RESEARCH ARTICLE

DOI: 10.4274/cjms.2024.2024-119 Cyprus J Med Sci 2025;10(Suppl 1):46-49



Early Prediction of Mortality due to Carbapenem-Resistant Gram-Negative Bacterial Infection in Intensive Care Units Using Machine Learning

● Buket Baddal¹, ● Cemile Bağkur², ● Bardia Arman³

¹Department of Medical Microbiology and Clinical Microbiology, Near East University Faculty of Medicine, Nicosia, North Cyprus ²DESAM Research Institute, Near East University, Nicosia, North Cyprus ³Artificial Intelligence Application and Research Center, Cyprus International University Faculty of Engineering, Nicosia, North Cyprus

Abstract

BACKGROUND/AIMS: The occurrence of hospital-acquired infections due to carbapenem-resistant Gram-negative bacteria (CR-GNB) is on the rise globally. Studies show that infections with multidrug-resistant Gram-negative bacteria are associated with high mortality mainly in intensive care units (ICUs). This study aims to develop machine learning (ML) algorithms to identify variables correlated with mortality and construct a prediction model for ICU mortality due to CR-GNB infections.

MATERIALS AND METHODS: Data from patients admitted to a private hospital between 2016 and 2023 were included. The dataset included patients from the ICU who had a positive culture of CR-GNB after 3 days of admission (n=788). Demographic data, vital signs, important blood test indicators, intubation, and catheterization history were collected. The proposed models included a classifier and a mortality prediction system, utilizing seven ML algorithms: Extreme Gradient Boosting (XGBoost), logistic regression, random forest (RF), k-nearest neighbors, support vector machine, naive bayes, and decision tree (DT).

RESULTS: Analysis showed that blood C-reactive protein, urea, creatinine, platelet-large cell ratio, along with patient age and presence of endotracheal intubation, were strong predictors of mortality in ICU patients. In terms of accuracy, XGBoost (96.2%) outperformed RF (93.7%) and DT (91.8%). The area under the receiver operating characteristic curve for these models was 0.98, 0.99, and 0.93, while F1 scores were 0.97, 0.95, and 0.94, respectively.

CONCLUSION: ML prediction models can predict patient mortality in ICUs due to CR-GNB and guide medical staff to identify high-risk groups in advance.

Keywords: Carbapenem-resistant Gram-negative bacteria, machine learning, intensive care unit, mortality prediction

INTRODUCTION

Carbapenem resistance is a global public health issue. The excessive application and escalating misuse of the carbapenem class of antibiotics have caused a significant rise in the occurrence of carbapenem-resistant Gram-negative bacterial (CR-GNB) infections.¹ This is particularly due to

the existence of β -lactamase genes, which are found on mobile genetic elements that can be disseminated among bacteria within a hospital environment.² Indeed, the World Health Organization 2017 priority list of pathogens ranked carbapenem-resistant *Pseudomonas aeruginosa*, carbapenem-resistant *Acinetobacter baumannii*, and carbapenemresistant *Enterobacteriaceae* within the top priority category, which

To cite this article: Baddal B, Bağkur C, Arman B. Early prediction of mortality due to carbapenem-resistant Gram-negative bacterial infection in intensive care units using machine learning. Cyprus J Med Sci. 2025;10(Suppl 1):46-49

ORCID IDs of the authors: B.B. 0000-0003-3319-2179; C.B. 0000-0003-0998-4176; B.A. 0009-0001-9449-2688.



Corresponding author: Buket Baddal E-mail: buket.baddal@neu.edu.tr ORCID ID: orcid.org/0000-0003-3319-2179 Received: 05.11.2024 Accepted: 13.12.2024 Publication Date: 04.06.2025

Copyright© 2025 The Author. Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License. was termed the critical category.³ Infections by these pathogens are associated with longer stays, added healthcare costs, and higher mortality, particularly in those within the intensive care unit (ICU).⁴

Infections in ICUs are central causes of morbidity and mortality due to the increased vulnerability of these patients to nosocomial infections. ICUs are termed the epicenter of multidrug-resistant (MDR)-GNB, which primarily arise from the frequent and irrational use of broad-spectrum antibiotics, that lead to the evolution of drug-resistant strains.⁵ ICUadmitted patients are also more prone to MDR-GNB infections due to numerous invasive medical procedures such as mechanical ventilation, catheterization, and intubation.⁶

The present study aims to develop machine learning (ML) algorithms to identify variables correlated with CR-GNB-associated mortality in patients admitted to the ICU; provide a prediction model for ICU mortality; and evaluate the performance of ML models in the prediction of mortality in individuals that require ICU admission.

