sensitivity to ensure timely identification of high-risk cases in this vulnerable population.

Materials and Methods

Data Source

A total of 5872 eligible participants with leukaemia were retrospectively enrolled from the Affiliated Hospital of Guangdong Medical University between January 2005 and June 2024. The diagnoses were made based on the criteria of the World Health Organization (WHO), the Society for Hematopathology, and the European Association for Hematopathology. Patients were excluded if more than 20% of their data were missing and with prior malignancy. After applying the criteria, 4310 eligible leukaemia patients were included (Figure 1). For participant allocation, we divided the data into training (70%) and validation (30%) cohorts to develop and test the predictive model effectively, ensuring a balance between learning complex patterns and validating model performance to prevent overfitting. Ethical approval for the study was obtained from the Institutional Review Board of the Affiliated Hospital of Guangdong Medical University (Approval No. PJKT2024-211). This study complies with the Declaration of Helsinki.

Data and Variables

The extracted variables included demographic and clinical characteristics of the patients, Age, Weight, Marital Status, Hypertension, coronary heart disease (CHD), Diabetes, Smoking, Alcohol, Infection History, Hepatitis, Family Cancer History, Family History, Gender, ABO blood, duration of hospital stay, discharge status, basic vital signs, and laboratory parameters. Blood and biochemical test results were collected on the first day of admission. In cases where multiple test results were available for a specific variable, the first measurement was used in the analysis.

Model Construction and Validation

Univariate logistic analysis, Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis, and Boruta algorithm were employed to determine the potential risk factors in the training dataset. In the univariate logistic analysis, variables with $p < 0.05$ were considered potential biomarkers. We chose LASSO regression analysis due to its ability to impose penalties on variables, helping to reduce the likelihood of overfitting. To determine the most predictive variables, we employed 10-fold cross-validation, focusing on those that minimized the cross-validated error. Boruta is a feature selection method based on random forests that determines the importance of each variable by comparing its Z-score with

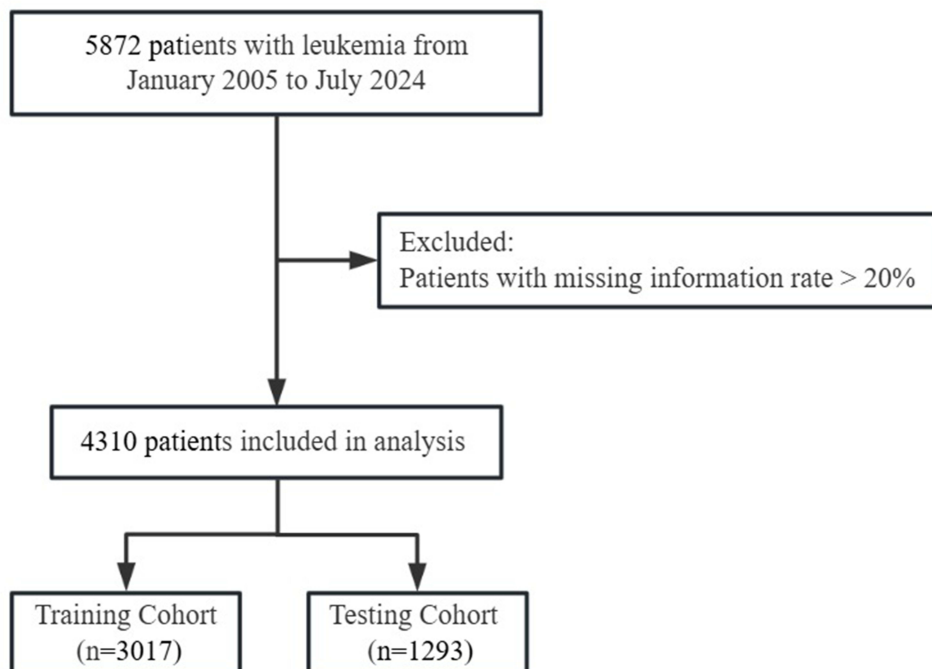


Figure 1 Flow diagram of patient selection.

that of its “shadow” counterparts. During the algorithm’s execution, all real features were duplicated and randomly shuffled to generate Z-scores. If a real feature’s Z-score consistently exceeded the maximum Z-score of the shadow features across multiple independent tests, it was deemed important and included in subsequent machine learning model construction. The identical variables obtained by the above three methods were subsequently included in the multivariate logistic regression analysis to determine the final model. A variance inflation factor (VIF) of ≤ 5 indicated no collinearity among the variables in the final model.

The important variables were incorporated into seven different machine learning algorithms for model construction, including logistic regression (LR), k-nearest neighbor (KNN), support vector machine (SVM), decision tree (DTree), random forest (RF), extreme gradient boosting (XGBoost), and artificial neural network (NNET). To prevent overfitting, we employed a 10-fold cross-validation strategy. The dataset was split into ten folds, with nine folds used for training and one-fold used for validation, cycling through all folds. Grid search was used for hyperparameter tuning, optimizing parameters such as the number of trees (n-estimators), maximum tree depth (max-depth), and minimum samples per leaf (min-samples-leaf). The optimal configuration was selected based on the highest AUC and F1 score from the validation set.

The model with the highest area under the receiver operating characteristic curve (ROC) was selected to determine model performance. Model discrimination was further assessed using sensitivity, specificity, recall, accuracy, and F1 score metrics. Decision curve analysis (DCA) was conducted to evaluate the clinical utility of the models. The best-performing model was subsequently used for further interpretive analysis. After training the model on the training cohort, all model parameters were fixed, and the model’s performance was further evaluated using the validation cohort.

To better understand the decision-making process of the best-performing model, Shapley additive explanations (SHAP) were employed. Based on cooperative game theory’s Shapley values, SHAP rationally allocates contributions to the model’s output among individual input features. SHAP also reveals feature importance and visually displays the direction and magnitude of each feature’s contribution to the predictive outcome, thereby providing a deep understanding of the model’s decision-making process. Restricted cubic spline (RCS) regression was used to describe the non-linear relationship. Possible nonlinear relationships between the change in important variables and Sepsis were examined by

a logistic regression model with RCS. In addition, we explore the interactions between predictors to gain insight into the potential relationships between predictors.

Statistical Analysis

Missing data (<20% of total) were imputed using the Multiple Imputation (MI) method. Additionally, to assess the impact of missing data handling on the model, we performed sensitivity analysis. Continuous variables were expressed as medians and interquartile ranges (IQR), while categorical variables were expressed as total numbers and percentages. Chi-square (χ^2) tests, Wilcoxon rank-sum tests, or Fisher's exact tests were used as appropriate. Propensity score matching (PSM) was performed using 1:1 optimal pair matching to minimize the influence of confounding factors and enhance the validity of our findings. The matching was considered optimal when the sum of the absolute pairwise distances in the matched sample was minimized. The analyses were performed using R software (version 4.2.2). Statistical significance was determined by a two-tailed *P* value of less than 0.05.

Results

Comparative Analysis of Clinical Profiles in Leukemia Patients With and Without Sepsis

Based on the inclusion and exclusion criteria, a total of 4310 patients with leukemia were enrolled and randomly assigned to the training (n=3017) and validation cohorts (n=1293) in a ratio of 7:3. The baseline characteristics table provides key insights into the demographic and clinical profiles of the study population. As a result, all the characteristics were equally distributed between the two cohorts (Table 1). A total of 3017 leukemia patients were included in the training cohort, of whom 917 experienced sepsis.

Table 1 Patient Demographics and Baseline Characteristics in the Training Cohort and Validation Cohort

Characteristic	Group			p-value
	Overall, N = 4,310	Train, N = 3,017	Validation, N = 1,293	
Age	46 (15, 66)	46 (15, 65)	46 (15, 66)	0.743
Hospital Days	14 (6, 24)	14 (6, 24)	15 (6, 24)	0.618
Weight	55 (45, 60)	55 (46, 60)	55 (45, 60)	0.232
Temperature	36.70 (36.50, 37.10)	36.70 (36.50, 37.10)	36.70 (36.50, 37.10)	0.483
Pulse Rate	90 (82, 104)	90 (82, 103)	90 (82, 104)	0.382
HR	90 (83, 100)	90 (83, 100)	90 (83, 100)	0.524
RR	20.0 (20.0, 21.0)	20.0 (20.0, 21.0)	20.0 (20.0, 21.0)	0.183
DBP	68 (62, 73)	68 (63, 73)	68 (62, 74)	0.462
SBP	120 (110, 127)	120 (110, 127)	119 (110, 127)	0.244
WBC	14 (4, 33)	14 (4, 33)	14 (4, 32)	0.648
RBC	2.76 (2.12, 3.63)	2.78 (2.12, 3.64)	2.74 (2.12, 3.58)	0.464
MCV	90 (84, 97)	90 (84, 98)	90 (84, 97)	0.206
MCHC	330 (318, 341)	330 (318, 340)	330 (319, 341)	0.215
MCH	29.8 (27.6, 32.1)	29.9 (27.6, 32.1)	29.8 (27.5, 32.1)	0.637
RDW	16.5 (14.6, 19.2)	16.4 (14.6, 19.2)	16.5 (14.6, 19.4)	0.562
Lymph	2 (1, 7)	2 (1, 8)	2 (1, 7)	0.578
Mono	1 (0, 3)	1 (0, 3)	1 (0, 3)	0.847
Neut	4 (1, 12)	4 (1, 13)	4 (1, 12)	0.852
Hct	25 (19, 33)	26 (19, 33)	25 (19, 33)	0.647
Eos	0.02 (0.00, 0.13)	0.02 (0.00, 0.13)	0.02 (0.00, 0.14)	0.467
Baso	0.01 (0.00, 0.07)	0.01 (0.00, 0.07)	0.01 (0.00, 0.07)	0.916

(Continued)

Table I (Continued).

