Early Prediction of the Carbapenem Resistance Gram-negative Bacteria Carriage in Intensive Care Unit using Machine-Learning

Qiqiang Liang

Zhejiang University School of Medicine Second Affiliated Hospital

Qinyu Zhao

Healsci technology co., ltd

Xin Xu

Zhejiang University School of Medicine Second Affiliated Hospital

Yu Zhou

Zhejiang University School of Medicine second Affiliated Hospital

Man Huang (✉ huangman@zju.edu.cn)

Zhejiang University School of Medicine <https://orcid.org/0000-0002-0842-2254>

Research

Keywords: Machine learning, Carbapenem resistance gram-negative bacteria, Multi-drug resistant bacteria prediction, multivariate logistic regression, Prevention and control of nosocomial infection

Posted Date: August 25th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-60222/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

The prevention and control of carbapenem-resistance gram-negative bacteria (CR-GNB) is the difficulty and focus for clinicians in the intensive care unit (ICU). This study constructs a CR-GNB carriage prediction model in order to predict the CR-GNB incidence in one week.

Methods

The database is comprised of nearly 10,000 patients. The model is constructed by the multivariate logistic regression model and three machine learning algorithms. Then we choose the optimal model and verify the accuracy by daily predicted and recorded the occurrence of CR-GNB of all patients admitted for 4 months.

Results

There are 1385 patients with positive CR-GNB cultures and 1535 negative patients in this study. Forty-five variables have statistical significant differences. We include the 17 variables in the multivariate logistic regression model and build three machine learning models for all variables. In terms of accuracy and the area under the receiver operating characteristic (AUROC) curve, the random forest is better than XGBoost and multivariate logistic regression model, and better than decision tree model (accuracy: 84% > 82% > 81% > 72%), (AUROC: 0.9089 > 0.8947 $\approx$ 0.8987 > 0.7845). In the 4-month prospective study, 81 cases were predicted to be positive in CR-GNB culture within 7 days, 146 cases were predicted to be negative, 86 cases were positive, and 120 cases were negative, with an overall accuracy of 84% and AUROC of 91.98%.

Conclusions

Prediction models by machine learning can predict the occurrence of CR-GNB colonization or infection within a week period, and can real-time predict and guide medical staff to identify high-risk groups more accurately.

Background:

Multidrug-resistant bacteria (MDRB) infection is a global healthcare crisis which not only increases the length and cost of hospitalization but also advances the use of novel antibiotics¹. This is a vicious loop in which novel antibiotics create selective pressure on bacteria and then induce the occurrence of MDRB^{2,3}. In recent years, the main MDRBs are carbapenem resistance gram-negative bacteria (CR-GNB), including carbapenem-resistant Enterobacteriaceae, carbapenem-resistant *Acinetobacter baumannii*, and carbapenem-resistant *Pseudomonas aeruginosa*, which are on the list of priority pathogens by the World Health Organization¹. The prevention and control of CR-GNB is the difficulty and focus for clinicians in the intensive care unit (ICU). Clinicians take many measures to prevent the occurrence of CR-GNB, which also reflects the dilemma from the side⁴⁻⁷. The staff's compliance and CR-GNB unknown transmission limit the prevention and control measures.

The prevention and control of CR-GNB are mainly carried out by cutting off the route of transmission and protecting the susceptible population. The CR-GNB transmission route is difficult to monitor in daily clinical work and identification of high-risk populations is easier comparatively. Many studies focus on the high-risk factors of CR-GNB colonization and infection included carbapenem antibiotic exposure, third-generation cephalosporins exposure, invasive manipulation, mechanical ventilation and so on⁸⁻¹⁰. Targeted prevention and control measures for patients with high-risk factors seem to have good potential, but no effective tools have been developed to predict the occurrence of drug-resistant bacteria.

