Machine Learning Algorithms to Predict the Mortality of Carbapenem Resistant Klebsiella Pneumoniae bacteremia

CURRENT STATUS: UNDER REVIEW

Antimicrobial Resistance and Infection Control  BMC

Qiqiang Liang  
Zhejiang University School of Medicine Second Affiliated Hospital

Fang Qian  
Zhejiang University School of Medicine Second Affiliated Hospital

Yibing Chen  
Zhejiang University School of Medicine Second Affiliated Hospital

Zhijun Xu  
Zhejiang University School of Medicine Second Affiliated Hospital

Zhijiang Xu  
Zhejiang University School of Medicine Second Affiliated Hospital

Man Huang  huangman@zju.edu.cn  
Zhejiang University School of Medicine  
Corresponding Author  
ORCiD: 0000-0002-0842-2254

DOI: 10.21203/rs.2.18733/v1

SUBJECT AREAS
Infectious Diseases  Artificial Intelligence and Machine Learning

KEYWORDS
Machine learning, Carbapenem Resistant Klebsiella Pneumoniae, Bloodstream infection, Prediction model, Risk factors
Abstract

Purpose To establish mortality prediction models in 14 days of Carbapenem-Resistant Klebsiella Pneumoniae bacteremia using Machine learning.

Materials and Methods It is a single-center retrospective study. We collect the relevant clinical information of all patients with Carbapenem-Resistant Klebsiella Pneumoniae (CRKP) bacteremia in the past 5 years using the local database. Data analysis and verification are carried out by multiple logical regression, decision tree, random forest, support vector machine (SVM), and XGBoost.

Result This study includes 187 patients with 40 related variables. In multiple logical regression, acute renal injury (P=0.003), Apache II score (P=0.036), immunodeficiency (P=0.025), severe thrombocytopenia (P=0.025) and septic shock (P=0.044) are the high-risk factors for 14 days mortality of CRKP bloodstream infections. According to the importance of those parameters, risk scoring is established to predict the survival rate of CRKP bacteremia. The analysis of the five models, with 70% training set and 30% test set, show the comprehensive performance of random forest (AUROC=0.953, precision=91.85%) is slightly better than that of XGBoost (AUROC=0.912, precision=86.41%) and SVM (AUROC=0.936, precision=79.89%) in predicting 14-day mortality of CRKP bacteremia. The multiple logical regression model (AUROC=0.825, precision=81.52%) is the second, and the decision tree model (AUROC=0.712, precision=79.89%) is not very ideal.

Conclusion Machine learning has good performances in predicting 14-day mortality of CRKP bacteremia than multiple logical regression. Acute renal injury, severe thrombocytopenia, and septic shock are the high-risk factors of CRKP bacteremia mortality.
Background

Carbapenem-Resistant Klebsiella pneumoniae (CRKP) infections have become the focus of global researchers and are one of the most serious hospital-acquired infections, especially in the intensive care unit (ICU)\(^1,\,2\). Among the different sites of CRKP infection, the mortality rate of bloodstream infection is the highest. The largest study in China shows that the mortality rate of CRKP bacteremia is \(32.9\%\)\(^3\) with the reported worldwide ranging from \(17\%\) to \(50\%\) \(^4\text{--}^6\).

Risk factors for death of CRKP bloodstream infections include long-term hospitalization, medication history of carbapenem and glucocorticoids, invasive operations, septic shock, and inappropriate empirical treatment, etc\(^4\text{--}^7,\,8\). High-risk factors can predict the prognosis, understand the disease severity and direct to take medication earlier. For patients with high-risk factors for death, we may prefer to use higher levels antibiotics, especially in the context of the increasing number of new antibiotics\(^9\). However, the results of risk factor analysis have great heterogeneity because of its regional epidemiological characteristics in each of retrospective studies. There are differences in human species, medication methods and treatment concepts of CRKP bloodstream infection, especially in the west countries and Asia countries\(^5,\,6,\,10,\,11\). It is difficult for us to apply foreign high-risk factors to Chinese patients and to popularize CRKP-related scores dominated by western countries. The establishment of own scores of CRKP bacteremia mortality may be of better significance, which is one of the purposes of this study.

