Journal of Multidisciplinary Healthcare

Dovepress
Taylor & Francis Group

Open Access Full Text Article

REVIEW

# Translational Potential and Explainability of Artificial Intelligence-Based Clinical Decision Support for Adults in Intensive Care: A Scoping Review

Yodha Pranata[1,2], Ristina Mirwanti[3], Yusuf Achmad Bahtiar[4], Hesti Purwanti[5]

[1]Master of Critical Care Program, Faculty of Nursing, Universitas Padjadjaran, Sumedang, West Java, Indonesia; [2]Data Science Program, Faculty of Science and Technology, Universitas Terbuka, Tangerang Selatan, Banten, Indonesia; [3]Department of Critical Care Nursing, Faculty of Nursing, Universitas Padjadjaran, Sumedang, West Java, Indonesia; [4]Department of Anesthesiology and Intensive Care, Ngudi Waluyo Regional General Hospital, Blitar, East Java, Indonesia; [5]Department of Internal Medicine, Ngudi Waluyo Regional General Hospital, Blitar, East Java, Indonesia

Correspondence: Ristina Mirwanti, Department of Critical Care Nursing, Faculty of Nursing, Universitas Padjadjaran, Sumedang, West Java, Indonesia, Email ristina.mirwanti@unpad.ac.id

Journal of Multidisciplinary Healthcare downloaded from https://www.dovepress.com/
For personal use only.

**Introduction:** Intensive care units require rapid, high-stakes decision-making. Although artificial intelligence (AI) offers superior predictive accuracy compared with traditional scoring methods, its "black-box" nature remains a barrier to clinical adoption.

**Objective:** This scoping review systematically mapped the translational potential and characteristics of explainable artificial intelligence (XAI) strategies in AI/ML-based clinical decision support tools for adult intensive care unit (ICU) settings.

**Methods:** Following PRISMA-ScR guidelines, we searched Web of Science, PubMed, Scopus, and EBSCOhost up to January 2026. Translational potential was staged using an ICU-adapted, nine-level Technology Readiness Level (TRL) framework, and explainability strategies were classified as post-hoc or inherently interpretable (glass-box) to assess methodological transparency and clinical readiness.

**Results:** A total of 808 records were identified, of which 29 studies met the inclusion criteria. The findings revealed a marked retrospective predominance (86.2%) and reliance on North American data, predominantly MIMIC (Medical Information Mart for Intensive Care). Tree-based ensembles (82.8%) and post-hoc SHAP explanations (86.2%) were dominant, with proposed clinical utility spanning three domains: therapeutic guidance, resource-allocation optimisation, and user-centric design. Most innovations were standalone, web-based prototypes requiring manual data entry (69.0%, TRL 4–5); a further 10.3% were shared only as open-source code, and only 17.2% reported integration with hospital systems. Only one study claimed clinical maturity (TRL 9), although its validation remained retrospective.

**Conclusion:** Accuracy is no longer the primary bottleneck; the constraint has shifted to "last-mile" integration and external validity. Current XAI relies almost entirely on post-hoc methods that risk an "illusion of clarity", while inherently interpretable, glass-box models remain a rare but promising alternative. Future research should prioritise external validation in independent settings, prospective evaluation of clinical impact, and explicit comparison between post-hoc and interpretable approaches.

**Keywords:** artificial intelligence, clinical decision support, explainable AI, intensive care unit, technology readiness level, translational research

## Introduction

The Intensive Care Unit (ICU) is a high-risk medical environment characterized by time-critical decision-making under considerable uncertainty, often compounded by significant operational pressure on the interdisciplinary healthcare team,[1] a structure built on close cooperation among physicians, nurses, and allied health professionals.[2] The abundance of patient data from various monitoring devices often leads to cognitive fatigue and a significant administrative burden for healthcare professionals.[3] Data-driven clinical decision support systems (CDSS) are essential for the effective processing

https://doi.org/10.2147/JMDH.S621469
Received: 1 May 2026
Accepted: 8 July 2026
Published: 17 July 2026



© 2026 Pranata et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v4.0) License (http://creativecommons.org/licenses/by-nc/4.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

and integration of clinical information.[4] Digital technologies are expected to reduce medical errors and optimise the use of limited ICU resources through real-time predictive analytics that complement clinical judgement.[3,5]

Prompt medical responses are required to manage critical conditions such as sepsis, acute kidney injury (AKI), acute respiratory distress syndrome (ARDS), and ischaemic stroke. Each hour of diagnostic delay can significantly worsen the clinical outcomes.[6,7] Artificial Intelligence (AI) models have demonstrated excellent performance, with AUROC values reaching up to 0.98, often surpassing conventional scoring systems such as SOFA and APACHE.[3,8] A clear gap remains between the model accuracy on paper and the tangible benefits in clinical settings. The "black-box" phenomenon continues to pose a major barrier for clinicians who require transparency before adopting algorithmic recommendations.[4,9]