In the past, the multivariate logistic regression is usually used to establish clinical prediction models. However, there are many new ideas to solve the problem with the upsurge of machine learning^{11,12}. In the field of intensive care medicine, machine learning has been successfully applied in the classification and prediction of the mortality of sepsis, acute kidney injury and the mortality in ICU hospital, which is more stable and efficient than the traditional modeling¹³⁻¹⁶. At present, the prediction models using machine learning is more accurate and competitive than the traditional multivariate logistic regression model due to the advantages of large samples and novel algorithms. It is unknown whether machine learning has the same effect in the prediction of CR-GNB.

In this study, we establish a CR-GNB prediction model using our center database by machine learning with the purpose of recognition whether they were colonized or infected with CR-GNB in a week. We also have verified the accuracy of the prediction model through a small prospective study.

Methods

Study Characteristics

This study is a retrospective control study in the model establishment and prospective study in the model validation. The second affiliated hospital of Zhejiang university school of medicine is an academic teaching hospital. The general ICU has 40 monitoring beds and treats 2000 patients every year

with a full range of diseases. The general ICU database uses MIMIC database structure to collect basic patients' information, medical advice, image examination, laboratory testing, nursing and doctor documents. The database is comprised of nearly 10,000 ICU patients, and the data have been updated daily.

Model Establishment

- 1) Participants: all the patients in the general ICU database. Patients with positive culture of CR-GNB is defined as CR-GNB carriage. Exclude standard: i) patients with hospital stay less than a week; ii) patients with a positive culture of CR-GNB within 48 hours of admission; iii) patients with only one positive culture result and considering as contamination during hospitalization. See Fig. 1 for the flow chart.
- 2) Variable selection: demographic data, vital signs, basic and primary diseases, important test indicators, operations histories, and antibiotic use records in the prior month. The time range of variable collection was determined according to the positive CR-GNB culture time. i) patients with positive CR-GNB culture within a week of admission, the variable collection period is limited in 48 hours of admission. ii) patients with positive culture more than a week, the collection period is limited between admission to a week before the positive time. iii) according to the previous investigation, 87% of patients would carry CR-GNB within 2 weeks at admission, so the variable collection period of negative patients is within 1 week of admission. We select 65 variables into our study, as detailed in Table 1.

Table 1
Basic characteristics of variables

	Positive	Negative	P
N	1385	1535	
AGE	58.27	58.60	0.63
Gender (Man)	60.20%	62.21%	0.000
Weight	68.32	57.97	0.081
Single	703	875	0.001
Hospital residence history (< a month)	135	256	0.000
Primary disease	1028	1171	0.087
Infectious Diseases	163	194	/
Trauma	369	305	/
Cardio-cerebrovascular accident	247	255	/
Postoperative diseases	151	227	/
Internal medicine diseases	98	187	/
Other	357	364	/
Malignant tumor	102	150	0.034
Chronic kidney disease	71	101	0.134
Diabetes Mellitus	108	114	0.550
Hematological Disease	62	65	0.630
Cardiovascular disease	52	66	0.549
Liver cirrhosis	7	21	0.02
Chronic lung disease	7	12	0.385
CRRT	144	214	0.04
Mechanical ventilation	836	1238	0.000
Tracheostomy	127	143	0.892
Invasive catheterization	343	674	0.000
History of cephalosporins (< a month)	450	906	0.000
History of carbapenems (< a month)	388	524	0.000
History of glucocorticoids (< a month)	310	400	0.021
Operation history (< six month)	386	791	0.000
Drainage tube	533	576	0.594
Fever (temperature > 38.5°C)	679	777	0.390
High APACHE II scores (> 20 points)	594	689	0.277
APACHE II scores	18.76	20.32	0.000
Charlson scores	2.18	2.28	0.180
Hypoleukaemia	97	186	0.000
Thrombopenia	176	257	0.003