The risk predictive model contains the weight of different risk factors and is more intuitive and more practical. There are few kinds of researches on the high-risk prediction model of CRKP infection with previous researches using the multiple
logistic regression methods\textsuperscript{6,12-14}. However, some studies have proved the effect of the machine learning is better than the multiple logical regression in the prediction of gram-negative bacteria and CRE colonization\textsuperscript{13-15}. Therefore, this study plans to use machine learning to build a risk predictive model of the mortality of CRKP bloodstream infection.

Materials and Methods

1. Study Design: It is a single-center retrospective study using the comprehensive ICU database of the second affiliated hospital of Zhejiang university school of medicine, an academic teaching hospital with 3500 beds. General ICU has 40 independent units and has established a complete intensive care unit database. Our database has been included more than 7000 samples and the structures are similar to the famous MIMIC database. We included all the patients with CRKP positive in blood culture from 2012.05 to 2019.05, but the contamination was excluded if there were no obvious signs of bacteremia. We collected the patients demography, Apache II score, Charlson score, medication before bloodstream infection, disease status, various indicators and medication at the time of bloodstream infection. The predicted outcome was mortality in 14 days after the onset of bloodstream infection.

2. Definition: The 14 days mortality was mainly considered the higher mortality rate of bloodstream infection in the short term and the higher sensitivity of prediction. The history of antibiotics and glucocorticoids refers to the use of cephalosporins, carbapenems, tigecycline and glucocorticoids for more than 3 days in the past month of the occurrence of bloodstream infection. Pseudomonas aeruginosa carriage history means the positive culture of
Pseudomonas aeruginosa in all samples in the past month. Multiple sites infection mainly refers to CRKP bacteremia complicated with pulmonary infection, abdominal infection, urinary tract infection, intracranial infection, biliary tract infection and so on. The standard of acute kidney injury conformed to the standard of KDIGO guideline, which was determined by the baseline and difference value of creatinine. Septic shock is defined by sepsis3.0. Acute liver injury was mainly caused by the analysis of the results of glutamic-pyruvic transaminase, glutamic-oxaloacetic transaminase, and total bilirubin. Severe thrombocytopenia refers to thrombocytopenia less than $30 \times 10^9$/L within 3 days after the occurrence of bacteremia. Gastrointestinal bleeding is defined as positive fecal and gastric drainage occult blood test within 3 days after the occurrence of bloodstream infection. The above data are all derived from the automatic collection of the database after the manual establishment of links and preset conditions.

3. Microbiological tests: The strains with mediated sensitive of tigecycline were determined by the broth dilution method. The drug susceptibility test was carried out by a laboratory physician with analysis instruments (VITEK2 AST-GN16 France). Minimum inhibitory concentration (MIC) determination and interpretation complied with standards of the Clinical and Laboratory Standards Institute (CLSI). Carbapenem resistance means MIC of imipenem or meropenem $\geq 2$ mg/L by CLSI. The tigecycline MIC definition follows by the criteria of the European Committee on Antimicrobial Susceptibility Testing (EUCAST), MIC $> 2$ mg/L.

4. Statistical methods: all statistics are done in R and Rstudio. Data packets include "VIM", "psych", "regplot", "mass", "vcd", "rpart", "rpart.plot",
"randomforest", "e1071", "xgboost" and "rattle". For the test indicators that may be collinear in clinical evaluation, principal component analysis are quantified to form dichotomous variables. the missing values are processed by multiple interpolations and we delete independent variables with the missing more than 60%. In univariate analysis, numerical variables are tested by independent sample T-test, dichotomous variables are completed by chi-square test, and P less than 0.05 was considered to be statistically significant. Odds ratio (OR) and 95% confidence interval (CI) were calculated to evaluate the strength of association. All samples are grouped, 70% into the training set and 30% into the test set. The evaluation parameters include sensitivity, specificity, negative likelihood ratio, positive likelihood ratio, and area under operating characteristic curve (AUROC). The independent variable screening method is the gradually decreasing method in the multiply logical regression. The decision tree is constructed and pruned according to the complexity parameters, and the pruned results are presented by the visual package. 500 trees are constructed in the Random forest, and the importance of measurable variables is presented by the corresponding visual package. The cost parameter of the support vector machine (SVM) is set as 0.01, which is adjusted by the performance of the model. The specific and task parameters of the XGBoost linear rise are adjusted according to the performance of the model.

