Research Article

Open Access

CrossMark

Machine learning prediction of ICU length of stay in Saudi Arabia. retrospective analytical study

Abstract

Background: Prolonged intensive care unit (ICU) stays are associated with increased morbidity, resource utilization, and cost. Early identification of patients at risk for extended ICU length of stay (LOS) can support clinical decision-making and improve resource management.

Objective: To develop and evaluate a machine learning model to predict ICU LOS using routine laboratory tests available early during admission.

Methods: We conducted a retrospective study using electronic health record data from adult ICU patients. Evaluated the predictive performance of four machine learning (ML) models to choose the best model, which was trained on a set of demographic and laboratory tests' results to predict LOS category. Model performance was assessed using accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, and specificity on a validation set.

Results: The XGBoost model demonstrated the highest accuracy (90%) and Kappa (79%) among the four evaluated models. On the testing data, XGBoost had an accuracy of 87.5%, sensitivity 88%, specificity 87.1%, and AUC of 95.3%. The top five important predictor variables were blood glucose, arterial partial pressure of oxygen (PaO2), arterial partial pressure of carbon dioxide (PaCO2), body mass index (BMI), and age. Diagnostic accuracy measures on the validation data were: Accuracy = 83.9%, sensitivity = 79.4%, specificity = 88%, and AUC = 92.5%

Conclusion: Machine learning can effectively predict ICU length of stay early in the course of admission. Such models could aid clinicians in identifying patients at risk for prolonged ICU stays, facilitating proactive discharge planning and ICU resource optimization. Future studies should focus on external validation and real-world implementation.

Ahmed F Mady,[1,2] Mohammed A Al-Odat,[1] Rayan A Alshaya,[1] Hend M Hamido,[3] Ahmed W Aletreby,[4] Anas A Mady,[5] Fares M Eladrousi,[5] Huda A Mhawish,[6] Jennifer Q Gano,[6] Waleed T Aletreby[1]

[1]Critical care Department, King Saud Medical City, Saudi Arabia
[2]Anesthesia Department, Faculty of Medicine, Tanta University, Egypt
[3]Women Health Hospital, King Saud Medical City, Saudi Arabia
[4]Medical Student, Faculty of Medicine, Alexandria University, Egypt
[5]Medical Student, Faculty of Medicine, Alfaisal University, Saudi Arabia
[6]Nursing Department, King Saud Medical City, Saudi Arabia

Correspondence: Waleed T Aletreby, Nursing Department, King Saud Medical City, Riyadh, Saudi Arabia

Received: December 12, 2025 | **Published:** January 22, 2026

Introduction

The intensive care unit (ICU) is a crucial part of any healthcare system, providing intensive management to critically ill patients.[1] Such services are expensive and resource-consuming, accounting for up to 30% of hospitals' budgets, and a large portion of any country's healthcare expenditure.[1,2] imposing financial pressures on healthcare systems, and resulting in imbalances between demand and resources.[2,3] This imbalance is aggravated by an exponential increase in the demand for critical care services, in view of population growth, improvement in treatment and diagnostic technologies, and an increase in the prevalence of elderly patients with chronic and complex medical conditions.[4]

Accordingly, hospitals are perpetually aiming to improve operational efficiency and reduce costs of critical care services through monitoring of different healthcare quality indicators, and conducting quality improvement projects to improve them.[5,6] Among the key operational indicators in the ICU is the average length of stay (LOS),[7] since if prolonged, LOS negatively impacts several aspects of healthcare services, such as resource utilization, risk of adverse events (such as infection), access to care for other patients, in addition to suffering of families.[1] It logically follows that if LOS can be accurately predicted as early as admission to ICU, it may be helpful in resource allocation, and operational optimization, in addition to other benefits such as providing realistic expectations to patients' families.[2]

