



AI-ASSISTED PREDICTION OF ICU MORTALITY USING ROUTINE CLINICAL PARAMETERS: A SYSTEMATIC REVIEW

Dr. (Major) Sourabh Kumar^{1*}, Dr. Kriti Prasad², Dr. Prabha Rashmi Lakra³,
Dr. Vashist Raj⁴, Dr. Vikas Kumar⁵, Dr. Shio Priye⁶, Dr. Ladhu Lakra⁷

^{1*}Assistant Professor Intensive Care, Department of Anaesthesiology, Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, India. Mail ID: sourabh_bhu0@yahoo.co.in, Contact: 9162976054

²Consultant & Senior IVF Specialist, Seeds of Innocens, Ranchi, Jharkhand, India. Mail ID: kritipras@gmail.com, Contact: 9599976112

³Senior Resident, Department of Anaesthesiology, Rajendra Institute of Medical Sciences, Ranchi. Mail ID: lakra.pr@gmail.com, Contact: 7045704145

⁴Junior Resident, Santosh Medical College, Ghaziabad, Uttar Pradesh, India. Mail ID: drvashistjaiswal@gmail.com, Contact: 8861056060

⁵Junior Resident, Department of Anaesthesiology, Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, India. Mail ID: vikaschauhan9119@gmail.com, Contact: 8789805651

⁶Professor, Department of Anaesthesiology, Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, India. Mail ID: shiopriye@gmail.com, Contact: 9661120481

⁷Professor & Head of Department, Department of Anaesthesiology, Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, India. Mail ID: dr.ladhulakra007@gmail.com, Contact: 9835160608

***Corresponding Author:** Dr. (Major) Sourabh Kumar,
Assistant Professor Intensive Care, Department of Anaesthesiology, Rajendra Institute of Medical Sciences, Ranchi, Jharkhand, India. Mail ID: sourabh_bhu0@yahoo.co.in, Contact: 9162976054

Abstract

Background: Accurate mortality prediction is critical in intensive care units (ICUs) for risk stratification, clinical decision-making, and resource allocation. Traditional prognostic scoring systems have limitations in capturing the complex and dynamic nature of critical illness. Artificial intelligence (AI) has emerged as a promising approach for improving mortality prediction using routinely collected clinical data.

Objective: To systematically evaluate the performance, clinical applicability, and limitations of AI-assisted mortality prediction models developed using routine clinical parameters in ICU populations.

Methods: A systematic review was conducted following PRISMA 2020 guidelines. Searches were performed in PubMed, Scopus, Web of Science, Embase, and Google Scholar for studies published between Feb 2020 - Nov 2022. Studies investigating AI or machine learning models for mortality prediction in adult ICU patients using routinely available clinical parameters were included.

Results: Eleven studies met the inclusion criteria and were included in the final synthesis. Commonly used algorithms included XGBoost, Gradient Boosting Machine, Random Forest, Vol. 29 No.03 (2022); JPTCP (2433 -2446)

CatBoost, artificial neural networks, and long short-term memory models. Reported AUROC values ranged from 0.79 to 0.92, with ensemble learning approaches demonstrating the highest predictive performance. Age, Glasgow Coma Scale score, blood urea nitrogen, serum creatinine, heart rate, respiratory rate, and oxygen saturation were consistently identified as important predictors. AI models generally outperformed conventional prognostic scoring systems while increasingly incorporating explainability techniques to improve transparency and clinical interpretability.

Conclusion: AI-assisted mortality prediction using routine clinical parameters demonstrates strong predictive performance and significant potential for supporting critical care decision-making.

Further external validation and prospective implementation studies are needed to facilitate clinical adoption.

Keywords: Artificial intelligence; Machine learning; Intensive care unit; Mortality prediction; Critical care; Routine clinical parameters; Explainable artificial intelligence.

Introduction

1.1 Burden of ICU Mortality

Intensive care units (ICUs) provide specialized care for patients with life-threatening illnesses and organ dysfunction. Despite substantial advances in monitoring technologies, supportive therapies, and evidence-based critical care practices, mortality among critically ill patients remains a major healthcare concern worldwide. Accurate estimation of mortality risk is essential for guiding clinical decision-making, prioritizing interventions, optimizing resource utilization, and facilitating communication with patients and their families. Consequently, reliable prognostic tools have become an integral component of modern critical care practice (1).

1.2 Limitations of Conventional Mortality Prediction Models

Traditional prognostic scoring systems, including the Acute Physiology and Chronic Health Evaluation (APACHE), Sequential Organ Failure Assessment (SOFA), and Simplified Acute Physiology Score (SAPS), have been widely adopted to estimate disease severity and mortality risk in ICU populations. Although these models have demonstrated clinical utility, they are constrained by their reliance on predefined variables, assumptions of linear relationships among predictors, and dependence on static measurements collected at specific time points. Such limitations may reduce their ability to capture the dynamic and multifactorial nature of critical illness, thereby affecting predictive accuracy in heterogeneous patient populations (2).