MATERIALS AND METHODS

This study was approved by the Ethics Committee of Near East University (NEU/2023/110-1685, date: 26.01.2023). Due to the retrospective nature of the study, informed consent was waived.

Data Collection, Study Design, and Population

This retrospective study evaluated patients from the ICU at Near East University Hospital between January 2016 and December 2023. The inclusion criteria were: (1) age \geq 18 years; (2) ICU admission; (3) a positive culture of CR-GNB after 3 days of ICU admission; (4) the presence of important blood test indicators. The exclusion criteria were: (1) outpatients; (2) patients transferred to the general ward within 3 days.

Demographic data including age, gender, as well as clinical features such as prognostic scores, vital signs, laboratory blood examination results, history of invasive catheterization, and endotracheal intubation were collected. In terms of blood test variables, serum albumin; alanine aminotransferase (serum pyruvic transaminase); aspartate aminotransferase (serum glutamate oxaloacetate transaminase); C-reactive protein (CRP); glucose; white blood cell count; neutrophil number (NEU#); neutrophil percentage (NEU%); lymphocyte number (LYM#); lymphocyte percentage (LYM%); monocyte number (MONO#); monocyte percentage (MONO%); eosinophil number (EOS#); eosinophil percentage (EOS%); basophil number (BASO#); basophil percentage (BASO%); red blood cell count; hemoglobin; mean corpuscular hemoglobin; mean corpuscular hemoglobin concentration; platelet count; red blood cell distribution width; mean platelet volume; plateletcrit; platelet-large cell ratio (P-LCR); nucleated red blood cell count; chlorine; creatinine; potassium; procalcitonin; sodium; urea and calcium were used.

Data Splitting and Preprocessing

The preprocessing of the dataset involved mapping categorical values for "gender", "invasive catheterization", "endotracheal intubation", and "CR-GNB infection" into their binary numerical representations: 0 and 1. Numerical feature missing values were handled by converting data into numeric format and missing values of numerical features were handled by filling them with zero. For the purpose of prediction the target variable "death" was separated from the rest of the predictor variables. Following this, the dataset was divided into a training set and a test set in the ratio of 80:20, using the function train_test_split. Standardization of the variables was performed prior to the evaluation of different ML models.

Statistical Analysis

The performance of several models including logistic regression (LR), decision trees (DT), k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), and naive bayes (NB) was compared using area under the receiver operating characteristic curve (AU-ROC) curves and classification metrics such as accuracy, precision, recall, and F1 score. This enabled the comparison of each model's performance by plotting the ROC curves, highlighting key features that contribute to mortality prediction. These model developments and statistical tests were performed in Python version 3.6.

RESULTS

Variable Importance and Feature Selection

The selected predictors of in-ICU mortality due to CR-GNB infection are shown in Figure 1A. From a total of 37 variables, the top six variables to correlate with mortality in descending order were CRP, urea, creatinine, age, endotracheal intubation, and P-LCR. Variables such as CRP, urea, and creatinine were moderately positively correlated with mortality, indicating that the higher these biomarkers are, the worse the outcome. Age and endotracheal intubation, although with weaker correlations, represented well-established risk factors of mortality in the ICU. The low correlation observed in the correlation matrix suggested minimal multicollinearity in the dataset, indicating that each variable provided unique information supporting model accuracy (Figure 1B).

Overall Performance of the ML Models

The comparison of different models is necessary to identify the most accurate and reliable approach for predicting biomarkers associated with mortality. The prognostic performance and accuracy of the prediction models for ICU mortality due to CR-GNB infection are summarized in Table 1. Overall, the XGBoost classifier had the highest accuracy with 95.6%, followed by the RF classifier (93.7%). Indeed, both models demonstrated positive performance, with precision, recall, and F1-scores above 0.93. The DT model also performed well with an accuracy of 91.8%, although it was lower than that of the ensemble models.