Characteristic	Group			p-value
	Overall, N = 4,310	Train, N = 3,017	Validation, N = 1,293	
Hb	76 (50, 103)	75 (51, 103)	77 (49, 103)	0.921
PLT	81 (29, 217)	82 (29, 224)	77 (27, 205)	0.076
MPV	9.20 (8.00, 10.20)	9.20 (8.00, 10.20)	9.10 (8.00, 10.20)	0.598
PDW	16.45 (12.00, 17.70)	16.40 (11.90, 17.70)	16.50 (12.50, 17.60)	0.526
PCTI	0.07 (0.03, 0.18)	0.07 (0.03, 0.18)	0.07 (0.03, 0.16)	0.061
ALT	21 (13, 35)	21 (13, 36)	21 (12, 35)	0.568
AST	25 (17, 39)	25 (17, 40)	25 (17, 39)	0.970
GGT	34 (19, 58)	34 (19, 58)	33 (19, 56)	0.460
TBIL	10 (8, 12)	10 (8, 13)	10 (8, 12)	0.807
DBIL	4.3 (3.2, 5.8)	4.3 (3.2, 5.8)	4.3 (3.2, 5.7)	0.555
IBIL	5.5 (3.7, 7.7)	5.5 (3.7, 7.6)	5.6 (3.8, 7.8)	0.229
ALB	38 (33, 41)	38 (33, 41)	38 (33, 41)	0.431
GLO	26 (24, 29)	26 (24, 29)	26 (24, 29)	0.194
TP	64 (59, 69)	64 (59, 69)	65 (59, 69)	0.192
A.G	1.44 (1.14, 1.70)	1.45 (1.17, 1.70)	1.42 (1.10, 1.70)	0.124
PA	156 (106, 209)	157 (105, 209)	154 (106, 208)	0.659
TBA	6 (4, 11)	6 (4, 11)	6 (3, 11)	0.106
CREA	67 (48, 83)	67 (48, 83)	67 (47, 84)	0.797
UREA	4.61 (3.80, 5.50)	4.60 (3.80, 5.45)	4.66 (3.73, 5.61)	0.484
UA	314 (260, 370)	314 (262, 371)	314 (255, 369)	0.643
TC	3.49 (2.83, 4.21)	3.49 (2.83, 4.22)	3.48 (2.84, 4.20)	0.858
TG	1.41 (0.96, 1.93)	1.40 (0.96, 1.93)	1.43 (0.95, 1.93)	0.664
HDL.C	0.83 (0.61, 1.03)	0.84 (0.61, 1.03)	0.83 (0.61, 1.02)	0.608
LDL.C	1.92 (1.44, 2.47)	1.92 (1.45, 2.48)	1.92 (1.40, 2.45)	0.430
LPa	113 (68, 185)	112 (67, 185)	115 (72, 187)	0.359
apoAI	0.89 (0.71, 1.05)	0.89 (0.71, 1.05)	0.89 (0.71, 1.05)	0.997
apoB	0.71 (0.58, 0.87)	0.71 (0.58, 0.87)	0.71 (0.57, 0.87)	0.852
Ca	2.16 (2.05, 2.27)	2.16 (2.05, 2.27)	2.16 (2.05, 2.28)	0.640
K	3.87 (3.56, 4.18)	3.87 (3.55, 4.18)	3.86 (3.57, 4.18)	0.911
Na	139.4 (136.7, 141.1)	139.4 (136.8, 141.1)	139.3 (136.3, 141.0)	0.093
Cl	103.0 (100.4, 105.3)	103.1 (100.5, 105.4)	103.0 (100.3, 105.0)	0.135
LDH	350 (219, 608)	355 (217, 607)	336 (225, 613)	0.776
ALP	83 (64, 123)	83 (64, 123)	82 (64, 121)	0.566
ChE	5.82 (4.17, 7.29)	5.82 (4.18, 7.28)	5.82 (4.10, 7.30)	0.861
HCY	11 (8, 14)	11 (8, 14)	11 (8, 14)	0.596
Cys.C	0.88 (0.71, 1.12)	0.87 (0.70, 1.12)	0.88 (0.71, 1.13)	0.356
CO2	22.4 (20.1, 23.9)	22.4 (20.1, 24.0)	22.4 (20.1, 23.9)	0.500
AG	13.8 (11.8, 15.8)	13.7 (11.7, 15.8)	13.8 (11.9, 15.9)	0.312
Glu	5.84 (4.84, 6.88)	5.81 (4.82, 6.83)	5.90 (4.89, 6.95)	0.129
TT	17.10 (15.90, 18.49)	17.10 (15.80, 18.47)	17.18 (15.99, 18.50)	0.772
PT	13.30 (12.30, 15.00)	13.20 (12.26, 14.90)	13.40 (12.32, 15.20)	0.011
PTR	1.09 (1.02, 1.22)	1.09 (1.02, 1.22)	1.11 (1.02, 1.25)	0.005
APTT	35 (31, 40)	35 (31, 40)	35 (31, 40)	0.209
APTTR	1.12 (1.02, 1.24)	1.12 (1.02, 1.24)	1.12 (1.01, 1.24)	0.653
PT.INR	1.08 (1.00, 1.22)	1.07 (1.00, 1.21)	1.08 (1.01, 1.24)	0.016
Fbg	3.32 (2.59, 4.06)	3.31 (2.63, 4.06)	3.34 (2.51, 4.09)	0.672
CRP	32 (14, 54)	31 (14, 53)	32 (14, 56)	0.684
NRBC	0.00 (0.00, 0.02)	0.00 (0.00, 0.02)	0.00 (0.00, 0.02)	0.227
α -HBDH	300 (200, 463)	299 (198, 461)	303 (203, 471)	0.480

(Continued)

Table I (Continued).

Characteristic	Group			p-value
	Overall, N = 4,310	Train, N = 3,017	Validation, N = 1,293	
AFU	18 (14, 22)	18 (14, 22)	17 (14, 22)	0.460
RET	0.05 (0.03, 0.08)	0.05 (0.03, 0.08)	0.05 (0.03, 0.08)	0.532
RET	5 (2, 14)	5 (2, 14)	6 (2, 14)	0.323
IRF	6 (0, 15)	5 (0, 15)	6 (0, 15)	0.857
PCT	0.29 (0.12, 0.68)	0.28 (0.11, 0.68)	0.30 (0.12, 0.68)	0.183
Sepsis				0.145
No	2,971 (68.9%)	2,100 (69.6%)	871 (67.4%)	
Yes	1,339 (31.1%)	917 (30.4%)	422 (32.6%)	
Gender				0.273
Male	2,554 (59.3%)	1,804 (59.8%)	750 (58.0%)	
Female	1,756 (40.7%)	1,213 (40.2%)	543 (42.0%)	
ABO				0.768
A	1,090 (25.3%)	763 (25.3%)	327 (25.3%)	
AB	256 (5.9%)	179 (5.9%)	77 (6.0%)	
B	1,157 (26.8%)	797 (26.4%)	360 (27.8%)	
O	1,807 (41.9%)	1,278 (42.4%)	529 (40.9%)	
Death				0.140
No	4,016 (93.2%)	2,800 (92.8%)	1,216 (94.0%)	
Yes	294 (6.8%)	217 (7.2%)	77 (6.0%)	
Marital Status				0.394
Single, divorced, and widowed	1,762 (40.9%)	1,246 (41.3%)	516 (39.9%)	
Married	2,548 (59.1%)	1,771 (58.7%)	777 (60.1%)	
Hypertension				0.141
No	4,014 (93.1%)	2,821 (93.5%)	1,193 (92.3%)	
Yes	296 (6.9%)	196 (6.5%)	100 (7.7%)	
CHD				0.573
No	4,229 (98.1%)	2,958 (98.0%)	1,271 (98.3%)	
Yes	81 (1.9%)	59 (2.0%)	22 (1.7%)	
Diabetes				0.636
No	4,149 (96.3%)	2,907 (96.4%)	1,242 (96.1%)	
Yes	161 (3.7%)	110 (3.6%)	51 (3.9%)	
Smoking				0.234
No	4,039 (93.7%)	2,836 (94.0%)	1,203 (93.0%)	
Yes	271 (6.3%)	181 (6.0%)	90 (7.0%)	
Alcohol				0.878
No	4,176 (96.9%)	2,924 (96.9%)	1,252 (96.8%)	
Yes	134 (3.1%)	93 (3.1%)	41 (3.2%)	
Infection History				0.611
No	4,099 (95.1%)	2,866 (95.0%)	1,233 (95.4%)	
Yes	211 (4.9%)	151 (5.0%)	60 (4.6%)	
Hepatitis				0.935
No	4,192 (97.3%)	2,934 (97.2%)	1,258 (97.3%)	
Yes	118 (2.7%)	83 (2.8%)	35 (2.7%)	
FamilyCancer History				0.731
No	4,284 (99.4%)	2,998 (99.4%)	1,286 (99.5%)	
Yes	26 (0.6%)	19 (0.6%)	7 (0.5%)	
Family History				0.763
No	4,172 (96.8%)	2,922 (96.9%)	1,250 (96.7%)	
Yes	138 (3.2%)	95 (3.1%)	43 (3.3%)	

(Continued)

Table 1 (Continued).

Characteristic	Group			p-value
	Overall, N = 4,310	Train, N = 3,017	Validation, N = 1,293	
Fecal OB				0.298
Negative	3,316 (76.9%)	2,308 (76.5%)	1,008 (78.0%)	
Positive	994 (23.1%)	709 (23.5%)	285 (22.0%)	
Urine TP				0.823
Negative	3,441 (79.8%)	2,406 (79.7%)	1,035 (80.0%)	
Positive	869 (20.2%)	611 (20.3%)	258 (20.0%)	
Urine URO				0.825
Negative	3,080 (71.5%)	2,153 (71.4%)	927 (71.7%)	
Positive	1,230 (28.5%)	864 (28.6%)	366 (28.3%)	
Urine BIL				0.619
Negative	3,225 (74.8%)	2,251 (74.6%)	974 (75.3%)	
Positive	1,085 (25.2%)	766 (25.4%)	319 (24.7%)	
Urine KET				0.795
Negative	3,711 (86.1%)	2,595 (86.0%)	1,116 (86.3%)	
Positive	599 (13.9%)	422 (14.0%)	177 (13.7%)	
Urine Glu				0.569
Negative	3,923 (91.0%)	2,751 (91.2%)	1,172 (90.6%)	
Positive	387 (9.0%)	266 (8.8%)	121 (9.4%)	
Urine OB				0.100
Negative	3,189 (74.0%)	2,254 (74.7%)	935 (72.3%)	
Positive	1,121 (26.0%)	763 (25.3%)	358 (27.7%)	

Abbreviations: HR, heart rate; RR, respiratory rate; DBP, diastolic blood pressure; SBP, systolic blood pressure; WBC, white blood cell count; RBC, red blood cell count; MCV, mean corpuscular volume; MCHC, mean corpuscular hemoglobin concentration; MCH, mean corpuscular hemoglobin; RDW, red cell distribution width; Lymph, lymphocyte count; Mono, monocyte count; Neut, neutrophil count; Hct, hematocrit; Eos, eosinophil count; Baso, basophil count; Hb, hemoglobin; PLT, platelet count; MPV, mean platelet volume; PDW, platelet distribution width; PCT1, plateletcrit; ALT, alanine transaminase; AST, aspartate transaminase; GGT, gamma-glutamyl transferase; TBIL, total bilirubin; DBIL, direct bilirubin; IBIL, indirect bilirubin; ALB, albumin; GLO, globulin; TP, total protein; A.G, albumin-to-globulin ratio; PA, prealbumin; TBA, total bile acid; CREA, creatinine; UREA, urea; UA, uric acid; TC, total cholesterol; TG, triglycerides; HDL.C, high-density lipoprotein cholesterol; LDL.C, low-density lipoprotein cholesterol; LPa, lipoprotein (a); apoA1, apolipoprotein A-1; apoB, apolipoprotein B; Ca, calcium; K, potassium; Na, sodium; Cl, chloride; LDH, lactate dehydrogenase; ALP, alkaline phosphatase; ChE, cholinesterase; HCY, homocysteine; Cys.C, cystatin C; CO₂, carbon dioxide; AG, anion gap; Glu, glucose; TT, thrombin time; PT, prothrombin time; PTR, prothrombin ratio; APTT, activated partial thromboplastin time; APTTR, activated partial thromboplastin time ratio; PT:INR, prothrombin international normalized ratio; Fbg, fibrinogen; CRP, C-reactive protein; NRBC, nucleated red blood cells; α -HBDH, alpha-hydroxybutyrate dehydrogenase; AFU, alpha-L-fucosidase; RET, reticulocyte count; IRF, immature reticulocyte fraction; PCT, procalcitonin; CHD, coronary heart disease; Urine OB, urine occult blood; Urine Glu, urine glucose; Urine TP, urine protein; Urine URO, urine urobilinogen; Urine BIL, urine bilirubin; Urine KET, urine ketones.