1.3 Emergence of Artificial Intelligence in Critical Care

Recent developments in artificial intelligence (AI) and machine learning (ML) have transformed predictive analytics across healthcare. AI-based methods can process large volumes of structured and unstructured clinical data, identify complex nonlinear relationships, and continuously improve prediction performance through iterative learning. These characteristics make AI particularly suitable for critical care settings, where patient outcomes are influenced by numerous interacting physiological and clinical variables that evolve over time. Consequently, AI has emerged as a promising approach for enhancing risk prediction and supporting clinical decision-making in ICUs (3). The increasing availability of electronic health records and large critical care databases, including the Medical Information Mart for Intensive Care (MIMIC) and the eICU Collaborative Research Database, has further accelerated the development of AI-driven predictive models. Multiple machine learning algorithms, such as random forests, gradient boosting machines, support vector machines, artificial neural networks, and deep learning architectures, have been applied to predict ICU mortality with encouraging results (4).

1.4 AI-Assisted Mortality Prediction in ICU Patients

Several studies have demonstrated that AI-assisted prediction models can achieve superior performance compared with conventional scoring systems. Dynamic machine learning approaches are capable of continuously updating mortality estimates by incorporating temporal changes in patient status throughout the ICU stay. Such models may provide more individualized and clinically relevant risk assessments than traditional static prediction tools. In particular, deep learning architectures and ensemble learning methods have shown strong predictive capabilities across diverse ICU populations, highlighting the growing potential of AI-assisted prognostication in critical care medicine (5).

Recent evidence indicates that machine learning models consistently outperform traditional prognostic systems in terms of discrimination and predictive accuracy. Algorithms such as Extreme Gradient Boosting (XGBoost), Random Forest, and CatBoost have demonstrated particularly strong performance across multiple datasets and clinical settings, supporting their potential role in future ICU decision-support systems (6).

1.5 Importance of Routine Clinical Parameters

A notable feature of many high-performing AI models is their reliance on routinely collected clinical parameters that are readily available during standard ICU care. These variables typically include demographic characteristics, vital signs, neurological assessments, and basic laboratory measurements. Age, heart rate, blood pressure, respiratory rate, oxygen saturation, Glasgow Coma Scale score, blood urea nitrogen, serum creatinine, and urine output have consistently emerged as important predictors of mortality across multiple studies (7).

The use of routine clinical parameters offers several advantages for clinical implementation. Because these data are collected as part of routine patient care, predictive models based on such variables do not require additional diagnostic procedures, specialized equipment, or expensive biomarkers. Furthermore, reliance on readily available clinical information enhances the feasibility of implementing AI-assisted mortality prediction systems across a wide range of healthcare environments, including resource-limited settings (8).

1.6 Research Gap and Rationale

Although the application of AI in critical care has expanded rapidly, several challenges continue to limit its translation into routine clinical practice. Existing studies vary considerably with respect to study populations, predictor selection, model development methodologies, validation procedures, and reporting standards. In addition, many predictive models have been developed using retrospective datasets and have undergone limited external validation, raising concerns regarding generalizability and real-world clinical applicability (9).

Previous reviews have examined the broader role of artificial intelligence in critical care; however, a focused synthesis specifically evaluating AI-assisted prediction of ICU mortality using routine clinical parameters remains limited. Given the practical importance of routinely available clinical variables for real-world implementation, a comprehensive evaluation of current evidence is warranted. Therefore, this systematic review aims to critically assess the performance, clinical applicability, and limitations of AI-assisted mortality prediction models developed using routine clinical parameters in ICU populations (10).

2. Methods

2.1 Study Design and Reporting Standards

This systematic review was conducted to evaluate the evidence regarding artificial intelligence-assisted prediction of intensive care unit (ICU) mortality using routine clinical parameters. The review methodology was developed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement to ensure methodological transparency, reproducibility, and comprehensive reporting of the review process (11).

2.2 Review Question and Eligibility Criteria

The review question was formulated using the Population, Intervention, Comparator, Outcome, Timing, and Setting (PICOTS) framework. Eligibility criteria and extracted study variables are summarized in Table 1.

Studies were included if they investigated AI- or machine learning-based mortality prediction models in adult ICU populations using routinely collected clinical parameters. Studies focusing exclusively on imaging, genomic data, pediatric populations, or non-ICU settings were excluded. Additional inclusion and exclusion criteria are presented in Table 1 (12).

Table 1. Eligibility Criteria and Data Extraction Framework for Study Selection

Domain	Criteria/Variables
Population	Adult patients (≥ 18 years) admitted to intensive care units (ICUs)
Intervention	Artificial intelligence (AI), machine learning (ML), or deep learning models developed for mortality prediction
Comparator	Traditional prognostic scores (APACHE, SOFA, SAPS) where available
Outcome	ICU mortality, in-hospital mortality, 28-day mortality, 30-day mortality, or 90-day mortality
Study Design	Original observational studies, retrospective cohort studies, prospective cohort studies
Publication Period	Feb 2020 - Nov 2022
Language	English
Inclusion Criteria	Studies evaluating AI-assisted mortality prediction using routinely collected clinical parameters in ICU patients
Exclusion Criteria	Reviews, editorials, conference abstracts, pediatric studies, imaging-only studies, genomic-only studies, non-ICU populations
Data Extracted	Author, year, country, dataset, sample size, AI algorithm, predictor variables, mortality outcome, validation method, AUROC, sensitivity, specificity, accuracy
Quality Assessment Tool	Prediction Model Risk of Bias Assessment Tool (PROBAST)

2.3 Information Sources and Search Strategy

A comprehensive literature search was conducted using PubMed, Scopus, Web of Science, Embase, and Google Scholar. Search terms related to artificial intelligence, machine learning, mortality prediction, intensive care, and routine clinical parameters were combined using Boolean operators. The search strategy was developed according to established recommendations for systematic evidence synthesis and database searching (13).

