Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning–Based Prediction of Mortality and Multidrug-Resistant Infection in Non-Surgical ICU Patients with Suspected Infection: A Second Analysis of A Prospective National Multicenter Cohort Study

Shuguang Yang
Peking University People's Hospital
Ting Wang
Peking University People's Hospital
Yao Sun
Peking University People's Hospital
Hua Zhang
Peking University Third Hospital
Fengxue Zhu
Peking University People's Hospital
Youzhong An
Peking University People's Hospital
Huiying Zhao
zhaohuiying@pku.edu.cn

Peking University People's Hospital

Research Article

Keywords: Non-Surgical, Suspected infection, Mortality, Multidrug-Resistant Infection

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11378194/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Multidrug-resistant infection raise global disease burden. We aimed to develop and validate machine learning predictive models to identify independent risk factors for mortality and multidrug-resistant infection in non-surgical patients with suspected infection upon ICU admission in mainland China, using data from a prospective national multicenter cohort study.

Methods

This secondary analysis used prospectively collected medical data from **non-surgical patients** with suspected infection admitted to ICUs across mainland China between July 2021 and December 2022, derived from a national multicenter cohort study. Patients retained the parent study's 7:3 random split into training and validation cohorts. Machine learning algorithms were applied to identify risk factors and develop predictive models for mortality and multidrug-resistant infection. Performance of logistic regression, random forest, extreme gradient boosting, and gradient boosting machine models was assessed using AUC, sensitivity, specificity, PPV, NPV, and Youden index.

Results

A total of 2270 patients from 67 hospitals were enrolled. Pulmonary (83.74%) were predominant, with an overall mortality of 9.74%. The logistic regression achieved an AUC of 0.93 for mortality prediction, identifying APACHE Ⅱ score($p < 0.01$), and *K. pneumoniae* infections($p = 0.03$) as independent risk factors. It also yielded the optimal AUC of 0.86 for multidrug-resistant infection prediction. Nine independent risk factors were confirmed, including: bloodstream infections($p = 0.01$), urinary infections($p = 0.03$), central system infections($p = 0.02$), *K. pneumoniae* infections($p < 0.01$), *Stenotrophomonas maltophilia* infections($p < 0.01$), *Escherichia coli* infections($p < 0.01$), *Pseudomonas aeruginosa* infections($p < 0.01$), *Enterococcus faecium* infections($p < 0.01$), and *Staphylococcus epidermidis* infections($p < 0.01$) for non-surgical ICU patients with suspected infections.

Conclusions

Among non-surgical ICU patients with suspected infection, the logistic regression model identified subgroup-specific predictors for mortality and MDR infection, which differed from the overall cohort. This model may assist risk stratification and antimicrobial stewardship in medical critically ill patients.

Introduction

Infection is a common among patients in the intensive care unit(ICU) and is a prerequisite for the development of sepsis or septic shock[1]. Among patients admitted to the ICU, the proportion with suspected or proven infection is 54%, and the rate of ICU-acquired infections is 22%[2]. Detailed global data on infection types, as well as the use and availability of diagnostic and treatment options, can help identify risk factors for infections. Moreover, timely and appropriate antimicrobial therapy can help achieve the best outcomes for patients with sepsis[3].

Clinicians must strike a balance between adequate pathogen coverage and rational use of broad-spectrum antibiotics to mitigate the emergence of antimicrobial resistance. The rising burden of multidrug-resistant (MDR) infection has become a major clinical challenge, contributing to prolonged hospital stays, elevated mortality and increased healthcare expenditure[4]. Excessive antibiotic consumption has resulted in high bacterial resistance rates across many regions, including China[5, 6]. Over recent decades, drug-resistant Enterobacteriaceae strains resistant to extended-spectrum cephalosporins and carbapenems have spread rapidly worldwide[7]. Furthermore, mortality and the prevalence of MDR infections are higher in older patients due to greater disease severity[8].

Although numerous national and international studies have described sepsis epidemiology[9–11]. Previous findings from our prospective national multicenter cohort characterized predictors of mortality and multidrug-resistant infection in the overall population of ICU patients with suspected infection. However, the risk factors and predictive models specifically for non-surgical patients remain poorly described. We therefore performed this secondary analysis of the parent cohort to develop and validate machine learning predictive models for mortality and multidrug-resistant infection, and to identify their independent risk factors, in non-surgical patients with suspected infection at ICU admission in mainland China.

Methods

Study design and sampling

This secondary analysis used data from a prospective national multicenter cohort of ICU patients across mainland China enrolled from July 2021 to December 2022 (Fig. 1). Ethical approval was obtained from the Ethics Committee of Peking University People's Hospital (No. 2021PHB020-001) and the ethics committees of all participating hospitals. The parent study was registered at ClinicalTrials.gov on July 14, 2021 (NCT04966390; https://www.clinicaltrials.gov).

Patients

The onset of a suspected infection episode was defined as in the parent cohort: the first occurrence of antibiotic administration (oral or parenteral) or body fluid culture collection, following these timing rules:

- • If antibiotics were given first, culture sampling must occur within 24 h.

- • If culture sampling preceded antibiotics, antimicrobial therapy must be initiated within 72 h.

The time of this initial event was defined as time zero for the episode[12]. The parent cohort's exclusion criteria were retained: (1) absence of antibiotic therapy or antibiotic duration < 3 days, (2) age < 18 years, and (3) incomplete clinical data. This secondary analysis further limited the study population to non-surgical patients from the parent cohort.

Microbiological testing and Definitions of infections

Microbiological testing and infection definitions were adopted from the parent cohort. Bacterial pathogens from clinical specimens (blood, urine, ascites, pleural fluid) were identified by standard microbiological assays. Antimicrobial susceptibility was tested by the Kirby–Bauer disc diffusion method on Mueller-Hinton agar following CLSI guidelines[13].

A sample was considered culture-positive if one or more fluid cultures yielded positive results. Patients could be included more than once for distinct infection episodes. MDR infections were defined as acquired nonsusceptibility to at least one agent in three or more antimicrobial categories[8]. If multiple cultures were performed, MDR status was defined by the presence of at least one positive culture for a MDR organism.

Diagnostic criteria for infection types were defined as follows:

Pneumonia: new or progressive pulmonary infiltrates on chest radiography plus at least two of: (1) fever (> 38.5°C) or hypothermia (< 36°C); (2) positive tracheobronchial or bronchoalveolar lavage (BAL) culture or significant bacterial counts in respiratory secretions; (3) leukocytosis ≥ 12×10^9/L[14–16].

Bloodstream infections

fever (> 37.3°C) or signs of systemic infection with at least one positive blood culture[17].

Intra-abdominal infections

clinical evidence (abdominal pain or tenderness) with positive drainage or puncture fluid culture[18].

Urinary infections

lower urinary tract symptoms with a quantitative count ≥ 10^5 colony-forming units (CFU)/mL of bacteria and a positive urine culture[19].

Skin/Soft tissue infections

a spectrum from mild to life-threatening, diagnosed by positive culture from lesional skin[20].

Central nervous system infections

anatomically classified (meningitis, encephalitis, or myelitis) with a positive cerebrospinal fluid culture[21].

Pelvic infections

uterine, tubal, or ovarian infections diagnosed by pelvic pain or effusion with positive pelvic drainage or vaginal/uterine secretion culture[22].