Several ICU LOS predictive models have been in use for decades now, the most common among which are "The Acute Physiology and Chronic Health Evaluation" (APACHE) and "Simplified Acute Physiology Score" (SAPS),[8] APACHE system went through several updates, and its most recent version (APACHE IV) was introduced in 2006, based on the data from 104 ICUs in the United States (⁹), whereas SAPS III was created using data from 300 ICUs around the world.[10] Both predictive systems depend on data obtained within the first 24 hours of ICU admission, including demographics, vital signs, basic laboratory investigations, mechanical ventilation, chronic comorbidities, diagnostic category, operative procedures and hospital days before ICU admission.[11,12] Despite the popularity of those two models, they are subject to several limitations, first, they were constructed based on multi-variable linear regression analyses, which despite easiness of interpretation, used data usually do not fulfill the assumptions of linear regression,[13] second, the models use algorithms that are protected by property rights and are not freely available,[14] but perhaps more importantly, their discriminative ability to predict ICU LOS has been repeatedly questioned. APACHE IV was deemed as a poor predictor of prolonged ICU LOS, both for the general ICU population,[8] as well as for specific diagnoses such as severe sepsis.[14] While SAPS III was reported to have just a satisfactory discriminative performance, with an area under the curve (AURC) of 0.75.[15]

Recently, advanced analytical methods emerged as powerful tools of prediction and decision making, collectively known as "machine learning" (ML).[16] ML is a group of powerful analytical tools that can study the association between a set of data (features) and the outcome, especially when the data itself doesn't fulfill the assumptions of traditional regression models.[17] They utilize flexible algorithms

that are able to learn from a subset of the data (training), and then generalize the prediction to the remainder of the data (testing).[18] ML methods have gained popularity in healthcare research, particularly with regard to diagnosis and prognosis, and have been utilized in different fields such as diabetes, malignancies, cardiology, and (intensive care.[16-19] Very few studies were conducted in Saudi Arabia utilizing machine learning techniques to predict ICU LOS of the general population, available studies either focused on a special diagnosis such as COVID-19,[20,21] prognostic prediction,[22] or used publicly available data not originating from the Saudi population.[23] In view of the scarcity of machine learning based studies from Saudi Arabia on ICU patients in general, we conducted this study intending to predict LOS of patients admitted to the ICU in general, regardless of their diagnosis, since we believe that not focusing on a particular diagnosis would be more generalizable.

The importance of this work is that eventually the best predictive model of ICU LOS may be finalized, and it could be continuously used to predict the LOS of newly admitted patients in the future, based on previous data.

Method

Study design, setting, and timeframe

This was a retrospective, observational analysis of data collected from patients admitted to the ICU of a large tertiary referral center in the central region of Saudi Arabia. The ICU includes 110 beds, divided into respiratory, medical, surgical, neuro-critical, burn, and maternity units. All ICU beds are fully equipped with invasive and non-invasive monitoring and ventilation capabilities. The ICU is operated around the clock by intensivists, with a 1:1 nurse-to-patient ratio.

The study included ICU patients admitted during the period between July 1st, 2024 and January 31st, 2025.

Inclusion and exclusion criteria

We included patients who fulfilled the following criteria, regardless of their diagnosis or medical condition:

a) Adults (age ≥ 18 years).

b) Admitted to the ICU for at least 48 hours.

c) Discharged alive from the ICU to the general ward of KSMC, or home. This includes patients who were discharged against medical advice.

Exclusion criteria were:

a) Patients are admitted to the maternity and burn units, as well as pregnant ladies.

b) Patients who died in the ICU, or were discharged to the ICU of another healthcare facility. Since the LOS of patients who died in ICU may not reflect actual operational performance, and most probably does not reflect the expected LOS of the associated condition of the patient (2), while that of patients discharged to other ICUs cannot be obtained.

i. Readmission to ICU within the same hospitalization episode, to maintain independence of data.

ii. Patients with missing values of predictor variables and/or outcome.

