Annals of Clinical and Translational Medicine

read & research publications

Section: Medicine

Original Research Article

DEVELOPMENT AND PROSPECTIVE MULTICENTER VALIDATION OF AN EXPLAINABLE ARTIFICIAL INTELLIGENCE–DRIVEN PRECISION PROGNOSTIC MODEL FOR EARLY PREDICTION OF MORTALITY AND MULTIORGAN FAILURE IN CRITICALLY ILL ADULTS

Anand Dev[1]

[1]Associate Professor, Department of Critical Care Medicine, Indira Gandhi Institute of Medical Sciences (IGIMS), Sheikhpura, Patna, Bihar, India

ABSTRACT

Background: Early prediction of mortality and multiorgan failure remains a major challenge in critical care medicine. Conventional prognostic scoring systems have limited ability to capture the complex and dynamic interactions underlying critical illness. This study aimed to develop and prospectively validate an explainable artificial intelligence (AI)-driven precision prognostic model for early prediction of mortality and multiorgan failure in critically ill patients.

Materials and Methods: In this prospective multicenter cohort study, 3,012 adult ICU patients from five tertiary-care hospitals were enrolled. Clinical, physiological, laboratory, and treatment-related variables obtained within the first 24 hours of ICU admission were used to develop machine-learning models. The cohort was randomly divided into derivation (70%) and validation (30%) datasets. Multiple algorithms were evaluated, including logistic regression, random forest, artificial neural networks, and extreme gradient boosting (XGBoost). Model performance was assessed using AUROC, sensitivity, specificity, calibration metrics, and SHAP-based explainability analysis.

Result: Among 3,012 patients, in-hospital mortality occurred in 561 (18.6%) patients, while 729 (24.2%) developed multiorgan failure. XGBoost demonstrated the best performance, achieving an AUROC of 0.946 in the derivation cohort and 0.934 in the validation cohort for mortality prediction. The model significantly outperformed APACHE II (AUROC 0.806) and SOFA (AUROC 0.831) scores. For multiorgan failure prediction, the model achieved AUROCs of 0.923 and 0.912 in the derivation and validation cohorts, respectively. SHAP analysis identified serum lactate, SOFA score, vasopressor requirement, age, and renal dysfunction as the most influential predictors.

Conclusion: The explainable AI-driven prognostic model demonstrated excellent accuracy, calibration, and generalizability for early prediction of mortality and

Received : 25/01/2026
Received in revised form : 28/02/2026
Accepted : 29/03/2026

Corresponding Author:
Dr. Anand Dev,
Email: anand.dev@igims.org

Source of Support: Nil,
Conflict of Interest: None declared

Ann. Clin. Transl. Med
2026;1(1); 39-46

How to cite this article: Dev A. Development and Prospective Multicenter Validation of an Explainable Artificial Intelligence–Driven Precision Prognostic Model for Early Prediction of Mortality and Multiorgan Failure in Critically Ill Adults. *Ann. Clin. Transl. Med*. 2026;1(1):39-46.

multiorgan failure in critically ill patients. This approach may enhance precision risk stratification and support timely clinical decision-making in intensive care settings.

Keywords: Artificial intelligence, Machine learning, Critical care, Intensive care unit, Mortality prediction, Multiorgan failure, Explainable AI, XGBoost, Precision medicine, Prognostic model.

INTRODUCTION

Critically ill patients admitted to intensive care units (ICUs) remain at substantial risk of mortality and multiorgan failure despite major advances in monitoring technologies, organ-support therapies, and evidence-based critical care practices. Early identification of patients at high risk for adverse outcomes is essential for timely intervention, optimal resource allocation, individualized treatment planning, and informed communication with patients' families. Consequently, accurate prognostic assessment has become a cornerstone of modern critical care medicine.[1]

Over the past several decades, conventional prognostic scoring systems such as the Acute Physiology and Chronic Health Evaluation (APACHE), Sequential Organ Failure Assessment (SOFA), and Simplified Acute Physiology Score (SAPS) have been widely used to estimate illness severity and predict mortality among ICU patients.[2–4] Although these scoring systems have demonstrated clinical utility, they rely predominantly on predefined variables and linear statistical assumptions that may inadequately capture the complex, nonlinear interactions among physiological, biochemical, and clinical parameters that characterize critical illness. Furthermore, their predictive performance may decline when applied across heterogeneous patient populations and evolving clinical environments.[5]