On the contrary, The KNN model exhibited had a reasonable performance with a reasonable accuracy of 78.5%, performing well in terms of precision while giving a slightly lower recall. The LR and the SVM were also less effective, with accuracy values at 69% and 71%, respectively. The worst performing model was NB, with an accuracy of 62% and lower F1-scores, giving a poor performance for the correct classification of survived patients. Overall, the tree-based models resulted in the best predictive performance for predicting mortality.

The prediction algorithm using RF had the highest predictive value, for mortality due to CR-GNB infection in ICU with an area under the curve (AUC) of 0.99, for mortality due to CR-GNB infection in ICU (Figure 2).

DISCUSSION

In this study, the predictive capabilities of several ML models were evaluated for ICU mortality due to CR-GNB infections. Among the models examined, XGBoost, RF, and DT were identified as the top performers. XGBoost demonstrated superior predictive accuracy at 96.2%, followed by RF at 93.7%, and DT at 91.8%. These models proved highly effective in identifying high-risk ICU patients, with XGBoost achieving an AU-ROC of 0.99, indicating excellent performance in distinguishing between mortality outcomes. Tree-based models, such as XGBoost and LightGBM, are widely recognized for their strong predictive capabilities and



Figure 1. (A,B) Importance of the selected variables and their correlation matrix.

LTM: Lymphocyte, WBC: White blood cell count, MCV: Mean corpuscular volume, P-LCR: Platelet-large cell ratio, CRP: C-reactive protein, ROC: Receiver operating characteristic curve, AUC: Area under the curve.

frequently outperform other ML models. Consistent with our findings, Jeon et al.7 reported that LightGBM (AU-ROC; 0.827) outperformed LR and conventional clinical scoring systems in predicting ICU mortality, reinforcing the value of ML models in critical care settings. Both studies identified similar key parameters influencing predictions, including CRP, urea, and creatinine levels, which emerged as crucial indicators of mortality risk.7 CRP. a well-established biomarker of inflammation. plays a central role in assessing mortality risk, particularly in septic conditions. Elevated CRP levels, particularly those measured on the third day of ICU admission, have been linked to increased mortality. Similarly, elevated urea and creatinine levels are critical predictors of mortality, particularly when a high urea-to-creatinine ratio is observed.⁸⁻¹⁰ In addition to CRP and other biochemical markers, P-LCR emerged as a significant predictor of mortality in our study. Elevated P-LCR, indicative of systemic inflammation, was also strongly correlated with mortality in Coronavirus disease-2019 patients in the study by Celik et al.¹¹, reflecting the role of inflammation-driven markers in ICU settings. This highlights the diverse applicability of both complex ML models and simpler clinical biomarkers in predicting ICU outcomes across different conditions.

A unique feature of the current study was the use of a local hospital database rather than publicly available datasets. Other studies, such as Iwase et al.'s¹² work on a broader ICU population, achieved similarly high predictive accuracy (AU-ROC; 0.945, using a RF model) but relied on more generalized datasets.

Study Limitations

The study's single-center data allow for specific healthcare applications but limits broader generalizability. The inclusion of multiple datasets with the same variables from different hospitals within the same region would improve the accuracy of the prediction models. The prediction models lack external validation, restricting their use in other clinical environments.

CONCLUSION

This study shows that ML prediction models can predict patient mortality in ICUs due to CR-GNB and guide medical staff to identify high-risk groups in advance. The localized approach in this study offers a more relevant predictive model for clinical decision-making in the hospital setting, potentially improving patient care by offering customized risk assessment for CR-GNB infections in ICUs. This focus on local data strengthens the real-world applicability of our findings, making them more directly actionable for improving patient outcomes in specific healthcare environments.

Table 1. Overall performance of the prediction models							
Model	Accuracy	Precision (class 0)	Precision (class 1)	Recall (class 0)	Recall (class 1)	F1-score (class 0)	F1-score (class 1)
XGBoost	95.6	0.92	0.97	0.94	0.96	0.93	0.97
Random forest	93.7	0.87	0.97	0.94	0.94	0.90	0.95
Decision tree	91.8	0.82	0.97	0.94	0.91	0.88	0.94
K-nearest neighbors	78.5	0.66	0.84	0.63	0.85	0.65	0.85
Support vector machine	71	0.67	0.71	0.12	0.97	0.21	0.82
Logistic regression	69	0.5	0.76	0.43	0.81	0.46	0.78
Naive bayes	62	0.42	0.77	0.57	0.64	0.48	0.7
XGBoost: Extreme gradient boosting.							