To minimize potential confounding factors, PSM was performed on variables such as age, weight, marital status, hypertension, CHD, diabetes, smoking, alcohol, infection history, hepatitis, family cancer history, family history, gender, and ABO blood, achieving a balanced distribution of these selected characteristics between the groups. [Supplementary Table 1](#) compares the baseline characteristics before and after matching. Post-PSM, no statistically significant differences were observed between the groups for the matched variables ($P > 0.05$), indicating that the matching process effectively reduced confounding bias.

The detailed demographic and baseline clinical characteristics of the PSM-adjusted training cohort are summarized in [Table 2](#), with 917 cases each in the sepsis and non-sepsis groups. The analysis revealed significant differences in various clinical parameters between the two groups, particularly in hospitalization days, pulse rate, heart rate, respiratory rate, diastolic pressure, systolic pressure, and several hematological indices. Patients with sepsis experienced longer hospital stays and exhibited elevated pulse rates, heart rates, respiratory rates, white blood cell counts, and lower hemoglobin levels compared to the non-sepsis group. Additionally, liver function tests, including ALT, AST, and GGT, demonstrated significant disparities between the groups. Biochemical markers also indicated lower albumin levels and higher creatinine levels in the sepsis cohort. Furthermore, lipid profiles showed lower HDL cholesterol and higher LDL cholesterol in the

Table 2 Patient Demographics and Baseline Characteristics in the Training Cohort

Characteristic	Overall, N = 1,834	Non-Sepsis, N = 917	Sepsis, N = 917	p-value
Age	26 (7, 54)	26 (7, 55)	26 (8, 53)	0.726
Hospital Days	16 (7, 26)	11 (5, 21)	22 (10, 30)	<0.001
Weight	53 (26, 59)	54 (29, 59)	52 (23, 59)	0.277
Temperature	36.80 (36.50, 37.20)	36.79 (36.50, 37.10)	36.80 (36.50, 37.50)	0.168
Pulse Rate	94 (84, 110)	91 (84, 106)	100 (84, 114)	<0.001
HR	92 (84, 108)	90 (84, 102)	95 (84, 110)	<0.001
RR	20.0 (20.0, 23.0)	20.0 (19.9, 22.0)	20.0 (20.0, 24.0)	<0.001
DBP	67 (62, 72)	68 (64, 72)	66 (60, 72)	<0.001
SBP	118 (108, 124)	119 (112, 125)	116 (105, 123)	<0.001
WBC	12 (4, 30)	16 (6, 35)	10 (3, 25)	<0.001
RBC	2.77 (2.14, 3.53)	2.87 (2.20, 3.79)	2.66 (2.10, 3.34)	<0.001
MCV	89 (83, 96)	89 (82, 96)	89 (83, 97)	0.201
MCHC	330 (318, 341)	329 (316, 340)	331 (320, 342)	0.005
MCH	29.6 (27.3, 31.8)	29.3 (26.9, 31.6)	29.8 (27.6, 31.8)	0.021
RDW	16.5 (14.6, 19.3)	16.3 (14.5, 19.2)	16.7 (14.7, 19.6)	0.074
Lymph	2 (1, 8)	3 (1, 10)	2 (1, 5)	<0.001
Mono	1 (0, 3)	1 (0, 4)	1 (0, 2)	<0.001
Neut	3 (1, 10)	4 (1, 16)	2 (1, 7)	<0.001
Hct	25 (19, 32)	26 (19, 34)	24 (19, 30)	<0.001
Eos	0.01 (0.00, 0.10)	0.04 (0.00, 0.18)	0.01 (0.00, 0.08)	<0.001
Baso	0.01 (0.00, 0.05)	0.02 (0.00, 0.09)	0.01 (0.00, 0.04)	<0.001
Hb	74 (51, 99)	77 (52, 104)	72 (50, 94)	<0.001
PLT	75 (27, 214)	78 (27, 225)	73 (27, 202)	0.247
MPV	9.20 (8.00, 10.20)	9.10 (8.00, 10.20)	9.20 (8.00, 10.20)	0.575
PDW	16.40 (11.90, 17.70)	16.40 (12.20, 17.70)	16.30 (11.60, 17.80)	0.819
PCTI	0.07 (0.03, 0.18)	0.07 (0.03, 0.18)	0.06 (0.03, 0.17)	0.098
ALT	22 (13, 37)	21 (12, 34)	23 (13, 41)	0.001
AST	26 (18, 42)	27 (19, 41)	25 (17, 44)	0.047
GGT	32 (17, 58)	31 (16, 55)	33 (18, 64)	0.016
TBIL	10 (7, 13)	10 (8, 12)	10 (7, 13)	0.721
DBIL	4.2 (3.0, 5.8)	4.3 (3.1, 5.5)	4.2 (2.9, 6.4)	0.462
IBIL	5.2 (3.4, 7.5)	5.5 (3.7, 7.5)	4.7 (3.1, 7.6)	<0.001
ALB	38 (33, 41)	39 (35, 42)	37 (32, 40)	<0.001
GLO	25.9 (22.9, 28.4)	25.8 (23.8, 28.1)	26.0 (22.0, 28.9)	0.441
TP	64 (58, 68)	64 (60, 69)	62 (57, 68)	<0.001
A.G	1.48 (1.20, 1.75)	1.50 (1.29, 1.76)	1.45 (1.16, 1.73)	0.005
PA	153 (100, 200)	158 (110, 204)	146 (89, 196)	<0.001
TBA	7 (4, 12)	7 (4, 11)	7 (4, 14)	0.384
CREA	59 (38, 78)	63 (40, 78)	56 (35, 77)	0.006
UREA	4.50 (3.52, 5.30)	4.52 (3.68, 5.15)	4.50 (3.35, 5.50)	0.955
UA	307 (251, 363)	314 (267, 364)	297 (229, 363)	<0.001
TC	3.43 (2.77, 4.13)	3.48 (2.85, 4.19)	3.34 (2.69, 4.07)	0.004
TG	1.38 (0.94, 1.93)	1.44 (0.97, 1.94)	1.33 (0.91, 1.92)	0.024
HDL.C	0.82 (0.58, 1.02)	0.85 (0.63, 1.02)	0.78 (0.54, 1.03)	0.002
LDL.C	1.88 (1.40, 2.43)	1.93 (1.47, 2.51)	1.84 (1.33, 2.38)	0.017
LPa	106 (59, 176)	111 (67, 181)	101 (53, 171)	<0.001
apoAI	0.86 (0.69, 1.04)	0.90 (0.75, 1.05)	0.82 (0.64, 1.03)	<0.001
apoB	0.70 (0.57, 0.86)	0.71 (0.58, 0.87)	0.69 (0.55, 0.85)	0.014
Ca	2.17 (2.06, 2.28)	2.19 (2.09, 2.30)	2.15 (2.02, 2.27)	<0.001
K.	3.89 (3.57, 4.24)	3.90 (3.63, 4.24)	3.87 (3.53, 4.24)	0.132
Na.	139.3 (136.6, 141.1)	139.7 (137.5, 141.1)	138.9 (136.0, 141.0)	<0.001

(Continued)

Table 2 (Continued).

Characteristic	Overall, N = 1,834	Non-Sepsis, N = 917	Sepsis, N = 917	p-value
Cl.	103.0 (100.4, 105.3)	103.1 (101.2, 105.4)	102.8 (99.8, 105.2)	0.009
LDH	351 (219, 610)	402 (252, 642)	292 (196, 561)	<0.001
ALP	89 (66, 142)	85 (67, 138)	94 (66, 145)	0.195
ChE	5.84 (4.26, 7.30)	6.23 (4.76, 7.50)	5.34 (3.89, 6.84)	<0.001
HCY	10 (7, 14)	10 (8, 14)	9 (6, 13)	<0.001
Cys.C	0.84 (0.67, 1.05)	0.84 (0.69, 1.07)	0.82 (0.65, 1.05)	0.040
CO2	22.0 (19.8, 23.7)	22.3 (19.9, 23.7)	21.8 (19.6, 23.8)	0.150
AG	13.9 (11.9, 16.2)	13.9 (12.2, 16.1)	13.9 (11.7, 16.3)	0.517
Glu	5.64 (4.72, 6.63)	5.71 (4.78, 6.66)	5.55 (4.68, 6.58)	0.173
TT	17.10 (15.80, 18.50)	17.40 (16.30, 18.62)	16.80 (15.50, 18.30)	<0.001
PT	13.10 (12.22, 14.70)	13.00 (12.20, 14.50)	13.30 (12.30, 14.90)	0.021
PTR	1.08 (1.01, 1.21)	1.08 (1.01, 1.19)	1.10 (1.02, 1.22)	0.019
APTT	35 (31, 40)	35 (32, 39)	36 (31, 40)	0.020
APTTTR	1.12 (1.03, 1.24)	1.12 (1.03, 1.21)	1.13 (1.02, 1.27)	0.017
PT.INR	1.07 (1.00, 1.20)	1.06 (1.00, 1.18)	1.08 (1.00, 1.22)	0.086
Fbg	3.30 (2.58, 4.04)	3.19 (2.56, 3.82)	3.42 (2.61, 4.25)	<0.001
CRP	29 (12, 59)	27 (12, 41)	33 (11, 94)	<0.001
NRBC	0.00 (0.00, 0.02)	0.00 (0.00, 0.02)	0.00 (0.00, 0.02)	0.680
α.HBDH	296 (203, 464)	316 (224, 470)	281 (183, 449)	<0.001
AFU	18 (14, 22)	18 (15, 22)	17 (14, 23)	0.175
RET	0.05 (0.03, 0.08)	0.05 (0.04, 0.08)	0.05 (0.03, 0.08)	0.001
RET.	4 (1, 13)	6 (2, 13)	3 (1, 12)	<0.001
IRF	5 (0, 15)	4 (0, 14)	6 (0, 16)	0.969
PCT	0.28 (0.11, 0.72)	0.28 (0.10, 0.59)	0.29 (0.12, 1.09)	<0.001
Gender				0.848
Male	1,116 (60.9%)	556 (60.6%)	560 (61.1%)	
Female	718 (39.1%)	361 (39.4%)	357 (38.9%)	
ABO				0.947
A	478 (26.1%)	236 (25.7%)	242 (26.4%)	
AB	111 (6.1%)	58 (6.3%)	53 (5.8%)	
B	502 (27.4%)	249 (27.2%)	253 (27.6%)	
O	743 (40.5%)	374 (40.8%)	369 (40.2%)	
Death				<0.001
No	1,669 (91.0%)	873 (95.2%)	796 (86.8%)	
Yes	165 (9.0%)	44 (4.8%)	121 (13.2%)	
Marital Status				0.852
Single, divorced, and windowed	974 (53.1%)	485 (52.9%)	489 (53.3%)	
Married	860 (46.9%)	432 (47.1%)	428 (46.7%)	
Hypertension				0.471
No	1,762 (96.1%)	884 (96.4%)	878 (95.7%)	
Yes	72 (3.9%)	33 (3.6%)	39 (4.3%)	
CHD				0.826
No	1,813 (98.9%)	907 (98.9%)	906 (98.8%)	
Yes	21 (1.1%)	10 (1.1%)	11 (1.2%)	
Diabetes				0.900
No	1,769 (96.5%)	885 (96.5%)	884 (96.4%)	
Yes	65 (3.5%)	32 (3.5%)	33 (3.6%)	
Smoking				0.316
No	1,768 (96.4%)	888 (96.8%)	880 (96.0%)	
Yes	66 (3.6%)	29 (3.2%)	37 (4.0%)	

