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.

Mortality prediction in a cardiac intensive care unit using machine learning and penalized regression: comparison with APACHE II and GRACE

Bruno Alcantara Gabardo[1]

bgabardo@hotmail.com

Universidade Federal do Paraná (UFPR)

Mariana Oliveira Borges[1]

Universidade Federal do Paraná (UFPR)

Giulio Cesar Gequelim[1]

Universidade Federal do Paraná (UFPR)

Ulisses Almicci Santos[1]

Universidade Federal do Paraná (UFPR)

Bruno Siegel Guerra[1]

Universidade Federal do Paraná (UFPR)

Gustavo Lenci Marques[1]

Universidade Federal do Paraná (UFPR)

Research Article

Keywords: Intensive Care Units. Cardiovascular Diseases. Mortality. Machine Learning. Prognosis

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11311561/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

Introduction: To evaluate and compare the prognostic performance of traditional scores and machine learning–based and penalized regression models for in-hospital mortality prediction in patients admitted to a heterogeneous cardiac intensive care unit.

Methods: This retrospective observational study included adult patients admitted to a tertiary university cardiac intensive care unit in south of Brazil. Mortality prediction models based on Extreme Gradient Boosting (XGBoost) and Firth's penalized logistic regression were developed using routinely collected clinical and laboratory data. Model performance

was compared with APACHE II in the overall cohort and with APACHE II and GRACE in the acute coronary syndrome subgroup using area under the receiver operating characteristic curve (AUC).

Results: A total of 1,855 patients were included, with an in-hospital mortality rate of 8.2%. In the overall population, XGBoost demonstrated superior discriminative performance (AUC 0.900) compared with APACHE II (AUC 0.840), while penalized logistic regression also showed high discrimination (AUC 0.889). In the acute coronary syndrome subgroup, both XGBoost and penalized regression outperformed GRACE. The advantage of XGBoost was most evident in regions of low false-positive rates. Renal dysfunction and hemodynamic instability were the main prognostic determinants.

Conclusion: Machine learning–based and penalized regression models, particularly XGBoost, demonstrated robust and clinically relevant performance for mortality prediction in a heterogeneous cardiac intensive care unit. These approaches may complement traditional prognostic scores by improving early risk stratification and supporting clinical decision-making in high-complexity settings.

Introduction

The management of cardiac emergencies has become increasingly complex due to population aging, the rising prevalence of comorbidities, and improved survival among patients with structural and ischemic heart disease. As a result, contemporary cardiac

intensive care units (CICUs) now care for patients with greater clinical heterogeneity, higher severity of illness, and frequent multiorgan dysfunction, requiring rapid and accurate risk stratification to guide early decision-making.

Cardiovascular diseases remain the leading cause of mortality worldwide and in Brazil, with acute coronary syndrome (ACS) representing one of the most frequent and lethal acute presentations[1]. Despite advances in reperfusion strategies and intensive care support, in-hospital mortality associated with ACS remains substantial, and its management imposes a significant economic burden on health systems, particularly in middle-income countries, where resource constraints further challenge

sustainability[2]. In this context, improving early prognostic assessment in CICUs is essential not only to optimize clinical outcomes but also to support efficient allocation of intensive care resources.

Traditionally, prognostic assessment in intensive care has relied on clinical scores such as the Acute Physiology and Chronic Health Evaluation II (APACHE II), while disease-specific scores such as the Global Registry of Acute Coronary Events (GRACE) score have been used in selected cardiac populations[3,4]. Although these tools are well validated and widely used, they are based on linear assumptions, fixed variable weights, and limited modeling of interactions, which may restrict their performance in contemporary CICUs characterized by marked diagnostic and physiological heterogeneity[5]. Moreover, APACHE II was developed for general intensive care populations and may not fully capture the complexity of critically ill cardiac patients, while GRACE is applicable primarily to relatively homogeneous ACS cohorts.

In recent years, machine learning (ML)–based approaches have emerged as promising alternatives for outcome prediction in critical care. By modeling nonlinear relationships and capturing complex interactions among multiple clinical and laboratory variables, ML algorithms may offer improved prognostic performance in high-complexity settings[6–8]. Several studies have demonstrated superior or comparable performance of ML models, particularly Extreme Gradient Boosting (XGBoost), for mortality prediction in heterogeneous critical care populations, including patients with sepsis and other high-risk conditions[9–11].