Result

1. Basic information: There are 195 patients with positive CRKP blood culture from 2014/06/ to 2019/06. Finally, we include 187 patients in the analysis because 8
patients meet the exclusion criteria. According to the 14-day mortality after CRKP bloodstream infection, the patients are divided into the survival and death group, of which 147 people in the survival group and 40 patients died. The 14-day mortality rate is 21.39%. The flow chart can be seen in Figure 1. The baseline characteristics and the treatments of the two groups are summarized in Table1. Comparison survival group, the death group is older (67 vs. 60; P=0.007; OR: 2.445; 95% CI: 1.801-6.423) and have higher percentage of prior tigecycline history (27.50% vs. 14.29%; P=0.049, OR: 1.925; 95% CI: 1.015-3.652), the history of glucocorticoid use (42.50% vs. 23.81%; P=0.019; OR=1.785; 95%CI:1.124-2.834), immunodeficiency (50.00% vs. 31.29%; P=0.028; OR=1.598; 95%CI:1.080-2.364) and severe thrombocytopenia (65.00% vs. 31.97%; P=0.001; OR=2.033; 95%CI:1.465-2.821). More of the dead patients were admitted into ICU by severe infectious disease (55.00% vs. 31.29%; P=0.006; OR=1.758; 95%CI:1.216-2.541), with more acute kidney injury (50.00% vs. 24.49%; P=0.002; OR=2.042; 95%CI:1.341-3.108) and septic shock (85.00% vs. 54.42%; P=0.001; OR=1.562; 95%CI:1.283-1.902).

In multivariate analysis, acute kidney injury (P=0.003), immunodeficiency (P=0.025), severe thrombocytopenia (P=0.025) and septic shock (P=0.044) are the high-risk factors for 14-day mortality in CRKP patients without obvious collinearity. Then we incorporate these variables into the multiple logistic regression model and establishe a preliminary score to predict the survival rate of CRKP bloodstream infection according to the contribution of the parameters. As see in Table 2, patients with scores higher than 20 are more likely to survive in our scoring system.

2.Model comparison: we take 130 patients as the training set and the remaining 57 patients as the test set. The decision tree model has two branches, including severe thrombocytopenia and acute kidney injury. The specific classification method can be
found in the poor accuracy of Figure 2. The effect of the random forest model is satisfactory and mainly includes severe thrombocytopenia, invasive operation, septic shock, acute kidney injury, multiple site infection, early use of carbapenem drugs in Figure 3. After the five models are established and verified by the test set, we find that the comprehensive performance established by random forest (AUROC=0.953, Precision=91.85%) is slightly better than that of XGBoost (AUROC=0.912, Precision=86.41%) and SVM (AUROC=0.936, Precision=79.89%), followed by multiple logical regression model (AUROC=0.825, Precision=81.52%), and decision tree model (AUROC=0.712, Precision=79.89%) in Figure 4. In the performance evaluation of different dimensions, random forest and XGBoost are the best in terms of accuracy. SVM has high specificity but low sensitivity, see Table 3; Overall, random forest, XGBoost, and SVM are better in predicting 14-day mortality of CRKP bloodstream infection.