Tree-based algorithms and deep learning models can now analyse both tabular data and medical images with expert-level accuracy.[8,10] High technical performance does not automatically translate to clinical readiness, as most innovations remain in the early development stage.[5,8] The translational potential of medical AI is defined as the ability of a model to move from a research prototype to an operational tool embedded within clinical workflows to improve patient outcomes.[5,7] To systematically operationalise this construct, the present review adapts the nine-level Technology Readiness Level (TRL) framework recently proposed for AI applications in the ICU setting, which stages AI maturity from initial problem identification (TRL 1) to full integration into routine clinical practice (TRL 9).[1] Crossing this research-to-practice boundary requires explainable AI (XAI) strategies, such as Shapley Additive exPlanations (SHAP) and LIME, which are crucial for providing visual transparency to clinicians and facilitating more informed decision-making.[8,11] Explainable AI (XAI) strategies for clinical decision support fall into two broad paradigms: models designed from the outset to be interpretable and post-hoc explanation methods applied to otherwise opaque "black-box" models.[12,13] Visual transparency for interdisciplinary ICU teams is currently delivered almost exclusively through the latter, with SHAP and LIME being the most widely adopted approaches; both extract explanations only after the underlying model has been trained and have been applied extensively to ICU prediction tasks,[14] including mortality prediction in ischaemic stroke patients.[8] Distinguishing between these two paradigms is important because post-hoc methods typically offer only an approximation of a model's reasoning, risking an illusion of explanatory depth or clarity rather than genuine insight into its decision mechanism.[13,15,16] This distinction is especially consequential in the ICU, where the accountability demanded of clinical decisions calls for a deeper understanding than a mere summary of model patterns or trends.[14]

XAI techniques are increasingly being applied in critical care; however, the literature remains fragmented. Żerdziński et al (2026)[5] mapped AI clinical maturity across five ICU domains but did not specifically examine the explainability methods. Yang et al (2026)[17] mapped 86 multimodal ICU AI studies but treated XAI only as a secondary theme, while Athukorala et al (2026)[14] focused specifically on XAI but restricted their scope to three clinical outcomes (sepsis onset, ICU readmission, and mortality). None of these reviews systematically evaluated whether XAI explanations genuinely improve clinician trust or introduce a new form of cognitive burden, namely, information overload. This review addresses this gap by evaluating the translational potential of XAI across clinical tasks at the primary study level. The principal operational barrier is "last-mile" integration, where accurate models remain difficult to convert into applications directly connected to electronic health records (EHR) because of interoperability challenges. This gap is further compounded by dataset bias and limited external validation across diverse populations.[5,9]

The Current literature is highly heterogeneous in terms of model types, methodologies, and input data. Therefore, a scoping review was selected as the most appropriate method for systematically mapping this rapidly evolving evidence landscape. This scoping review aimed to systematically map the translational potential and characteristics of explainable artificial intelligence (XAI) strategies in AI/ML-based clinical decision support tools for adult ICU settings. This review is essential for identifying real-world implementation barriers to ensure that AI innovations do not remain mere statistical achievements but become safe, transparent, and reliable clinical solutions for frontline healthcare professionals.

# Method

## Study Design

A scoping review design was employed to map the literature on translational potential and explainability strategies in AI/ML-based decision-support tools for adult ICU settings. The review was conducted in a systematic, transparent, and

reproducible manner and reported in accordance with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) guidelines.[18] We documented all stages of the scoping review process, including identification and deduplication of records, title and abstract screening, full-text eligibility assessment, data charting, and narrative synthesis of the findings.

The main methodology was established before screening and analysis. Inclusion and exclusion criteria based on the PCC framework, definitions of translational potential indicators, data extraction plans, and synthesis approaches were all pre-specified. Two independent reviewers (YP and YAB) screened and extracted data. Any disagreements were resolved through discussion until a consensus was reached, thereby minimising selection and extraction bias. The research question underpinning this review is: "What is the extent of translational potential and the nature of explainability strategies reported for AI/ML-based clinical decision support tools in adult ICU settings?"

## Eligibility Criteria

Eligibility criteria were established using the population, concept, and context (PCC) to align with the study objective of mapping translational potential and explainability strategies in AI/ML-based decision-support tools for adult ICU settings.

**Population**: Adult patients (aged ≥18 years) admitted to the ICU, including medical, surgical, or mixed ICUs. Studies were considered relevant if the population was within the context of adult critical care.

**Concept**: AI/ML-based clinical decision support tools incorporating components of explainable AI (XAI) or interpretable machine learning for clinical decision support. This includes risk prediction, early warnings, alerts, triage, and clinical recommendations. The concept encompasses explainability strategies identified through the search, such as SHAP, LIME, counterfactual explanations, feature attribution/importance, saliency/Grad-CAM, and attention-based explanations. These are commonly communicated through visualisation methods that are easily understood by clinicians, such as force plots, beeswarm plots or summaries of key features. Studies employing inherently interpretable (glass-box) modelling approaches (eg, Explainable Boosting Machines), incidentally captured within the broader AI/ML search terms, were also eligible and characterised during data extraction according to whether they represented post-hoc explainability methods or inherently interpretable models. Evidence of translational potential, such as clinical actionability, availability of prototypes or web calculators, and implementation artefacts, was also included in this concept. This evidence was subsequently mapped to the Technology Readiness Level (TRL) framework (Table 1) to provide an objective assessment of the maturity of the included studies.

**Context**: Clinical practice and workflow integration in adult ICU settings. Context covers actual implementation, such as bedside use, EHR/EMR integration, dashboards, or automated notifications, as well as proposed or simulated implementations, including web-based prototypes and feasibility evaluations within clinical workflows.