2.4 Study Selection

All retrieved records were imported into reference management software, and duplicate records were removed prior to screening. Title and abstract screening was performed initially, followed by full-text assessment of potentially eligible studies. Disagreements regarding study eligibility were resolved through discussion and consensus.

The study identification, screening, eligibility assessment, and final inclusion process are illustrated in the PRISMA flow diagram (Figure 1) (14).

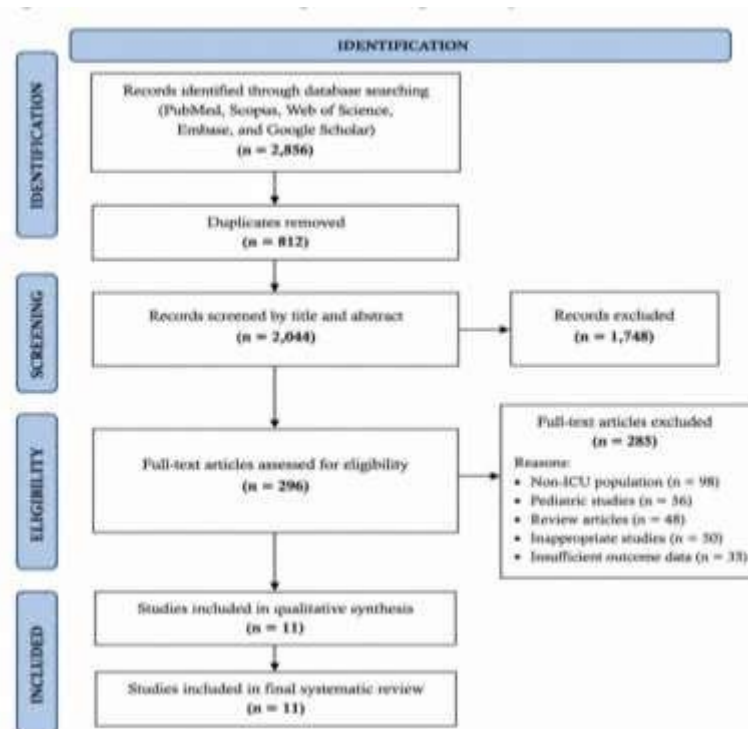


Figure 1. PRISMA 2020 Flow Diagram Showing the Study Selection Process

2.5 Data Extraction

Data were extracted using a predefined standardized framework. Information collected from each study included publication characteristics, study design, patient population, data source, predictor variables, machine learning algorithms, mortality outcomes, validation methods, and performance metrics. The variables extracted from each eligible study are summarized in Table 1 (15).

2.6 Quality Assessment and Data Synthesis

Methodological quality and risk of bias were evaluated using the Prediction Model Risk of Bias Assessment Tool (PROBAST). The assessment focused on participant selection, predictor variables, outcome definitions, and analytical methods.

Due to anticipated heterogeneity in study populations, predictor selection, machine learning algorithms, and reported outcomes, findings were synthesized narratively rather than through quantitative meta-analysis. The overall methodological approach followed recommendations for systematic reviews of prediction model studies (16, 17).

3. Results

3.1 Characteristics of Included Studies

A total of 11 studies published between Feb 2020 - Nov 2022 were included in the final synthesis. The included studies evaluated the application of artificial intelligence (AI) and machine learning (ML) techniques for mortality prediction in diverse intensive care unit (ICU) populations. Most studies employed retrospective cohort designs using large critical care databases such as the Medical Information Mart for Intensive Care (MIMIC-III), MIMIC-IV, and the eICU Collaborative Research Database. Additional studies utilized institution-specific ICU registries and multicenter cohorts.

The study populations included general ICU patients as well as disease-specific cohorts comprising patients with sepsis, critically ill influenza, and heart failure. Sample sizes varied substantially across studies, ranging from 336 patients to more than 200,000 ICU admissions. Mortality outcomes included ICU mortality, in-hospital mortality, 30-day mortality, and 90-day mortality. The detailed characteristics of all included studies are presented in Table 2.