Study objectives

The primary objective was the identification of the most accurate ML model capable of predicting ICU LOS, along with its area under the curve (AUC) of receiver operator characteristics (ROC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy.

Additional objectives included predictions of the model on unseen data (data that does not include the outcome), reported with the same diagnostic accuracy measures as the trained model, and the five most predictor variables (as identified by the ML model) will also be presented.

Data management

The outcome of ICU LOS was dichotomized into two categories (7 days or less, and more than 7 days) (8). Predictive variables included: Biological sex, age, presence or absence of comorbidities (regardless of their count), mechanical ventilation upon ICU admission, body mass index, in addition to the following laboratory values, taken as the worst value in the first 24 hours in ICU:

a) Hemoglobin (gm / dl)

b) White blood cell count (count * 10^9 / L)

c) Platelet count (count * 10^3 / microliter)

d) Serum creatinine (mg/ dl)

e) Blood glucose (mg/ dl).

f) Serum albumin (gm/ L)

g) Arterial blood Ph (unitless).

h) Arterial blood partial pressure of oxygen (PaO_2) (mm/Hg)

i) Arterial blood partial pressure of carbon dioxide ($PaCO_2$) (mm/Hg)

j) Arterial blood bicarbonate (HCO_3) (mmol/L)

k) The chosen variables were intended to be routine investigations usually done upon ICU admission, to maximize generalizability. The list of predictors was finalized after reviewing the available literature on the topic.

l) The data were retrospectively recorded from the patients' electronic medical records, and divided into two main parts:

m) Between 1 July 2024 and 31 December 2024: Data used to choose the most accurate model, and also constituted the data for training and testing of the best model, accordingly, this data included the binary outcome.

n) Between 1 January and 31 January 2025: Data used to validate the model, and the outcome was not included when analyzed by the model, however, the actual outcome was stored separately, for comparison with the predicted outcomes.

The inclusion and exclusion criteria were applied by the study personnel, and data of included patients were retrieved and recorded on a pre-prepared spreadsheet. All recorded data were anonymized, with no personal identifications of the patients.

Statistical and analytical method

The analysis was conducted using the statistical language of R,[24] with several packages about the different ML techniques used. R-Studio is an integrated development environment (IDE) for the

R language, a programming language used for statistical computing and data analysis. The libraries used were: (caret), (e1071), (nnet), (randomForest), (xgboost), and (pROC).

Analysis was done in the following steps:

Best classifier selection: A simple script with minimal fine-tuning parameters was run using data from July 1, 2024 to December 31, 2024. The script split the data into training and testing subsets, in a ratio of 70% to 30% respectively. The best model was chosen based on overall accuracy and the highest Kappa values of prediction performed on the testing subset. Kappa is a statistical metric for categorical variables, which evaluates the degree of agreement between predicted and actual classification, taking into account chance agreement. The script included the following ML classifiers:

i. Support Vector Machine Learning (SVM): It is an ML algorithm that attempts to find the optimal separation plane hyperplane between the predictor data to achieve the best classification. SVM models are known to operate adequately with either linear or non-linear data.[2,25]

ii. Artificial Neural Network (ANN): A powerful modelling algorithm that allows non-linearity between predictor variables, and accordingly, is considered a non-parametric technique that learns from the data without being restricted to specific assumptions.[26]

iii. Random Forest (RF): As its name indicates, RF is a model that relies on a huge number of decision trees to achieve the best classification,[2] and thus is considered an "ensemble" model, that is to say it combines predictions from multiple individual models to produce a stronger, more accurate, and more robust overall prediction.[19]

iv. Extreme Gradient Boosting (XGBoost): Is another "ensemble" model, in which decision trees are introduced into the model one at a time, and after fitting, are used to correct mistakes of prediction made by the previous model.[19] It has been widely utilized in a range of difficult classification problems, with very satisfactory results.[27]

Running best model: Once the best model of prediction was identified, it was run again with the same data, with fine-tuning using specific fine-tuning parameters suitable for the identified best model.