## Inclusion and Exclusion Criteria

This review included primary quantitative research, such as retrospective or prospective cohort studies, cross-sectional studies, model development and validation studies, and other observational studies that applied XAI/IML methods in the context of ICU decision support. Studies were required to demonstrate at least one indicator of translational potential, defined as the provision of source code, web-based tools, dashboards, calculators, or the implementation of workflow/EHR integration. This eligibility criterion reflected the review's specific focus on studies demonstrating explicit translational orientation, rather than model development and validation alone; this scope was necessary to characterize the spectrum of translational readiness—from early-stage outputs (eg, open-source code) to clinically mature applications—using the Technology Readiness Level (TRL) framework[1] (Table 1). Studies reporting only model development and internal or external validation without any translational output, while technically classifiable at lower TRL stages (eg, TRL 3–5), were excluded because they fell outside this review's translational focus; the implications of this scoping decision are discussed in the Limitations.

Exclusion criteria comprised:

- Secondary research, such as systematic reviews, meta-analyses, and scoping reviews.
- Non-empirical publications, including editorials, commentaries, letters to the editor, or opinion pieces.
- AI models that are entirely black-box without clear explainability or interpretability components.
- Studies that do not focus on medical decision support or lack clinical actionability relevance in the ICU context.
- No restrictions were applied to the year of publication to provide a comprehensive overview of the development of evidence.

**Table 1** Definition of Technology Readiness Level (TRL) Framework for AI Applications in Intensive Care[1]

| TRL | Clinical Definition | Explanation |
|---|---|---|
| 1 | Problem identification | Studies that identify or describe a clinical problem but do not propose an AI-based solution to address it. |
| 2 | Proposal of solution | Studies that propose an AI-based approach for a specific clinical problem without proceeding to model development. |
| 3 and 4 | Model prototyping and development | Studies in which an AI model was developed using a dataset, with its performance assessed solely through internal validation. |
| 5 | Model validation | Studies in which model performance is assessed on an independent dataset that differs temporally, geographically, or in the clinical domain from the development data, with evaluation conducted offline and without real-time deployment in a clinical environment. |
| 6 | Real-time testing | Studies in which the model processes real-time clinical data and is embedded within the existing workflow, although its output remains concealed from the clinical staff ("silent" deployment). |
| 7 | Workflow integration | Studies assessing the practical feasibility of deploying the model within routine clinical care, such as pilot implementations, user experience evaluations, or technical feasibility testing, in which the model output becomes visible to the staff within the workflow. |
| 8 | Clinical testing | Studies evaluating the clinical effectiveness of the integrated model typically compared it with a control group, standard practice, benchmark performance, or pre- versus post-implementation outcomes, with model output visible to staff. |
| 9 | Integration into clinical practice | Studies reporting models that have been fully adopted and used routinely in clinical practice over a sustained period. |

## Search Strategy

This scoping review adhered to the PRISMA-ScR. A systematic literature search was conducted in Web of Science, PubMed/MEDLINE, Scopus, and EBSCOhost, covering all publications from the inception of each database until January 2026. The search strategy was developed based on the PCC framework, utilising free-text terms and Boolean operators (AND/OR). The population (P) targeted adult patients in the ICU: ("intensive care unit" OR ICU OR "critical care" OR "critically ill" OR "critical illness") AND (adult OR adults OR "adult patient*" OR "18 years" OR "aged 18 years"). The concept (C) included AI/ML for clinical decision support with explainability or interpretability components: ("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR algorithm* OR "predictive model*") AND ("clinical decision support" OR CDSS OR "risk stratification" OR alert* OR "early warning") AND ("explainable AI" OR XAI OR SHAP OR LIME OR counterfactual* OR "feature attribution" OR saliency OR "Grad-CAM" OR PDP OR ICE). The context (C) focused on integration into clinical practice and workflow in the ICU: ("clinical workflow" OR implementation OR deployment OR bedside OR EHR OR EMR OR usability) AND ("intensive care" OR ICU OR "critical care"). The complete search string, including all synonyms from the PCC table and syntax adjustments for each database, is presented in Supplementary Table 1. Additional searches were performed using backward and forward citation-tracking. All search results were exported to Mendeley (Elsevier, Netherlands) for deduplication and screening, which was conducted by YP and YAB.

## Data Extraction

Data were extracted using a pre-designed worksheet in Microsoft Excel, comprising three linked tables (Table 2, Supplementary Tables 2A and B). Table 2 captures the Country, study design, population, dataset(s) used, sample size, and validation strategy, graded as low/moderate/high using pre-specified criteria (Table 2 footnote). Supplementary Table 2A captures the AI/ML architecture and interpretability paradigm, distinguishing post-hoc explainability methods (for example, SHAP and LIME) from inherently interpretable, "glass-box" designs (for example, Explainable Boosting