Table 2. Characteristics of Included Studies

Author	Year	Dataset	Population	Sample Size	Mortality Outcome
Thorsen-Meyer et al.	2020	Danish Multicenter ICU Database	General ICU	11,492 patients (14,190 admissions)	90-day Mortality
Hu et al.	2020	Multicenter Hospital Cohort	Critically Ill Influenza Patients	336	30-day Mortality
Kong et al.	2020	MIMIC-III	Sepsis Patients	16,688	In-hospital Mortality
Baker et al.	2020	MIMIC-III	General ICU	51,279 ICU stays	ICU Mortality
Yu et al.	2020	MIMIC-III	General ICU	22,049	ICU Mortality
Zhang et al.	2020	MIMIC-III	Sepsis-3 Patients	4,559	30-day Mortality
Alghatani et al.	2021	MIMIC Database	General ICU	44,626	ICU Mortality
Li et al.	2021	MIMIC-III	Heart Failure ICU Patients	1,177	In-hospital Mortality
Pang et al.	2022	MIMIC-IV	General ICU	14,110	ICU Mortality
Safaei et al.	2022	eICU-CRD	General ICU	>200,000 ICU Admissions	ICU Mortality
Li et al.	2022	eICU-CRD	Heart Failure ICU Patients	2,798	In-hospital Mortality

3.2 Artificial Intelligence Algorithms Used for Mortality Prediction

A wide variety of machine learning algorithms were employed across the included studies. The distribution of AI algorithms is shown in Figure 2, while detailed model characteristics are summarized in Table 3. Extreme Gradient Boosting (XGBoost) was the most frequently utilized algorithm and was consistently associated with high predictive performance. Random Forest models were also commonly used and demonstrated robust discrimination across multiple datasets. Other frequently reported methods included artificial neural networks, long short-term memory (LSTM) networks, CatBoost models, and ensemble machine-learning approaches. Several studies adopted hybrid architectures combining multiple algorithms to improve predictive accuracy. Several studies incorporated interpretable machine-learning frameworks and feature-importance analyses to improve model transparency and facilitate clinical interpretation. Internal validation was performed in nearly all studies, whereas temporal validation was reported less frequently.

Table 3. AI Model Characteristics and Validation Methods

Study	Algorithm	Validation Method	Explainability
Baker et al.	CNN-BiLSTM	Ten-fold Cross-validation	No
Hu et al.	XGBoost	Internal Validation	SHAP
Kong et al.	Gradient Boosting	Five-fold Cross-	Calibration Analysis

	Machine	validation	
Yu et al.	Bidirectional LSTM	Train/Validation/Test Split	No
Thorsen-Meyer et al.	Dynamic Deep Learning Model	Temporal Validation	SHAP
Zhang et al.	XGBoost	Internal Validation	Clinical Impact Analysis
Alghatani et al.	XGBoost / Random Forest	Train/Test Split	No
Li et al. (2021)	XGBoost-Based Model	Internal Validation	Nomogram
Pang et al.	XGBoost	Internal Validation	SHAP
Safaei et al.	CatBoost / E-CatBoost	Ten-fold Cross-validation	SHAP + LIME
Li et al. (2022)	XGBoost	Internal Validation	SHAP

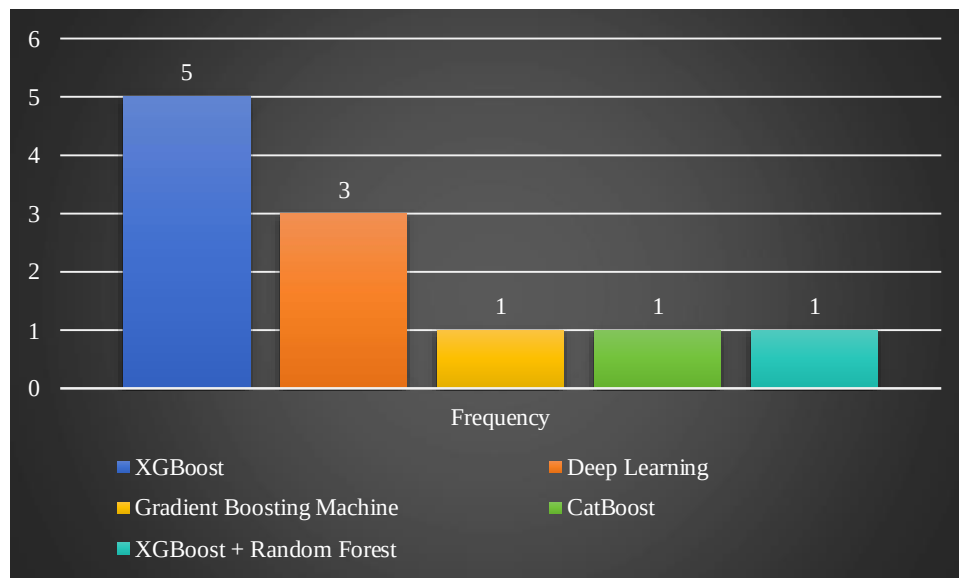


Figure 2. Distribution of Artificial Intelligence Algorithms Used for ICU Mortality Prediction

3.3 Routine Clinical Parameters Used in Mortality Prediction Models

The clinical variables used for mortality prediction are summarized in Table 4, while the frequency of predictor utilization across studies is illustrated in Figure 3.