CRRT: Continuous Renal Replacement Therapy; GCS : Glasgow coma scale; SBP: Systolic pressure; WBC: White blood cell count; NEP: neutrophil percentage; HCT: Hematokrit; CRP: C-Reactive Protein; PCT: procalcitonin; PLT: Platelet count; HB: hemoglobin; CR: Creatinine; CTnl: troponin; CK-MB: Creatine kinase isoenzyme MB; ALT: Alanine transaminase; AST: Aspartate aminotransferase; TBIL: Total bilirubin; DBIL: Direct bilirubin; IBIL: Indirect bilirubin; TP: Total protein; ALB: Albumin ; PT: Prothrombin time; APTT: Activated partial thromboplastin time; FBG: Fibrinogen; GLU: Blood glucose; LAC: Lactic acid;

The maximum and minimum values of the subscript of some parameters are the maximum or minimum values in the specified collection node, which are determined according to the clinical significance.

	Positive	Negative	P
Acute kidney injury	301	424	0.001
Acute liver failure	148	255	0.000
Pulse _{max}	109.11	115.79	0.000
Respiratory rate _{max}	24.58	27.03	0.000
Temperature _{max}	38.33	38.47	0.000
GCS _{min}	8.38	10.64	0.000
SBP _{max}	166.16	170.75	0.000
WBC _{max}	14.427	16.46	0.000
WBC _{min}	8.67	7.23	0.000
NEU _{max}	89.58	91.51	0.000
HCT _{max}	44.04	48.02	0.009
CRP _{max}	118.32	132.79	0.000
PCT _{max}	5.51	7.36	0.000
PLT _{min}	137.96	114.23	0.000
HB _{min}	78.06	73.66	0.000
CR _{max}	104.40	131.68	0.000
CTnI _{max}	3.73	12.00	0.086
CK-MB _{max}	42.61	62.94	0.000
ALT _{max}	99.73	148.24	0.003
AST _{max}	197.71	354.71	0.001
TBIL _{max}	27.07	35.63	0.000
DBIL _{max}	8.78	12.44	0.000
IBIL _{max}	18.98	24.66	0.000
TP _{max}	61.96	65.71	0.000
ALB _{min}	28.59	26.58	0.000
PT _{max}	17.44	18.80	0.000
APTT _{max}	51.78	57.47	0.000
FBG _{max}	4.94	5.39	0.000
D-Dimer _{max}	7043.89	8052.42	0.000
GLU _{max}	12.72	14.49	0.001
LAC _{max}	3.41	4.39	0.000
CRRT: Continuous Renal Replacement Therapy; GCS : Glasgow coma scale; SBP: Systolic pressure; WBC: White blood cell count; NEP: neutrophil percentage; HCT: Hematokrit; CRP: C-Reactive Protein; PCT: procalcitonin; PLT: Platelet count; HB: hemoglobin; CR: Creatinine; CTnI: troponin; CK-MB: Creatine kinase isoenzyme MB; ALT: Alanine transaminase; AST: Aspartate aminotransferase; TBIL: Total bilirubin; DBIL: Direct bilirubin; IBIL: Indirect bilirubin; TP: Total protein; ALB: Albumin ; PT: Prothrombin time; APTT: Activated partial thromboplastin time; FBG: Fibrinogen; GLU: Blood glucose; LAC: Lactic acid;			
The maximum and minimum values of the subscript of some parameters are the maximum or minimum values in the specified collection node, which are determined according to the clinical significance.			

3) Statistical methods: we use MySQL and Navicat to complete data collection with structured query language. Data analysis are accomplished through R and Rstudio with related packets. Multivariate logistic regression and three machine learning algorithms, as decision tree, random forest, and XGBoost, are selected to establish the models. The main R packets include "glm", "rpart", "randomForest", "xgboost" and "rattle". We delete independent variables with the missing more than 50% and the other missing values are processed by multiple interpolations. In univariate analysis, numerical variables are tested by independent sample T-test, dichotomous variables are completed by the chi-square test, and a P value less than 0.05 is considered as statistical significance. All samples are grouped into 70% training set, 15% validation set, and 15% test set. Multivariate logistic regression uses the step-by-step decreasing method to adjust the parameters. Five hundred trees are constructed and the exhaustive method is used to adjust the parameters in the Random forest, and the importance of measurable variables is presented by the corresponding visual package. The specific and task parameters of the XGBoost linear rise are adjusted according to the performance of the model. The evaluation parameters include sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic (AUROC) curves.