Data collection

Data for this secondary analysis were extracted from the parent cohort's electronic data capture (EDC) system (https://edc2.cttq.com), including demographics, comorbidities, diagnoses, laboratory, imaging, inflammatory and microbiological variables, as well as data on infections and in-hospital mortality. Research staff received standardized training, and participating hospitals could adequately perform standard microbiological and antimicrobial susceptibility assays.

Statistical analysis

Statistical analyses for this secondary analysis were conducted using SPSS version 26 (IBM, SPSS Inc, Chicago, USA). Continuous variables with normal distribution were expressed as mean ± standard deviations(SD) and categorical variables were presented as counts and percentages(%). A two-independent-sample *t* test was used for variance analysis, and Pearson's chi-squared test was used for comparing rates or proportions. Statistical significance was set at $p < 0.05$.

Model development

R software PMS package (version 4.0.3) was used to construct prediction models using four machine learning algorithms: logistic regression(LR), random forest(RF), extreme gradient boosting(XGBoost), and gradient boosting machine(GBM). Patients retained the parent study's predefined 7:3 split into training (70%) and validation (30%) cohorts; no new randomization was conducted in this subgroup. Model performance was assessed by the area under the curve of receiver operating characteristic (AUC), sensitivity, specificity, positive predictive value(PPV), negative predictive value(NPV), and Youden index. Internal validation was performed with 1000 bootstrap iterations. An AUC > 0.60 was defined as acceptable predictive performance, and values approaching 1 indicated better discrimination. Variables with $p < 0.05$ in the final model were regarded as independent risk factors. A nomogram would be constructed if logistic regression yielded superior predictive performance.

Results

Baseline characteristics

Among 2614 patients with suspected infections recruited from 72 hospitals across 16 provinces in China, a total of 2270 patients were finally enrolled in this study (Fig. 1). Baseline clinical characteristics of the included patients are summarized in Table S1. The mean APACHE II score was 15.70 (range, 8–34), and the mean patient age was 64.58 years (range, 18–103 years). The average ICU length of stay was 18.44 days. Most patients (n = 1973) were admitted to tertiary hospitals (Figure S1) and resided in North China (Figure S2). Pneumonia was the predominant infection type (83.74%), followed by bloodstream infections (17.05%) (Figure S3). *Klebsiella pneumoniae* was the most commonly isolated pathogen among all infectious strains (Figure S4).

Risk factors and prediction model of mortality

In this study, 221 patients(9.74%) with suspected infections died during their hospital stay in non-surgical ICUs. Baseline clinical characteristics of these patients are shown in Table 1. Univariate analysis demonstrated that age($p < 0.01$), region of China($p < 0.01$), hospital grades($p < 0.01$), Acute Physiology and Chronic Health Evaluation II (APACHE Ⅱ) score($p < 0.01$), diabetes mellitus($p = 0.01$), coronary atherosclerotic heart disease(CHD)($p < 0.01$), bloodstream infections($p < 0.01$), microbiological findings(*Klebsiella pneumoniae*[$p < 0.01$], *Acinetobacter baumannii*[$p < 0.01$], *Stenotrophomonas maltophilia*[$p = 0.02$], and *Aspergillus*[$p = 0.01$]), and MDR infections($p < 0.01$) were associated with mortality among patients with suspected infections upon admission to non-surgical ICUs.

Table 1
Baseline of mortality for patients with suspected infections when they admitted to non-surgical ICU

	Non-survivors	Survivors	χ^2/t	*p*
	n = 221	n = 2049		
Age	70.25 ± 16.66	63.97 ± 16.17	-5.34	< 0.01
Female	69(31.22)	740(36.12)	2.08	0.15
Region			99.81	< 0.01*
North China	20(9.05)	603(29.43)		
East China	41(18.55)	477(23.28)		
Central China	52(23.53)	443(21.62)		
Southwest China	89(40.27)	326(15.91)		
South China	8(3.62)	94(4.59)		
Northwest China	11(4.98)	75(3.66)		
Northeast China	0(0.00)	31(1.51)		
Hospital grades			15.74	< 0.01*
Grade Ⅰ A	210(95.02)	1763(86.04)		
Grade Ⅰ B	1(0.45)	110(5.37)		
Grade Ⅰ C	3(1.36)	55(2.68)		
Grade Ⅱ A	7(3.17)	121(5.91)		
APACHE Ⅱ[a]	23.39 ± 4.33	14.87 ± 3.24	-28.41	< 0.01
Hypertension	37(16.74)	261(12.74)	2.81	0.09
Diabetes Mellitus	31(14.03)	175(8.54)	7.28	0.01
CHD[b]	36(16.29)	193(9.42)	10.38	< 0.01
Cerebrovascular disease	26(11.76)	193(9.42)	1.26	0.26
COPD[c]	22(9.95)	198(9.66)	0.02	0.89
Immunosuppression	2(0.90)	41(2.00)	1.29	0.26

[a]APACHE: Acute Physiology and Chronic Health Evaluation, [b]CHD: Coronary atherosclerotic heart disease, [c]COPD: Chronic obstructive pulmonary disease. [d]ICU LOS: Intensive care unit length of stay. * Comparison between two groups, *group and*group. The difference between groups was statistically significant ($p < 0.01$ or $p < 0.05$).

	Non-survivors	Survivors	χ^2/t	*p*
	n = 221	n = 2049		
Pneumonia	194(87.78)	1707(83.31)	2.93	0.09
Bloodstream infections	56(25.34)	331(16.19)	11.90	< 0.01
Intro-abdominal infections	20(9.05)	188(9.18)	0.01	0.95
Urinary infections	23(10.41)	166(8.10)	1.39	0.24
Skin infections	7(3.17)	40(1.95)	1.45	0.23
Central system infections	6(2.71)	36(1.76)	1.01	0.32
Pelvic infections	1(0.45)	7(0.34)	0.07	0.79
Klebsiella pneumoniae	44(19.91)	175(8.54)	29.58	< 0.01
Acinetobacter baumannii	35(15.84)	140(6.83)	22.73	< 0.01
Stenotrophomonas maltophilia	7(3.17)	25(1.22)	5.44	0.02
Escherichia coli	9(4.07)	75(3.66)	0.10	0.76
Pseudomonas aeruginosa	11(4.98)	113(5.51)	0.11	0.74
Staphylococcus aureus	2(0.90)	43(2.10)	1.46	0.23
Staphylococcus epidermidis	3(1.36)	20(0.98)	0.29	0.59
Enterococcus faecium	4(1.81)	24(1.17)	0.67	0.41
Candida albicans	11(4.98)	77(3.76)	0.80	0.37
Aspergillus	7(3.17)	21(1.02)	7.52	0.01
Virus	1(0.45)	8(0.39)	0.02	0.89
Multidrug resistant infections	73(33.03)	310(15.13)	45.58	< 0.01
ICU LOS[d](day)	15.76 ± 14.02	18.73 ± 16.37	2.55	0.16

[a]APACHE: Acute Physiology and Chronic Health Evaluation, [b]CHD: Coronary atherosclerotic heart disease, [c]COPD: Chronic obstructive pulmonary disease. [d]ICU LOS: Intensive care unit length of stay. * Comparison between two groups, *group and*group. The difference between groups was statistically significant ($p < 0.01$ or $p < 0.05$).