Despite this growing body of evidence, important knowledge gaps remain. Data evaluating the performance of ML-based prognostic models specifically in heterogeneous cardiac intensive care unit populations are still limited, particularly in middle-income countries. Furthermore, direct comparisons between ML models and widely used traditional scores, such as APACHE II and GRACE, within the same CICU setting are scarce. Addressing these gaps is essential to determine whether ML-based approaches can meaningfully complement established prognostic tools in real-world cardiac intensive care practice.

Therefore, the objective of this study was to evaluate and compare the prognostic performance of machine learning–based models and traditional prognostic scores for in-hospital mortality prediction in patients admitted to a heterogeneous cardiac intensive care unit of a tertiary university hospital. Specifically, we compared models based on Extreme Gradient Boosting and Firth's penalized logistic regression with APACHE II in the overall cohort and with APACHE II and GRACE in the acute coronary syndrome

subgroup, aiming to assess their potential clinical applicability for early risk stratification in contemporary cardiac intensive care.

Methods

Study Design and Setting

This retrospective observational study was conducted in a cardiac intensive care unit (CICU) of a tertiary university hospital in south of Brazil. The CICU provides care for patients with a wide range of acute cardiovascular conditions, including acute coronary syndromes, acute decompensated heart failure, arrhythmias, cardiogenic shock, and postoperative cardiac surgery, characterizing a clinically heterogeneous population.

The study was approved by the Research Ethics Committee (Comitê de Ética em Pesquisa, CEP) of the Hospital de Clínicas da Universidade Federal do Paraná (HC-UFPR), Curitiba, Brazil (CAAE: 59074421.2.0000.0096). The requirement for informed consent was waived by the committee due to the retrospective nature of the study and the use of anonymized data. The study was conducted in accordance with national regulations and the Declaration of Helsinki.

Study Population

All consecutive adult patients (≥ 18 years) admitted to the CICU between January 2018 and December 2023 were eligible for inclusion. Patients were excluded if essential clinical or laboratory data required for model development were missing. The acute coronary syndrome (ACS) subgroup was defined according to standard clinical, electrocardiographic, and biochemical criteria documented in the medical records.

Outcome

The primary outcome was all-cause in-hospital mortality during the index hospitalization.

Data Collection and Variables

Clinical, demographic, and laboratory data were extracted from the institutional electronic medical record system. Variables were selected based on clinical relevance, availability in routine practice, and prior evidence of association with mortality in critically ill and cardiology patients.

Collected data included age, sex, body mass index (BMI), comorbidities, source of admission, primary diagnosis, and need for renal replacement therapy or cardiopulmonary resuscitation within the first 24 hours. Physiological and laboratory variables included heart rate, respiratory rate, body temperature, arterial blood pressure, arterial blood gas parameters (pH and PaO_2), hemoglobin, hematocrit, leukocyte count, serum sodium and potassium, troponin levels, and serum creatinine values at admission,

as well as minimum and maximum creatinine values recorded during the CICU stay.

The APACHE II score was calculated using the worst physiological values within the first 24 hours of admission, according to the original methodology described by Knaus, Wagner and Zimmerman (1985)[3].

For patients admitted with ACS, the GRACE score was calculated following established definitions by Fox et al. (2006)[4].

Handling of missing data

Variables with missing values were assessed prior to model development. Patients with missing data in key predictors required for score calculation or model input were excluded from the analysis. Given the retrospective design and the objective of evaluating

real-world applicability using routinely collected data, no data imputation procedures were performed.

Statistical Analysis

Continuous variables were summarized as mean and standard deviation or median and interquartile range, as appropriate. Categorical variables were presented as absolute and relative frequencies. Statistical analyses were conducted using appropriate statistical and machine learning libraries, with significance set at a two-sided p-value <

0.05 when applicable.

Penalized Logistic Regression (Firth Method)

To identify variables independently associated with mortality and reduce small-sample bias, a multivariable logistic regression model using Firth's penalized likelihood approach was constructed[12,13]. Variables with clinical relevance and/or p values < 0.20 in univariate analysis were considered for inclusion in the multivariable model. Results were reported as odds ratios (ORs) with 95% confidence intervals (95% CIs).