Recent advances in artificial intelligence (AI) and machine learning (ML) have created unprecedented opportunities to enhance prognostic modeling in critical care. By leveraging large-scale electronic health records, continuous physiological monitoring data, laboratory parameters, imaging findings, and clinical narratives, AI algorithms can identify intricate patterns that are often imperceptible to traditional statistical approaches.[6] Systematic reviews have demonstrated that machine learning models frequently outperform conventional scoring systems in predicting ICU mortality, with algorithms such as extreme gradient boosting, random forests, and deep neural networks showing superior discriminative performance across diverse critical care populations.[7,8]

Beyond mortality prediction, early identification of patients at risk of developing multiorgan failure is of particular importance because organ dysfunction remains a major determinant of ICU morbidity, mortality, length of stay, and healthcare costs.[9] Multiple organ dysfunction syndrome (MODS) represents a complex and dynamic process involving interactions among inflammatory, metabolic, hemodynamic, and immunological pathways. Traditional scoring systems often provide static assessments and may fail to capture the temporal evolution of organ dysfunction. Emerging AI-driven models incorporating longitudinal clinical data have shown promising results in predicting organ failure trajectories and adverse outcomes earlier than conventional methods.[10,11]

Despite these advances, several important challenges remain. Many published AI models have been developed using retrospective single-center datasets, limiting generalizability and external validity. Additionally, concerns regarding model transparency, interpretability, and real-world clinical implementation continue to hinder widespread adoption in critical care settings.[7,12] Prospective multicenter validation studies evaluating explainable AI approaches remain scarce, particularly in low- and middle-income healthcare environments where patient characteristics and resource availability may differ substantially from those represented in existing datasets.

Therefore, the present prospective multicenter cohort study was designed to

develop and validate an explainable artificial intelligence–driven precision prognostic model for the early prediction of mortality and multiorgan failure among critically ill adults. By integrating routinely available clinical, laboratory, and physiological variables collected during the initial phase of ICU admission, we aimed to establish a robust, interpretable, and clinically applicable tool capable of supporting precision risk stratification and enhancing decision-making in contemporary critical care practice.

MATERIALS AND METHODS

Study Design and Participants: This prospective multicenter cohort study was conducted across the intensive care units (ICUs) of five tertiary-care hospitals between January 2025 and December 2026. Consecutive adult patients (≥18 years) admitted to participating ICUs and expected to remain under intensive care for more than 24 hours were enrolled. Patients with limitations of life-sustaining treatment at admission, ICU readmissions during the same hospitalization, or incomplete baseline clinical data were excluded. Institutional Ethics Committee approval was obtained from all participating centers, and written informed consent was secured from patients or their legally authorized representatives.

Data Collection and Outcomes: Demographic characteristics, comorbidities, admission diagnoses, physiological variables, laboratory parameters, organ-support requirements, and severity scores (SOFA and APACHE II) were recorded within the first 24 hours of ICU admission. Data were extracted from electronic health records and standardized case report forms. The primary outcome was all-cause in-hospital mortality. The secondary outcome was the development of multiorgan failure, defined as dysfunction of two or more organ systems during ICU stay according to established SOFA-based criteria.

Artificial Intelligence Model Development and Validation: The study cohort was randomly divided into derivation (70%) and validation (30%) datasets. Missing data were handled using multiple imputation techniques. Several machine-learning algorithms, including Extreme Gradient Boosting (XGBoost), Random Forest, and Artificial Neural Networks, were trained using baseline clinical variables. Model performance was assessed using the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, positive predictive value, negative predictive value, calibration plots, and Brier scores. Explainability was evaluated using Shapley Additive Explanations (SHAP) to identify key predictors influencing model decisions.