Prospective Study

1) Study duration: 09/01/2019 to 12/31/2019.

2) Protocol: i) determination of the optimal prediction model through model evaluation; ii) application of the model for all patients admitted to ICU to make daily predictions. iii) termination of prediction with the following conditions: a) when prediction results suggest high risk (CR-GNB carriage will positive within 7 days), stop prediction and enter the 7-day observation period; b) when patients leave the ICU ward, including transfer, discharge or death. iv) analysis of the CR-GNB carriage within 7 days compare with the predicted results and calculation of the relevant evaluation indexes and evaluation the model performance. v) all the predicted results are kept secret to clinicians, and this study does not interfere with clinical decision-making.

Results:

Prediction models

A total of 3303 patients are included in the study, of which 1768 patients are culture positive. According to the exclusion criteria, we include 1385 CR-GNB carriage patients and 1535 negative patients. Among the positive CR-GNB carriages, carbapenem-resistant *Acinetobacter baumannii* accounts for 59.16%, carbapenem-resistant *Klebsiella pneumoniae* accounts for 21.21%, and carbapenem-resistant *Pseudomonas aeruginosa* accounts for 18.25%. In terms of the distribution of cultural sites, sputum culture accounts for 80.95%. As far as the types of diseases, the center is dominated by multiple injuries, especially severe craniocerebral trauma, followed by various infectious diseases, and cardio-cerebrovascular accidents. See Table 5 for details.

In univariate analysis, there are 45 variables with statistical differences as Table 1. In the multivariate logistic regression model, a stable model is obtained by incorporating 17 variables, see as Table 2. These risk factors includes gender, invasive catheterization, mechanical ventilation and hospital residence history over the past month, vital signs including systolic blood pressure, respiratory rate and glasgow coma scale, laboratory indicators including white blood cell count, hematocrit, C-reaction protein, direct bilirubin, total protein, and fibrinogen. The use of cephalosporins and carbapenems is also a high-risk factor. In the models established by three common machine learning methods, random forest presents the important characteristics of model parameters, as Fig. 2, in which hospital residence history over the past month, total protein and respiratory rate are the main risk factors which have much overlap with the multivariate logistic regression model.

Table 2
Parameters in the multivariate logistic regression model

Parameters	P
Gender	0.000
Invasive catheterization	0.000
Drainage tube	0.006
Mechanical ventilation	0.000
Hospital residence history (< a month)	0.000
History of cephalosporins (< a month)	0.000
Hypoleukaemia	0.000
Respiratory rate _{max}	0.000
GCS _{min}	0.000
SBP _{max}	0.014
APACHE II scores	0.000
WBC _{max}	0.000
HCT _{max}	0.003
CRP _{max}	0.005
DBIL _{max}	0.000
TP _{max}	0.000
FBG _{max}	0.005

The multivariate logistic regression model is well fitted and has high prediction accuracy (84% in validation and 81% in test). This traditional model has great specificity (91.27% in validation and 91.73% in test) and a positive predictive value (87.21% in validation and 85.14% in test). Except for the decision tree model, the fitting effect of random forest and XGBoost are similar, the sensitivity of XGBoost (77.23% in validation and 74.47% in test) is similar with that of random forest (70.30% in validation and 77.66% in test) and the specificity of random forest is the best (92.86% in validation and 87.97% in test). In terms of accuracy and AUROC curve, random forest (0.9182 in validation and 0.9089 in test) is better than XGBoost (0.9127 in validation and 0.8947 in test), and multivariate logistic regression model (0.9073 in validation and 0.8983 in test), and is generally better than decision tree model (0.8178 in validation and 0.7845 in test), see as Table 3, Fig. 3 and Fig. 4.