**Table 2** Study Characteristics

| No | Study (Author, Year, Country) | Country | Study Design | Target Population | Dataset(s) Used | Sample Size (N) | Validation Strategy |
|---|---|---|---|---|---|---|---|
| 1 | Zhu et al (2025)[19] | China | Multicenter retrospective and prospective cohort study | Adult ICU patients with sepsis | MIMIC-III, MIMIC-IV, eICU-CRD, and Local Chinese ICU | 12,408 patients | High: Multicentre, external, and prospective validation. |
| 2 | X. Ge et al (2025)[20] | China | Multicenter retrospective cohort study | Adult ICU patients with sepsis-associated AKI | MIMIC-IV, eICU-CRD, and FAH-HMU ICU (China) | 28,819 patients | High: Multicentre external and international validations. |
| 3 | M. Chen et al (2026)[21] | China | Multicenter retrospective cohort study | Hemodialysis (HD) patients in the ICU | SYSU (China), MIMIC-IV, and eICU-CRD | 2,895 patients | High: Multicentre external and international validations. |
| 4 | Ma et al (2025)[22] | China | Retrospective cohort study | Geriatric patients (≥65) undergoing non-cardiac surgery | MIMIC-IV (Development) and MIMIC-III (Validation) | 2,113 patients | Moderate (Temporal/ External Database Validation) |
| 5 | Y. Ge et al (2025)[23] | China | Retrospective cohort study | Adult ICU patients with a primary diagnosis of asthma | MIMIC-IV v2.2 (Development) and MIMIC-IV v3.1 (Validation) | 4,385 patients | Moderate: Temporal external validation (same hospital system). |
| 6 | R. Yang et al (2025)[24] | China | Multicenter retrospective cohort study | Critically ill patients with acute myocardial infarction (AMI) | MIMIC-III, MIMIC-IV, and eICU database | 12,170 patients | High: Multicentre external and international validations. |
| 7 | Hu et al (2023)[25] | China | Retrospective cohort study | Critically ill patients with ischemic stroke | MIMIC-IV and eICU-CRD | 5,613 patients | High: Independent external database validation. |
| 8 | S. Q. Wang et al (2025)[26] | China | Single-center retrospective cohort study | Adult patients with extensive burns (≥50% TBSA) | Two Burn Critical Care Units in Eastern China | 237 patients | Low: Internal validation only (single-centre setting). |
| 9 | Gu & Lu (2025)[27] | China | Retrospective cohort study | Critically ill cancer patients in the ICU | eICU (Development) and MIMIC-IV (Validation) | 10,828 patients | High: Independent external database validation. |
| 10 | Z. J. Wang et al (2025)[28] | China | Multicenter retrospective cohort study | Critically ill patients with liver cirrhosis | MIMIC-IV, eICU-CRD, and QUAH (China) | 3,303 patients | High: Multicentre external and international validations. |
| 11 | Liu et al (2025)[29] | China | Multicenter prospective observational study | Adult critically ill patients (aged ≥18) | Sichuan Provincial People's Hospital registries | 1,306 patients | Moderate: Multicentre internal validation within the same hospital. |
| 12 | Cheng et al (2026)[30] | China | Multicenter retrospective cohort study | Ischemic stroke patients with impaired consciousness (GCS≤8) | MIMIC-IV and three top-tier hospitals in China | 707 patients | High: Multicentre external and regional validations. |
| 13 | Y. Wang et al (2026)[31] | China | Multicenter retrospective cohort study | Traumatic spinal injury (TSI) post-surgery patients | MIMIC-IV, eICU-CRD, and Xinjiang Medical University | 1,166 patients | High: Multicentre external and international validations. |
| 14 | Bai et al (2026)[32] | China | Multicenter retrospective cohort study | Acute pancreatitis with acute kidney injury (AP-AKI) | MIMIC-III, IV, eICU (Training) and 2 Chinese Hospitals (Validation) | 2,714 patients | High: Multicentre external validation across independent hospitals. |
| 15 | C. Yang et al (2026)[33] | China | Retrospective cohort study | Critically ill cancer patients in the ICU | MIMIC-IV (USA) and Chongqing University Cancer Hospital | 2,092 patients | High: Independent, external, and geographically distinct validation. |
| 16 | Liang et al (2026)[34] | China | Multicenter retrospective cohort study | ICU patients with primary traumatic fractures | MIMIC-IV, Maoming Hospital, and Hainan Medical University | 4,321 patients | High: Multicentre external validation across independent hospitals. |
| 17 | Wei et al (2025)[35] | China | Multicenter retrospective cohort study | Adult ICU patients with acute pancreatitis | MIMIC-III, MIMIC-IV, and Beijing Chaoyang Hospital | 1,802 patients | High: Multicentre external validation across independent hospitals. |

(*Continued*)