Making un-seen predictions: The set of data between January 1, and January 31, 2025 was used by the final model without including the outcome variable. Since this data was not included in the training or testing data, it is called "Un-seen" (sometimes called validation data). Predictions made by the model for this data were compared to the actual outcomes (since it is also retrospective), to generate diagnostic accuracy measures.

Ethical considerations

This study involves the application of ML analytical techniques on retrospective data, without actual involvement in the management of patients. Furthermore, the utilized results of laboratory investigations are already available, and were not specifically taken for the study. Accordingly, it was approved by the local IRB with waiver of consent (IRB reference: H1RI-16-Apr25-02). The study observes the research subjects' rights, as outlined by the Declaration of Helsinki, under the ultimate responsibility of the primary investigator to maintain data privacy and confidentiality.

Results

Training and testing data

During the last six months of 2024, there were 1937 admissions to the ICU, of those 1176 fulfilled the inclusion criteria, while 761 were excluded for various reasons (Figure 1).
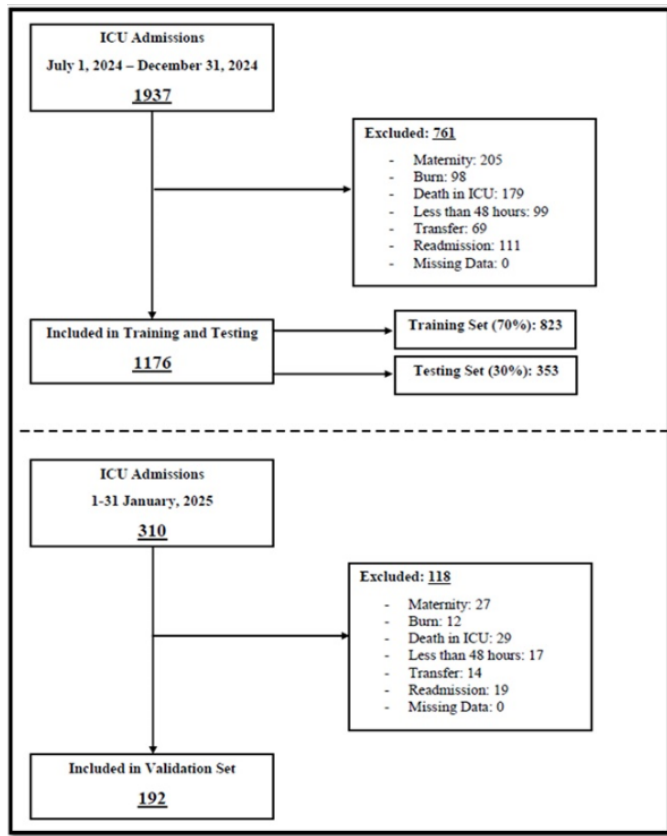

Figure 1 Study enrollment flow diagram.

Included patients had a mean age of 64.2 ± 14 years, and included 532 (45.2%) females, 275 (23.3%) mechanically ventilated patients upon ICU admission, and 556 (47.3%) who stayed in ICU more than seven days. Table S1 shows details of enrolled patients, and comparisons according to LOS category. It shows that the group with LOS more than seven days had significantly higher age, percentage with comorbidities, percentage mechanically ventilated, mean blood glucose, $PaCO_2$, and lower PaO_2.

After running a simple script to choose the best model based on predictions on the "Testing" subset according to overall accuracy and Kappa (Table S2), the results indicated that XGBoost is the model that best fits the data (Table S3, Figure S1).

Accordingly, the model was fine-tuned by creating a data frame that includes all possible combinations of (hyperparameters, using values that balance over-fitting (higher values) and generalization (lower values) (Table S4).