Four machine learning algorithms were employed to construct mortality predictive models using the above significant variables. The entire cohort was randomly divided into a training set (1589 patients) and a validation set (681 patients). In the validation set, the AUC values for patients with suspected infections upon admission to non-surgical ICUs were 0.93(0.89–0.95) for LR, 0.87(0.84–0.93) for RF, 0.88(0.84–0.93) for XGB, and 0.85(0.82–0.89) for GBM(Figure.2). Threshold-specific performance

metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Youden's index, are summarized in Table 2. Given its optimal and robust discriminative performance, the LR model was selected aas the final predictive model. A visual nomogram was further established based on the LR algorithm to facilitate clinical mortality risk stratification (Fig. 3). The model identified two independent risk factors: APACHE II score (1.91, 95CI% 1.75– 2.08, $p < 0.01$), and *Klebsiella pneumoniae* infection(2.34, 95CI% 1.10–4.97, $p = 0.03$)(Table S2). Decision curve analysis(DCA) plots for model clinical utility are displayed in the Additional file(Figure S5).

Table 2
Analysis of the different indicators among different machine learning models of mortality

Prediction models	AUC*	Sensitivity	Specificity	PPV*	NPV*	Youden's index
Logistic regression (LR)	0.93	0.76	0.94	0.33	0.99	0.70
Random forest (RF)	0.87	0.36	0.99	0.79	0.96	0.35
Extreme gradient boosting (XGBoost)	0.88	0.64	0.95	0.40	0.98	0.59
Gradient boosting machine (GBM)	0.85	0.47	0.88	0.26	0.95	0.35

*AUC: area under the curve of receiver operating characteristic, PPV: positive predictive value, NPV: negative predictive value.

Risk factors and prediction model of multidrug resistant infections

The overall incidence of MDR infection was 18.87% across all infection episodes. Univariate analysis identified 23 significant correlates of MDR infection among non-surgical ICU patients with suspected infection (Table 3), including: age($p < 0.01$), female($p = 0.03$), region($p < 0.01$), hospital grades($p < 0.01$), APACHE Ⅱ($p < 0.01$), Cerebrovascular disease($p < 0.01$), infections site(pneumonia[$p < 0.01$], bloodstream infections[$p < 0.01$], urinary infections[$p < 0.01$], skin infections[$p = 0.01$], and central system infections[$p < 0.01$]), and the presence of specific pathogens (*Klebsiella pneumoniae*[$p < 0.01$], *Acinetobacter baumannii*[$p < 0.01$], *Stenotrophomonas maltophilia*[$p < 0.01$], *Escherichia coli*[$p < 0.01$], *Pseudomonas aeruginosa*[$p < 0.01$], *Staphylococcus aureus*[$p < 0.01$], *Staphylococcus epidermidis*[$p < 0.01$], and *Enterococcus faecium*[$p < 0.01$]).

Table 3
Baseline of multidrug-resistant infections for patients with suspected infections when they admitted to non-surgical ICU

	Multidrug resistance	**Non-multidrug resistance**	χ^2/t	*p*
	n = 383	**n = 1887**		
Age	68.79 ± 15.10	63.73 ± 16.44	-5.67	< 0.01
Female	118(30.81)	691(36.62)	4.69	0.03
Region			65.49	< 0.01
North China	51(13.32)	572(30.31)		
East China	103(26.89)	415(21.99)		
Central China	88(22.98)	407(21.57)		
Southwest China	105(27.41)	310(16.43)		
South China	21(5.48)	81(4.29)		
Northwest China	15(3.92)	71(3.77)		
Northeast China	0(0.00)	31(1.64)		
Hospital grades			13.08	< 0.01
Grade □A	335(87.47)	1638(86.80)		
Grade □B	9(2.35)	102(5.41)		
Grade □C	7(1.83)	51(2.70)		
Grade □A	32(83.55)	96(5.09)		
APACHE □[a]	16.78 ± 4.67	15.48 ± 4.07	-5.55	< 0.01
Hypertension	60(15.67)	238(12.61)	2.60	0.11
Diabetes Mellitus	37(9.66)	169(8.96)	0.19	0.66
CHD[b]	42(10.97)	187(9.91)	0.39	0.53
Cerebrovascular disease	54(14.10)	165(8.74)	10.48	< 0.01
COPD[c]	29(7.57)	191(10.12)	2.37	0.12
Immunosuppression	8(2.09)	35(1.85)	0.09	0.76

[a]APACHE: Acute Physiology and Chronic Health Evaluation, [b]CHD: Coronary atherosclerotic heart disease, [c]COPD: Chronic obstructive pulmonary disease. * Comparison between two groups, *group and*group. The difference between groups was statistically significant ($p < 0.01$ or $p < 0.05$).

	Multidrug resistance	Non-multidrug resistance	χ^2/t	*p*
	n = 383	n = 1887		
Pneumonia	349(91.12)	1552(82.25)	18.43	< 0.01
Bloodstream infections	120(31.33)	267(14.15)	66.47	< 0.01
Intro-abdominal infections	39(10.18)	169(8.96)	0.58	0.45
Urinary infections	73(19.06)	116(6.15)	69.55	< 0.01
Skin infections	15(3.92)	32(1.70)	7.74	0.01
Central system infections	17(4.44)	25(1.32)	17.00	< 0.01
Pelvic infections	0(0.00)	8(0.42)	1.63	0.20
Klebsiella pneumoniae	130(33.94)	89(4.72)	311.98	< 0.01
Acinetobacter baumannii	148(38.64)	27(1.43)	619.62	< 0.01
Stenotrophomonas maltophilia	18(4.70)	14(0.74)	35.88	< 0.01
Escherichia coli	43(11.23)	41(2.17)	73.25	< 0.01
Pseudomonas aeruginosa	71(18.54)	53(2.81)	152.53	< 0.01
Staphylococcus aureus	15(3.92)	30(1.59)	8.87	< 0.01
Staphylococcus epidermidis	15(3.92)	8(0.42)	38.72	< 0.01
Enterococcus faecium	22(5.74)	6(0.32)	76.95	< 0.01
Candida albicans	21(5.48)	67(0.32)	3.19	0.07
Aspergillus	5(1.31)	23(1.22)	0.02	0.89

[a]APACHE: Acute Physiology and Chronic Health Evaluation, [b]CHD: Coronary atherosclerotic heart disease, [c]COPD: Chronic obstructive pulmonary disease. * Comparison between two groups, *group and*group. The difference between groups was statistically significant ($p < 0.01$ or $p < 0.05$).

In the validation set, the LR model yielded an AUC of 0.86 (95% CI: 0.82–0.89) for MDR infection prediction, followed by RF (AUC = 0.86, 95% CI: 0.83–0.90), XGBoost (AUC = 0.85, 95% CI: 0.81–0.88), and GBM (AUC = 0.85, 95% CI: 0.82–0.89) (Fig. 4). The LR model was selected as the final predictive model due to its optimal AUC and Youden's index (Table 4). A MDR-prediction nomogram was established based on nine independent risk factors (Fig. 5): bloodstream infections(1.68, 95%CI 1.14–2.48, $p = 0.01$), urinary infections(1.77, 95%CI 1.07–2.94, $p = 0.03$), central system infections(3.16, 95%CI 1.17–8.48, $p = 0.02$), *Klebsiella pneumoniae* infections(17.81, 95%CI 11.68–27.16, $p < 0.01$), *Stenotrophomonas maltophilia* infections(10.37, 95%CI 3.24–33.18, $p < 0.01$), *Escherichia coli* infections(8.32, 95%CI 4.31–16.06, $p < 0.01$), *Pseudomonas aeruginosa* infections(14.28, 95%CI 8.01–

25.43, $p < 0.01$), *Enterococcus faecium* infections(23.85, 95%CI 8.06–70.57, $p < 0.01$), and *Staphylococcus epidermidis* infections(16.46, 95%CI 5.52–49.12, $p < 0.01$). Decision curve analysis was performed to evaluate the clinical applicability of the model, with results shown in Additional file(Figure S6).