Table 3
Evaluation Result of machine learning and multiple logistic regression model validation set and test set

	Validation set				Test set							
	Sensitivity	specificity	PPV	NPV	Accuracy	AUROC	Sensitivity	specificity	PPV	NPV	Accuracy	AUROC
Decision tree	67.60	81.78	77.84	72.73	0.75	0.8178	62.86	80.35	74.58	70.23	0.72	0.7845
Random forest	70.30	92.86	88.75	79.59	0.83	0.9182	77.66	87.97	82.02	84.78	0.84	0.9089
XGBoost	77.23	89.68	85.71	83.09	0.84	0.9127	74.47	87.97	81.40	82.98	0.82	0.8947
multivariate logistic regression	74.26	91.27	87.21	81.56	0.84	0.9073	67.02	91.73	85.14	79.74	0.81	0.8983
PPV: positive likelihood ratio; NPV: negative predictive value; AUROC: area under the receiver operating characteristic;												

According to the effect of the model, we finally choose the random forest as the optimal prediction model for the prospective study.

Prospective Study

In the 4-month prospective study, a total of 673 patients are treated in our center, and the total number of prediction period by the model is 4553. Among them, 431 patients with hospitalization days less than 7 days and 36 patients with positive in 48 hours as admission are excluded. It is finally predicted that 81 cases are positive and 146 cases are negative. There are 86 positive cases and 120 negative cases after 7 days of observation. The sensitivity is

72.28%, lower than the performance of the test set, but the specificity is as high as 93.65%, better than that of the previous test set. Therefore our model is computationally efficient and achieves high accuracy of 84% and AUROC of 91.98%. It is equivalent to the accuracy of the validation set and the test set. In the verification stage, the proportion of sputum samples was still the largest, and the proportion of *Pseudomonas aeruginosa* was the highest, more detail can be seen in Tables 4 and 5.

Table 4
Prospective study results

Duration	2019.09.01-2019.12.31					
Total patients	673					
Total number of predictions	4553					
Number of patients hospitalized < 7 days	431					
Numbers of positive active surveillance patients at admission	36					
Predictive positive patients	81					
Predictive negative patients	146					
Actual positive patient	86					
Actual negative patient	120					
	Sensitivity	specificity	PPV	NPV	Accuracy	AUROC
Evaluation Result	72.28	93.65	90.12	80.82	0.84	0.9198
PPV: positive likelihood ratio; NPV: negative predictive value; AUROC: area under the receiver operating characteristic;						

Table 5
The positive culture sites and bacterial distribution

		retrospective control study	Prospective study
Total cases		1385	86
Bacterial distribution	CRAB	820(59.16%)	26(30.23%)
	CRKP	294(21.21%)	26(30.23%)
	CRPA	253(18.25%)	31(36.05%)
	Other CRE	22(1.59%)	3(3.49%)
Sites distribution	Sputum	1122(80.95%)	73(84.88%)
	Blood	97(7.00%)	0
	Stool	49(3.54%)	9(10.47%)
	Urine	40(2.89%)	0
	Others	77(5.56%)	4(4.65%)
CRAB: carbapenem-resistant <i>Acinetobacter baumannii</i> ; CRKP: carbapenem-resistant <i>Klebsiella pneumoniae</i> ; CRPA: carbapenem-resistant <i>Pseudomonas aeruginosa</i> ; CRE: carbapenem-resistant <i>Enterobacteriaceae</i> ;			

Discussion:

In this study, we build the prediction model of CR-GNB carriage within a week at admission by machine learning. We analyze to further verify the accuracy of the model through a 4-month prospective and consecutive predictions.