Predictions of the final model on the "Testing" data resulted in the confusion matrix and diagnostic accuracy measures shown in Table 1. The overall diagnostic accuracy was 87.5% (95% CI: 83.6 – 90.8) and p value < 0.001 (compared to non-informative rate = 0.55). Sensitivity was 88% (95% CI: 81.9 – 92.6), specificity of 87.1% (95% CI: 81.6 – 91.5), the PPV was 84.8% (95% CI: 79.3 - 89), and the NPV was

89.9% (95% CI: 85.3 – 93.2). Analysis of the predicted probabilities for each class yielded an AUC of 95.3% (95% CI: 92.5 – 97.3; p < 0.001) (Table 2, Figure 2).
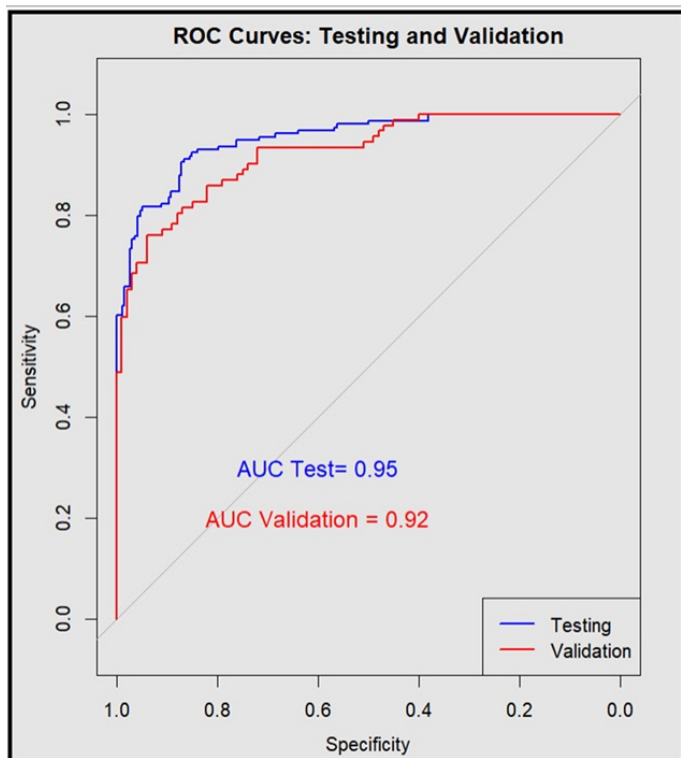
Figure 2 ROC curve of prediction on "Testing" and "Validation" data.

ROC, Receiver operator characteristics;a AUC, Area under the curve.

Table 1 XGBoost confusion matrix and diagnostic accuracy of prediction on "Testing" data

A: Confusion Matrix of predictions on "Testing Data"		
Predictions	Actual	
	≥ 7 days	< 7 days
≥ 7 days	139	25
< 7 days	19	169
B: Diagnostic accuracy measures:		
Measure	Value	95% CI	P value
Accuracy	87.5%	83.6 – 90.8	< 0.001
Sensitivity	88%	81.9 – 92.6	----------
Specificity	87.1%	81.6 – 91.5	----------
PPV	84.8%	79.3 – 89	----------
NPV	89.9	85.3 – 93.2	----------
AUC	95.3%	92.5 – 97.3	< 0.001

Mcnemar's p value = 0.5, indicating the model does not exhibit a systematic bias toward one class over the other.

PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve.

Table 2 XGBoost confusion matrix and diagnostic accuracy of prediction on "Validation" data

A: Confusion Matrix of predictions on "Validation Data"		
Predictions	Actual	
	≥ 7 days	< 7 days
≥ 7 days	73	12
< 7 days	19	88
B: Diagnostic accuracy measures:		
Measure	Value	95% CI	P value
Accuracy	83.9%	77.9 – 88.8	< 0.001
Sensitivity	79.4%	69.6 – 87.1	----------
Specificity	88%	80 – 93.6	----------
PPV	85.9%	78 – 91.3	----------
NPV	82.2%	75.5 – 87.4	----------
AUC	92.5%	87.8 – 95.8	< 0.001

Mcnemar's p value = 0.7, indicating the model does not exhibit a systematic bias toward one class over the other.

PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve.

The XGBoost model indicated that the top five most important predictor variables in order were blood glucose, PaO_2, $PaCO_2$, BMI, and age (Figure S2).

Validation data

During January 2025 there were 310 admissions to the ICU, 118 were excluded for various reasons, and 192 were included in the data analysis (Figure 1). Included patients had a mean age of 67.4 ± 13.8 years, and included 80 (41.7%) females, 43 (22.4%) mechanically ventilated patients upon ICU admission, and 92 (47.9%) who stayed in ICU more than seven days. Table S5 shows details of the patients in the validation set, and comparisons according to LOS category. Comparisons show that the group with LOS more than seven days had a significantly higher percentages of mechanically ventilated patients, white blood cell count, and blood glucose, while having a significantly lower serum albumin, and PaO_2. The XGBoost final model was applied to this data (without including the outcome) according to the script detailed in Table S6, to validate the model. We used the model to make predictions both as a classification (binary) and as a probability of being in each group. Then we compared those predictions to the actual (separately) recorded data to produce the confusion matrix and diagnostic accuracy measures shown in Table 2. Predictions of the final model on the "Validation" data had an overall diagnostic accuracy of 83.9% (95% CI: 77.9 – 88.8) and p value < 0.001 (compared to non-informative rate = 0.52). Sensitivity was79.4 % (95% CI: 69.6 – 87.1), specificity of 88% (95% CI: 80 – 93.6), the PPV was 85.9% (95% CI: 78 – 91.3), and NPV was 82.2% (95% CI: 75.5 – 87.4), while area under the ROC curve was 92.5% (95% CI: 87.8 – 95.8) (Table 1, Figure 2). Notably, both confusion matrices had statistically non-significant McNemar's tests, indicating the model does not exhibit a systematic bias toward one class over the other (p values of 0.5 and 0.7).

Discussion

In this ML analytical study we identified XGBoost as the best model to fit our data, with an overall accuracy of 87.5% and 83.9% on "Testing" and "Validation" data, respectively. Performance of the model on "Testing" data achieved an excellent[28] AUC of 95.3%, as well as on the "Validation" data with an AUC of 92.5%. This outstanding performance of the XGBoost model is echoed by others, as it similarly achieved the highest AUC for the prediction of COVID-19 patients' LOS,[19] and prediction of ICU LOS using vital sign.[23] The top five identified important predictor variables by the model were not surprising, as blood glucose, PaO_2, $PaCO_2$, BMI, and age are all components of the conventionally used APACHE IV prediction model. Additionally, they were all found in previous research to be associated with LOS in the ICU. For example, admission blood glucose was associated with increased LOS regardless of the diagnosis or medical specialty.[29] PaO_2 and $PaCO_2$ were among the predictors of ICU LOS in a similar ML study,[30] and BMI in another,[31] while age was identified as an independent predisposing factor of prolonged ICU stay.[32]

The ability to predict ICU LOS with high accuracy can positively impact all stakeholders of the healthcare system. Administratively, it is an effective method to address capacity management, allocation of resources, and staffing issues.[16] Clinically, predicted LOS can be an important indicator to optimize interventions and use of medical devices to ensure access to critical medical needs in a timely fashion.[16,33] Equally important, predicted LOS can be referenced during family counseling to address relatives' queries and anticipations, which may facilitate clinical decision making.[2] And obviously, accurate prediction of LOS is of significant importance to insurance companies and payors.[13,33]

Our model performed well in predicting ICU LOS, although with higher values for all diagnostic accuracy measures on the "Testing" data compared to the "Validation" data, which is not unusual but in fact expected when the model performs on unseen data.[2] Yet, those slightly lower diagnostic accuracy measures remain quite adequate and satisfactory, since all of which were at or above 80%, with an overall accuracy of 83.9%, providing reasonable confidence in the predictions. To our best knowledge, this is the first study to use ML to predict LOS of the general ICU Saudi population, regardless of the diagnosis. Our model utilized results of routine laboratory tests that are commonly performed for any patient admitted to the ICU, rather than sophisticated tests which may not be available in resource-limited hospitals, enhancing the utility and applicability of the model. Once the model is finalized using the initial data, it can be saved to be used repeatedly with the addition of new data, which may improve its performance.