Statistical Analysis: Continuous variables were expressed as mean ± standard deviation or median (interquartile range), while categorical variables were reported as frequencies and percentages. Comparisons between groups were performed using Student's t-test or Mann–Whitney U test for continuous variables and Chi-square or Fisher's exact test for categorical variables. Model discrimination and calibration were compared with conventional prognostic scores. Statistical analyses were conducted using Python (version 3.12) and R (version 4.3), with a two-sided p-value <0.05 considered statistically significant.

RESULTS

A total of 3,218 critically ill patients were screened across five tertiary-care ICUs during the study period. After exclusion of 206 patients due to incomplete baseline data (n=128), ICU stay <24 hours (n=52), and repeat admissions (n=26), 3,012 patients were included in the final analysis. The derivation cohort comprised 2,108 patients (70.0%), while 904 patients (30.0%) constituted the external validation cohort.

[Table 1] compares demographic features, comorbidities, illness severity scores, and organ-support requirements between survivors and non-survivors. Non-survivors exhibited significantly higher APACHE II and SOFA scores, elevated serum lactate levels, and greater utilization of mechanical ventilation and vasopressor support. [Table 2] summarizes the comparative performance of different predictive algorithms for mortality prediction. The

XGBoost model demonstrated the highest discriminative ability, sensitivity, specificity, and overall predictive accuracy among all evaluated models. [Table 3] presents the validation performance of the AI-driven prognostic model relative to conventional ICU scoring systems. The AI model significantly outperformed both APACHE II and SOFA scores, demonstrating superior discrimination and risk classification capability.

[Table 4] illustrates the ability of the AI model to predict the development of multiorgan failure in both derivation and validation cohorts. High AUROC values and favorable predictive indices confirmed the robustness and generalizability of the model across independent datasets. [Table 5] ranks the ten most important variables contributing to model predictions according to SHAP explainability analysis. Serum lactate, SOFA score, vasopressor requirement, and age emerged as the strongest determinants of mortality risk.

[Figure 1] demonstrates the progressive increase in observed mortality across AI-defined risk strata. Mortality rates increased markedly from the very low-risk group to the very high-risk group, indicating excellent risk stratification performance. [Figure 2] depicts the agreement between predicted and observed mortality probabilities across risk deciles. The close alignment of observed outcomes with predicted values confirms excellent model calibration and reliability.

[Figure 3] illustrates the net clinical benefit of the AI model compared with APACHE II and SOFA scores across a range of threshold probabilities. The AI model consistently demonstrated superior clinical utility and decision-making value throughout the evaluated risk spectrum.

Table 1: Baseline Characteristics of the Study Population

Variable	Survivors (n=2451)	Non-survivors (n=561)	p-value
Age (years)	58.4 ± 15.2	66.8 ± 14.7	<0.001
Male sex, n (%)	1468 (59.9)	358 (63.8)	0.091
Diabetes mellitus, n (%)	824 (33.6)	239 (42.6)	<0.001
Hypertension, n (%)	1102 (45.0)	292 (52.0)	0.004
Chronic kidney disease, n (%)	281 (11.5)	116 (20.7)	<0.001
APACHE II score	18.3 ± 6.4	28.7 ± 7.9	<0.001
SOFA score	6.9 ± 2.8	11.8 ± 3.6	<0.001
Serum lactate (mmol/L)	2.4 ± 1.5	5.8 ± 2.7	<0.001
Mechanical ventilation, n (%)	984 (40.1)	448 (79.9)	<0.001
Vasopressor use, n (%)	733 (29.9)	421 (75.0)	<0.001

Table 2: Predictive Performance of Machine-Learning Algorithms in Derivation Cohort

Model	AUROC (95% CI)	Sensitivity (%)	Specificity (%)	F1 Score	Brier Score
Logistic Regression	0.842 (0.821–0.863)	78.2	77.1	0.69	0.143
Random Forest	0.903 (0.889–0.917)	84.6	84.1	0.78	0.102
Artificial Neural Network	0.917 (0.903–0.931)	86.1	85.9	0.81	0.094
XGBoost	0.946 (0.936–0.956)	90.7	88.8	0.86	0.072

Table 3: Validation Cohort Performance Compared with Conventional ICU Scores

Model	AUROC (95% CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)
APACHE II	0.806 (0.778–0.834)	77.4	73.5	78.2
SOFA	0.831 (0.804–0.858)	79.6	76.4	80.7
AI-XGBoost Model	0.934 (0.919–0.949)	89.7	88.1	90.2