Discussion

In this study, some machine learning methods are used to construct the prediction mortality model of 14-day of CRKP bloodstream infection. The effect of random forest is better than that of XGBoost and SVM with limited sample sizes, and is generally better than multiple logical regression models. The purpose of establishing a 14-day mortality prediction model is that CRKP bloodstream infection deteriorates rapidly into severe multiple organ dysfunction syndromes, resulting in death. The establishment of a relatively short-term risk prediction model can help clinicians identify high-risk groups when bloodstream infection occurs, and guide the use of more active monitoring measures and anti-infection programs. Recently, artificial intelligence has been used to predict the incidence of
hospitalization and colonization of CRO$^{17}$. Although there is no research on machine learning to predict the mortality rate of CRKP bloodstream infection, this study partially proves the superiority of the machine learning algorithm$^{18,19}$. The machine learning algorithms we chose are currently recognized, mature and common, including random forest, SVM and XGBoost. With small samples, the random forest can spread the branches of the decision tree through its algorithm, determine the important variables according to the Gini score, and present an effective model. SVM and XGBoost, as introduced algorithms recently, can get more accurate prediction through more complex operation mechanism, and can fully demonstrate their superior performance with a large amount of data. There are similar conclusions in the comparative study of other clinical medical models$^{20-22}$. The modeling results are almost the same accuracy. XGBoost is similar to the random forest in performance and SVM has higher specificity. The prediction ability of the multivariate logical regression is not as good as three machine learning algorithms, but it is more intuitive and easy to understand. In this study, the risk score based on multivariate logical regression can predict the 14-day survival rate of 82.5% of patients using five highly independent and important factors. The unexpected result is the decision tree model, because the inclusion of too few branch points, the prediction result is not ideal.

In this study, the 14-day mortality rate of CRKP bloodstream infection was 21.4%, which was lower than the previously published epidemiological results of 31.1% in China$^{3}$. This may be related to the regulation of antibiotics use in carbapenems and tigecycline recently, and to the approves of polymyxin B and ceftazidime-avibactam in the domestic market. More clinical experience of clinicians in CRKP bloodstream
infection and earlier use of appropriate antibiotics may be associated with lower mortality. The multiple logistic regression model included five major risk factors including age, septic shock, acute kidney injury, severe thrombocytopenia, and immunodeficiency. Septic shock and acute kidney injury are often mentioned in previous retrospective studies. At present, the most famous research is the INCREMENT study, which divides the numerical variables into groups through the classification decision tree and predicts the model through the hierarchical logical regression model, and finds that five indicators can predict CRKP mortality, namely, severe infection/septic shock, Pitt score more than 6, Charles scores more than 2, biliary tract organ blood stream infection, inappropriate treatment, etc. There are few reports of severe thrombocytopenia but it is consistent with the clinical manifestations of CRKP bloodstream infection, including earlier digestive tract bleeding, and faster thrombocytopenia rather than a decrease of myelosuppression. We found that septic shock, acute renal insufficiency, severe thrombocytopenia, these three independent variables are present in different models, and occupy an important position in the model.

In this study, there are advantages and disadvantages to the use of the local database. Different databases include different amounts of dimensions. In the critical area, a large number of articles have been published on the public database MIMIC, but the data may not apply to the Chinese population. CRKP bloodstream infection is gradually increasing in recent years, but the epidemic trend is quite different from that of foreign countries. The CRKP samples in the database of MIMIC are very small, so it is impossible to carry out. Our department database has been established for two years, integrates the data stored in various
data centers of the hospital in the past 5 years, and carries on the data mining of sepsis and acute kidney injury. The use of the local database to obtain samples of high accuracy, a wide selection of independent variables and flexible and complete extraction of independent variables. However, the local database may have some shortcomings, such as data bias, limited extrapolation, small sample size, unstable model. The models developed in recent years have no clear visualization tools, which makes it difficult to understand and explain the logic of the model. We have considered this issue in the data modeling phase and are actively working with other centers for external validation of the model, which will be the next phase of our team's research.

Conclusion

Machine learning is superior to the traditional multiple logistic regression model in predicting the 14-day mortality of CRKP bloodstream infection. The risk factors of CRKP bacteremia mortality includes septic shock, acute kidney injury and severe thrombocytopenia which can be explained clinically.

Acknowledgements

Ethics approval and consent to participate: Our research has been approved by the ethics committee of the second affiliated hospital of Zhejiang University School of Medicine.

Consent for publication: Yes.

Availability of data and material: Please contact author for data requests.

Funding: No funding.

Author contributions: Qiqiang Liang write manuscript and analyze data and
modeling; Fang Qian assist at data cleaning and sorting; Yibing Chen assist at data secondary audit, independent variable screening, and chart drawing; Yu Zhou establish the database and collect original data. Zhijun Xu assist at data analysis and revision of script. Xu Zhijiang provides technical support related to microbiology. Huang Man corrects the data of the paper;

Acknowledgements: Thanks to the clinicians in the general ICU for their support and guidance in data collection.