Limitations

Despite promising results, this study has several limitations. First, the data used were derived from a single-center ICU cohort, which may limit the generalizability of the findings to other institutions with different patient populations or care practices. Second, although the model showed high diagnostic accuracy, it was trained on retrospective data, and prospective validation is necessary to confirm its real-world utility. Third, the binary classification of ICU length of stay into <7 days and ≥7 days, while clinically practical, may oversimplify the complexity and continuous nature of LOS. Lastly, potential confounders such as ICU staffing levels, care protocols, and discharge policies were not accounted for in the model.

Conclusion

This study demonstrates that XGBoost machine learning can be used to accurately predict prolonged ICU length of stay early in the admission. By leveraging the results of routine investigations, the model achieved high diagnostic performance and has the potential to support clinicians in identifying patients at risk for prolonged ICU stays. Such predictive insights could enhance discharge planning, resource allocation, and overall ICU efficiency. Future work should focus on external validation across multiple centers and integration into clinical workflows to evaluate the model's impact on decision-making and patient outcomes.

References

1. Mady AF, Al–Odat MA, Alshaya R, Hussien S, Aletreby A. Mortality rates in early versus late intensive care unit readmission. *Saudi J Med Med Sci.* 2023;11(2):143–149.

2. Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. *Diagnostics (Basel).* 2021;11(12):2242.

3. Alharbi KK, Arbaein TJ, Alzhrani AA, et al. Factors affecting the length of stay in the intensive care unit among adults in Saudi Arabia: a cross–sectional study. *J Clin Med.* 2023;12(21):6787.

4. Ramadan OE, Mady AF, Al–Odat MA, et al. Diagnostic accuracy of ePOS score in predicting DNR labeling after ICU admission: a prospective observational study (ePOS–DNR). *J Intensive Med.* 2023;4(2):216–221.

5. Kılıç M, Yüzkat N, Soyalp C, Gülhaş N. Cost analysis on intensive care unit costs based on the length of stay. *Turk J Anaesthesiol Reanim.* 2019;47(2):142–145.

6. Hignett S, Lang A, Pickup L, et al. More holes than cheese: what prevents the delivery of effective, high quality and safe health care in England? *Ergonomics.* 2018;61(1):5–14.

7. Pari V. Collaboration for Research Implementation, Training in Critical Care, Asia Africa (CCAA). Development of a quality indicator set to measure and improve quality of ICU care in low– and middle–income countries. *Intensive Care Med.* 2022;48(11):1551–1562.

8. Widyastuti Y, Zaki WA, Widodo U, Jufan AY, Pratomo BY. Predictive accuracy of the APACHE IV scores on mortality and prolonged stay in the intensive care unit of Dr Sardjito Hospital. *Med J Malaysia.* 2022;77(Suppl 1):53–58.

9. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med.* 2006;34(5):1297–1310.

10. Ghazaly HF, Aly AAA, Sayed MH, et al. APACHE IV, SAPS III, and SOFA scores for outcome prediction in a surgical/trauma critical care unit: an analytical cross–sectional study. *Ain–Shams J Anesthesiol.* 2023;15:101.

11. Bloria SD, Chauhan R, Sarna R, Gombar S, Jindal S. Comparison of APACHE II and APACHE IV score as predictors of mortality in patients with septic shock in intensive care unit: a prospective observational study. *J Anaesthesiol Clin Pharmacol.* 2023;39(3):355–359.