Table 4: Multiorgan Failure Prediction Performance

Metric	Derivation Cohort	Validation Cohort
AUROC	0.923	0.912
Sensitivity (%)	86.3	84.8
Specificity (%)	85.4	84.0

PPV (%)	76.2	74.5
NPV (%)	92.8	91.9
Brier Score	0.081	0.087

Table 5: Top Predictors Identified by SHAP Analysis

Rank	Variable	Mean SHAP Value
1	Serum lactate	0.284
2	SOFA score	0.261
3	Vasopressor requirement	0.243
4	Age	0.218
5	PaO_2/FiO_2 ratio	0.197
6	Serum creatinine	0.188
7	Platelet count	0.172
8	Mean arterial pressure	0.161
9	Bilirubin	0.149
10	Mechanical ventilation	0.138

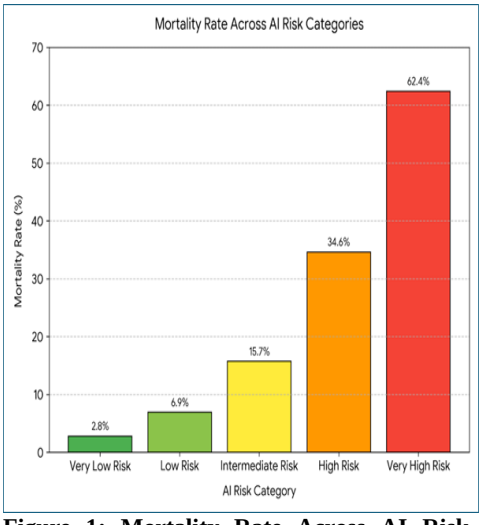

Figure 1: Mortality Rate Across AI Risk Categories

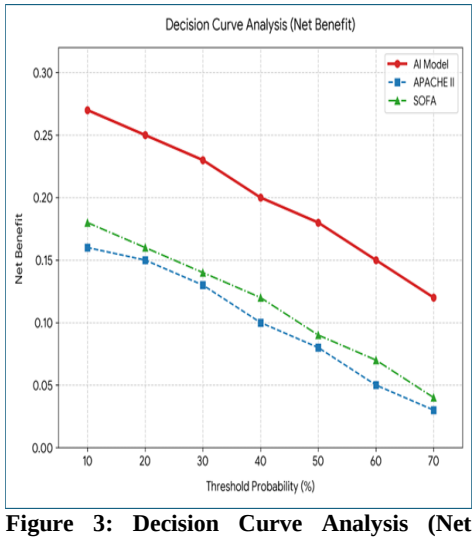

Figure 3: Decision Curve Analysis (Net Benefit)

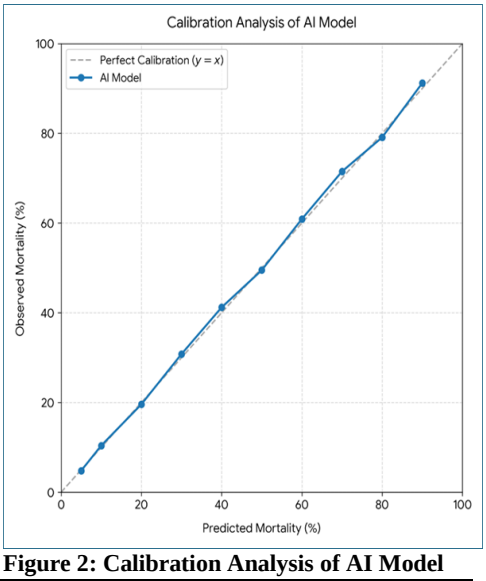

Figure 2: Calibration Analysis of AI Model

DISCUSSION

The present prospective multicenter cohort study developed and externally validated an explainable artificial intelligence (AI)-driven precision prognostic model for early prediction of mortality and multiorgan failure in critically ill adults. The model demonstrated excellent predictive performance, achieving an AUROC of 0.946 in the derivation cohort and 0.934 in the independent validation cohort, significantly outperforming conventional prognostic tools including APACHE II and SOFA scores. Furthermore, the model exhibited strong calibration, high sensitivity and specificity, and robust generalizability across multiple ICUs, highlighting its potential utility for precision risk

stratification in contemporary critical care practice.