Despite differences in patient populations and model architectures, several routinely collected clinical parameters consistently emerged as important predictors of mortality. Demographic variables, particularly age and sex, were included in most models. Frequently utilized physiological variables included heart rate, respiratory rate, blood pressure, oxygen saturation, temperature, and Glasgow Coma Scale (GCS) score. Common laboratory predictors included blood urea nitrogen (BUN), serum creatinine, lactate concentration, hemoglobin, and electrolyte measurements.

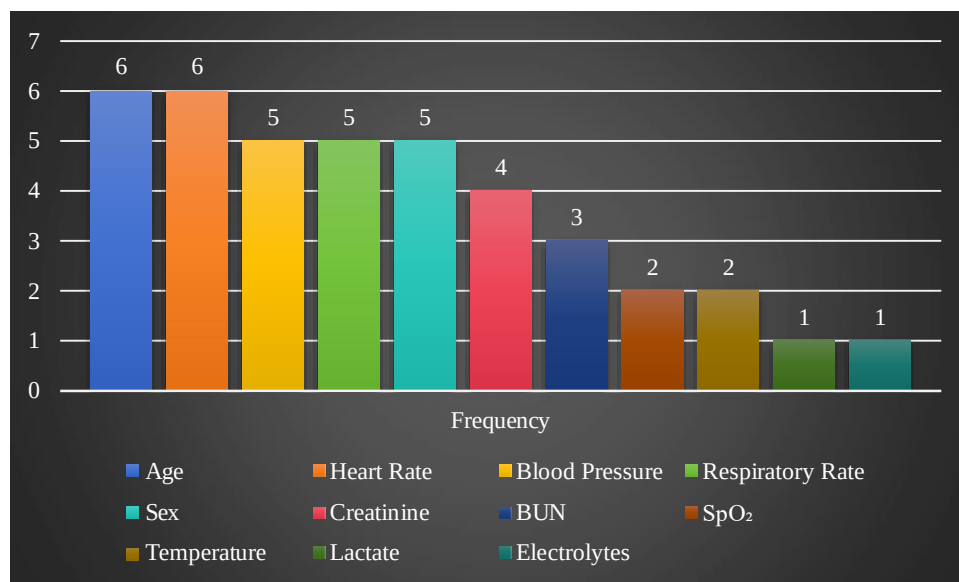
Among all predictors, age, GCS score, BUN, creatinine, heart rate, respiratory rate, and oxygen saturation were most consistently identified as high-importance variables. Markers of renal dysfunction, neurological impairment, and physiological instability contributed substantially to mortality risk estimation across multiple studies.

The consistent use of these variables across studies highlights the feasibility of developing clinically applicable prediction models using routinely available ICU data without requiring specialized biomarkers or advanced diagnostic testing.

Table 4. Routine Clinical Parameters Used Across Included Studies

Study	Demographics	Vital Signs	Laboratory Variables
Baker et al.	Age, Sex	HR, BP, RR, SpO ₂ , Temperature	—
Zhang et al.	Age, Sex	HR, BP, RR	Lactate, Creatinine
Alghatani et al.	Age, Sex	HR, BP, RR, Temperature	—
Pang et al.	Age, Sex	HR, BP, RR, SpO ₂	BUN, Creatinine
Safaei et al.	Age	HR, RR	BUN, Creatinine
Li et al. (2022)	Age, Sex	HR, BP	BUN, Creatinine, Electrolytes

HR = Heart Rate; BP = Blood Pressure; RR = Respiratory Rate; SpO₂ = Peripheral Oxygen Saturation; BUN = Blood Urea Nitrogen.

**Figure 3. Frequency of Routine Clinical Predictors Used in AI-Based ICU Mortality Prediction Models**

3.4 Predictive Performance of AI Models

The predictive performance of the included models is presented in Table 5, whereas a comparative overview of AUROC values is provided in Figure 4.

Overall, AI-assisted mortality prediction models demonstrated good to excellent discriminatory performance across diverse ICU populations. Reported AUROC values generally ranged from approximately 0.79 to 0.92. Ensemble learning algorithms consistently achieved high predictive accuracy. XGBoost-based models frequently reported AUROC values above 0.85, while CatBoost-based models demonstrated similarly strong performance across multiple ICU datasets.