With the worldwide spread of CR-GNB in ICU, clinicians invest a lot of times and resources in nosocomial prevention and control measures, including colonization supervision, contact isolation, hand hygiene, antibiotic control and so on^{2, 6}. However, in the face of heavy clinical work, the implementation rate of these measures has been criticized. The normalization of nosocomial prevention and control measures is not only the compliance of medical staff but also the problem of cost and benefits. Control the source of infection, cut off the route of transmission, and protect the vulnerable population are three classical pathways. Current research has found more and more dormant sources of infection, including fiberoptic bronchoscope, ICU flume^{17, 18}. The route of transmission also has more possibility of analysis with the assistance of next-generation sequencing¹⁹. Many pieces of researches focus on protecting susceptible people because of its simplicity and effectiveness^{5, 20, 21}. Some studies pay attention to the identification of high-risk factors, which are determined by building models²¹⁻²⁴. Thomas and his colleagues predicted the infection of MDR by multivariate logical regression model by the public database and found that the main risk factors of MDR infection were the high use of antibiotics previously, the site and degree of infection in

the previous three months²⁵. The researchers included a total of 120,000 samples, but there were a few variables and the models could only vaguely indicate which patients were likely to develop into the infections. Katherine and his teams used a multiple logical regression model and a decision tree model to assess the risk of extended Spectyrum-β lactamase (ESBLs) bacteremia through the data of 1288 cases of gram-negative bacteremia. A total of 14 variables were included. The prediction effect of the decision tree and multivariate logistic regression model for predicting ESBL infection was similar²¹. However, the research sample size and variables are limited, and the incoordination between the number of positive cases (15%) and that of control cases (85%) restricted the performance of machine learning. Wang et al used 512 patients to predict MDR carriages and found that males, high CRP levels, and high Pitt scores were high-risk predictors, and a line chart was used to predict the occurrence of MDR²⁴. In our study, invasive procedures include endotracheal intubation, intravenous intubation, drainage tube, and the hospital residence history over the past month are high-risk factors for CR-GNB carriages, which have been identified in other studies^{9, 22}. According to the history of residence in other hospitals, our center adopts preemptive isolation and active surveillance which can reduce the incidence of carbapenem-resistant Enterobacteriaceae (CRE)²⁶. However, these kinds of literatures only provide information on which patients may develop MDR colonization or infection, but the exact time is unknown. As a result, the prevention and control of nosocomial infection measures are faced with the problems of when to implement and when to remove, as well as the cost-effectiveness, clinical burnout. Therefore, our center developed a CR-GNB prediction model in a week for ICU patients, which aims to carry out more targeted prevention including single isolation in advance, special management, and other measures. At present, this study has completed the verification of clinical applicability, and the subsequent clinical randomized controlled trial will be conducted to verify its clinical practicability.

We set one week as the forecast period in several aspects. First of all, too short or too long prediction periods will affect the performance of the prediction model. Secondly, ICU hospitalization as a high-risk factor for CR-GNB, the longer the ICU hospital stay, the higher the incidence of CR-GNB. According to the average hospitalization days, about 6 days in our center, most of the patients with hospitalization in a week are postoperative patients who are not the beneficiaries of this study. Lastly, according to the pre-experimental results, the peak time of CR-GNB carriages in our center was 5 days, then decreased slowly. One week as the prediction node can balance and the comparability in the trial group and the control group. Besides, sputum samples were still dominant in the colonization and infection of CR-GNB in this study, which was related to the types of diseases, including severe craniocerebral trauma and community-acquired pneumonia, which required long-term mechanical ventilatio. and the proportion of. These patients sent sputum samples for examination are higher than that of other parts and stay longer than the general postoperative patient trial. Therefore they are more likely to be included and caused bias in the study.