Machine Learning Model

A machine learning model based on Extreme Gradient Boosting (XGBoost) was developed for mortality prediction. XGBoost was selected due to its ability to model nonlinear relationships, capture complex interactions between variables, and handle heterogeneous data structures[6].

Model training was performed using a supervised learning approach. Hyperparameter tuning was conducted using grid search with five-fold cross-validation on the training dataset. The following hyperparameters were optimized: number of

estimators (n_estimators), maximum tree depth (max_depth), learning rate (eta), subsample ratio, and column subsample ratio. Early stopping was applied to prevent overfitting.

Model validation and performance evaluation

Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC), with AUCs compared descriptively across models[14]. Additional performance metrics included sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy.

Performance was evaluated in the overall CICU population and, separately, in the subgroup of patients with ACS. Diagnostic analysis of residuals was performed for the penalized logistic regression model to assess model adequacy.

Analyses were conducted for the overall CICU population and separately for the subgroup of patients admitted with ACS.

Variable importance and interpretability

For the XGBoost model, variable importance was assessed using the gain metric, reflecting the contribution of each variable to improving model performance. This analysis was used to explore the relative importance of clinical and laboratory variables in mortality prediction.

Software

All statistical analyses and model development were performed using R (R Foundation for Statistical Computing, Vienna, Austria), version 4.0.2, within the RStudio integrated development environment (RStudio, PBC, Boston, MA, USA).

Results

Study Population

A total of 1,855 adult patients were included in the final analysis, with an in-hospital mortality rate of 8.2% (153 deaths). The median age was 66 years (interquartile range [IQR], 57–74), and 49.1% of patients were male. Most admissions were urgent (79.7%),
reflecting the high clinical severity of the cohort. Acute coronary syndrome (ACS) was the most frequent reason for admission (721 patients, 38.9%).

Baseline demographic, clinical, hemodynamic, reasons for CICU admission,
source of admission, and assistential characteristics and laboratory characteristics of the study population are presented in Table 1.

Clinical Characteristics and Outcomes

Non-survivors exhibited a significantly higher severity of illness compared with survivors. Prognostic scores were markedly higher among patients who died, both for APACHE II (median 22 [IQR 14–26] vs. 11 [IQR 8–14]; $p < 0.001$) and GRACE (166.4 ±

49.6 vs. 132.9 ± 34.9; p < 0.001). Length of CICU stay was also longer among non-survivors (median 5 [IQR 2–11] vs. 3 [IQR 2–5] days; p < 0.001).

Markers of organ dysfunction were more prevalent among non-survivors, particularly renal dysfunction and hemodynamic instability, as well as more pronounced metabolic and acid–base derangements at admission.

Model Performance in the Overall CICU Population

In the overall cohort, both predictive models developed in this study demonstrated good discriminative performance for in-hospital mortality (Figure 1). The XGBoost model

achieved the highest AUC (0.900), outperforming the traditional APACHE II score (AUC 0.840; p = 0.04). The Firth's penalized logistic regression model showed intermediate performance (AUC 0.889), also superior to APACHE II (p = 0.04). Overlap between the ROC curves of XGBoost and penalized regression was observed, indicating comparable overall discrimination.

Notably, the XGBoost model demonstrated consistently superior performance in regions of low false-positive rates (1–specificity < 0.30), suggesting improved clinical utility for identifying high-risk patients without excessive false alarms.

Model Performance in the Acute Coronary Syndrome Subgroup

In the subgroup of patients admitted with ACS, all evaluated models showed high discriminative ability (Figure 2). The Firth's penalized logistic regression model achieved the highest AUC (0.928), closely followed by the XGBoost model (AUC 0.920). The APACHE II score maintained good discrimination (AUC 0.886), while the GRACE score demonstrated acceptable performance (AUC 0.857).

These findings indicate that machine learning–based models perform at least as well as disease-specific prognostic scores in ACS patients, while maintaining broader applicability across the heterogeneous CICU population.

Model Interpretability and Variable Importance

Variable importance analysis of the XGBoost model revealed a predominance of parameters related to renal function, including maximum, minimum, and admission creatinine levels, as the most influential predictors of mortality (Figure 3). Variables associated with metabolic and respiratory instability, such as arterial pH, PaO_2, and

respiratory rate, also ranked highly. Clinical and contextual variables, including reason for admission and occurrence of cardiac arrest, further contributed to the model's predictive performance.