(Continued)

Table 2 (Continued).

Characteristic	Overall, N = 1,834	Non-Sepsis, N = 917	Sepsis, N = 917	p-value
Alcohol				0.512
No	1,796 (97.9%)	900 (98.1%)	896 (97.7%)	
Yes	38 (2.1%)	17 (1.9%)	21 (2.3%)	
Infection History				0.847
No	1,720 (93.8%)	859 (93.7%)	861 (93.9%)	
Yes	114 (6.2%)	58 (6.3%)	56 (6.1%)	
Hepatitis				0.810
No	1,762 (96.1%)	882 (96.2%)	880 (96.0%)	
Yes	72 (3.9%)	35 (3.8%)	37 (4.0%)	
Family Cancer History				0.223
No	1,817 (99.1%)	911 (99.3%)	906 (98.8%)	
Yes	17 (0.9%)	6 (0.7%)	11 (1.2%)	
Family History				0.295
No	1,757 (95.8%)	883 (96.3%)	874 (95.3%)	
Yes	77 (4.2%)	34 (3.7%)	43 (4.7%)	
Fecal OB				0.079
Negative	1,437 (78.4%)	703 (76.7%)	734 (80.0%)	
Positive	397 (21.6%)	214 (23.3%)	183 (20.0%)	
Urine TP				0.131
Negative	1,499 (81.7%)	762 (83.1%)	737 (80.4%)	
Positive	335 (18.3%)	155 (16.9%)	180 (19.6%)	
Urine URO				<0.001
Negative	1,371 (74.8%)	637 (69.5%)	734 (80.0%)	
Positive	463 (25.2%)	280 (30.5%)	183 (20.0%)	
Urine BIL				<0.001
Negative	1,399 (76.3%)	664 (72.4%)	735 (80.2%)	
Positive	435 (23.7%)	253 (27.6%)	182 (19.8%)	
Urine KET				0.015
Negative	1,565 (85.3%)	801 (87.4%)	764 (83.3%)	
Positive	269 (14.7%)	116 (12.6%)	153 (16.7%)	
Urine Glu				0.522
Negative	1,662 (90.6%)	835 (91.1%)	827 (90.2%)	
Positive	172 (9.4%)	82 (8.9%)	90 (9.8%)	
Urine OB				0.012
Negative	1,413 (77.0%)	729 (79.5%)	684 (74.6%)	
Positive	421 (23.0%)	188 (20.5%)	233 (25.4%)	

Abbreviations: HR, heart rate; RR, respiratory rate; DBP, diastolic blood pressure; SBP, systolic blood pressure; WBC, white blood cell count; RBC, red blood cell count; MCV, mean corpuscular volume; MCHC, mean corpuscular hemoglobin concentration; MCH, mean corpuscular hemoglobin; RDW, red cell distribution width; Lymph, lymphocyte count; Mono, monocyte count; Neut, neutrophil count; Hct, hematocrit; Eos, eosinophil count; Baso, basophil count; Hb, hemoglobin; PLT, platelet count; MPV, mean platelet volume; PDW, platelet distribution width; PCT1, plateletcrit; ALT, alanine transaminase; AST, aspartate transaminase; GGT, gamma-glutamyl transferase; TBIL, total bilirubin; DBIL, direct bilirubin; IBIL, indirect bilirubin; ALB, albumin; GLO, globulin; TP, total protein; A/G, albumin-to-globulin ratio; PA, prealbumin; TBA, total bile acid; CREA, creatinine; UREA, urea; UA, uric acid; TC, total cholesterol; TG, triglycerides; HDL.C, high-density lipoprotein cholesterol; LDL.C, low-density lipoprotein cholesterol; LpA, lipoprotein (a); apoAI, apolipoprotein A-I; apoB, apolipoprotein B; Ca, calcium; K, potassium; Na, sodium; Cl, chloride; LDH, lactate dehydrogenase; ALP, alkaline phosphatase; ChE, cholinesterase; HCY, homocysteine; Cys.C, cystatin C; CO₂, carbon dioxide; AG, anion gap; Glu, glucose; TT, thrombin time; PT, prothrombin time; PTR, prothrombin ratio; APTT, activated partial thromboplastin time; APTTR, activated partial thromboplastin time ratio; PT:INR, prothrombin international normalized ratio; Fbg, fibrinogen; CRP, C-reactive protein; NRBC, nucleated red blood cells; α -HBDH, alpha-hydroxybutyrate dehydrogenase; AFU, alpha-L-fucosidase; RET, reticulocyte count; IRF, immature reticulocyte fraction; PCT, procalcitonin; CHD, coronary heart disease; Urine OB, urine occult blood; Urine Glu, urine glucose; Urine TP, urine protein; Urine URO, urine urobilinogen; Urine BIL, urine bilirubin; Urine KET, urine ketones.

sepsis group. Overall, the sepsis group displayed more severe pathological conditions across most physiological and laboratory characteristics. These findings suggest distinct clinical profiles between sepsis and non-sepsis patients, which may facilitate early identification and management of individuals with sepsis.

Identifying Key Predictors of Sepsis Risk in Leukemia Patients

To identify key variables associated with sepsis occurrence in leukemia patients, univariate logistic analysis, Boruta algorithm, and LASSO regression analysis were employed for feature selection.

Univariate logistic regression analysis revealed several variables significantly associated with sepsis risk, including urine indicators such as Urine URO ($p < 0.001$), Urine BIL ($p < 0.001$), and Urine OB ($p = 0.013$). Physiological metrics like temperature ($p < 0.001$), pulse rate ($p < 0.001$), diastolic pressure (DBP, $p < 0.001$), and systolic pressure (SBP, $p < 0.001$) were also significant. Hematological indices, including white blood cell count (WBC, $p < 0.001$), red blood cell count (RBC, $p < 0.001$), hemoglobin (Hb, $p < 0.001$), and C-reactive protein (CRP, $p < 0.001$), showed notable differences. Additionally, mean platelet volume (MPV, $p = 0.029$) and the activated partial thromboplastin time ratio (APTTR, $p = 0.020$) were significant as well. These results may provide crucial clinical insights for the early identification and management of sepsis patients ([Supplementary Table 2](#)). [Figure 2A](#) illustrates the Z-scores of each variable, demonstrating how this analysis enhanced model optimization by concentrating on the most relevant features. A total of 45 variables were identified as important, including vital signs such as Temp, PR, HR, RR, DBP, and SBP; hematological indices including WBC, RBC, Lymph, Mono, Neut, Hct, Eos, Baso, and Hb; liver function tests like ALT and AST; biochemical markers such as ALB and TP; lipid profiles including HDL.C and LpA; and coagulation factors like Fbg and CRP. This grouping underscores the multifaceted nature of sepsis risk assessment, indicating their strong explanatory power in predicting sepsis occurrence. In the LASSO regression analysis, the optimal tuning parameter λ , determined through 10-fold cross-validation, was 0.032, based on the one standard error of the minimum criteria. At this λ value, the following variables were identified as risk factors for sepsis: Urine URO, Temperature, Pulse Rate, SBP, RBC, Lymph, Neut, TP, TBA, ChE, TT, CRP, and PCT ([Figure 2B](#) and [C](#)).

Following the identification of 13 common biomarkers through the three selection methods ([Figure 2D](#)), we incorporated these variables into a multivariate logistic regression analysis to determine the most reliable predictors of sepsis risk. The analysis indicated that SBP, RBC, Lymph, Neut, TBA, TT, CRP, and PCT emerged as significant independent risk factors for sepsis in leukemia patients ([Table 3](#)). To ensure the robustness of our model, we conducted a multicollinearity analysis, which yielded an average VIF of 1.038. This result suggests that there is no significant multicollinearity among the selected variables, thereby reinforcing the reliability of our findings ([Supplementary Table 3](#)). These variables were incorporated as primary predictors in the subsequent model construction and analysis, providing a solid foundation for developing predictive models.