This study uses the central database, which has both advantages and disadvantages. On the one hand, the local database has with the diversified and integrated variables, which is difficult to achieve in the public database. The central database has been updated in real-time and provides the feasibility for continuous prediction. Also, the real-time updated database can continuously carry out iterative learning for the model to incorporate data and keep pace with the times. On the other hand, the public database has more data and complete diseases than the single-center database, based by stable models and multi-center researches. However, the problem is also very prominent. The variables are fixed and may simplify the number of variables to ensure the integrity, followed by a slow update and unable to achieve timely prediction. The number adopted in this study is limited, and as single-center research, follow-up promotion will take a long time.

Conclusion:

The prediction model by machine learning can predict the occurrence of CR-GNB carriage within one week, with a success rate of 84%. This model can predict the high-risk groups of CR-GNB carriers in real-time and help guide medical staff to take more targeted prevention and control of nosocomial measures.

Abbreviations

CR-GNB
carbapenem-resistance gram-negative bacteria;
AUROC
area under the receiver operating characteristic;
MDRB
multidrug-resistant bacteria;
ICU
intensive care unit

Declarations

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 materials:

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

Competing interest:

None.

Funding:

No funding.

Author contributions:

Qiqiang Liang write manuscript and analyze data; Qinyu Zhao assist at data cleaning and modeling; Xin Xu assist at data secondary audit, independent variable screening, and chart drawing; Yu Zhou and Jie Xu establish the database and collect original data. Huang Man corrects the data of the paper;

Acknowledgements:

Not applicable.

References

1. Willyard C. The drug-resistant bacteria that pose the greatest health threats. *Nature* 2017; 543: 15. 2017/03/03. DOI: 10.1038/nature.2017.21550.
2. Martin-Loeches I, Leone M, Madach K, et al. Antibiotic therapy in the critically ill - expert opinion of the Intensive Care Medicine Scientific Subcommittee of the European Society of Anaesthesiology. *European journal of anaesthesiology* 2017; 34: 215-220. DOI: 10.1097/EJA.0000000000000595.
3. Zilberberg MD, Nathanson BH, Sulham K, et al. Carbapenem resistance, inappropriate empiric treatment and outcomes among patients hospitalized with Enterobacteriaceae urinary tract infection, pneumonia and sepsis. *BMC infectious diseases* 2017; 17: 279. 2017/04/19. DOI: 10.1186/s12879-017-2383-z.
4. Palmore TN and Henderson DK. Managing transmission of carbapenem-resistant enterobacteriaceae in healthcare settings: a view from the trenches. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2013; 57: 1593-1599. 2013/08/13. DOI: 10.1093/cid/cit531.
5. Karampatakis T, Tsergouli K, Iosifidis E, et al. Impact of active surveillance and infection control measures on carbapenem-resistant Gram-negative bacterial colonization and infections in intensive care. *J Hosp Infect* 2018; 99: 396-404. Article. DOI: 10.1016/j.jhin.2018.05.010.
6. Gandra S and Ellison RT, 3rd. Modern trends in infection control practices in intensive care units. *Journal of intensive care medicine* 2014; 29: 311-326. 2013/06/12. DOI: 10.1177/0885066613485215.
7. Asensio A, Cantero M, Shaw E, et al. Control strategies for carbapenemase-producing Enterobacteriaceae at different levels of the healthcare system. *Enfermedades infecciosas y microbiología clínica* 2014; 32 Suppl 4: 61-66. 2014/12/30. DOI: 10.1016/s0213-005x(14)70176-4.
8. Frencken JF, Wittekamp BHJ, Plantinga NL, et al. Associations Between Enteral Colonization With Gram-Negative Bacteria and Intensive Care Unit-Acquired Infections and Colonization of the Respiratory Tract. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017 2017/12/01. DOI: 10.1093/cid/cix824.
9. Giacobbe DR, Del Bono V, Bruzzi P, et al. Previous bloodstream infections due to other pathogens as predictors of carbapenem-resistant Klebsiella pneumoniae bacteraemia in colonized patients: results from a retrospective multicentre study. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2017; 36: 663-669. DOI: 10.1007/s10096-016-2843-1.
10. 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: 726-734. DOI: 10.1016/S1473-3099(17)30228-1.
11. Komorowski M, Celi LA, Badawi O, et al. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018; 24: 1716-1720. 2018/10/24. DOI: 10.1038/s41591-018-0213-5.
12. Bailly S, Meyfroidt G and Timsit JF. What's new in ICU in 2050: big data and machine learning. *Intensive Care Med* 2018; 44: 1524-1527. 2017/12/28. DOI: 10.1007/s00134-017-5034-3.
13. 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: 541-550. 2019/03/28. DOI: 10.1017/ice.2019.42.