Overall, the importance pattern indicates that the XGBoost model prioritized markers of global disease severity and multiorgan dysfunction, consistent with clinical reasoning in critically ill cardiac patients.

Model Calibration

Calibration analysis demonstrated good agreement between observed and predicted mortality probabilities for the XGBoost model across the spectrum of risk (Figure 4). The calibration curve closely followed the line of identity, indicating reliable estimation of absolute mortality risk in this cohort.

Summary of Model Performance

A comparative summary of discrimination and performance metrics for all evaluated models is presented in Table 2. Overall, machine learning–based modeling using XGBoost achieved the highest discriminative performance in the overall CICU population, while traditional prognostic scores retained relevant clinical utility, particularly APACHE II in the full cohort and GRACE in the ACS subgroup.

Discussion

In this retrospective study conducted in a heterogeneous cardiac intensive care unit, we evaluated and compared the prognostic performance of traditional risk scores and modern predictive approaches, including machine learning–based models and penalized regression, for in-hospital mortality prediction. The main findings demonstrate

that models based on Extreme Gradient Boosting (XGBoost) achieved superior or comparable discriminative performance relative to established prognostic scores, particularly APACHE II and GRACE, both in the overall cohort and in the clinically relevant subgroup of patients with acute coronary syndrome (ACS). These results reinforce the potential role of machine learning–based approaches as complementary tools for risk stratification in contemporary cardiac intensive care settings.

From a clinical perspective, the superior performance of XGBoost in regions of low false-positive rates is particularly relevant. In critical care, false-positive risk classification may lead to unnecessary interventions, excessive monitoring, and inefficient resource allocation. An AUC of 0.900 indicates that in approximately 90% of randomly selected patient pairs, the XGBoost model correctly assigns a higher mortality risk to patients who die than to those who survive. Although the numerical difference compared with APACHE II may appear modest, this improvement occurs precisely in the most clinically meaningful portion of the receiver operating characteristic curve, where decision thresholds are typically applied[14–16].

The limitations of traditional prognostic scores help explain these findings. APACHE II assumes linear and additive effects, applies fixed weights, and does not model conditional interactions among variables. Moreover, it was developed for general intensive care populations rather than specifically for cardiac intensive care units[3]. As a result, although APACHE II demonstrated consistent and acceptable

performance in this study—confirming its historical value—it may be less adaptable to the complex clinical phenotypes commonly encountered in modern CICUs. In contrast, XGBoost is able to model nonlinear relationships and capture interactions between multiple variables,

allowing prognostic estimates to be dynamically adjusted according to the combined clinical context[5,6,8]. For example, a mean arterial pressure of 65 mmHg may have markedly different prognostic implications depending on concurrent renal dysfunction or acid–base imbalance—relationships that traditional scores are unable to distinguish.

In the subgroup of patients with ACS, the GRACE score demonstrated adequate discriminative performance, consistent with its original validation studies[4]. However, its applicability was restricted to this relatively homogeneous population, underscoring the limitations of disease-specific scores when applied to heterogeneous CICU populations. In contrast, both XGBoost and penalized logistic regression maintained high performance in this subgroup, highlighting the robustness of these approaches across different clinical scenarios.

Beyond model performance metrics, the analysis of prognostic determinants provides important insights into the pathophysiology of critically ill cardiac patients. Early hypotension emerged as a key prognostic factor, reflecting systemic circulatory dysfunction and impaired tissue perfusion. In CICU patients, hypotension often represents a complex interplay of ventricular dysfunction, autonomic dysregulation, inflammatory response, and limited cardiovascular reserve, and should be interpreted as part of a broader phenotype of early clinical instability[17,18].

Renal dysfunction was identified as a central determinant of mortality in both modeling approaches. Patients who died exhibited significantly higher creatinine levels throughout hospitalization, including both minimum and maximum values, which remained independently associated with mortality after multivariable adjustment. In critically ill cardiac patients, acute kidney injury frequently reflects the integration of

multiple pathophysiological mechanisms—such as hypoperfusion, venous congestion, neurohormonal activation, and systemic inflammation—representing an early manifestation of multisystem failure. These findings are consistent with robust evidence linking renal dysfunction to adverse outcomes in acute coronary syndrome, heart failure, and cardiogenic shock[18,19].