Table 4
Analysis of the different indicators among different machine learning models of multidrug resistant infections

Prediction models	AUC*	Sensitivity	Specificity	PPV*	NPV*	Youden's index
Logistic regression (LR)	0.86	0.51	0.88	0.43	0.91	0.39
Random forest (RF)	0.86	0.44	0.85	0.63	0.93	0.29
Extreme gradient boosting (XGBoost)	0.85	0.57	0.90	0.52	0.96	0.47
Gradient boosting machine (GBM)	0.85	0.83	0.79	048	0.95	0.62

*AUC: area under the curve of receiver operating characteristic, PPV: positive predictive value, NPV: negative predictive value.

Discussion

Our secondary analysis revealed notable differences in risk factors between non-surgical patients and the overall parent cohort. In the full cohort, surgery, APACHE II score and bloodstream infection predicted mortality, whereas only APACHE II score and *Klebsiella pneumoniae* infection were independent mortality predictors among non-surgical patients. For MDR infection, the parent model identified APACHE II score, bloodstream infection, urinary infection, *K. pneumoniae*, *A. baumannii*, and *E. faecium*. In contrast, the non-surgical model incorporated bloodstream infection, urinary infection, central nervous system infection, and a broader spectrum of pathogens, including *K. pneumoniae*, *S. maltophilia*, *E. coli*, *P. aeruginosa*, *E. faecium*, and *S. epidermidis*, while *A. baumannii* was no longer retained. These distinctions likely reflect divergent host characteristics, infection sources and pathogen exposure between surgical and medical critically ill patients, supporting the value of subgroup-specific risk stratification.

The mortality rate was 30–50% for patients suffering from sepsis all around world, with rapid onset, deteriorating progress, and difficult to reverse[23]. APACHEII scores help identify disease severity with a point score from 0 to 71, comparing acute physiological score, the patient's age, and chronic health status within the first 24 hour of admission[24]. However, although the APACHE II score can predict mortality in patients with most acute medical conditions, it tends to underestimate the mortality risk among those with sepsis and stroke. The potential reasons for this underestimation may include a greater severity of illness among patients admitted to the internal medical intensive care unit and delayed medical seeking behavior, multidrug-resistant organism[25]. In our study, APACHE II score is an independent risk factor of mortality for patients within suspected infections when they admitted to non-surgical ICU.

Klebsiella pneumoniae is one of the most common gram-negative bacteremia[26], and the reported mortality rate of *Klebsiella pneumoniae* varied widely from 11% to 81% in most studies[27]. It is important to accurately estimating the mortality rate of *Klebsiella pneumonia* for clarifying the clinical characteristics of this infectious disease, providing accurate prognostic information, and optimizing prevention and control strategies. Although numerous studies and clinical guidelines have been implemented to sustain progress in combating antimicrobial resistance, the prevalence of extended-spectrum β-lactamases(ESBL)-producing and carbapenem-resistant *Klebsiella pneumoniae*(CRKP) has risen sharply. The incidence of CRKP strains has increased significantly over time. Most studies have demonstrated that patients with ESBL-producing or CRKP bacteremia exhibit substantially higher mortality rates than those infected with non-ESBL-producing or non-CRKP strains[27–29]. *Klebsiella pneumonia is also an* independent risk factor of mortality for patients within suspected infections when they admitted to non-surgical ICU in this study.

The emergence of MDR infections is a threat in ICUs and high rates of drug resistant pathogens have been reported globally. There is an alarming increase in drug resistant Gram-negative infections[30].In 2017, the World Health Organization classified carbapenem-resistant *Enterobacteriaceae*, carbapenem-resistant *Pseudomonas aeruginosa*(CRPA), and carbapenem-resistant *Acinetobacter baumannii*(CRAB) as critical-priority pathogens[31]. Carbapenem resistance may result from the hydrolysis of the β-lactam ring by specific carbapenemases, or from altered membrane permeability caused by efflux pump overexpression or porin mutations, accompanied by the production of extended-spectrum β-lactamases (ESBLs). Carbapenemases belong to Ambler class A or class D serine β-lactamases, as well as Ambler class B metallo-β-lactamases (MBLs)[32, 33]. Patients with infections of *Klebsiella pneumonia*, *Stenotrophomonas maltophilia*, *Escherichia coli*, and *Pseudomonas aeruginosa* are more prone to developing multidrug-resistant bacterial infections in non-surgical ICU in our study.

Machine learning is a branch of artificial intelligence that learns from data through computational modeling. Advanced models can capture high-order relationships between covariates and clinical outcomes in large medical datasets, making them widely applicable to complex medical problems and generally outperforming traditional statistical methods, especially for big healthcare data analysis[34]. Notably, machine learning algorithms exempt strict statistical assumptions such as independent observations and multicollinearity. Multiple algorithms including random forests, Bayesian networks and support vector machines are available for predictive modeling. Still, it remains uncertain which algorithms are superior for disease prediction. In our study, LR as conventional statistical methods, performed best among all machine learning models.

Most previous predictive models for ICU infections were limited to a single pathogen (e.g. CRE, CRAB) or a single infection site(e.g. ventilator-associated pneumonia, bloodstream infection)[17, 27, 35]. This study has the following strengths: broad coverage to simultaneously predict all-cause mortality and MDR infection; multicenter representativeness with data collected from 72 ICUs across 16 provinces in China, which improves the external validity of the model; and high clinical practicability, as the selected variables (APACHE II score, infection site, and common pathogens) are all routinely recorded clinical

data in ICUs. However, consistent with most existing similar models, the present model still has certain limitations in predicting MDR infections, with an AUC of 0.86. This may be partly attributed to the failure to incorporate antibiotic exposure history, local prevalence of drug-resistant pathogens, and host genetic factors.

Several limitations still need to be addressed in our study. Firstly, we were unable to calculate the precise rate of suspected infections in the source population due to incomplete screening data for all ICU admissions and the results may not be applicable to surgical ICUs, pediatric ICUs, or emergency departments. Secondly, it is difficult to differentiate between community-acquired and hospital-acquired infections in newly admitted patients with suspected infections. Thirdly, MDR was defined as acquired non-susceptibility to at least one agent in three or more antimicrobial categories, without distinguishing between inducible resistance and intrinsic resistance. Despite these limitations, this is the first nationwide study in China to employ machine learning algorithms for identifying independent risk factors for both mortality and MDR infections among patients with suspected infections on ICU admission.