The RF Model Outperforms Other Machine Learning Models in Predicting Sepsis in Leukemia Patients

The performance of seven machine learning models in predicting sepsis in leukemia patients was compared, with their ROC curves and DCA results presented ([Figure 3](#)). In the training cohort, the SVM model exhibited excellent predictive performance, achieving an AUC of 0.972, indicating high accuracy. In contrast, the AUC values for the other models were as follows: RF at 0.765, XGBoost at 0.762, NNET at 0.727, LOGI at 0.722, and DCTree at 0.685, while the KNN model performed the poorest with an AUC of 0.665 ([Figure 3A](#)). DCA assessed the clinical utility of the models, further confirming that the RF model provided the highest net benefit across most threshold ranges, particularly within the intermediate range, highlighting its superiority in predicting sepsis ([Figure 3B](#)). The RF model exhibited the best overall performance, with the highest AUC and net gain, indicating that it was superior in predictive power for sepsis, hence the choice of the RF model as the best model. To further evaluate the model's performance, results from the validation cohort are depicted in [Figure 3C](#). The RF model achieved an AUC of 0.700, demonstrating its ability to generalize effectively to new data. Similarly, the net benefit of the RF model in the validation cohort, shown in [Figure 3D](#), reflects trends akin to those observed in the training group, reaffirming its robust predictive capability for sepsis among leukemia patients across both cohorts.

Further detailed performance metrics for each model, including sensitivity, specificity, recall, accuracy, and F1 score ([Table 4](#)). The RF model demonstrated the highest performance, achieving an accuracy of 0.710, a sensitivity of 0.727, a specificity of 0.694 and a F1 score of 0.715, indicating its strong ability to differentiate between sepsis and non-sepsis

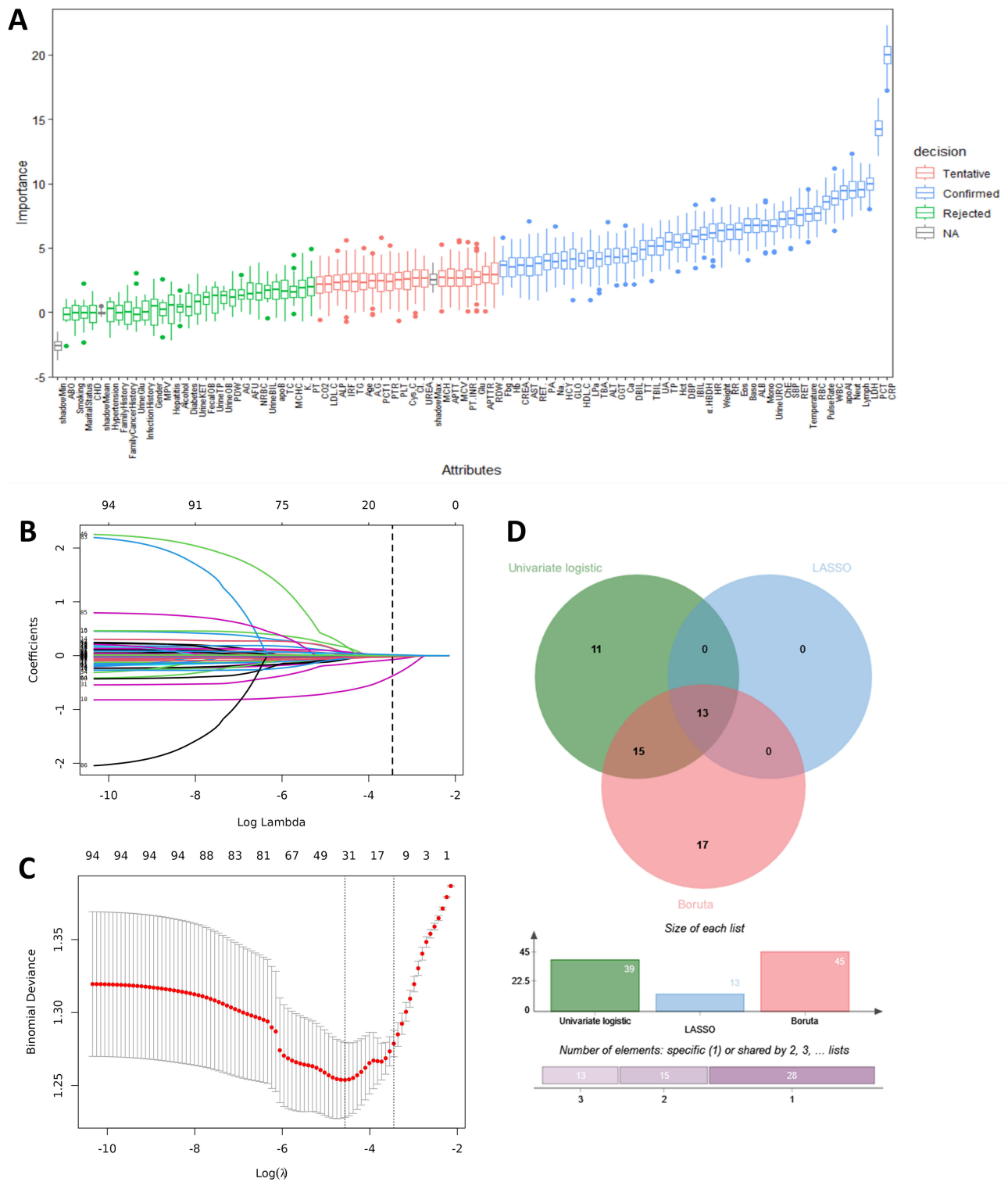


Figure 2 Feature selection process for variables included in the prediction model. **(A)** Variable importance based on the Boruta algorithm, where attributes are classified as “Tentative” (green), “Confirmed” (blue), “Rejected” (red), and “NA” (grey). **(B)** LASSO regression path showing the coefficients of variables across different values of the regularization parameter (λ). **(C)** Cross-validation error plot for selecting the optimal λ in LASSO. The vertical dashed line represents the optimal λ where the minimal cross-validation error is achieved. **(D)** Venn diagram comparing variables selected by three different methods: univariate logistic regression (green), LASSO (blue), and the Boruta algorithm (red), showing the overlap of selected variables.

Table 3 Multivariate Logistic Regression Analysis for Sepsis in the Training Cohort

Characteristic	OR	95% CI	p-value
Urine URO			
Negative	—	—	
Positive	0.43	0.34, 0.54	0.053
Temperature	1.07	0.93, 1.22	0.345
Pulse Rate	1.00	1.00, 1.01	0.155
SBP	0.99	0.98, 1.00	0.004
RBC	0.85	0.76, 0.94	0.002
Lymph	1.00	0.99, 1.00	0.003
Neut	0.99	0.99, 1.00	<0.001
TP	0.99	0.97, 1.00	0.057
TBA	1.01	1.00, 1.01	0.008
ChE	0.97	0.92, 1.02	0.246
TT	0.94	0.91, 0.97	<0.001
CRP	1.01	1.01, 1.01	<0.001
PCT	1.08	1.04, 1.13	<0.001

Abbreviations: OR, odds ratio; CI, confidence interval; systolic blood pressure (SBP); red blood cells (RBC), lymphocytes (Lymph); neutrophils (Neut); total protein (TP); total bile acid (TBA); cholinesterase (ChE); thrombin time (TT); C-reactive protein (CRP); procalcitonin (PCT).

patients. The XGBoost model maintained balanced metrics, with an accuracy of 0.689. The SVM model followed closely, achieving an accuracy of 0.686, with a sensitivity of 0.710 but a lower specificity of 0.661. The NNET model showed moderate performance, with an accuracy of 0.675 and a sensitivity of 0.623, while the KNN model had the lowest overall performance, with an accuracy of 0.593 and a sensitivity of 0.541. The Logistic Regression model exhibited an accuracy of 0.658, demonstrating balanced sensitivity and specificity. The DCTree model reached a specificity of 0.792 but had a low sensitivity of 0.508, leading to an overall accuracy of 0.650. In the validation cohort, the performance metrics exhibited similar trends, further validating the models' effectiveness. The RF model achieved an accuracy of 0.758, with a sensitivity of 0.655 and a specificity of 0.714, affirming its reliability in distinguishing sepsis cases ([Supplementary Table 4](#)). We also performed sensitivity analyses to assess the impact of missing data processing on the model, confirming the robustness of the chosen imputation method ([Supplementary Table 5](#)). The analysis showed that the results produced by no imputation were consistent with those imputed using the MI method.

Overall, the RF model consistently outperformed other machine learning models in both predictive accuracy and clinical utility, making it the most reliable tool for predicting sepsis in leukemia patients across multiple datasets.

SHAP Analysis Quantifies Feature Contributions to Sepsis Risk in Leukemia Patients

The SHAP analysis was employed to interpret the predictions of the best-performing model the RF. SHAP analysis highlighted the impact of 8 key features on the sepsis prediction model in leukemia patients, ranked by SHAP values ([Figure 4A](#)). [Figure 4B](#) provides a detailed visualization of the multidimensional influence of each feature, with the SHAP values clearly showing how both high and low levels of these variables contribute to the overall prediction model. CRP demonstrates the strongest predictive power, with elevated levels significantly increasing the risk of sepsis. PCT follows closely, contributing substantially to the model, indicating that higher levels of PCT are also associated with an increased risk. Immune-related markers, such as Neut and Lymph, further play important roles in the prediction, where lower neutrophil and lymphocyte counts are linked to a greater sepsis risk. TT, RBC, and TBA also have notable impacts, with abnormal values correlating with heightened risk. Lastly, SBP adds to the prediction, where lower blood pressure suggests a higher likelihood of sepsis, possibly reflecting cardiovascular instability. This analysis highlights the critical roles of inflammatory, immune, and coagulation markers in predicting sepsis risk.