12. Siddiqui SS, Narkhede AM, Kulkarni AP, et al. Evaluation and validation of four scoring systems: the APACHE IV, SAPS III, MPM0 II, and ICMM in critically ill cancer patients. *Indian J Crit Care Med.* 2020;24(4):263–269.

13. Peres IT, Hamacher S, Oliveira FLC, Bozza FA, Salluh JIF. Prediction of intensive care unit length of stay: a concise review. *Rev Bras Ter Intensiva.* 2021;33(2):183–187.

14. Chattopadhyay A, Chatterjee S. Predicting ICU length of stay using APACHE–IV in persons with severe sepsis: a pilot study. *J Epidemiol Res.* 2016;2(1):1–8.

15. Mungan İ, Bektaş Ş, Altınkaya **Çavuş** M, Sarı S, Turan S. The predictive power of SAPS–3 and SOFA scores and their relations with patient outcomes in the surgical intensive care unit. *Turk J Surg.* 2019;35(2):124–130.

16. Chen Q, Zhang B, Yang J, et al. Predicting intensive care unit length of stay after acute type A aortic dissection surgery using machine learning. *Front Cardiovasc Med.* 2021;8:675431.

17. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart.* 2018;104(14):1156–1164.

18. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD. Predicting intensive care unit readmission with machine learning using electronic health record data. *Ann Am Thorac Soc.* 2018;15(7):846–853.

19. Saadatmand S, Salimifard K, Mohammadi R, et al. Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID–19 patients. *Ann Oper Res.* 2022:1–29.

20. Alabbad DA, Almuhaideb AM, Alsunaidi SJ, et al. Machine learning model for predicting the length of stay in the intensive care unit for COVID–19 patients in the eastern province of Saudi Arabia. *Inform Med Unlocked.* 2022;30:100937.

21. Elhazmi A, Al–Omari A, Sallam H, et al. Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID–19 patients admitted to the ICU. *J Infect Public Health.* 2022;15(7):826–834.

22. Alanazi A, Aldakhil L, Aldhoayan M, Aldosari B. Machine learning for early prediction of sepsis in intensive care unit patients. Medicina (Kaunas). 2023;59(7):1276.

23. Alghatani K, Ammar N, Rezgui A, Shaban–Nejad A. Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. *JMIR Med Inform.* 2021;9(5):e21347.

24. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.

25. Zoppis I, Mauri G, Dondi R. Kernel methods: support vector machines. In: Encyclopedia of Bioinformatics and Computational Biology. Elsevier; 2018.

26. Zhang Z. A gentle introduction to artificial neural networks. *Ann Transl Med.* 2016;4(19):370.

27. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM; 2016:785–794.

28. **Çorbacıoğlu** ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: a guide to interpreting the area under the curve value. *Turk J Emerg Med.* 2023;23(4):195–198.

29. Björk M, Melin EO, Frisk T, Thunander M. Admission glucose level was associated with increased short–term mortality and length of stay irrespective of diagnosis, treating medical specialty or concomitant laboratory values. *Eur J Intern Med.* 2020;75:71–78.

30. Abd–Elrazek MA, Eltahawi AA, Abd Elaziz MH, Abd–Elwhab MN. Predicting length of stay in hospital intensive care units using general admission features. *Ain Shams Eng J.* 2021;12(4):3691–3702.

31. Hempel L, Sadeghi S, Kirsten T. Prediction of intensive care unit length of stay in the MIMIC–IV dataset. *Appl Sci.* 2023;13(12):6930.

32. **Çevik** B, Geyik FD. Prolonged stay in intensive care unit: retrospective analysis of predisposing factors and outcome. *J Turk Soc Intens Care.* 2019;17(2):96–101.

33. Yildirim AE, Canayaz M. Machine learning–based prediction of length of stay in the neonatal intensive care unit using ensemble methods. *Neural Comput Appl.* 2024;36:14433–14448.