Competing interests: The authors declare that they have no competing interests.

Reference

1. Brink AJ. Epidemiology of carbapenem-resistant Gram-negative infections globally. *Current opinion in infectious diseases*. 2019;32(6):609-616.

2. Willyard C. The drug-resistant bacteria that pose the greatest health threats. *Nature*. 2017;543(7643):15.

3. Wang X, Wang Q, Cao B, et al. Impact of Combination Therapy vs Monotherapy on Mortality from Carbapenem-Resistant Enterobacteriaceae Bacteremia: A Retrospective Observational Study from a Chinese Network. *Antimicrobial agents and chemotherapy*. 2018.

4. Iacchini S, Sabbatucci M, Gagliotti C, et al. Bloodstream infections due to carbapenemase-producing Enterobacteriaceae in Italy: results from nationwide surveillance, 2014 to 2017. *Euro Surveill*. 2019;24(5).

5. Satlin MJ, Chen L, Patel G, et al. Multicenter Clinical and Molecular Epidemiological Analysis of Bacteremia Due to Carbapenem-Resistant Enterobacteriaceae (CRE) in the CRE Epicenter of the United States. *Antimicrobial agents and chemotherapy*. 2017;61(4).
6. Gutierrez-Gutierrez B, Salamanca E, de Cueto M, et al. Effect of appropriate combination therapy on mortality of patients with bloodstream infections due to carbapenemase-producing Enterobacteriaceae (INCREMENT): a retrospective cohort study. *Lancet Infectious Diseases*. 2017;17(7):726-734.

7. Liu P, Li X, Luo M, et al. Risk Factors for Carbapenem-Resistant Klebsiella pneumoniae Infection: A Meta-Analysis. *Microbial drug resistance*. 2018;24(2):190-198.

8. Trecarichi EM, Pagano L, Martino B, et al. Bloodstream infections caused by Klebsiella pneumoniae in onco-hematological patients: clinical impact of carbapenem resistance in a multicentre prospective survey. *American journal of hematology*. 2016;91(11):1076-1081.

9. Plazak ME, Tamma PD, Heil EL. The antibiotic arms race: current and emerging therapy for Klebsiella pneumoniae carbapenemase (KPC) - producing bacteria. *Expert opinion on pharmacotherapy*. 2018:1-13.

10. Garcia-Arenzana N, Redondo-Bravo L, Espinel-Ruiz MA, et al. Carbapenem-Resistant Enterobacteriaceae Outbreak in a Medical Ward in Spain: Epidemiology, Control Strategy, and Importance of Environmental Disinfection. *Microbial drug resistance*. 2019.

11. Papadimitriou-Oliveris M, Fligou F, Bartzavali C, et al. Carbapenemase-producing Klebsiella pneumoniae bloodstream infection in critically ill patients: risk factors and predictors of mortality. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*. 2017;36(7):1125-1131.

12. Song JY, Jeong IS. Development of a risk prediction model of carbapenem-resistant Enterobacteriaceae colonization among patients in intensive care
units. *American journal of infection control*. 2018;46(11):1240-1244.

13. Lodise TP, Bonine NG, Ye JM, Folse HJ, Gillard P. Development of a bedside tool to predict the probability of drug-resistant pathogens among hospitalized adult patients with gram-negative infections. *BMC Infect Dis*. 2019;19(1):718.

14. Goodman KE, Simner PJ, Klein EY, et al. Predicting probability of perirectal colonization with carbapenem-resistant Enterobacteriaceae (CRE) and other carbapenem-resistant organisms (CROs) at hospital unit admission. *Infect Control Hosp Epidemiol*. 2019;40(5):541-550.

15. Flechet M, Falini S, Bonetti C, et al. Machine learning versus physicians’ prediction of acute kidney injury in critically ill adults: a prospective evaluation of the AKIpredictor. *Critical Care*. 2019;23(1):282.

16. Peri AM, Doi Y, Potoski BA, Harris PNA, Paterson DL, Righi E. Antimicrobial treatment challenges in the era of carbapenem resistance. *Diagnostic microbiology and infectious disease*. 2019.