Conclusion

his secondary analysis of a national multicenter cohort developed and validated a logistic regression model for predicting in-hospital mortality and multidrug-resistant infection among non-surgical ICU patients with suspected infection. Distinct risk factor profiles were identified in this medical subgroup compared with the overall cohort. APACHE II score and *Klebsiella pneumoniae* infection were independent predictors of mortality, while multiple infection sites and a broader spectrum of pathogens were associated with MDR infection. These findings highlight the heterogeneity of risk factors between surgical and non-surgical critically ill patients. The present model may facilitate individualized risk stratification, support early recognition of high-risk patients, and guide antimicrobial stewardship and infection control strategies for non-surgical critically ill patients with suspected infection.

Abbreviations

ICU: intensive care unit, APACHE: acute physiology and chronic health evaluation, CHD: coronary atherosclerotic heart disease, COPD: chronic obstructive pulmonary disease, ICU LOS: intensive care unit length of stay, MDR: multidrug resistant, EDC: electronic data capture. LR: logistic regression, RF: random forest, XGBoost: extreme gradient boosting, GBM: gradient boosting machine, ROC: receiver operating characteristic, AUC: area under the curve of receiver operating characteristic, PPV positive predictive value, NPV negative predictive value, DCA decision curve analysis.

Declarations

Author contributions

SY and HZ designed the study. SY, YS and TW participated in the literature search, analysis of data, as well as manuscript writing. YS participated in the literature search and data analysis and YA revised the

manuscript. HZ participated in the data analysis and FZ revised the manuscript. HZ had made contributions to the acquisition, analysis of data. HZ is responsible for ensuring that all listed authors have approved the manuscript before submission.

Funding

Beijing Natural Science Foundation(L244067) and Beijing Key Clinical Specialty Outstanding Project.

Data availability statement

The data are available from the corresponding author on reasonable request. But the datasets are not publicly available due to privacy or ethical restrictions.

Ethics approval and consent to participate

This study was performed in line with the principles of Declaration of Helsinki. Approval was obtained from the medical ethics committee of Peking University People's Hospital (2021PHB020-001, Beijing, China). Written informed consent was obtained from all participants prior to the enrollment of this study.

Consent publication

Not applicable.

Competing interests

The authors declared no potential conflict of interest concerning the research, authorship, and/or publication of this article.

Acknowledgements

Thank you for all patients for their participation in this study and all participants for accomplishing this study. We thank Hua Rong for her electronic information technology. For continuous support, assistance, and cooperation, we thank Chen Li(QiLu Hospital of Shandong University), Li Kong(Affiliated Hospital of Shandong University of Traditional Chinese Medicine), Tingfa Zhou(Lin Yi People'Hospital), Haiming Jiang(The Second of Clinical Medicine of Binzhou Medical university), Xugang Li(Ri Zhao LinYi People'Hospital), Yun Jin(Zi Bo First Hospital), Qiang Sun(Affiliated Hospital of Jining Medical University), Long Qin(Tsinghua University Yuquan Hospital), Qin Dong(The First Affiliated Hospital of Xi'An Jiao Tong University), Xiaochuang Wang(The Second Affiliated Hospital of Xi'An Jiao Tong University), Yong Li(Bao Ji People's Hospital), Chuan Guo(School of Clinical Medicine& The First Affiliated Hospital of Chengdu Medical College), Xiaobo Huang, Xiaoqin Zhang(Sichuan Academy of Medical Sciences Sichuan Provincial People's Hospital), Chuan Zhang(Cheng Du third People's Hospital), Hongde Chen(Cheng Du Tenth People's Hospital), Yu Du(Sichuan University Fourth Hospital), Hang Xu(The First Affiliated Hospital of ShiHeZi University), Xianying Lei(Affiliated Hospital of Southwest Medical University), Junjing Luo(The First Affiliated Hospital of Xinxiang medical university), Weidong

Guo(Xin Xiang Central Hospital), Yinjiang Change(Pu Yang People's Hospital), Yiyi Sun(He Bi People's Hospital), Hua Li(The Second Affiliated Hospital of Henan Chinese Medical University), Zhengrong Mao(The First Affiliated Hospital of Henan Chinese Medical University), Li Chen(Affiliated Hospital of North Sichuan Medical College), Faming He(Chest Hospital of Zhengzhou University),Chaogui Zhang(Yi Bin Second People's Hospital), Yonghui Fan(General Hospital of Pingmei Shenma Group), Jianghua Zhu(The First Affiliated Hospital of Ningbo Hospital),Peili Chen(Shang Qiu First People's Hospital), Yong Cui(Chongqing University Three Gorges Hospital), Guolong Cai(Zhe Jiang Hospital), Ken Chen(Karamay Central Hospital of Xinjiang), Jian Chen(Affiliated Hospital of Xinjiang Chinese Medical University), Xiangcheng Zhang(Huai An First People's Hospital), Tianzeng Zhang(An Yang District Hospital), Yuling Liu(Jiaozuo Coal Industry [Group] Co., Ltd. Central Hospital), Chao Jia(Mian Yang Central Hospital), Lifang Wu(Da Feng People's Hospital), Jiangquan Yu(Northern Jiangsu People's Hospital), Fang Huang(The First Affiliated Hospital of Soochow University), Yishan Zheng(Nan Jing Sencod Hospital), Xingui Dai(The First Affiliated Hospital of Xiangnan University), Jiangliang Zhou(Huai Hua First People's Hospital), Xinju Xu(Jiao Zuo Second People's Hospital), Mingfan Yang(Huo Qiu First People's Hospital), Liujiu Tian(An Qing First People's Hospital), Chunli Yang(The First Affiliated Hospital of Nanchang University), Guohui Yang, Jiangquan Fu(Affiliated Hospital of Guizhou Medical University), Donghao Wang(Tianjin Medical University Cancer Institute & Hospital), Youjie Qiao, Jing Wang(Tian Jing People's Hospital), Bin Chen(The Second Affiliated Hospital of Tianjin Medical University), Baohua Wang, Aiguo Zhang, Yinhua Wang(North China of Science and Technology University Affiliated Hospital), Huisheng Qi(Tangshan Municipal Worker's Hospital), Hongmei Qin(Yu Lin First People's Hospital), Baoshan Li(Yun Cheng Central Hospital), Liang Luo(Wu Xi Second People's Hospital) Jian Luo, Chungen Nie(Ji An People's Hospital), Shichang Deng, Yonghong Deng(Yu Du People's Hospital).

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801–10.
2. Vincent JL, Sakr Y, Singer M, Martin-Loeches I, Machado FR, Marshall JC, Finfer S, Pelosi P, Brazzi L, Aditianingsih D, et al. Prevalence and Outcomes of Infection Among Patients in Intensive Care Units in 2017. JAMA. 2020;323(15):1478–87.
3. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, McIntyre L, Ostermann M, Prescott HC, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med. 2021;49(11):e1063–143.
4. van Langeveld I, Gagnon RC, Conrad PF, Gamelli RL, Martin B, Choudhry MA, Mosier MJ. Multiple-Drug Resistance in Burn Patients: A Retrospective Study on the Impact of Antibiotic Resistance on Survival and Length of Stay. J burn care research: official publication Am Burn Association. 2017;38(2):99–105.
5. Papadimitriou-Olivgeris M, Fligou F, Bartzavali C, Zotou A, Spyropoulou A, Koutsileou K, Vamvakopoulou S, Sioulas N, Karamouzos V, Anastassiou ED, et al. Carbapenemase-producing

Klebsiella pneumoniae bloodstream infection in critically ill patients: risk factors and predictors of mortality. Eur J Clin Microbiol Infect diseases: official publication Eur Soc Clin Microbiol. 2017;36(7):1125–31.