Figure 2. ROC curves and AUCs for ICU mortality were obtained using machine learning methods.

AUC: Area under the curve, SVM: Support vector machine, XGBoost: Extreme gradient boosting, LYM: Lymphocytes, MVC: Mean corpuscular volume, P-LCR: Platelet large cell ratio, CRP: C-reactive protein, WBC: White blood cell, ICU: Intensive care unit.

MAIN POINTS

- This study focuses on machine learning (ML) models to identify variables correlated with mortality due to carbapenem-resistant Gram-negative bacterial (CR-GNB) infection in intensive care units (ICUs).
- Analysis results indicated that blood C-reactive protein, urea, creatinine, platelet-large cell ratio, along with patient age and presence of endotracheal intubation, were strong predictors of mortality in ICU patients.
- In terms of accuracy, XGBoost had the highest accuracy among random forest and decision tree.
- ML prediction models can predict patient mortality in ICUs due to CR-GNB and can guide medical staff to identify high-risk groups in advance.

ETHICS

Ethics Committee Approval: This study was approved by the Ethics Committee of Near East University (NEU/2023/110-1685, date: 26.01.2023).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Concept: B.B., C.B., Design: B.B., C.B., Data Collection and/or Processing: C.B., B.A., Analysis and/or Interpretation: B.B., C.B., B.A., Literature Search: B.B., C.B., B.A., Writing: B.B., C.B., B.A.

DISCLOSURES

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study had received no financial support.

REFERENCES

- 1. Nordmann P, Poire L. Epidemiology and diagnostics of carbapenem resistance in gram-negative bacteria. Clin. Infect Dis. 2019; 69(7): 521-8.
- Logan LK, Weinstein RA. The epidemiology of carbapenem-resistant enterobacteriaceae: The impact and evolution of a global menace. J Infect Dis. 2017; 215(1): 28-36.
- World health Organization. WHO publishes list of bacteria for which new antibiotics are urgently needed. 2017. Available from: https://www.who. int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-newantibiotics-are-urgently-needed
- 4. van Loon K, Voor In 't Holt AF, Vos MC. A systematic review and meta-analyses of the clinical epidemiology of carbapenem-resistant enterobacteriaceae. Antimicrob Agents Chemother. 2018; 62(1): 1730-17.
- Breijyeh Z, Jubeh B, Karaman R. Resistance of gram-negative bacteria to current antibacterial agents and approaches to resolve it. Molecules. 2020; 25(6): 1340.
- Sengupta S, Barman P, Lo J. Opportunities to overcome implementation challenges of infection prevention and control in low-middle income countries. Curr Treat Options Infect Dis. 2019; 11(3): 267-80.
- Jeon ET, Lee HJ, Park TY, Jin KN, Ryu B, Lee HW, et al. Machine learning-based prediction of in-ICU mortality in pneumonia patients. Sci Rep. 2023; 13(1): 11527.
- Devran O, Karakurt Z, Adıgüzel N, Güngör G, Moçin OY, Balcı MK, et al. C-reactive protein as a predictor of mortality in patients affected with severe sepsis in intensive care unit. Multidiscip Respir Med. 2012; 7(1): 47.
- 9. Qu R, Hu L, Ling Y, Hou Y, Fang H, Zhang H, et al. C-reactive protein concentration as a risk predictor of mortality in intensive care unit: a multicenter, prospective, observational study. BMC Anesthesiol. 2020; 20(1): 292.
- 10. van der Slikke EC, Star BS, de Jager VD, Leferink MBM, Klein LM, Quinten VM, et al. A high urea-to-creatinine ratio predicts long-term mortality independent of acute kidney injury among patients hospitalized with an infection. Sci Rep. 2020; 10(1): 15649.
- Çelik O, Laloğlu E, Çelik N. The role of platelet large cell ratio in determining mortality in COVID-19 patients. Medicine (Baltimore). 2024; 103(18): 38033.
- 12. Iwase S, Nakada TA, Shimada T, Oami T, Shimazui T, Takahashi N, et al. Prediction algorithm for ICU mortality and length of stay using machine learning. Sci Rep. 2022; 12(1): 12912.