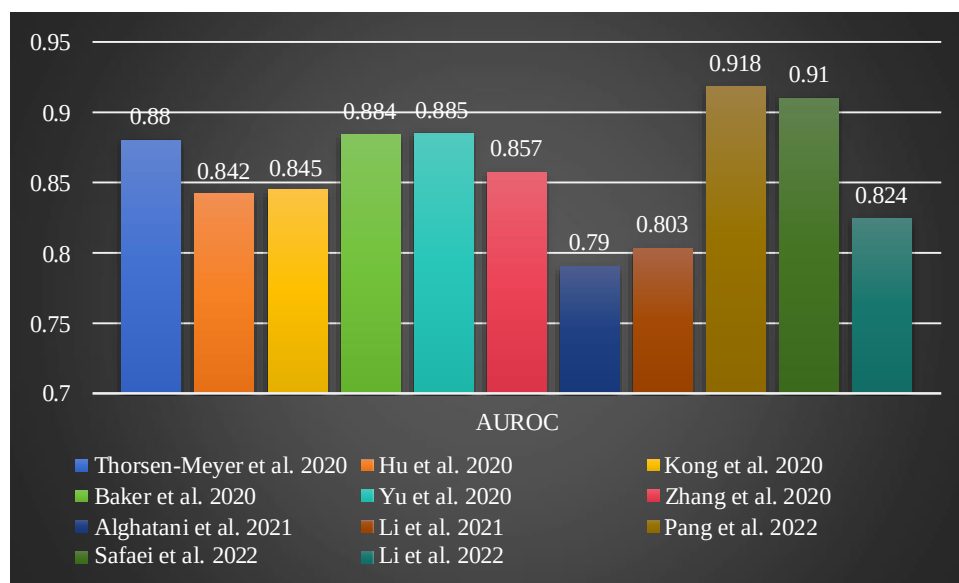
Several studies reported superior performance compared with conventional statistical approaches. Deep learning architectures, particularly LSTM and hybrid neural network models, showed advantages in capturing temporal clinical patterns and dynamic patient deterioration. Disease-specific models developed for sepsis and heart failure populations also demonstrated strong predictive performance, suggesting that AI-based approaches can be effectively adapted to specialized ICU populations.

Sensitivity and specificity values varied according to patient population, outcome definition, and model complexity; however, most high-performing models achieved balanced discrimination with clinically acceptable predictive accuracy.

Table 5. Predictive Performance Metrics of Included AI Models

Study	Best AI Model	AUROC	Sensitivity	Specificity
Thorsen-Meyer et al.	Dynamic Deep Learning Model	0.88	—	—
Hu et al.	XGBoost	0.842	—	—
Kong et al.	Gradient Boosting Machine	0.845	0.771	0.755
Baker et al.	CNN-BiLSTM	0.884	0.792	0.802
Yu et al.	Bidirectional LSTM	0.885	—	—
Zhang et al.	XGBoost	0.857	—	—
Alghatani et al.	XGBoost	0.790	0.164	0.982
Li et al. (2021)	XGBoost-Based Model	0.803	0.530	0.950
Pang et al.	XGBoost	0.918	0.822	0.846
Safaei et al.	CatBoost	0.910	—	—
Li et al. (2022)	XGBoost	0.824	0.595	—

AUROC = Area Under the Receiver Operating Characteristic Curve.

**Figure 4. Comparative Predictive Performance (AUROC) of AI Models Across Included Studies**

3.5 Comparison with Conventional Prognostic Scoring Systems

Comparisons between AI-based models and traditional prognostic scoring systems are summarized in Table 6. Across the included studies, machine learning models generally outperformed conventional severity scores, including the Acute Physiology and Chronic Health Evaluation (APACHE), Sequential Organ Failure Assessment (SOFA), and Simplified Acute Physiology Score (SAPS). AI models demonstrated superior discrimination, particularly when dynamic physiological and laboratory variables were incorporated into model development.

Several studies reported statistically significant improvements in AUROC values compared with conventional scoring systems. Ensemble learning approaches, including XGBoost and CatBoost, consistently achieved higher predictive accuracy than traditional risk assessment tools. Furthermore, AI models demonstrated enhanced capability for identifying complex nonlinear relationships among clinical variables, enabling more individualized mortality risk prediction.

The superiority of AI-based approaches was observed in both general ICU cohorts and disease-specific populations, supporting the growing role of machine learning in critical care prognostication.

Table 6. Comparison of AI Models with Traditional Prognostic Scoring Systems

Study	AI Model	AI AUROC	Comparator Score	Comparator AUROC
Hu et al.	XGBoost	0.842	APACHE II	0.720
Kong et al.	Gradient Boosting Machine	0.845	SAPS-II	0.770
Yu et al.	Bidirectional LSTM	0.885	SAPS-II	0.775
Zhang et al.	XGBoost	0.857	SAPS-II	0.797
Li et al. (2021)	XGBoost-Based Model	0.803	GWTG-HF Score	0.751
Pang et al.	XGBoost	0.918	Logistic Regression	0.872
Safaei et al.	CatBoost	0.910	Conventional ICU Severity Models	Lower Performance

APACHE II = Acute Physiology and Chronic Health Evaluation II; SAPS-II = Simplified Acute Physiology Score II; GWTG-HF = Get with The Guidelines–Heart Failure.

3.6 Summary of Findings

The findings of this systematic review demonstrate that AI-assisted mortality prediction models using routinely collected clinical parameters can achieve high predictive accuracy across diverse ICU populations. Ensemble learning algorithms, particularly XGBoost and CatBoost, were the most consistently high-performing approaches. Routine clinical variables obtained during standard ICU care were sufficient to support accurate mortality prediction, emphasizing the practicality and scalability of these models. Furthermore, increasing attention toward model interpretability and transparency may facilitate future integration of AI-assisted decision-support systems into routine critical care practice.