14. 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: e1002703. 2018/11/28. DOI: 10.1371/journal.pmed.1002703.
15. 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: 905-914. DOI: 10.1016/s2213-2600(18)30300-x.
16. Nemati S, Holder A, Razmi F, et al. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit Care Med* 2018; 46: 547-553. 2017/12/30. DOI: 10.1097/CCM.0000000000002936.
17. Smith ZL, Dua A, Saeian K, et al. A Novel Protocol Obviates Endoscope Sampling for Carbapenem-Resistant Enterobacteriaceae: Experience of a Center with a Prior Outbreak. *Dig Dis Sci* 2017; 62: 3100-3109. Article. DOI: 10.1007/s10620-017-4669-9.
18. Smismans A, Ho E, Daniels D, et al. New environmental reservoir of CPE in hospitals.
19. Snitkin ES, Won S, Pirani A, et al. Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak. *Science translational medicine* 2017; 9. DOI: 10.1126/scitranslmed.aan0093.
20. Magiorakos AP, Burns K, Rodríguez Baño J, et al. Infection prevention and control measures and tools for the prevention of entry of carbapenem-resistant Enterobacteriaceae into healthcare settings: guidance from the European Centre for Disease Prevention and Control. *Antimicrobial Resistance & Infection Control* 2017; 6. DOI: 10.1186/s13756-017-0259-z.
21. Goodman KE, Lessler J, Harris AD, et al. A methodological comparison of risk scores versus decision trees for predicting drug-resistant infections: A case study using extended-spectrum beta-lactamase (ESBL) bacteremia. *Infect Control Hosp Epidemiol* 2019; 40: 400-407. 2019/03/05. DOI: 10.1017/ice.2019.17.
22. Liu P, Li X, Luo M, et al. Risk Factors for Carbapenem-Resistant *Klebsiella pneumoniae* Infection: A Meta-Analysis. *Microbial drug resistance* 2018; 24: 190-198. Article. DOI: 10.1089/mdr.2017.0061.
23. Maseda E, Salgado P, Anillo V, et al. Risk factors for colonization by carbapenemase-producing enterobacteria at admission to a Surgical ICU: A retrospective study. *Enfermedades infecciosas y microbiología clinica* 2016 2016/03/27. DOI: 10.1016/j.eimc.2016.02.017.
24. Wang L, Huang X, Zhou J, et al. Predicting the occurrence of multidrug-resistant organism colonization or infection in ICU patients: development and validation of a novel multivariate prediction model. *Antimicrobial Resistance & Infection Control* 2020; 9. DOI: 10.1186/s13756-020-00726-5.
25. Lodise TP, Bonine NG, Ye JM, et al. 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: 718. 2019/08/16. DOI: 10.1186/s12879-019-4363-y.
26. Liang Q, Yan C, Xu Z, et al. Preemptive isolation and active surveillance in the prevention and control of nosocomial infection reduce the incidence of carbapenem-resistant Enterobacteriaceae.

Figures

Figure 1

Flow chart of study. CR-GNB: carbapenem-resistance gram-negative bacteria;

Figure 2

Important characteristics of model parameters by random forest included all variables.

Figure 3

Area under the receiver operating characteristic curve of validation set for three machine learning and multivariate logistic regression.

Figure 4

Area under the receiver operating characteristic curve of test set for three machine learning and multivariate logistic regression.