17. Goodman KE, Simner PJ, Klein EY, et al. Predicting probability of perirectal colonization with carbapenem-resistant Enterobacteriaceae (CRE) and other carbapenem-resistant organisms (CROs) at hospital unit admission. *Infection control and hospital epidemiology*. 2019:1-10.

18. Ghassemi M, Celi LA, Stone DJ. State of the art review: the data revolution in critical care. *Crit Care*. 2015;19:118.

19. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. 2018;15(141).

20. Taylor RA, Moore CL, Cheung KH, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. *PloS one*. 2018;13(3):e0194085.
21. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PloS one.* 2017;12(4).

22. Huang C, Murugiah K, Mahajan S, et al. Enhancing the prediction of acute kidney injury risk after percutaneous coronary intervention using machine learning techniques: A retrospective cohort study. *PLoS Med.* 2018;15(11):e1002703.

23. Liang Q, Huang M, Xu Z. Early use of polymyxin B reduces the mortality of carbapenem-resistant *Klebsiella pneumoniae* bloodstream infection. *Braz J Infect Dis.* 2019.

24. Xu L, Sun X, Ma X. Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant *Klebsiella pneumoniae*. *Annals of clinical microbiology and antimicrobials.* 2017;16(1):18.

25. Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. *J Crit Care.* 2018;45(- 1557-8615 (Electronic)):156-162.

26. Meyer A, Zverinski D, Pfahringer B, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. *The Lancet Respiratory Medicine.* 2018;6(12):905-914.

Tables
Table 1: Baseline characteristics of patients with CRKP bloodstream infection

| Demographics                        | Mortality; N=40 | Survival; N=147 | P       | OR and 95%CI |
|-------------------------------------|-----------------|-----------------|---------|--------------|
| Gender Male%                        | 31 (77.50%)     | 106 (72.11%)    | 0.551   |              |
| Age (y)                             | 67              | 60              | 0.007   | 2.445 (1.801-6.423) |
| Apache II score                     | 20              | 18              | 0.126   |              |
| Charlson score                      | 2.5             | 2.7             | 0.556   |              |
| High-risk time (day)                | 21              | 22              | 0.88    |              |
| Length of ICU stay (day)            | 22              | 26              | 0.41    |              |
| Length of stay (day)                | 27              | 54              | 0.001   | 3.453 (2.147-7.456) |
| **Primary disease**                 |                 |                 |         |              |
| Trauma                              | 9 (22.50%)      | 51 (34.69%)     | 0.143   |              |
| Infection                           | 22 (55.00%)     | 46 (31.29%)     | 0.006   | 1.758 (1.216-2.541) |
| Surgical diseases                   | 9 (22.50%)      | 28 (19.05%)     | 0.627   |              |
| Cerebrovascular accident            | 8 (20.00%)      | 43 (29.25%)     | 0.244   |              |
| Hypertension                        | 23 (57.50%)     | 65 (44.22%)     | 0.136   |              |
| Cardiac insufficiency               | 9 (22.50%)      | 26 (17.69%)     | 0.489   |              |
| Respiratory insufficiency           | 14 (35.00%)     | 39 (26.53%)     | 0.292   |              |
| Malignant tumor                     | 6 (15.00%)      | 22 (14.97%)     | 0.996   |              |
| Autoimmune diseases                 | 1 (2.50%)       | 4 (2.72%)       | 1       |              |
| Chronic renal insufficiency         | 9 (22.50%)      | 21 (14.29%)     | 0.209   |              |
| Recent surgical history             | 20 (50.00%)     | 96 (65.31%)     | 0.068   |              |
| Perior medication history           |                 |                 |         |              |
| Perior tigecycline history          | 11 (27.50%)     | 21 (14.29%)     | 0.049   | 1.925 (1.015-3.652) |
| Perior carbapenems history          | 26 (65.00%)     | 75 (51.02%)     | 0.116   |              |
| Perior cephalosporins history       | 30 (75.00%)     | 99 (67.35%)     | 0.354   |              |
| Perior glucocorticoid history       | 17 (42.50%)     | 35 (23.81%)     | 0.019   | 1.785 (1.124-2.834) |
| **Condition of bloodstream infection** |                 |                 |         |              |
| Immunodeficiency                    | 20 (50.00%)     | 46 (31.29%)     | 0.028   | 1.598 (1.08-2.364) |
| Tracheotomy                         | 15 (37.50%)     | 63 (42.86%)     | 0.521   |              |
| Invasive operation                  | 28 (70.00%)     | 74 (50.34%)     | 0.027   | 1.391 (1.074-1.801) |
| Acute hepatic injury                | 12 (30.00%)     | 40 (27.21%)     | 0.727   |              |
| Acute kidney injury                 | 20 (50.00%)     | 36 (24.49%)     | 0.002   | 2.042 (1.341-3.108) |
| Pseudomonas aeruginosa carriage     | 10 (25.00%)     | 39 (26.53%)     | 0.845   |              |
| Multiple sites infection            | 25 (62.50%)     | 98 (66.67%)     | 0.622   |              |
| Complicated pulmonary infections    | 22 (55.00%)     | 74 (50.34%)     | 0.601   |              |
| Tigecycline resistance              | 17 (42.50%)     | 55 (37.41%)     | 0.558   |              |
| Severe thrombocytopenia             | 26 (65.00%)     | 47 (31.97%)     | 0.001   | 2.033 (1.465-2.821) |
| Gastrointestinal hemorrhage         | 12 (30.00%)     | 51 (34.69%)     | 0.578   |              |
| Septic shock                        | 34 (85.00%)     | 80 (54.42%)     | 0.001   | 1.562 (1.283-1.902) |
| **Treatment protocols**             |                 |                 |         |              |
| Combinations with Carbapenems       | 30 (75.00%)     | 110 (74.83%)    | 0.982   |              |
| Combinations with Tigecycline       | 20 (50.00%)     | 83 (56.46%)     | 0.466   |              |
| Combinations with Polymyxin B       | 13 (32.50%)     | 35 (23.80%)     | 0.265   |              |
| Combinations with Amikacin          | 12 (30.00%)     | 46 (31.29%)     | 0.875   |              |
| Combinations with Fosfomycin        | 0               | 12 (8.16%)      | 0.073   |              |