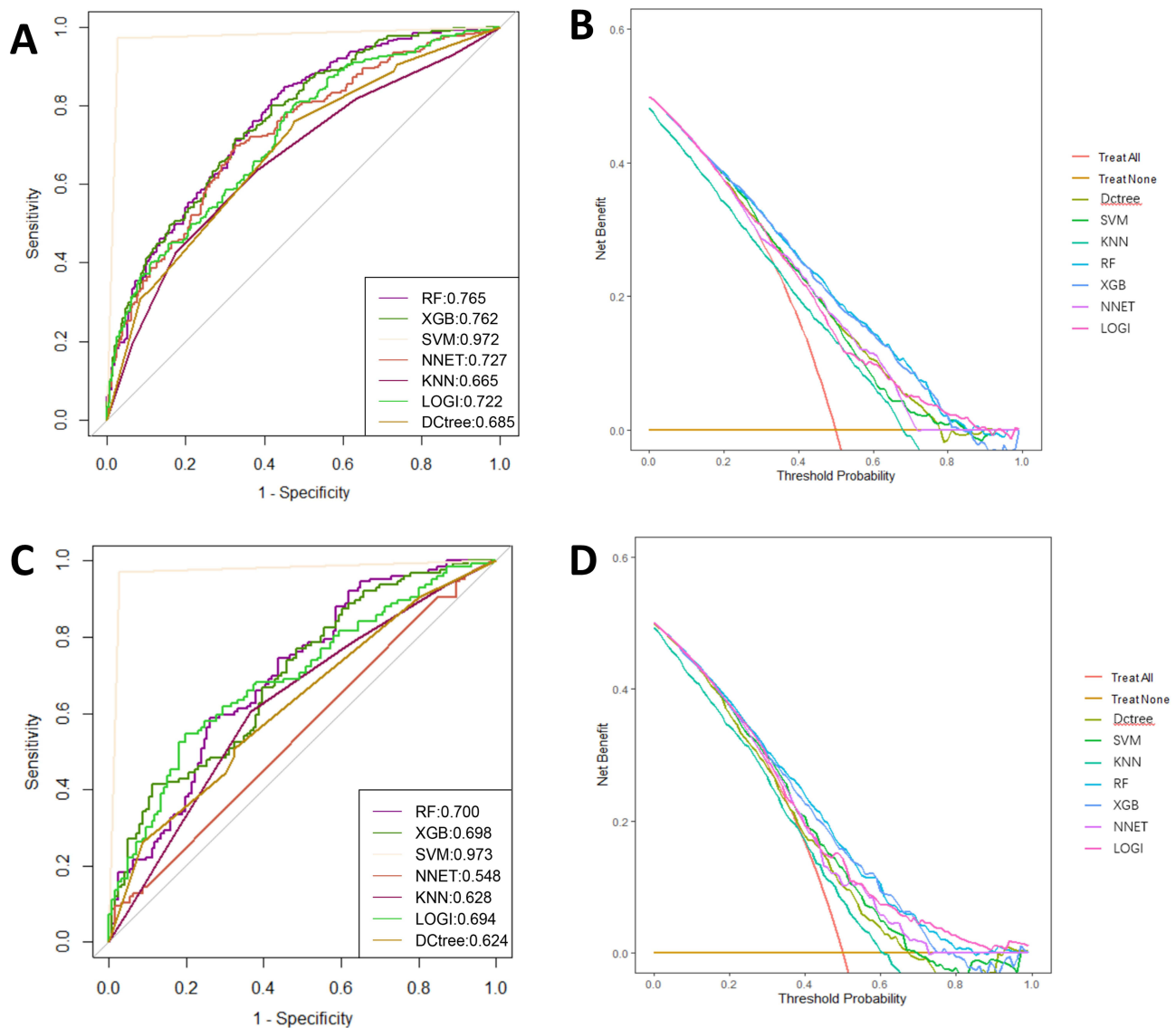


Figure 3 Predictive performance of the model in training and validation cohorts. **(A and C)** Receiver Operating Characteristic (ROC) curves for the training **(A)** and validation **(C)** cohorts. The curves illustrate the discriminatory ability of different predictive models, with the area under the ROC curve (AUC) values displayed for each model: Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), Neural Network (NNET), K-Nearest Neighbors (KNN), Logistic Regression (LOGI), and Decision Tree (DCTree). **(B and D)** Decision Curve Analysis (DCA) for the training **(B)** and validation **(D)** cohorts. The curves show the net benefit of each model across various threshold probabilities, comparing them with the “Treat All” and “Treat None” strategies, indicating their potential clinical utility.

Nonlinear Relationship Between Model Factors and Sepsis Risk in Leukemia Patients and Interactions Between Model Factors

The RCS analysis, as shown in [Figure 5A–H](#), visualizes the nonlinear relationships between sepsis risk and various continuous clinical variables in leukemia patients. In the analysis of sepsis risk factors, we utilized standardized odds ratios to compare the relative impact of various variables on sepsis risk ([Supplementary Table 6](#)). The results show that CRP levels above 24.068 mg/L significantly increase the risk of sepsis, with a standardized OR of 3.22 (95% CI: 2.58, 4.03, $p < 0.001$), while lower CRP levels (<24.068 mg/L) are associated with a decreased sepsis risk (OR=0.85, 95% CI: 0.74, 0.98, $p = 0.029$). Additionally, lower of Lymph and Neut (Lymph <5, Neut <5) are associated with a reduced sepsis risk, with standardized ORs of 0.68 (95% CI: 0.61, 0.77, $p < 0.001$) and 0.76 (95% CI: 0.67, 0.86, $p < 0.001$), respectively. Elevated PCT levels (≥ 0.324 $\mu\text{g/L}$) are also significantly associated with a higher risk of sepsis (OR=4.91, 95% CI: 2.77, 8.71, $p < 0.001$). Similarly, higher TBA levels (≥ 6.012 $\mu\text{mol/L}$) increase the risk of sepsis (OR=1.37, 95% CI: 1.18, 1.59, $p < 0.001$), while lower TT values (<17 seconds) reduce the risk (OR=0.83, 95% CI: 0.71, 0.96, $p = 0.015$). Furthermore,

Table 4 Comparison of Training Cohort Results of the Machine Learning Models

Model	Accuracy	Sensitivity	Specificity	Recall	F1 score
RF	0.710	0.727	0.694	0.727	0.715
XGB	0.689	0.689	0.689	0.689	0.689
SVM	0.686	0.710	0.661	0.710	0.693
NNET	0.675	0.623	0.727	0.623	0.657
KNN	0.593	0.541	0.645	0.541	0.571
LOGI	0.658	0.639	0.678	0.639	0.652
DCtree	0.650	0.508	0.792	0.508	0.592

Abbreviations: RF, Random Forest; XGB, eXtreme Gradient Boosting; SVM, Support Vector Machine; NNET, Neural Network; KNN, K-Nearest Neighbors; LOGI, Logistic Regression; DCtree, Decision Tree.

RBC counts above $2.48 \times 10^6/\mu\text{L}$ are associated with a significantly decreased risk of sepsis (OR=0.68, 95% CI: 0.60, 0.77, $p < 0.001$). Lastly, lower SBP (< 128.63 mmHg) is associated with a decreased risk of sepsis (OR=0.68, 95% CI: 0.61, 0.75, $p < 0.001$).

In addition to the nonlinear associations, pairwise correlation analysis among the eight variables included in the final model revealed key insights into their interactions ([Supplementary Figure 1](#)). CRP and PCT show a significant positive correlation ($r = 0.231$, $p < 0.001$), as do Neut and Lymph ($r = 0.153$, $p < 0.001$). TT exhibits a moderate positive correlation with TBA ($r = 0.159$, $p < 0.001$) and weak but significant positive correlations with Neut ($r = 0.100$, $p < 0.001$), Lymph ($r = 0.092$, $p < 0.001$), and RBC ($r = 0.104$, $p < 0.001$). TBA showed positive correlations with Lymph ($r = 0.057$, $p < 0.05$). SBP exhibits a negative correlation with PCT ($r = -0.066$, $p < 0.01$) and TBA ($r = -0.053$, $p < 0.05$), while showing weak but significant positive correlations with Neut ($r = 0.048$, $p < 0.05$) and RBC ($r = 0.124$, $p < 0.001$). RBC and CRP exhibit a weak negative correlation ($r = -0.077$, $p = 0.001$).

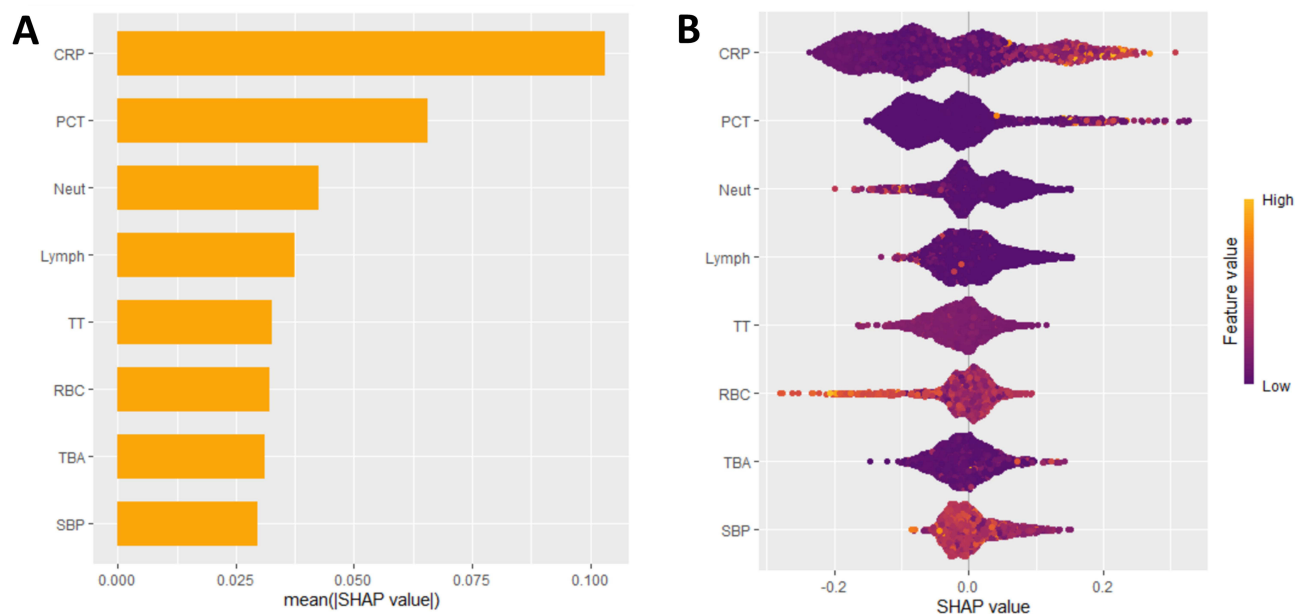


Figure 4 SHAP (SHapley Additive exPlanations) analysis of the model. **(A)** Bar plot showing the mean SHAP values for the top features ranked by their contribution to sepsis prediction in leukemia patients. Features include C-reactive protein (CRP), procalcitonin (PCT), neutrophil count (Neut), lymphocyte count (Lymph), thrombin time (TT), red blood cell count (RBC), total bile acid (TBA), and systolic blood pressure (SBP). Higher mean SHAP values indicate greater importance of the feature in the model. **(B)** SHAP summary plot visualizing the distribution of SHAP values for each feature. Each dot represents an individual data point, with the x-axis showing the SHAP value (feature's impact on the model output) and the color representing the feature value (yellow for higher values, purple for lower values).