4. Discussion

4.1 Principal Findings

This systematic review evaluated the current evidence regarding AI-assisted prediction of ICU mortality using routinely collected clinical parameters. Across the 11 included studies published between Feb 2020 - Nov 2022, machine learning models consistently demonstrated strong predictive performance, with most studies reporting good to excellent discrimination for mortality outcomes. A wide range of machine learning approaches was investigated; however, ensemble learning methods, particularly Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine, Random Forest, and CatBoost, emerged as the most effective algorithms for mortality prediction across diverse ICU populations. The consistent superiority of these approaches suggests that machine learning algorithms are capable of identifying complex nonlinear relationships among clinical variables that may not be adequately captured by conventional statistical techniques (9, 10).

A notable finding of this review was the consistent predictive value of routinely available clinical parameters. Variables such as age, heart rate, respiratory rate, blood pressure, Glasgow Coma Scale score, blood urea nitrogen, serum creatinine, and oxygen saturation were repeatedly identified as important predictors of mortality. These variables are routinely collected during standard ICU care and therefore do not require additional investigations or specialized testing. The ability of AI models to achieve high predictive accuracy using existing clinical information highlights their potential applicability within routine critical care practice and supports their feasibility for integration into electronic health record–based decision-support systems (10,11).

4.2 Why Routine Clinical Parameters Are Effective Predictors

The predictive utility of routine clinical parameters likely reflects their ability to capture multiple dimensions of physiological instability and organ dysfunction. Demographic characteristics provide information regarding baseline patient vulnerability, whereas vital signs offer real-time indicators of hemodynamic and respiratory status. Laboratory measurements such as blood urea nitrogen, serum creatinine, lactate concentration, and electrolyte levels provide objective markers of renal dysfunction, metabolic disturbance, and disease severity.

Critically ill patients frequently experience rapid physiological deterioration involving multiple organ systems. By simultaneously analyzing large numbers of routinely collected variables, machine learning algorithms can identify complex interactions and temporal patterns that may not be apparent using traditional analytical approaches. This capability enables the generation of individualized mortality risk estimates that more accurately reflect the multifactorial nature of critical illness.

An additional advantage of routine-parameter models is their practical applicability. Because these variables are already available in most ICU information systems, implementation of AI-assisted prediction tools would require minimal disruption to existing clinical workflows. This enhances the likelihood of successful adoption across a wide range of healthcare settings, including resource-constrained environments (10, 11).

4.3 Comparison with Conventional Prognostic Scoring Systems

The findings of this review indicate that AI-based models generally outperform traditional prognostic scoring systems, including APACHE II, SAPS-II, and other established risk assessment tools. Several studies reported higher AUROC values for machine learning models compared with conventional severity scores, suggesting improved discriminatory performance for mortality prediction.

Traditional scoring systems are based on predefined statistical relationships and fixed predictor structures. Although these tools have demonstrated substantial clinical utility and remain widely used, they may be limited in their ability to model complex nonlinear interactions among variables. In contrast, machine learning algorithms can process large volumes of multidimensional data and automatically identify patterns associated with patient outcomes.

The superiority of AI-based approaches was observed across both general ICU cohorts and disease-specific populations, including patients with sepsis, influenza, and heart failure. Nevertheless, conventional prognostic scores continue to provide important clinical value because of their simplicity, transparency, and widespread acceptance among healthcare professionals. Future strategies may therefore benefit from combining established clinical scoring systems with advanced machine learning approaches to optimize risk stratification (6,8,9).

4.4 Explainability and Clinical Applicability of AI Models

Although machine learning models demonstrated strong predictive performance, successful implementation in clinical practice depends not only on accuracy but also on transparency and interpretability. Historically, many AI models have been criticized for functioning as “black boxes,” limiting clinician confidence and reducing willingness to incorporate model predictions into clinical decision-making.

Several studies included in this review incorporated interpretability techniques such as SHapley Additive exPlanations (SHAP), feature-importance analyses, and model visualization approaches. These methods provide insights into the relative contribution of individual predictors and facilitate a better understanding of how mortality predictions are generated. Improved interpretability may increase clinician trust, support regulatory acceptance, and enhance the safe implementation of AI-assisted decision-support systems.

Furthermore, explainable models may help clinicians identify modifiable risk factors and prioritize interventions for high-risk patients. As artificial intelligence becomes increasingly integrated into healthcare systems, transparency and accountability are likely to become essential components of future mortality prediction models (1, 10, 11).

4.5 Challenges and Future Research Directions

Despite encouraging findings, several challenges continue to limit the widespread implementation of AI-assisted mortality prediction systems. First, most studies relied on retrospective datasets and internal validation procedures, raising concerns regarding model generalizability across different institutions, healthcare systems, and patient populations. Second, substantial heterogeneity was observed in predictor selection, outcome definitions, feature engineering strategies, and model development methodologies, making direct comparison between studies difficult.

Another important limitation is the relative scarcity of prospective clinical evaluations. While many models demonstrated strong predictive performance in retrospective analyses, few studies assessed their impact on real-world clinical decision-making or patient outcomes. Consequently, the practical benefits of AI-assisted mortality prediction remain incompletely understood.

Future research should prioritize multicenter external validation, prospective implementation studies, and standardized reporting frameworks. Additional investigation is needed to determine whether AI-assisted risk prediction improves patient outcomes, optimizes resource allocation, and enhances healthcare efficiency. Emerging developments such as federated learning, real-time clinical decision support systems, and continuously updated predictive models may further improve the safety, reliability, and scalability of AI applications in critical care (1, 12, 17).