**Table 2** (Continued).

| No | Study (Author, Year, Country) | Country | Study Design | Target Population | Dataset(s) Used | Sample Size (N) | Validation Strategy |
|---|---|---|---|---|---|---|---|
| 18 | Ding et al (2024)[36] | China | Multicenter retrospective cohort study | Adult ICU patients with traumatic brain injury (TBI) | eICU database and MIMIC-III | 1,085 patients | High: Multicentre external validation (USA-based). |
| 19 | Kim et al (2023)[37] | Korea | Retrospective cohort study | Heart failure patients at high risk for cardiac arrest | MIMIC-IV and eICU Collaborative Research Database | 12354 patients | High: Multicentre external validation across the USA datasets. |
| 20 | Sjoding et al (2021)[10] | USA | Retrospective cohort study | Patients with acute respiratory failure or sepsis | MIMIC-IV and University of Pennsylvania data | 2,518 images | High: Independent external validation in different hospital systems. |
| 21 | J. Chen et al (2025)[38] | China | Multicenter retrospective and prospective cohort study | Adult pneumonia patients in the ICU | MIMIC-IV, MIMIC-III, eICU, and Fudan University | 25,783 patients | High: Multicentre, external, and prospective validation. |
| 22 | Liu et al, (2023)[39] | China | Multicenter retrospective cohort study | Patients with wild mushroom poisoning | 5 county hospitals in Enshi, Hubei Province | 567 patients | High: Multicentre external validation (independent hospital cohort). |
| 23 | M. Yang et al (2024)[40] | China | Retrospective cohort study | Sepsis patients requiring vasopressors within 24h | MIMIC-IV (Development) and eICU-CRD (Validation) | 12,878 patients | High: Multicentre external validation (international). |
| 24 | S. Wang et al (2025)[41] | China | Multicenter retrospective cohort study | Adult and pediatric ICU patients with septic shock | Three Chinese Hospital Cohorts | 4,872 patients | High: Multicentre and cross-specialty external validation required. |
| 25 | (Rong et al (2025)[42] | USA | Retrospective cohort study | Hospitalized COVID-19, ARDS, and Sepsis patients | Texas Health Resources (THR) and MIMIC-IV | 11,999 patients | High: Independent external validation across distinct disease profiles. |
| 26 | Lin et al (2025)[43] | Taiwan | Single-center retrospective cohort study | Patients with suspected pneumonia-induced sepsis | Tri-Service General Hospital (EMR data) | 1,492 patients | Moderate: Chronological internal validation; lacks an external dataset. |
| 27 | Magunia et al (2021)[44] | Germany | Multicenter retrospective and prospective cohort study | Adult ICU patients with PCR-confirmed COVID-19 | Germany-wide electronic registry (27 hospitals) | 1,186 patients | High: Multicentre prospective validation in 27 hospitals. |
| 28 | Jia et al (2021)[45] | UK | Single-center retrospective cohort study | Adult ICU patients requiring invasive ventilation | MIMIC-III clinical database | 2,299 patients | Low: Internal split-sample validation only. |
| 29 | McWilliams et al (2019)[46] | UK | Multicenter retrospective cohort study | General Intensive Care Unit (GICU) patients | Bristol Royal Infirmary and MIMIC-III | 9,462 patients | High: Multicentre external validation between the UK and USA cohorts. |

**Notes**: The validation Strategy classification criteria are as follows: Low: internal validation only via random split-sample or standard cross-validation within a single-centre dataset, with no temporal or geographic separation between the development and test cohorts. Moderate: either (a) temporal/chronological validation within a single data source or site (test cohort separated by time but drawn from the same underlying population) or (b) validation across different versions/releases of the same source registry (eg MIMIC-III vs MIMIC-IV, which originate from the same institution, Beth Israel Deaconess Medical Center, Boston, with overlapping enrolment periods, and therefore do not constitute a genuinely independent external cohort). High: validation performed on data from a genuinely independent health system, institution, or country (eg a separate hospital cohort or cross-database validation between MIMIC and eICU-CRD, which draws from a distinct multi-hospital network), with or without a prospective component.

**Abbreviations**: AKI, acute kidney injury; AMI, acute myocardial infarction; AP-AKI, acute pancreatitis with acute kidney injury; ARDS, acute respiratory distress syndrome; eICU-CRD, eICU Collaborative Research Database; EMR, electronic medical record; GCS, Glasgow coma scale; GICU, general intensive care unit; HD, hemodialysis; ICU, intensive care unit; MIMIC, Medical Information Mart for Intensive Care; PCR, polymerase chain reaction; TBI, traumatic brain injury; TBSA, total body surface area; THR, Texas Health Resources; TSI, traumatic spinal injury.

Machines); clinical domain, coded using the non-mutually exclusive Prognostic/Early Warning (P), Diagnostic/Detection (D), Monitoring/Dynamic Assessment (M), Treatment/Decision Support (T), and Implementation/Readiness (I) taxonomy of Żerdziński et al (2026),[5] reflecting that individual studies commonly address more than one clinical function; clinical task and predicted outcome; clinical task cluster; and key predictors. Supplementary Table 2B captures model

performance, proposed clinical utility, a single dominant Translational Output category per study (eg interactive web platform/calculator, open-source repository, hospital system integration, or clinically mature application), and Technology Readiness Level (TRL)—a construct not directly reported by the primary studies but derived by the review team from each study's validation strategy and translational output, using the nine-level ICU-adapted framework of Berkhout et al (2025)[1] (definitions in Table 1)—with the classification basis documented per study for transparency.

Two reviewers (YP and YAB) independently extracted data using this worksheet. A fourth reviewer (HP) checked for completeness and quality of the data. Clinical relevance was assessed by an intensive care/anesthesiology physician (YAB). A second reviewer (RM) provided methodological supervision, including inter-variable consistency checks (eg, ensuring that the TRL classification was consistent with the reported validation strategy and translational output for each study) and resolution of complex cases. References were managed using Mendeley. Discrepancies between the two independent extractors were resolved through discussion and consensus with reference to the pre-specified eligibility criteria and TRL classification rules (Table 1). A formal chance-corrected agreement statistic (eg, Cohen's kappa) was not calculated.