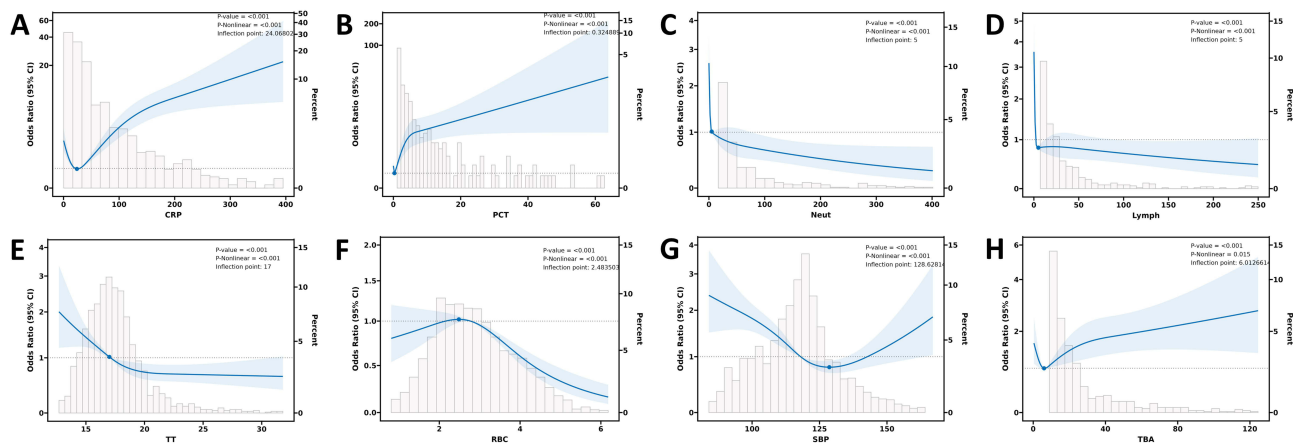


Figure 5 Restricted cubic spline (RCS) analysis of the continuous variables included in the model. (A–H) The RCS curves illustrate the nonlinear relationships between the risk of sepsis in leukemia patients and the following variables: C-reactive protein (CRP), Procalcitonin (PCT), Neutrophil count (Neut), Lymphocyte count (Lymph), Thrombin time (TT), Red blood cell count (RBC), Total bile acid (TBA), and Systolic blood pressure (SBP).

Discussion

In this study, we compared multiple machine learning models to evaluate their performance in predicting sepsis in leukemia patients. Based on a comprehensive analysis of multiple metrics, including ROC, DCA, accuracy, sensitivity, specificity, recall, and F1 score, the RF model was selected for its superior overall performance and its ability to effectively balance predictive accuracy and clinical utility.¹² These metrics underscore RF's ability to balance sensitivity and specificity, minimizing missed sepsis cases while maintaining predictive accuracy. Key features, including CPR, PCT, Neut, Lymph, TT, RBC, TBA, and SBP, were identified as significant predictors of sepsis through SHAP analysis. These findings highlight the importance of inflammatory, immune, and coagulation markers in predicting sepsis risk in leukemia patients. The nonlinear relationships between these variables and sepsis risk, as revealed by RCS analysis, provide valuable insights into how changes in these parameters affect patient outcomes.

Previous studies on sepsis diagnosis have identified some biomarkers, including CPR, PCT, RBC, neutrophil-to-lymphocyte ratio (NLR), and prothrombin time (PT).^{13,14} PCT and CRP are among the most widely used, showing good sensitivity for sepsis screening and monitoring, making them valuable tools for early diagnosis.^{13,15} PCT, while more specific for bacterial infections, can also be influenced by other factors, reducing its utility in some contexts. Similarly, while CRP is a common marker of systemic inflammation, its elevation in various pathological states limits its effectiveness as a sepsis biomarker.^{16,17} These challenges are particularly pronounced in leukemia patients, where severe immunosuppression can blunt inflammatory responses, potentially masking sepsis-related elevations in these biomarkers.^{18,19} This does not negate their relevance in sepsis, as these markers are commonly associated with disseminated intravascular coagulation (DIC) and other coagulopathies in septic patients.²⁰ The significant positive correlation between CRP and PCT ($r = 0.231$, $p < 0.001$) aligns with their roles as systemic inflammatory markers. Both are widely used in sepsis detection, and their correlation underscores their complementary contribution to the inflammatory response in septic patients.²¹ Interestingly, the weak negative correlation between RBC and CRP ($r = -0.077$, $p = 0.001$) suggests a potential inverse relationship between systemic inflammation and oxygen-carrying capacity. This finding may reflect the anemia of inflammation commonly observed in sepsis patients, where elevated inflammatory markers are associated with reduced RBC counts.^{22,23}

For inflammatory and immune-related predictors, our final model includes Neut and Lymph counts, both of which provide valuable insights into the systemic inflammatory response and immune status of leukemia patients. NLR, while conceptually relevant, was not explicitly included as a separate predictor because its components (Neut and Lymph) are directly incorporated in the model.²⁴ This approach ensures that the underlying inflammatory dynamics are captured while avoiding redundancy. NLR is a marker of acute inflammation, especially useful for predicting bacterial infections leading to sepsis.^{25,26} However, as sepsis progresses, most patients quickly exhibit signs of severe immune suppression.²⁷ In leukemia patients, who experience secondary bone marrow suppression and severely reduced leukocyte production,²⁸

NLR may not accurately reflect their true immune status, limiting its reliability in this population. Similarly, the positive correlation between Neut and Lymph ($r = 0.153$, $p < 0.001$) highlights the interplay between different components of the immune system, although the suppressed immune response often observed in leukemia may moderate this relationship.²⁹

The protective effects of higher RBC counts and SBP further emphasize the role of cardiovascular stability and oxygen transport in preventing sepsis.^{30,31} SBP also showed weak but significant positive correlations with Neut ($r = 0.048$, $p < 0.05$) and RBC ($r = 0.124$, $p < 0.001$), highlighting the interplay between immune response and cardiovascular function.³² Conversely, the negative correlations observed between SBP and PCT ($r = -0.066$, $p < 0.01$) and SBP and TBA ($r = -0.053$, $p < 0.05$) reflect the hemodynamic instability and liver impairment characteristic of sepsis progression.³³ Elevated TT indicates impaired clotting efficiency, which is common in sepsis due to DIC.³⁴ This prolonged TT may reflect the consumption of clotting factors and fibrinolysis, contributing to the higher risk of sepsis-related complications in these patients.²⁰ Higher TBA levels may reflect liver dysfunction, a common complication in sepsis, where impaired bile acid metabolism leads to accumulation in the bloodstream. This can further contribute to systemic inflammation and immune dysregulation, exacerbating the severity of sepsis.^{35,36} TBA's positive correlation with Lymph ($r = 0.057$, $p < 0.05$) reinforces its role as a marker of liver dysfunction and its downstream effects on immune regulation. TT exhibited a moderate positive correlation with TBA ($r = 0.159$, $p < 0.001$), suggesting a potential link between coagulation dysfunction and liver impairment in sepsis.³⁷ This connection is well-documented in the pathophysiology of sepsis, where DIC and hepatic dysfunction are common complications.³⁸ Additionally, TT showed weaker but significant positive correlations with Neut ($r = 0.100$, $p < 0.001$), Lymph ($r = 0.092$, $p < 0.001$), and RBC ($r = 0.104$, $p < 0.001$). These findings further emphasize the multifaceted impact of coagulation abnormalities on immune cells and oxygen transport capacity in septic patients.³⁹ During model development, we considered a wide range of potential predictors, including coagulation markers such as D-dimer and PT. However, after rigorous feature selection using SHAP and other statistical methods, these coagulation markers were not retained in the final model due to their relatively lower contributions to predictive performance within our dataset. Although these markers are known to play a role in the coagulopathy associated with sepsis, their signal in this specific leukemia cohort may have been overshadowed by other variables.

In this study, RF demonstrated clear advantages over traditional methods like logistic regression in predicting sepsis risk among leukemia patients. While logistic regression is valued for its simplicity and interpretability, its reliance on linear assumptions and limited ability to capture complex interactions among predictors constrained its performance.⁴⁰ By contrast, RF achieved higher accuracy, recall, and F1 score due to its ability to model nonlinear relationships and interactions. Furthermore, the use of SHAP analysis addressed RF's interpretability limitations, providing meaningful insights into the importance of individual predictors, thereby enhancing its clinical utility.⁴¹ Although individual biomarkers may contribute to sepsis prediction, their accuracy is often limited.⁴² The underlying mechanisms of sepsis in leukemia are complex, involving multiple biological pathways and immune responses, making it difficult for a single biomarker to capture the entire pathological process.^{43,44} SHAP analysis identified eight key predictors of sepsis in leukemia patients, these variables provide a holistic view of sepsis risk by capturing inflammation, immune function, coagulation abnormalities, and hemodynamic stability. For example, CRP and PCT are established inflammatory biomarkers, while Neut and Lymph reflect immune status, critical in immunosuppressed leukemia patients. TT highlights coagulopathy, a hallmark of sepsis, and TBA reflects liver dysfunction, a frequent complication. This comprehensive approach ensures that the model's predictions are grounded in clinically relevant markers, supporting its potential utility in risk stratification and management. Combining multiple biomarkers offers a more comprehensive view of the dynamic changes in immune function, inflammatory response, and pathogen activity, thereby enhancing the accuracy of early sepsis detection and risk assessment.⁴⁵ Recently, combined biomarker models have emerged as a key area of research for sepsis prediction, and their integration into machine learning models has shown improved sensitivity and specificity.⁴⁶

The eight biomarkers selected in this study are easily accessible in clinical settings and reflect various aspects of inflammation, immune function, and pathological changes, providing a more accurate assessment of sepsis risk in leukemia patients compared to single biomarkers. Their combined use significantly improved the model's predictive accuracy, offering clinicians better tools for identifying high-risk patients. This study has limitations. The data were sourced from a single medical center, which may introduce regional bias and limit the generalizability of the model.

Additionally, the retrospective nature of the study, while potentially introducing inherent biases, was mitigated through rigorous preprocessing techniques such as multiple imputation and cross-validation. Nonetheless, external validation using multi-center data from diverse geographical regions is essential to ensure the robustness and broader applicability of the model. Collaborative efforts with other institutions are underway to address these limitations and improve its clinical utility across varied patient populations.

Conclusion

The RF model demonstrated strong predictive performance for sepsis in leukemia patients, using eight key biomarkers. This model provides a valuable tool for early sepsis detection, enhancing clinical decision-making in high-risk patients.

Declaration of Figures Authenticity

All figures submitted have been created by the authors who confirm that the images are original with no duplication and have not been previously published in whole or in part.

Data Sharing Statement

The datasets from Guangdong Medical University Hospital used and analyzed in this study are available from the corresponding author upon reasonable request.

Ethics Approval and Consent to Participate

This study was reviewed and approved by the Ethics Committee of the Affiliated Hospital of Guangdong Medical University (Approval No. PJKT2024-211). The data were anonymized, eliminating the need for informed consent. This study complies with the Declaration of Helsinki.

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

This study was supported by the Big Data Platform of Affiliated Hospital of Guangdong Medical University.

Disclosure

The authors declared no competing interests for this work.