Table 2: The results of multiple logistic regression model and the scores based on

|                      | Yes | No | Multivariate analysis |
|----------------------|-----|----|-----------------------|
| Septic shock         | 2   | 5  | 0.044                 |
| Acute kidney injury  | 0   | 5  | 0.003                 |
| Severe thrombocytopenia | 1 | 5  | 0.025                 |
| Immunodeficiency     | 5   | 9  | 0.025                 |
| age                  | 12-0.12X |     | 0.03                  |
| Total scores         | Death<20 | Survival≥20 |               |
the weights of different independent variables; when the patient score was more than 20, the survival probability of patients with CRKP bloodstream infection was more than 50%.

| Model                  | AUROC  | Precision | Sensitivity | Specificity | Positive likelihood ratio | Negative likelihood ratio |
|------------------------|--------|-----------|-------------|-------------|---------------------------|--------------------------|
| Multivariate Logit Regression | 0.825  | 81.52%    | 67.5%       | 85.42%      | 4.63                      | 0                        |
| Decision tree          | 0.712  | 79.89%    | 30%         | 93.75%      | 4.80                      | 0                        |
| RandomForest           | 0.953  | 91.85%    | 67.5%       | 98.61%      | 48.6                      | 0                        |
| SVM                    | 0.936  | 79.89%    | 7.50%       | 100%        | 0                         | 0                        |
| XGBoost                | 0.912  | 86.41%    | 70%         | 90.97%      | 7.75                      | 0                        |

Table 3: Prediction effect analysis of five prediction models, including AUROC, accuracy, sensitivity, specificity, positive likelihood ratio and negative likelihood ratio.

Figures

Figure 1
The flow chart of the study.

Figure 2
The decision tree model with two branches of severe thrombocytopenia and acute

Figure 3
Visualization Analysis of different variables weight of Random Forest in the Model

Figure 4
AUROC curves of five models. Different colors represent different models.