## Data Analysis and Synthesis

Data were synthesised narratively across four domains aligned with the review objectives: (1) model architecture; (2) explainability (XAI) strategies and visualisations, including whether each model relied on post-hoc explainability methods applied to black-box architectures (for example, SHAP, LIME) or was inherently interpretable by design, such as Explainable Boosting Machines, which require no separate explainer; (3) translational potential; and (4) proposed clinical utility. For the translational domain, studies were grouped by the type of output reported (web-based prototypes or online calculators, hospital system integration, open-source repository/public source code, or clinically mature applications) and each was mapped to its corresponding Technology Readiness Level using the framework of Berkhout et al (2025)[1] (Table 1), grounding translational maturity in an established scale rather than in narrative judgement alone. Categorisation decisions were made through repeated discussions among the review team to ensure consistency of the decisions. Table 2 summarises the study characteristics, and the full extraction data are reported in Supplementary Tables 2A and B.

# Result

## Study Selection

A total of 808 records were identified from four databases (Web of Science, n = 52; PubMed/MEDLINE, n = 527; Scopus, n = 157; EBSCOhost, n = 72). After removing duplicates (n = 173), 635 records were screened by title and abstract, and 588 records were subsequently excluded. Full texts were retrieved and assessed for eligibility in 47 reports (no reports were unobtainable); 18 reports were subsequently excluded, with reasons provided. A total of 29 studies met the eligibility criteria and were included in this scoping review. The study selection process is illustrated in Figure 1.

## Characteristics of Included Studies

This review included 29 studies published between 2019 and 2026, with a marked rise in recent years; 20 (69.0%) appeared in 2025–2026, reflecting the growing global interest in the use of XAI in critical care. Most studies were conducted in Asia, particularly China (22 studies, 75.9%), followed by the United States and the United Kingdom (two each), and one each from South Korea, Taiwan, and Germany (Table 2).

Public US-based databases dominated the evidence base: 23 studies (79.3%) drew on MIMIC-III/IV, and 13 (44.8%) used eICU-CRD for training or external validation. Encouragingly, 20 studies (69.0%) supplemented these with local clinical data from the researchers' institutions or regional collaborations, strengthening external validity and local relevance. Retrospective cohort designs predominated (25 studies, 86.2%), with only four incorporating prospective or real-world observational components. Seventeen studies were explicitly multicentre in design, and the sample sizes ranged widely from 237 to 28,819 patients.