From a clinical standpoint, these results suggest that early and serial assessment of renal function should be prioritized in patients admitted with cardiac emergencies, particularly those with hemodynamic instability and requiring vasoactive support. Although the lack of systematically recorded urine output data precluded formal classification of acute kidney injury according to KDIGO or AKIN criteria[19,20], the use of serial laboratory measurements allowed identification of a robust and clinically meaningful association between renal dysfunction and mortality.

The comparison between Firth's penalized logistic regression and XGBoost highlights their complementary roles in prognostic stratification. Penalized regression enabled the identification and quantification of variables independently associated with mortality, providing direct clinical interpretability. In contrast, XGBoost demonstrated superior discriminative capacity by integrating complex interactions and nonlinear relationships among multiple clinical and laboratory variables. The concordance of key determinants identified by both approaches reinforces the physiological plausibility of the models, while the superior performance of XGBoost underscores its potential as a clinical decision-support tool in high-complexity environments[5,7].

Several previous studies support these findings. Analyses using large critical care databases, such as MIMIC-IV, have demonstrated superior performance of machine

learning models—including XGBoost and Random Forest—compared with traditional statistical approaches for mortality prediction[9–11]. Our study extends this evidence by focusing specifically on a heterogeneous cardiac intensive care population in a middle-income country, a context that remains underrepresented in the literature.

This study has limitations. Its retrospective, single-center design limits generalizability, and the absence of external validation precludes direct extrapolation of model performance to other institutions. In addition, missing data—particularly regarding urine output—restricted the assessment of certain clinical variables. Nevertheless, the study provides relevant real-world evidence from a large Brazilian CICU cohort and demonstrates the feasibility and potential clinical utility of machine learning–based prognostic models in this setting.

Conclusions

In conclusion, machine learning–based models, particularly XGBoost, demonstrated robust and clinically relevant performance for mortality prediction in a heterogeneous cardiac intensive care unit. These approaches may complement traditional prognostic scores by improving early risk stratification and supporting clinical decision-making, without replacing clinical judgment. Future prospective, multicenter studies with external validation are warranted to confirm these findings and to facilitate safe and effective clinical implementation.

Abbreviations

ACS: acute coronary syndrome; APACHE II: Acute Physiology and Chronic Health Evaluation II; AUC: area under the receiver operating characteristic curve; BMI: body mass index; CAAE: Certificado de Apresentação de Apreciação Ética; CI: confidence interval; CICU: cardiac intensive care unit; GRACE: Global Registry of Acute Coronary Events; ICU: intensive care unit; IQR: interquartile range; ML: machine learning; OR: odds ratio; PaO_2: partial pressure of arterial oxygen; ROC: receiver operating characteristic;

SD: standard deviation; XGBoost: Extreme Gradient Boosting; AKIN: Acute Kidney Injury Network; KDIGO: Kidney Disease Improving Global Outcomes.

Declarations

Ethics approval and consent to participate

The study was approved by the Research Ethics Committee (Comitê de Ética em Pesquisa, CEP) of the Hospital de Clínicas da Universidade Federal do Paraná (HC-UFPR), Curitiba, Brazil (CAAE: 59074421.2.0000.0096). The requirement for informed consent was waived by the committee due to the retrospective nature of the study and the use of anonymized data. The study was conducted in accordance with national regulations and the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The de-identified dataset supporting the findings of this study is available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

Bruno Alcantara Gabardo and Giulio Cesar Gequelim conceived and designed the study, collected the data, performed the statistical analyses, and drafted the manuscript. Mariana Oliveira Borges contributed to data collection. Ulisses Almicci dos Santos and Bruno Siegel Guerra contributed to the statistical analyses. Gustavo Lenci Marques supervised the study.

Acknowledgements

Not applicable.

References

1. Santana NM, Santos IS, Goulart AC, Lotufo PA, Benseñor IM, Correia LCL, et al. Sarcopenia e obesidade como preditores prognósticos em pacientes idosos hospitalizados com infarto agudo do

miocárdio. Einstein (São Paulo). 2019;17(4):eAO4632. 10.31744/einstein_journal/2019AO4632.

2. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013–2020. Geneva: World Health Organization; 2013.
3. Knaus WA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
4. Fox KA, Dabbous O, Goldberg RJ, Pieper KS, Eagle KA, Van de Werf F, et al. Global Registry of Acute Coronary Events Investigators. Prediction of risk of death and myocardial infarction in the six months after presentation with acute coronary syndrome: prospective multinational observational study (GRACE). BMJ. 2006;333(7578):1091. 10.1136/bmj.38985.646481.55.
5. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer; 2019.
6. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. 2016:785–94. 10.1145/2939672.2939785
7. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD, et al. Machine learning and decision support in critical care. Proc IEEE. 2016;104(2):444–66.
8. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2:719–31. 10.1038/s41551-018-0305-z.
9. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-day mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGBoost. J Transl Med. 2020;18(1):462. 10.1186/s12967-020-02620-5.
10. Kwon YS, Baek MS. Development and validation of a quick sepsis-related organ failure assessment-based machine-learning model for mortality prediction in patients with suspected infection in the emergency department. J Clin Med. 2020;9(3):875.
11. Yu C, Li Y, Yin M, Gao J, Xi L, Lin J, et al. Automated machine learning in predicting 30-day mortality in patients with non-cholestatic cirrhosis. J Pers Med. 2022;12(11):1930.
12. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27–38.
13. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002;21(16):2409–19.
14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
15. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol. 2004;159(9):882–90.
16. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA. 2019;321(20):2003–17.

17. Khanna AK, Kellner C, Beer BN, Besch L, Dettling A, Bertoldi LF, et al. Association of systolic, diastolic, mean, and pulse pressure with morbidity and mortality in septic ICU patients: a nationwide observational study. Ann Intensive Care. 2023;13:101. 10.1186/s13613-023-01101-4.
18. Sundermeyer J, Kellner C, Beer BN, Besch L, Dettling A, Bertoldi LF, et al. Kidney injury in patients with heart failure-related cardiogenic shock: results from an international, multicentre cohort study. Eur J Heart Fail. 2025 May;28. 10.1002/ejhf.3701.
19. Khwaja A. KDIGO Clinical Practice Guidelines for Acute Kidney Injury. Nephron Clin Pract. 2012;120(4):c179–84. 10.1159/000339789.
20. Mehta RL, Kellum JA, Shah SV, Molitoris BA, Ronco C, Warnock DG, et al. Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. Crit Care. 2007;11(2):R31. 10.1186/cc5713.

Tables

Tables 1 and 2 are available in the Supplementary Files section.

Figures

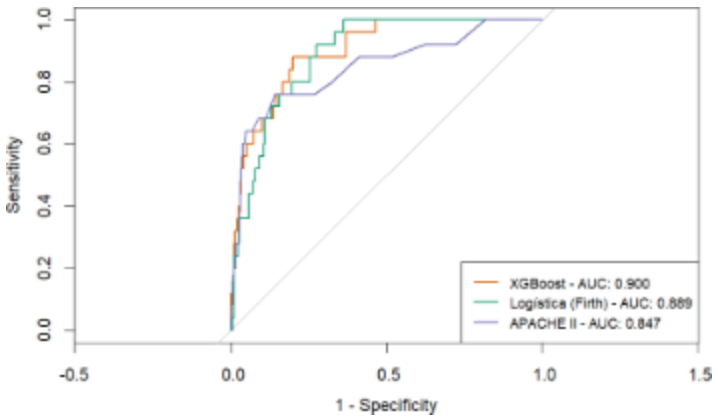


Figure 1

Receiver operating characteristic curves comparing mortality prediction models in the overall cardiac intensive care unit population

The XGBoost model demonstrated the highest discriminative performance (AUC = 0.900), followed by Firth's penalized logistic regression (AUC = 0.889) and the APACHE II score (AUC = 0.840). Both models developed in this study outperformed the traditional prognostic score. Overlap between the ROC curves of XGBoost and penalized logistic regression was observed, with superior performance of XGBoost particularly in regions of low false-positive rates.