6. Rojas LJ, Salim M, Cober E, Richter SS, Perez F, Salata RA, Kalayjian RC, Watkins RR, Marshall S, Rudin SD, et al. Colistin Resistance in Carbapenem-Resistant Klebsiella pneumoniae: Laboratory Detection and Impact on Mortality. Clin Infect diseases: official publication Infect Dis Soc Am. 2017;64(6):711–8.
7. Bertagnolio S, Dobreva Z, Centner CM, Olaru ID, Donà D, Burzo S, Huttner BD, Chaillon A, Gebreselassie N, Wi T, et al. WHO global research priorities for antimicrobial resistance in human health. Lancet Microbe. 2024;5(11):100902.
8. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, Harbarth S, Hindler JF, Kahlmeter G, Olsson-Liljequist B, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol infection: official publication Eur Soc Clin Microbiol Infect Dis. 2012;18(3):268–81.
9. Incidence of severe sepsis. and septic shock in German intensive care units: the prospective, multicentre INSEP study. Intensive Care Med. 2016;42(12):1980–9.
10. Baykara N, Akalın H, Arslantaş MK, Hancı V, Çağlayan Ç, Kahveci F, Demirağ K, Baydemir C, Ünal N. Epidemiology of sepsis in intensive care units in Turkey: a multicenter, point-prevalence study. Crit Care (London England). 2018;22(1):93.
11. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, Kadri SS, Angus DC, Danner RL, Fiore AE, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009–2014. JAMA. 2017;318(13):1241–9.
12. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):762–74.
13. Humphries R, Bobenchik AM, Hindler JA, Schuetz AN. Overview of Changes to the Clinical and Laboratory Standards Institute Performance Standards for Antimicrobial Susceptibility Testing, M100, 31st Edition. J Clin Microbiol. 2021;59(12):e0021321.
14. Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH. Radiographic and CT Features of Viral Pneumonia. Radiographics: Rev publication Radiological Soc North Am Inc. 2018;38(3):719–39.
15. Almirall J, Serra-Prat M, Bolíbar I, Balasso V. Risk Factors for Community-Acquired Pneumonia in Adults: A Systematic Review of Observational Studies. Respir Int Rev Thorac Dis. 2017;94(3):299–311.
16. Ottosen J, Evans H. Pneumonia: challenges in the definition, diagnosis, and management of disease. Surg Clin North Am. 2014;94(6):1305–17.

17. Timsit JF, Ruppé E, Barbier F, Tabah A, Bassetti M. Bloodstream infections in critically ill patients: an expert statement. Intensive Care Med. 2020;46(2):266–84.
18. Sartelli M, Chichom-Mefire A, Labricciosa FM, Hardcastle T, Abu-Zidan FM, Adesunkanmi AK, Ansaloni L, Bala M, Balogh ZJ, Beltrán MA, et al. The management of intra-abdominal infections from a global perspective: 2017 WSES guidelines for management of intra-abdominal infections. World J Emerg surgery: WJES. 2017;12:29.
19. Rowe TA, Juthani-Mehta M. Urinary tract infection in older adults. Aging health 2013, 9(5).
20. Moffarah AS, Al Mohajer M, Hurwitz BL, Armstrong DG. Skin and Soft Tissue Infections. Microbiol Spectr 2016, 4(4).
21. He T, Kaplan S, Kamboj M, Tang YW. Laboratory Diagnosis of Central Nervous System Infection. Curr Infect Dis Rep. 2016;18(11):35.
22. Mitchell C, Prabhu M. Pelvic inflammatory disease: current concepts in pathogenesis, diagnosis and treatment. Infect Dis Clin North Am. 2013;27(4):793–809.
23. Fleischmann C, Thomas-Rueddel DO, Hartmann M, Hartog CS, Welte T, Heublein S, Dennler U, Reinhart K. Hospital Incidence and Mortality Rates of Sepsis. Deutsches Arzteblatt Int. 2016;113(10):159–66.
24. Kotwal SR, Patel DV, Bamaniya CL. Prognostic Value of the Acute Physiology and Chronic Health Evaluation II (APACHE II) Score in Secondary Peritonitis: A Narrative Review. Cureus. 2026;18(3):e105546.
25. Magira EE, Islam S, Niederman MS. Multi-drug resistant organism infections in a medical ICU: Association to clinical features and impact upon outcome. Med intensiva. 2018;42(4):225–34.
26. Cubero M, Grau I, Tubau F, Pallarés R, Domínguez M, Liñares J, Ardanuy C. Molecular Epidemiology of Klebsiella pneumoniae Strains Causing Bloodstream Infections in Adults. Microb drug Resist (Larchmont NY). 2018;24(7):949–57.
27. Meng H, Han L, Niu M, Xu L, Xu M, An Q, Lu J. Risk Factors for Mortality and Outcomes in Hematological Malignancy Patients with Carbapenem-Resistant Klebsiella pneumoniae Bloodstream Infections. Infect drug Resist. 2022;15:4241–51.
28. Lee JS, Choi JY, Chung ES, Peck KR, Ko KS. Variation in the formation of persister cells against meropenem in Klebsiella pneumoniae bacteremia and analysis of its clinical features. Diagn Microbiol Infect Dis. 2019;95(3):114853.
29. Kim D, Park BY, Choi MH, Yoon EJ, Lee H, Lee KJ, Park YS, Shin JH, Uh Y, Shin KS, et al. Antimicrobial resistance and virulence factors of Klebsiella pneumoniae affecting 30 day mortality in patients with bloodstream infection. J Antimicrob Chemother. 2019;74(1):190–9.
30. Vincent JL, Rello J, Marshall J, Silva E, Anzueto A, Martin CD, Moreno R, Lipman J, Gomersall C, Sakr Y, et al. International study of the prevalence and outcomes of infection in intensive care units. JAMA. 2009;302(21):2323–9.
31. Nordmann P, Poirel L. Epidemiology and Diagnostics of Carbapenem Resistance in Gram-negative Bacteria. Clin Infect diseases: official publication Infect Dis Soc Am. 2019;69(Suppl 7):S521–8.

32. Potter RF, D'Souza AW, Dantas G. The rapid spread of carbapenem-resistant Enterobacteriaceae. Drug Resist updates: reviews commentaries Antimicrob anticancer Chemother. 2016;29:30–46.
33. Meletis G. Carbapenem resistance: overview of the problem and future perspectives. Therapeutic Adv Infect disease. 2016;3(1):15–21.
34. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA. 2018;319(13):1317–8.
35. Bin Ghaffar W, Nazir S, Siddiqui S, Abdul Ghaffar MB, Khan MF, Latif A, Cheema Z, Hanif S, Sohaib M. Association Between the Site of Infection and Mortality Analysis in Critically Ill Surgical Patients. Cureus. 2023;15(12):e50033.