References

1. Bauer M, Gerlach H, Vogelmann T, Preissing F, Stiefel J, Adam D. Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019- results from a systematic review and meta-analysis. *Crit Care*. 2020;24(1):239. doi:10.1186/s13054-020-02950-2
2. Jantunen E, Hämäläinen S, Pulkki K, Juutilainen A. Novel biomarkers to identify complicated course of febrile neutropenia in hematological patients receiving intensive chemotherapy. *Eur J Haematol*. 2024;113(4):392–399. doi:10.1111/ejh.14264
3. Malik IA, Cardenas-Turanzas M, Gaeta S, et al. Sepsis and acute myeloid leukemia: a population-level study of comparative outcomes of patients discharged from Texas Hospitals. *Clin Lymphoma Myeloma Leuk*. 2017;17(12):e27–e32. doi:10.1016/j.clml.2017.07.009
4. Alam ST, Dongarwar D, Lopez E, et al. Disparities in mortality among acute myeloid leukemia-related hospitalizations. *Cancer Med*. 2023;12(3):3387–3394. doi:10.1002/cam4.5084
5. Bleakley G, Cole M. Recognition and management of sepsis: the nurse's role. *Br J Nurs*. 2020;29(21):1248–1251. doi:10.12968/bjon.2020.29.21.1248
6. Cho DS, Schmitt RE, Dasgupta A, Ducharme AM, Doles JD. Acute and sustained alterations to the bone marrow immune microenvironment following polymicrobial infection. *Shock*. 2022;58(1):45–55. doi:10.1097/SHK.0000000000001951
7. Vago L, Gojo I. Immune escape and immunotherapy of acute myeloid leukemia. *J Clin Invest*. 2020;130(4):1552–1564. doi:10.1172/JCI129204
8. Peseski AM, McClean M, Green SD, Beeler C, König H. Management of fever and neutropenia in the adult patient with acute myeloid leukemia. *Expert Rev Anti Infect Ther*. 2021;19(3):359–378. doi:10.1080/14787210.2020.1820863

9. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* 2020;46(3):383–400. doi:10.1007/s00134-019-05872-y
10. Angione C, Silverman E, Yaneske E. Using machine learning as a surrogate model for agent-based simulations. *PLoS One.* 2022;17(2):e0263150. doi:10.1371/journal.pone.0263150
11. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med.* 2021;181(8):1065–1070. doi:10.1001/jamainternmed.2021.2626
12. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform.* 2023;24(2). doi:10.1093/bib/bbad002
13. Huang YH, Chen CJ, Shao SC, et al. Comparison of the diagnostic accuracies of monocyte distribution width, procalcitonin, and C-reactive protein for sepsis: a systematic review and meta-analysis. *Crit Care Med.* 2023;51(5):e106–e114. doi:10.1097/CCM.0000000000005820
14. Póvoa P, Coelho L, Dal-Pizzol F, et al. How to use biomarkers of infection or sepsis at the bedside: guide to clinicians. *Intensive Care Med.* 2023;49(2):142–153. doi:10.1007/s00134-022-06956-y
15. Kubo K, Sakuraya M, Sugimoto H, et al. Benefits and harms of procalcitonin- or C-reactive protein-guided antimicrobial discontinuation in critically ill adults with sepsis: a systematic review and network meta-analysis. *Crit Care Med.* 2024;52(10):e522–e534. doi:10.1097/CCM.0000000000006366
16. Paudel R, Dogra P, Montgomery-Yates AA, et al. Procalcitonin: a promising tool or just another overhyped test? *Int J Med Sci.* 2020;17(3):332–337. doi:10.7150/ijms.39367
17. Rizo-Téllez SA, Sekheri M, Filep JG. C-reactive protein: a target for therapy to reduce inflammation. *Front Immunol.* 2023;14:1237729. doi:10.3389/fimmu.2023.1237729
18. Schlaweck S, Radcke A, Kampmann S, Becker BV, Brossart P, Heine A. The immunomodulatory effect of different FLT3 inhibitors on dendritic cells. *Cancers.* 2024;16(21):3719. doi:10.3390/cancers16213719
19. Chen S, Zhu H, Jounaidi Y. Comprehensive snapshots of natural killer cells functions, signaling, molecular mechanisms and clinical utilization. *Signal Transduct Target Ther.* 2024;9(1):302. doi:10.1038/s41392-024-02005-w
20. Giustozzi M, Ehrlicher H, Bongiovanni D, et al. Coagulopathy and sepsis: pathophysiology, clinical manifestations and treatment. *Blood Rev.* 2021;50:100864. doi:10.1016/j.blre.2021.100864
21. Pierrakos C, Velissaris D, Bisdorff M, Marshall JC, Vincent JL. Biomarkers of sepsis: time for a reappraisal. *Crit Care.* 2020;24(1):287. doi:10.1186/s13054-020-02993-5
22. Erickson YO, Samia NI, Bedell B, Friedman KD, Atkinson BS, Raife TJ. Elevated procalcitonin and C-reactive protein as potential biomarkers of sepsis in a subpopulation of thrombotic microangiopathy patients. *J Clin Apher.* 2009;24(4):150–154. doi:10.1002/jca.20205
23. Dulmovits BM, Tang Y, Papoin J, et al. HMGB1-mediated restriction of EPO signaling contributes to anemia of inflammation. *Blood.* 2022;139(21):3181–3193. doi:10.1182/blood.2021012048
24. Prüfer S, Weber M, Sasca D, et al. Distinct signaling cascades of TREM-1, TLR and NLR in neutrophils and monocytic cells. *J Innate Immun.* 2014;6(3):339–352. doi:10.1159/000355892
25. Wu H, Cao T, Ji T, Luo Y, Huang J, Ma K. Predictive value of the neutrophil-to-lymphocyte ratio in the prognosis and risk of death for adult sepsis patients: a meta-analysis. *Front Immunol.* 2024;15:1336456. doi:10.3389/fimmu.2024.1336456
26. Zahorec R. Neutrophil-to-lymphocyte ratio, past, present and future perspectives. *Bratisl Lek Listy.* 2021;122(7):474–488. doi:10.4149/BLL_2021_078
27. Venet F, Monneret G. Advances in the understanding and treatment of sepsis-induced immunosuppression. *Nat Rev Nephrol.* 2018;14(2):121–137. doi:10.1038/nrneph.2017.165
28. Thol F, Döhner H, Ganser A. How I treat refractory and relapsed acute myeloid leukemia. *Blood.* 2024;143(1):11–20. doi:10.1182/blood.2023022481
29. Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer.* 2006;6(3):193–203. doi:10.1038/nrc1816
30. Tang Y, Sorenson J, Lanspa M, Grissom CK, Mathews VJ, Brown SM. Systolic blood pressure variability in patients with early severe sepsis or septic shock: a prospective cohort study. *BMC Anesthesiology.* 2017;17(1):82. doi:10.1186/s12871-017-0377-4
31. Piagnerelli M, Boudjeltia KZ, Vanhaeverbeek M, Vincent JL. Red blood cell rheology in sepsis. *Intensive Care Med.* 2003;29(7):1052–1061. doi:10.1007/s00134-003-1783-2
32. Gong X, Xia L, Su Z. Friend or foe of innate lymphoid cells in inflammation-associated cardiovascular disease. *Immunology.* 2021;162(4):368–376. doi:10.1111/imm.13271
33. Strnad P, Tacke F, Koch A, Trautwein C. Liver - guardian, modifier and target of sepsis. *Nat Rev Gastroenterol Hepatol.* 2017;14(1):55–66. doi:10.1038/nrgastro.2016.168
34. Zhang Y, Khalid S, Jiang L. Diagnostic and predictive performance of biomarkers in patients with sepsis in an intensive care unit. *J Int Med Res.* 2019;47(1):44–58. doi:10.1177/0300060518793791
35. Wang Y, Deng K, Lin P, et al. Elevated total bile acid levels as an independent predictor of mortality in pediatric sepsis. *Pediatr Res.* 2024. doi:10.1038/s41390-024-03438-3
36. Leonhardt J, Dorresteyn MJ, Neugebauer S, et al. Immunosuppressive effects of circulating bile acids in human endotoxemia and septic shock: patients with liver failure are at risk. *Critical Care.* 2023;27(1):372. doi:10.1186/s13054-023-04620-5
37. Yan J, Li S, Li S. The role of the liver in sepsis. *Int Rev Immunol.* 2014;33(6):498–510. doi:10.3109/08830185.2014.889129
38. Anderko RR, Gómez H, Canna SW, et al. Sepsis with liver dysfunction and coagulopathy predicts an inflammatory pattern of macrophage activation. *Intensive Care Med Exp.* 2022;10(1):6. doi:10.1186/s40635-022-00433-y
39. Unar A, Bertolino L, Patauner F, Gallo R, Durante-Mangoni E. Pathophysiology of disseminated intravascular coagulation in sepsis: a clinically focused overview. *Cells.* 2023;12(17):2120. doi:10.3390/cells12172120
40. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinf.* 2018;19(1):270. doi:10.1186/s12859-018-2264-5
41. Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. *Clin Transl Sci.* 2024;17(11):e70056. doi:10.1111/cts.70056
42. Faix JD. Biomarkers of sepsis. *Crit Rev Clin Lab Sci.* 2013;50(1):23–36. doi:10.3109/10408363.2013.764490

43. Xiao YP, Cheng YC, Chen C, Xue HM, Yang M, Lin C. Identification of the shared gene signatures of HCK, NOG, RNF125 and biological mechanism in pediatric acute lymphoblastic leukaemia and pediatric sepsis. *Mol Biotechnol.* 2023;67:80–90. doi:10.1007/s12033-023-00979-6
44. Xia S, Zhao YC, Guo L, et al. Do antibody-drug conjugates increase the risk of sepsis in cancer patients? A pharmacovigilance study. *Front Pharmacol.* 2022;13:967017. doi:10.3389/fphar.2022.967017
45. Barichello T, Generoso JS, Singer M, Dal-Pizzol F. Biomarkers for sepsis: more than just fever and leukocytosis—a narrative review. *Crit Care.* 2022;26(1):14. doi:10.1186/s13054-021-03862-5
46. Rojas JC, Fahrenbach J, Makhni S, et al. Framework for integrating equity into machine learning models: a case study. *Chest.* 2022;161(6):1621–1627. doi:10.1016/j.chest.2022.02.001

Journal of Inflammation Research

Publish your work in this journal

The Journal of Inflammation Research is an international, peer-reviewed open-access journal that welcomes laboratory and clinical findings on the molecular basis, cell biology and pharmacology of inflammation including original research, reviews, symposium reports, hypothesis formation and commentaries on: acute/chronic inflammation; mediators of inflammation; cellular processes; molecular mechanisms; pharmacology and novel anti-inflammatory drugs; clinical conditions involving inflammation. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-inflammation-research-journal>

Dovepress

Taylor & Francis Group