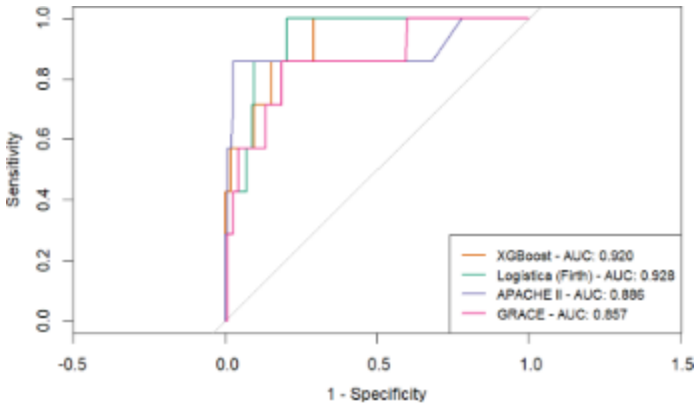


Figure 2

Receiver operating characteristic curves comparing mortality prediction models in patients with acute coronary syndrome

In the subgroup of patients admitted with acute coronary syndrome, Firth's penalized logistic regression achieved the highest discriminative performance (AUC = 0.928), closely followed by the XGBoost model (AUC

= 0.920). The APACHE II score showed good discrimination (AUC = 0.886), while the GRACE score demonstrated acceptable performance (AUC = 0.857). These results indicate that machine learning–based and penalized regression models perform at least as well as traditional prognostic scores in this clinical subgroup.

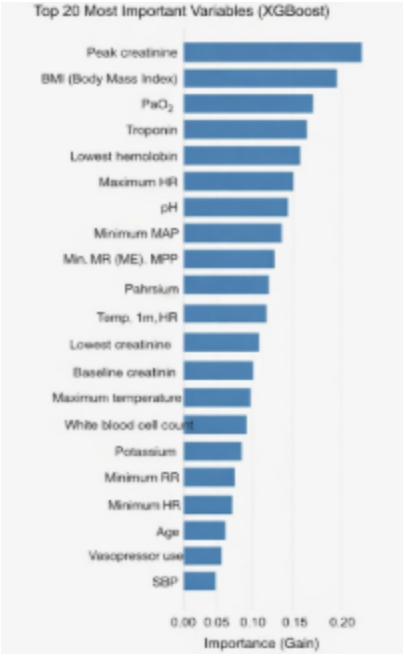


Figure 3

Variable importance in the XGBoost model for in-hospital mortality prediction

Variable importance analysis based on gain revealed a predominance of predictors related to organ dysfunction. Renal function parameters, including admission, minimum, and maximum creatinine levels, were the most influential variables, followed by markers of metabolic and respiratory instability such as arterial pH, PaO_2, and respiratory rate. Clinical and contextual variables, including reason for admission and occurrence of cardiac arrest, also contributed to the model's predictive performance, supporting the clinical plausibility of the machine learning–based approach.

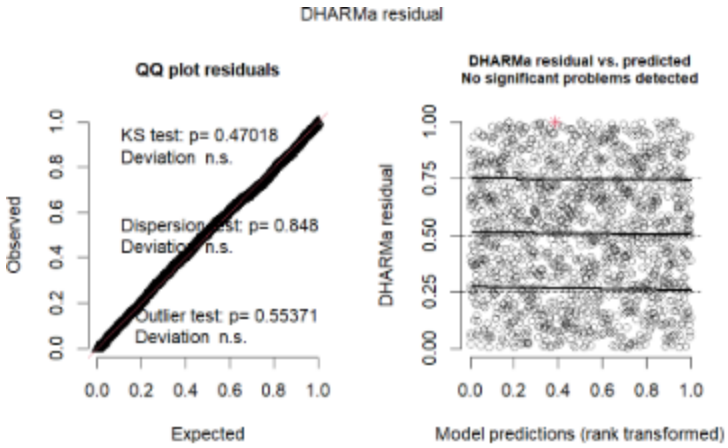


Figure 4

Calibration plot of the XGBoost model for in-hospital mortality prediction

The calibration plot shows the relationship between predicted and observed in-hospital mortality probabilities for the XGBoost model.

Predicted risks are well aligned with observed outcomes across risk strata, with the calibration curve closely following the line of identity, indicating good agreement and reliable estimation of absolute mortality risk in the overall cardiac intensive care unit population.

Supplementary Files

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

- Tables.docx