Figures

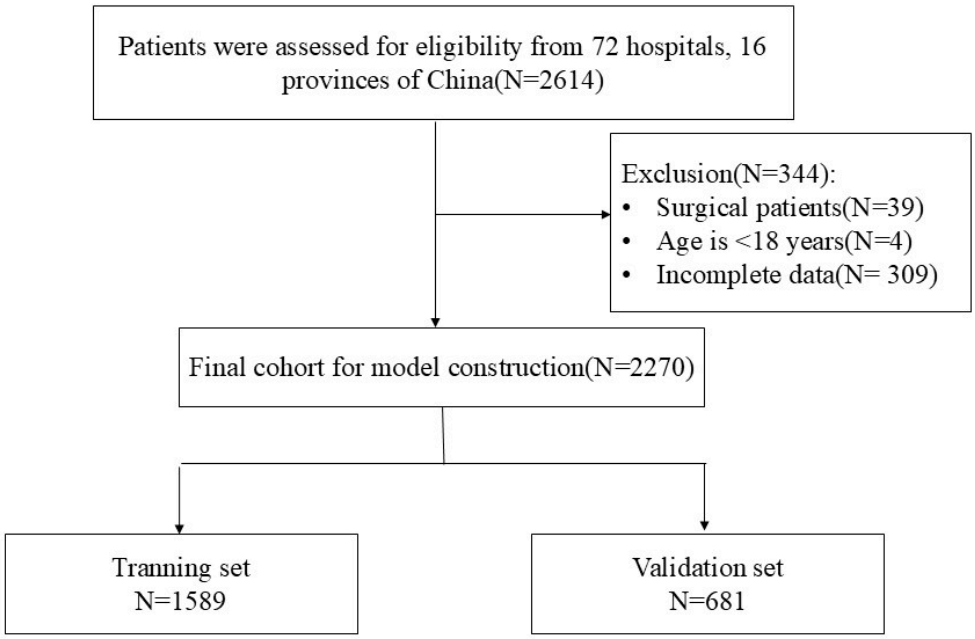


Figure 1

Flow chart of hospital recruitment and patient enrollment. Provinces in mainland China were divided into seven geographical regions. A total of 2614 patients from 72 hospitals were included in this study

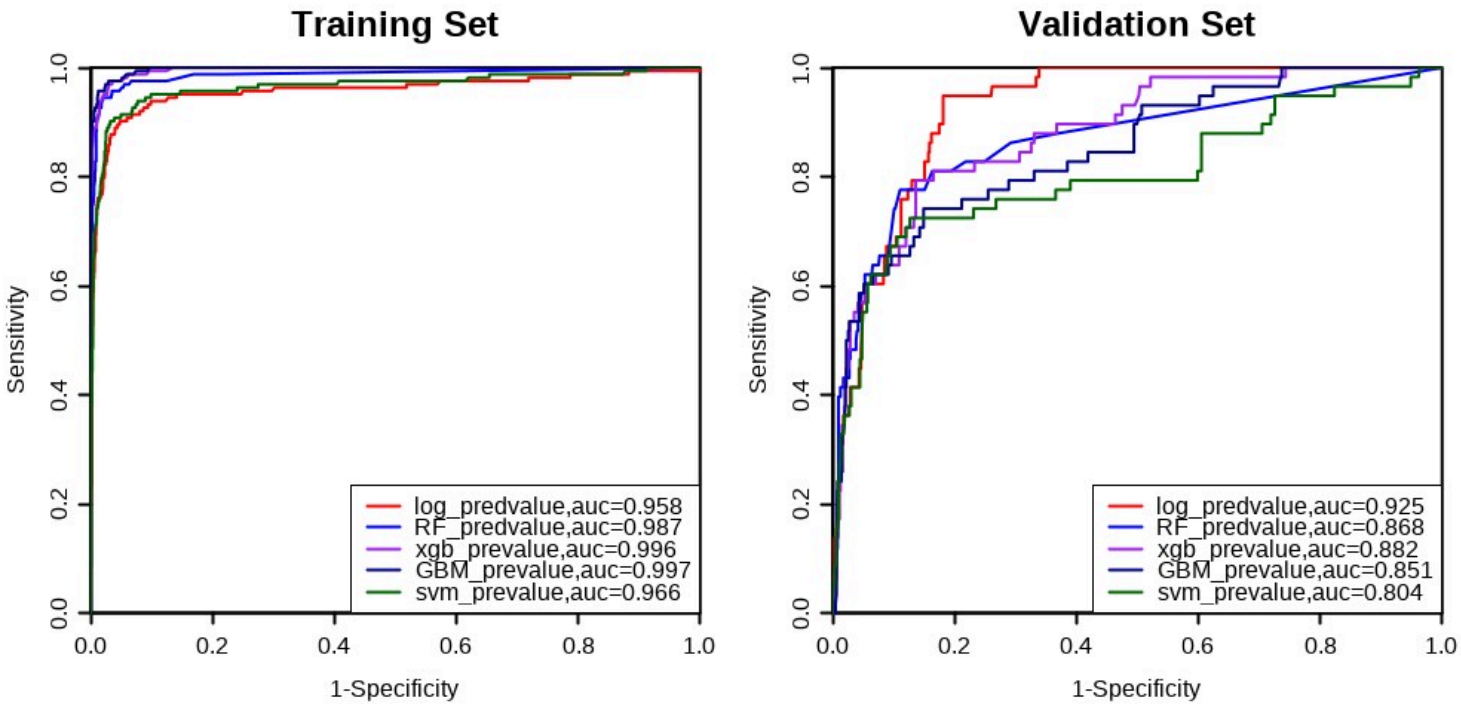


Figure 2

Receiver operating characteristic (ROC) curves of the prediction model for mortality of patients with suspected infections in non-surgical ICU conducted by machine learning algorithms

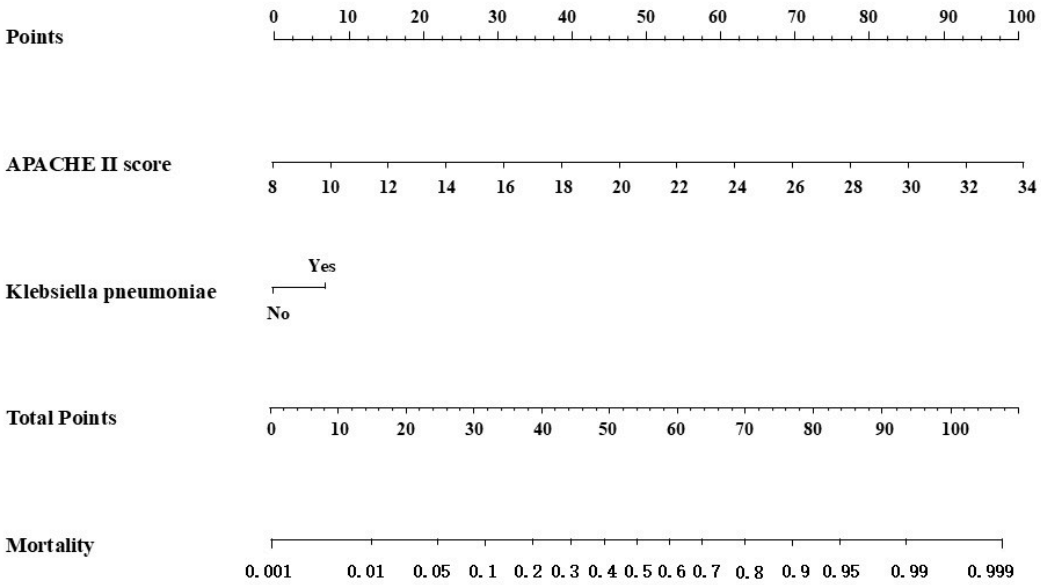


Figure 3

Nomogram for mortality. To estimate the probability of mortality, mark patient value at each axis, draw a straight line perpendicular to the point axis, and calculate the points for all variables. Then mark the sum on the total point axis and the points met the risk axis