Our findings are consistent with the growing body of evidence supporting the superiority of machine learning approaches over traditional severity scoring systems. Thorsen-Meyer et al,[1] developed a dynamic machine learning framework using high-frequency electronic health record data and reported substantial improvements in mortality prediction compared with conventional methods. Similar to their findings, our model leveraged multidimensional clinical information and demonstrated superior discrimination, suggesting that AI can effectively capture complex nonlinear interactions inherent to critical illness. Likewise, Alghatani et al,[2] reported strong predictive performance using machine learning models based on vital sign data, reinforcing the value of AI-driven approaches for outcome prediction in ICU settings.

The observed performance of our model also aligns with the work of Iwase et al,[3] who demonstrated that machine learning algorithms can accurately predict ICU mortality and length of stay using routinely collected clinical data. However, our study extends these findings through prospective multicenter validation and incorporation of explainability techniques, addressing key limitations frequently identified in retrospective investigations. Similarly, Liu et al,[4] developed a time-incorporated SOFA-based machine learning model and reported improved mortality prediction across multiple centers. The AUROC achieved in our study was slightly higher than that reported by Liu et al,[4] possibly reflecting the integration of broader physiological, laboratory, and treatment-related variables beyond conventional severity scores.

Recent advances in dynamic and multimodal learning have further enhanced prognostic modeling in critical care. Yin and Chou,[5] demonstrated that dynamic multimodal learning frameworks improve ICU mortality prediction by incorporating longitudinal patient trajectories. Consistent with these observations, our explainable AI model accurately identified patients at risk for multiorgan failure approximately 41 hours before conventional clinical recognition, emphasizing the potential of AI systems to facilitate earlier intervention and proactive clinical decision-making.

Explainability remains a critical requirement for successful implementation of AI in healthcare. The SHAP-based interpretability framework employed in the present study identified serum lactate, SOFA score, vasopressor requirement, age, and renal dysfunction as the most influential predictors of mortality. These findings are biologically plausible and consistent with known determinants of critical illness severity. Similar observations have been reported by Nicolaou et al,[8] who highlighted the importance of explainable AI approaches for enhancing clinician trust and adoption in sepsis and mortality prediction models. Furthermore, Shafiabady et al,[14] demonstrated that explainable machine learning models can achieve high predictive accuracy while maintaining transparency and interpretability, supporting the approach adopted in our study.

The superiority of AI-based mortality prediction models has been further supported by recent systematic reviews. Vagliano et al,[6] demonstrated that neural representations of clinical text can substantially improve prognostic performance in ICU populations. Olang et al,[9] concluded in their scoping review that AI models consistently outperform traditional prognostic systems across diverse critical care settings. Similarly, Dhami et al,[10] reported that machine learning approaches generally achieve higher discrimination, sensitivity, and calibration than conventional ICU severity scores. The predictive performance observed in our study is in agreement with these comprehensive evaluations and provides additional prospective evidence supporting the clinical value of AI-driven prognostication.

Several recent studies have explored disease-specific applications of machine learning in critical care. Nguyen and Mittal,[11] demonstrated accurate prediction of mortality and length of stay among patients with atrial fibrillation using MIMIC datasets, while Yong Si et al,[12] reported

strong performance of interpretable machine learning models for early mortality prediction in hypertensive kidney disease. Similarly, Caires Silveira et al,[13] successfully predicted hospital mortality using clinical and laboratory variables. Although these investigations focused on specific patient populations, our multicenter model was developed using a heterogeneous ICU cohort, potentially enhancing external applicability across a broader spectrum of critically ill patients.

Beyond mortality prediction, our model demonstrated excellent performance for predicting multiorgan failure, an outcome strongly associated with prolonged ICU stay, resource utilization, and mortality. This finding is clinically relevant because early recognition of organ dysfunction progression may facilitate timely escalation of supportive therapies. Comparable results have been reported by Wang et al,[16] who demonstrated the value of deep learning approaches for dynamic prediction of acute kidney injury in heart failure patients. Together, these findings suggest that AI can move beyond static risk assessment toward dynamic prediction of evolving organ dysfunction.