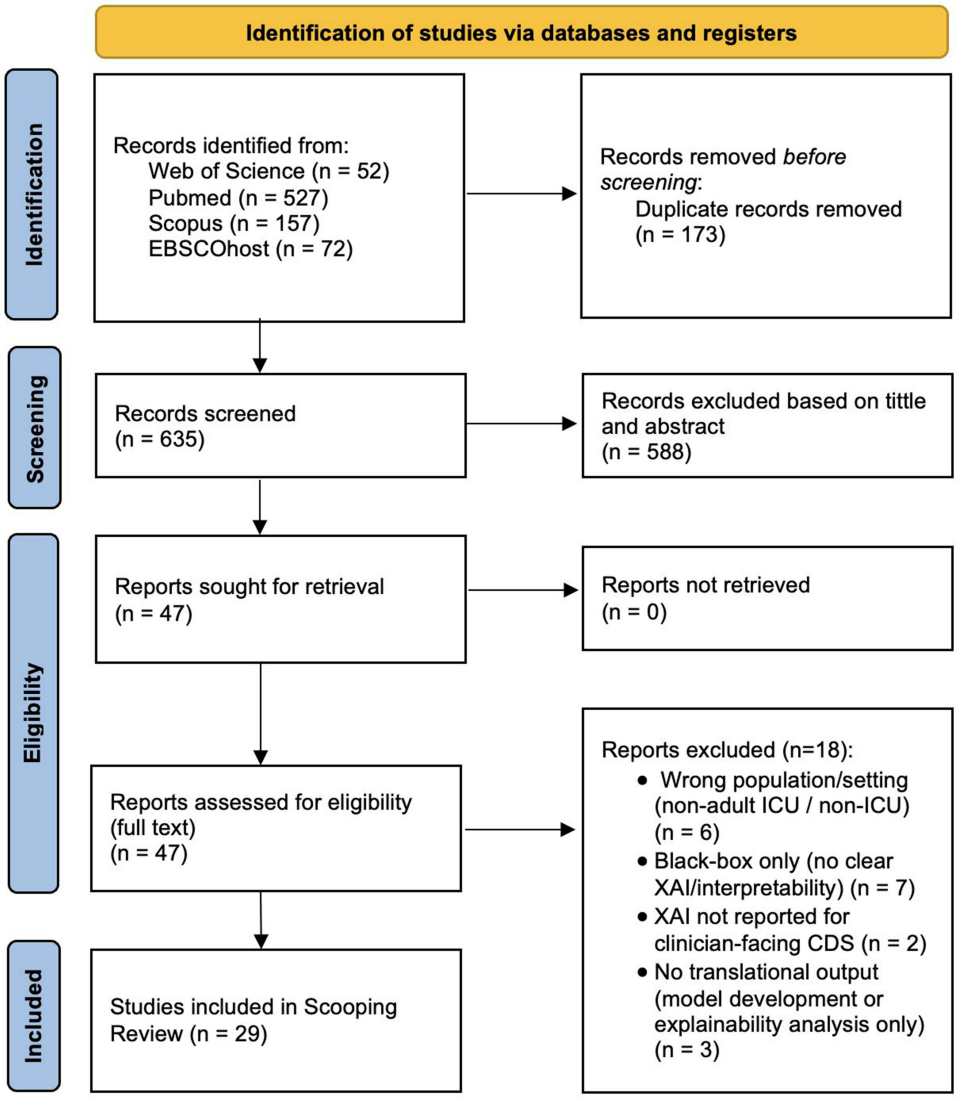


**Figure 1** Prisma Flow Diagram.[47]

## Model Architecture

Across the included studies, model development consistently sought to balance the predictive accuracy and clinical interpretability. Tree-based ensemble methods dominated, accounting for 24 of 29 studies (82.8%), including XGBoost,[23,29,32,34,36,38,39] CatBoost,[22,30,33,35] LightGBM,[20,37,43] Random Forest,[24,46] and several stacking or fusion variants.[25,27,28,31] The TOPSIS-based classification fusion model of S. Wang et al (2025) was similarly classified here, as its core remains that of a gradient boosting decision tree.[41] This preference reflects the suitability of tree-based models for the non-linear, incomplete tabular data that are typical of ICU settings (Table 2 and Supplementary Table 2A).

Deep learning architectures were used in only four studies (13.8% of the total). Two applied transformer-based models to temporal or encounter-level data (SepsisFormer[48] and TECO[42]) while two applied convolutional neural networks (CNNs) to different data types: Sjoding et al (2021)[10] used a DenseNet-121 CNN on chest radiographs for ARDS detection, the corpus's only true imaging application, whereas Jia et al (2021)[45] applied a CNN to ventilator waveform and physiological data to predict extubation readiness. Only one study, the Explainable Boosting Machine of Magunia et al (2021),[44] used a genuinely glass-box architecture, indicating that, among the studies identified through this search strategy, transparency in ICU AI models still overwhelmingly depends on post-hoc explanation methods applied

after model training rather than interpretable-by-design architectures. Overall, these findings suggest that within this corpus, the field continues to prioritise black-box predictive performance, with interpretability added downstream rather than built into the model design.

## Explainability (XAI) Strategies and Visualisation

Transparency in ICU-based AI decision support was overwhelmingly achieved through post-hoc explanations applied to otherwise opaque models, allowing clinicians to audit the reasoning behind a prediction after training the model itself was trained. Shapley Additive exPlanations (SHAP) dominated this landscape, employed in 25 of the 29 included studies (86.2%),[20–22,24,25,35,38,48] owing to its consistent, game-theoretic attribution of each feature's contribution to the model output. SHAP was typically applied at both the population level through summary and beeswarm plots highlighting influential predictors such as lactate,[28] age,[30] and anion gap,[22,23] and at the individual level, through force and waterfall plots that traced how specific predictor values shifted a single patient's risk estimate.[27,32,33]

Beyond SHAP, three studies adopted explanation techniques that were suited to their specific data and objectives. McWilliams et al (2019)[46] used permutation feature importance to rank predictors of ICU discharge readiness based on their impact on model performance, whereas Jia et al (2021)[45] combined DeepLIFT with counterfactual explanations (DiCE) to identify the feature changes needed to achieve extubation readiness. For imaging data, Sjoding et al (2021)[10] applied Grad-CAM to generate saliency heat maps over chest radiographs, highlighting the regions most contributory to automated ARDS detection.

One study departed from this post-hoc paradigm entirely: Magunia et al (2021)[44] employed an Explainable Boosting Machine, a generalised additive model that is inherently interpretable by design and therefore requires no separate explainer to trace each predictor's contribution to the outcome. The near-total reliance on post-hoc explanation observed within the identified corpus indicates that transparency in ICU AI has, to date, been layered onto black-box models after the fact, rather than built into model architecture from the outset, a pattern with direct implications for how much genuine interpretability current tools can offer at the bedside.

## Translational Potential

Beyond reporting model performance, most included studies showed some orientation toward real-world use, ranging from standalone prototypes to systems already embedded in the hospital infrastructure. Translational outputs were grouped into four categories (open-source repository or public source code, web-based prototypes or online calculators, hospital system integration, and clinically mature applications) and mapped to the corresponding Technology Readiness Level (TRL) using the framework of Berkhout et al (2025)[1] (Table 1), providing an objective basis for comparing translational maturity across the corpus. Figure 2 presents this synthesis as a three-column flow diagram (clinical task cluster, dominant XAI method, and translational output) ordered by ascending mean TRL, showing at a glance how the explainability approach and translational maturity shift across ICU applications.

### Open-Source Repository or Public Source Code

Three studies (10.3%) made their models available as open-source code, also corresponding to TRL 4–5. Liu et al (2023)[39] released a ready-to-use application for triaging wild mushroom poisoning, S. Wang et al (2025)[41] provided code for a septic shock mortality model, and Jia et al (2021)[45] published the source code underlying their ventilator-weaning readiness model. Public code of this kind supports independent audit and replication, a prerequisite for building trust ahead of wider clinical adoption.[29,45]

### Web-Based Prototypes or Online Calculators

This was the most common translational pathway, reported in 20 of the 29 studies (69.0%), corresponding to TRL 4–5. These tools typically allow clinicians to enter routine bedside variables and receive an immediate risk estimate through a web interface, extending testing beyond the original development site.[19,28] Their clinical scope was broad, spanning sepsis,[19] sepsis-associated acute kidney injury,[20] postoperative mortality in geriatric patients,[22] asthma,[23] liver cirrhosis,[28] ischaemic stroke,[25,30] pneumonia,[38] and imaging-based ARDS detection.[10] Most, however, remained