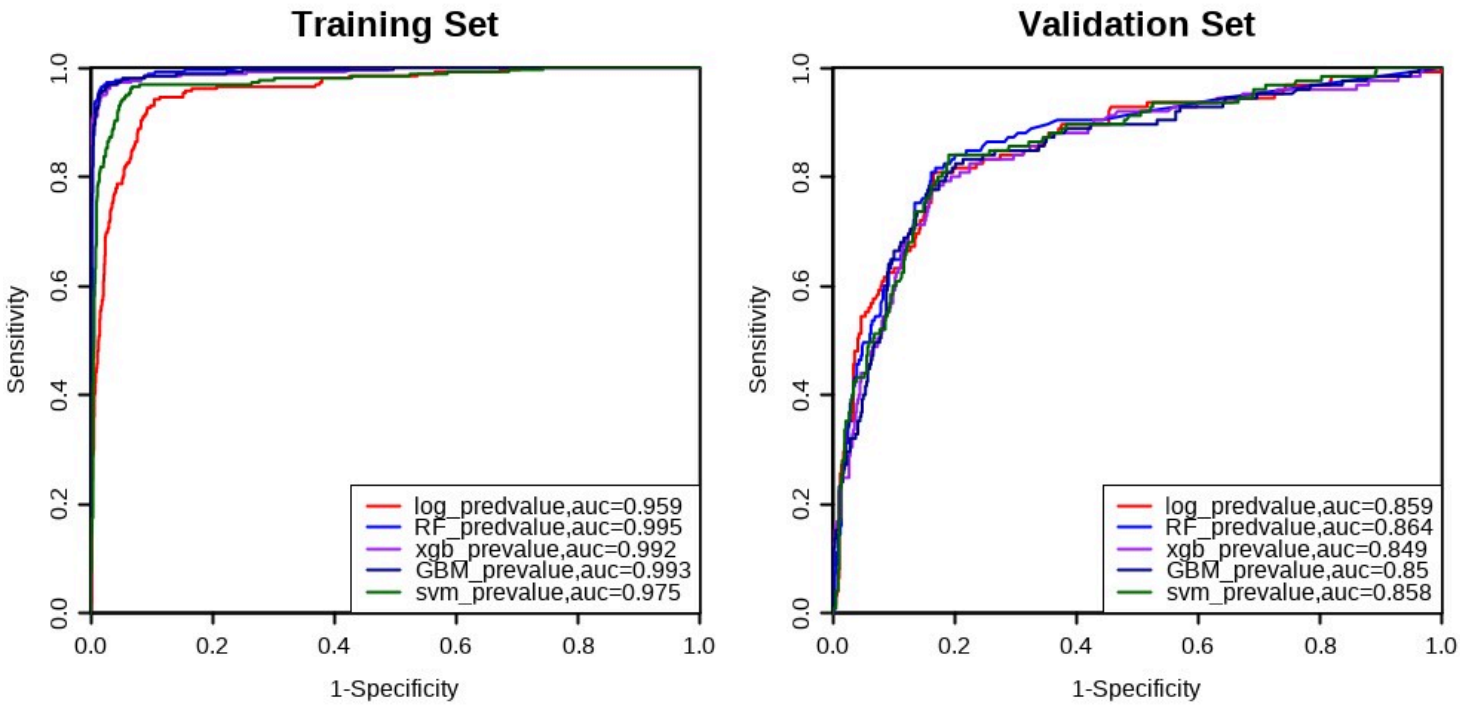


Figure 4

Receiver operating characteristic (ROC) curves of the prediction model for multidrug resistance of patients with suspected infections in non-surgical ICU conducted by machine learning algorithms

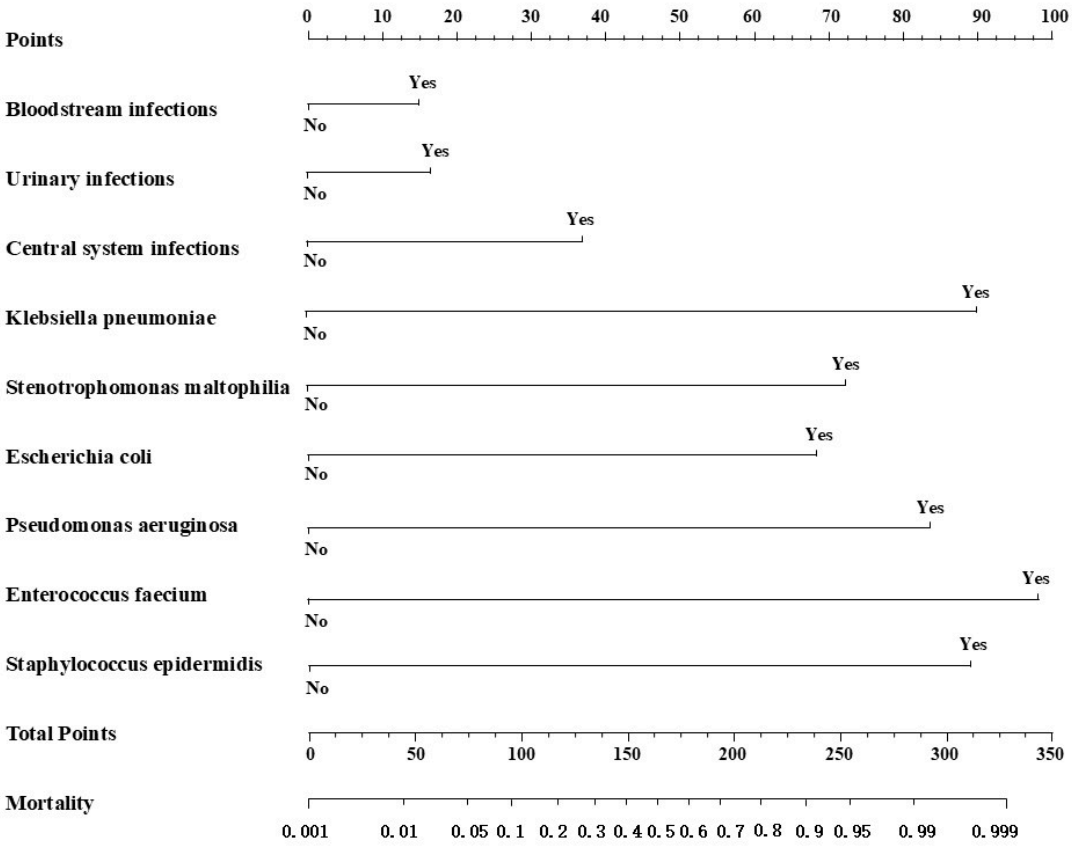


Figure 5

Nomogram for multidrug resistance. The value of variable was given a score on the point scale axis. To estimate the risk of the time of first postoperative defecation, a total score could be calculated by each axis and could be projected to the lower total point scale

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile.docx