4.6 Strengths of the Evidence Base

The studies included in this review collectively demonstrate the growing maturity of artificial intelligence applications in critical care. The availability of large critical care databases, advances in computational methodologies, and increasing emphasis on model interpretability have contributed to substantial progress in mortality prediction research. Importantly, the consistent identification of routine clinical parameters as key predictors across multiple patient populations strengthens the robustness of current evidence and supports the feasibility of translating these approaches into routine clinical practice (9, 10).

Overall, the available evidence indicates that AI-assisted mortality prediction using routinely collected clinical parameters represents a promising strategy for enhancing risk stratification, supporting clinical decision-making, and improving the delivery of critical care services. Nevertheless, further external validation and prospective clinical evaluation remain necessary before widespread implementation can be recommended (1, 9).

4. Conclusion This systematic review demonstrates the considerable potential of artificial intelligence (AI) for improving mortality prediction in intensive care unit (ICU) patients using routinely collected clinical parameters. The evidence synthesized from studies published between Feb 2020 - Nov 2022 indicates that machine learning models can achieve high predictive accuracy across diverse ICU populations and frequently outperform conventional prognostic scoring systems. Among the evaluated approaches, ensemble learning algorithms, particularly XGBoost, Gradient Boosting Machine, Random Forest, and CatBoost, consistently exhibited superior predictive performance, while deep learning architectures showed promise for modeling complex temporal patterns in critically ill patients.

A key finding of this review is the consistent importance of routinely available clinical variables, including demographic characteristics, vital signs, neurological assessments, and laboratory measurements. These parameters were repeatedly identified as major contributors to mortality prediction and were sufficient to support accurate risk stratification without the need for specialized biomarkers or advanced diagnostic investigations. Consequently, AI-assisted prediction models based on routine clinical data offer a practical and scalable solution for integration into existing ICU information systems and clinical workflows.

The findings also highlight the growing emphasis on model interpretability and transparency, which are essential for promoting clinician confidence and facilitating the safe implementation of AI-based decision-support tools. Nevertheless, important challenges remain, including limited external validation, variability in model development methodologies, and the lack of prospective clinical implementation studies. Addressing these limitations will be critical to ensuring the reliability, generalizability, and clinical utility of future AI applications in critical care.

In conclusion, AI-assisted mortality prediction using routine clinical parameters represents a promising strategy for enhancing risk stratification and supporting clinical decision-making in the ICU. Future research should focus on multicenter external validation, real-time clinical integration, and prospective evaluation of patient-centered outcomes to facilitate the translation of these technologies from research settings into routine clinical practice.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The study was conducted using the institutional resources available at Rajendra Institute of Medical Sciences (RIMS), Ranchi, Jharkhand, India. There is no role of any external funding body in study design, data collection or analysis, interpretation of results, or the decision to submit for publication.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest with respect to the research, authorship, and publication of this article.

References

1. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit. *Lancet Digit Health*. 2020;2(4):e179-e191.
2. Hu CA, Chen CM, Fang YC, et al. Using a machine learning approach to predict mortality in critically ill influenza patients. *BMJ Open*. 2020;10:e033898.
3. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med Inform Decis Mak*. 2020;20:251.
4. Baker S, Xiang W, Atkinson I. Continuous and automatic mortality risk prediction using vital signs in the ICU: A hybrid neural network approach. *Sci Rep*. 2020;10:21282.
5. Yu K, Zhang M, Cui T, Hauskrecht M. Monitoring ICU mortality risk with a long short-term memory recurrent neural network. *Pac Symp Biocomput*. 2020;25:103-114.
6. Zhang Z, Ho KM, Hong Y. Predicting 30-day mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGBoost. *J Transl Med*. 2020;18:462.
7. Alghatani K, Ammar N, Rezgui A, Shaban-Nejad A. Predicting ICU length of stay and mortality using patient vital signs. *JMIR Med Inform*. 2021;9:e21347.
8. Li F, Xin H, Zhang J, et al. Prediction model of in-hospital mortality in ICU patients with heart failure. *BMJ Open*. 2021;11:e044779.
9. Pang K, Li L, Ouyang W, et al. Establishment of ICU mortality risk prediction models with machine learning algorithm using MIMIC-IV database. *Diagnostics (Basel)*. 2022;12(5):1068.
10. Safaei N, Safaei B, Seyedekrami S, et al. E-CatBoost: An efficient machine learning framework for predicting ICU mortality. *PLoS One*. 2022;17:e0262895.

11. Li J, Liu S, Hu Y, et al. Predicting mortality in ICU patients with heart failure using an interpretable machine learning model. *J Med Internet Res*. 2022;24:e38082.
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
13. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICOS framework to improve searching in systematic reviews. *BMC Med Inform Decis Mak*. 2007;7:16.
14. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: An efficient and complete method to develop literature searches. *J Med Libr Assoc*. 2018;106(4):531-541.
15. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
16. Higgins JPT, Thomas J, Chandler J, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Chichester: Wiley; 2019.
17. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58.