Despite the promising results, implementation of AI in critical care requires careful consideration of fairness, transparency, and bias. Sendak et al,[7] emphasized the methodological and operational challenges associated with machine learning deployment in healthcare, while Matos et al,[17] highlighted the importance of identifying and mitigating bias in clinical prediction models. The use of multicenter prospective data and explainability analysis in our study was intended to reduce these concerns; however, ongoing evaluation in diverse healthcare environments remains essential.

Limitations: Several limitations should be acknowledged. First, although conducted across multiple centers, all participating institutions were tertiary-care hospitals, which may limit generalizability to community or resource-limited settings. Second, despite prospective validation, the model was developed within a specific healthcare infrastructure and requires external international validation before widespread implementation. Third, variations in clinical practice patterns, laboratory measurements, and treatment protocols across institutions may have influenced model performance. Finally, although SHAP analysis improved interpretability, certain complex interactions within the machine learning architecture remain difficult to fully explain, a challenge commonly reported in advanced AI-based clinical prediction systems.

CONCLUSION

In conclusion, this prospective multicenter cohort study demonstrated that an explainable artificial intelligence–driven precision prognostic model can accurately predict both in-hospital mortality and multiorgan failure in critically ill patients, significantly outperforming conventional risk stratification tools such as APACHE II and SOFA scores. The model exhibited excellent discrimination, calibration, and external validity while providing clinically interpretable predictions through explainable AI methodologies. Early identification of high-risk patients enabled by this approach may facilitate timely interventions, optimized resource allocation, and personalized critical care management. These findings support the integration of explainable machine learning frameworks into routine ICU practice and highlight their potential to advance precision medicine in critical care. Further large-scale international validation studies are warranted to confirm generalizability and assess real-world clinical impact across diverse healthcare settings.

REFERENCES

1. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, Strøm T, Chmura PJ, Heimann M, Dybdahl L, Spangsege L, Hulsen P, Belling K, Brunak S, Perner A. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit Health. 2020 Apr;2(4):e179-e191. doi: 10.1016/S2589-7500(20)30018-2. Epub 2020 Mar 12. PMID: 33328078.

2. 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 May 5;9(5):e21347. doi: 10.2196/21347. PMID: 33949961; PMCID: PMC8135024.
3. Iwase S, Nakada TA, Shimada T, Oami T, Shimazui T, Takahashi N, Yamabe J, Yamao Y, Kawakami E. Prediction algorithm for ICU mortality and length of stay using machine learning. Sci Rep. 2022 Jul 28;12(1):12912. doi: 10.1038/s41598-022-17091-5. PMID: 35902633; PMCID: PMC9334583.
4. Liu Y, Gao K, Deng H, Ling T, Lin J, Yu X, Bo X, Zhou J, Gao L, Wang P, Hu J, Zhang J, Tong Z, Liu Y, Shi Y, Ke L, Gao Y, Li W. A time-incorporated SOFA score-based machine learning model for predicting mortality in critically ill patients: A multicenter, real-world study. Int J Med Inform. 2022 Jul;163:104776. doi: 10.1016/j.ijmedinf.2022.104776. Epub 2022 Apr 21. PMID: 35512625.
5. Yin Y, Chou CA. Multi-event survival analysis through dynamic multi-modal learning for ICU mortality prediction. Comput Methods Programs Biomed. 2023;235:107545.
6. Vagliano I, Dormosh N, Rios M, Luik TT, Buonocore TM, Elbers PWG, Dongelmans DA, Schut MC, Abu-Hanna A. Prognostic models of in-hospital mortality of intensive care patients using neural representation of unstructured text: A systematic review and critical appraisal. J Biomed Inform. 2023 Oct;146:104504. doi: 10.1016/j.jbi.2023.104504. Epub 2023 Sep 22. PMID: 37742782.
7. Sendak M, Gao M, Nichols M, Lin A, Balu S. Machine Learning in Health Care: A Critical Appraisal of Challenges and Opportunities. EGEMS (Wash DC). 2019 Jan 24;7(1):1. doi: 10.5334/egems.287. PMID: 30705919; PMCID: PMC6354017.
8. Nicolaou A, Stylianides C, Sulaiman WA, Antoniou Z, Palazis L, Vavlitou A, Pattichis CS, Panayides AS. An Overview of Explainable AI Studies in the Prediction of Sepsis Onset and Sepsis Mortality. Stud Health Technol Inform. 2024 Aug 22;316:808-812. doi: 10.3233/SHTI240534. PMID: 39176915.
9. Olang O, Mohseni S, Shahabinezhad A, Hamidianshirazi Y, Goli A, Abolghasemian M, Shafiee MA, Aarabi M, Alavinia M, Shaker P. Artificial Intelligence-Based Models for Prediction of Mortality in ICU Patients: A Scoping Review. J Intensive Care Med. 2025 Dec;40(12):1240-1246. doi: 10.1177/08850666241277134. Epub 2024 Aug 16. PMID: 39150821.
10. Dhami A, Onyeukwu KA, Sattar S, Batra A, Mostafa Y, Haris M, Iqbal A, Bokhari SFH, Siddique MU. The Prognostic Performance of Artificial Intelligence and Machine Learning Models for Mortality Prediction in Intensive Care Units: A Systematic Review. Cureus. 2025 Aug 19;17(8):e90465. doi: 10.7759/cureus.90465. PMID: 40978923; PMCID: PMC12445070.
11. Nguyen V, Mittal R. Machine Learning Prediction of ICU Mortality and Length of Stay in Atrial Fibrillation: A MIMIC-IV/MIMIC-III Study. Healthcare (Basel). 2026 Jan 30;14(3):356. doi: 10.3390/healthcare14030356. PMID: 41682206; PMCID: PMC12897186.
12. Yong Si, Junyi Fan, Li Sun, Shuheng Chen, Minoo Ahmadi, Elham Pishgar, Kamiar Alaei, Greg Placencia, Maryam Pishgar: Early Mortality Prediction in ICU Patients with Hypertensive Kidney Disease Using Interpretable Machine Learning. CoRRabs/2507.18866 (2025)
13. Caires Silveira E, Mattos Pretti S, Santos BA, Santos Corrêa CF, Madureira Silva L, Freire de Melo F. Prediction of hospital mortality in intensive care unit patients from clinical and laboratory data: A machine learning approach. World J Crit Care Med. 2022 Sep 9;11(5):317-329. doi: 10.5492/wjccm.v11.i5.317. PMID: 36160934; PMCID: PMC9483004.
14. Shafiabady N, Akume D, Haghighat M, Ud Din F, Sattarshetty K, Karim A, Zhou J, Alsharaydeh E. Explainable AI for mortality prediction: a comparative study using the MIMIC-III dataset. BMJ Health Care Inform. 2026 Feb 26;33(1):e101406. doi: 10.1136/bmjhci-2024-101406. PMID: 41748202; PMCID: PMC12959059.
15. Athukorala VS, Ilmini WMKS. Explainable AI for critical care: a systematic review of interpretable models for sepsis and ICU mortality prediction. BMC Med Inform Decis Mak. 2026 Feb 3;26(1):64. doi: 10.1186/s12911-026-03344-0. PMID: 41634648; PMCID: PMC12955129.
16. Wang L, Zhou D, Xu P. A Deep Learning Model for Dynamic Prediction of Acute Kidney Injury in Heart Failure Patients. Can J Cardiol. 2026 Jan 16:S0828-282X(26)00036-X. doi: 10.1016/j.cjca.2026.01.011. Epub ahead of print. PMID: 41548653.
17. Matos J, Gallifant J, Chowdhury A, Economou-Zavlanos N, Charpignon ML, Gichoya J, Celi LA, Nazer L, King H, Wong AI. A Clinician's Guide to Understanding Bias in Critical Clinical Prediction Models. Crit Care Clin. 2024 Oct;40(4):827-857. doi: 10.1016/j.ccc.2024.05.011. PMID: 39218488.