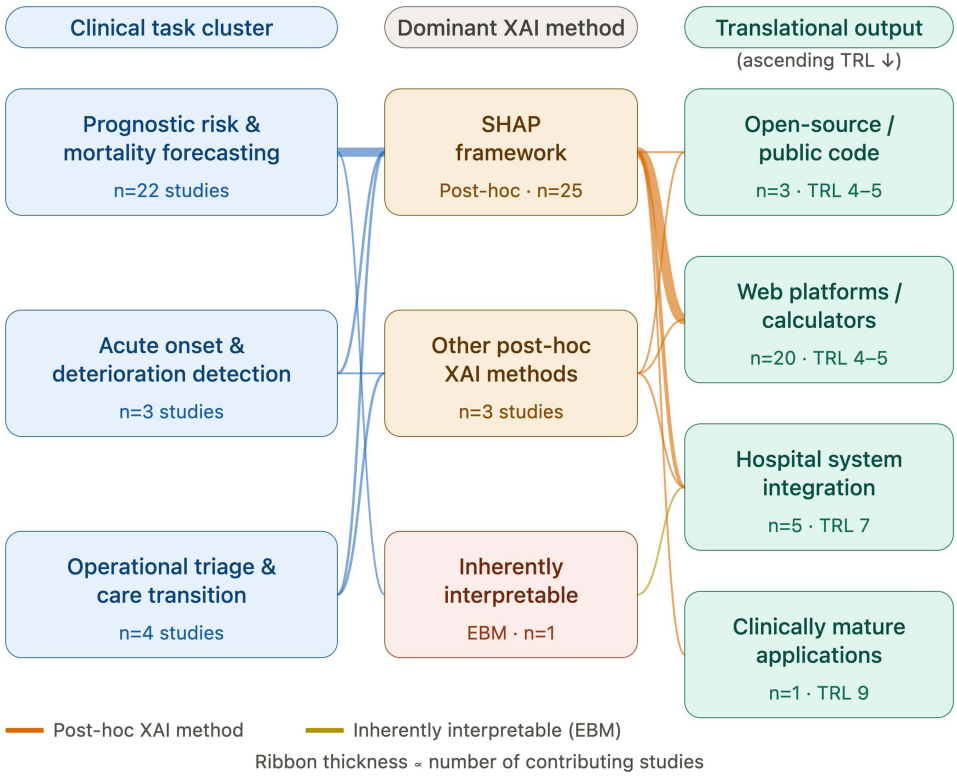


**Figure 2** The Translational Landscape of Explainable AI in Intensive Care.

standalone tools requiring manual data entry, positioning them as early indicators of translational intent rather than as evidence of workflow integration,[22,35] and reducing this dependence on manual input will likely be necessary before such tools can function efficiently in time-pressured ICU settings.[29,36]

## Hospital System Integration

Five studies (17.2%) progressed to direct integration with hospital systems, corresponding to TRL 7. Magunia et al (2021)[44] demonstrated this on the largest scale by embedding their model within a nationwide REDCap-based registry that spanned 27 hospitals in Germany. At the single-institution level, Yang et al (2024)[40] deployed a bedside application connected to the hospital EHR to screen patients for clinical trial enrolment, whereas Lin et al (2025)[43] embedded an AI-CDSS interface directly into the clinical workflow to support rapid sepsis confirmation. Rong et al (2025)[42] benchmarked their continuously updating mortality-risk model against a proprietary EHR deterioration tool to demonstrate real-time monitoring feasibility, and McWilliams et al (2019)[46] proposed a dashboard "nudge" to prompt earlier ICU discharge decisions, although this remained conceptual. Collectively, these examples suggest that direct workflow embedding, although achievable, remains confined to a small number of studies.

## Clinically Mature Application

Only one study (3.4%) reported translational output at the highest maturity level. Kim et al (2023)[37] described their cardiac arrest early warning system as a clinically mature application, stating that it was already "in use and verified for daily clinical practice" and "feasible for routine clinical use". However, the evidence underpinning this claim remained largely retrospective, drawn from internal validation on MIMIC-IV and external validation on eICU-CRD, with no accompanying technical details on how the system had been operationally embedded within the hospital infrastructure. This gap between the reported maturity claim and the retrospective nature of the supporting evidence illustrates a broader pattern across the reviewed literature, in which translational language can outpace the operational details needed to substantiate it.[37]

## Proposed Clinical Utility

Most studies in this review employed retrospective designs, in which the identified benefits were interpreted as proposed clinical utilities, thereby bridging the gap between algorithm validation and real-world implementation.[19,20] The analysis identified three main domains as the focus of development.

### Therapeutic Guidance & Proactive Escalation

XAI facilitates a shift from reactive to more preventive and targeted care. Early warning models provide valuable lead time, for example, by predicting cardiac arrest risk several hours before the event or by assisting clinicians in determining patient readiness for extubation.[37,45] Identification of subphenotypes also creates opportunities for more precise therapies, such as optimising heparin administration in patients with sepsis and guiding resuscitation strategies in cases of septic shock.[19,41]

### Resource Allocation Optimization

In hospitals, AI serves as an objective triage tool that enhances operational efficiency. Applications include automated notification systems on clinical dashboards to expedite patient discharge, mortality risk stratification to support care planning for elderly patients, and early triage in cases of poisoning to help control medical costs.[22,34,39]

### User-Centric Design and Usability

The successful implementation of AI depends heavily on clinician comfort and trust as end-users. Automated integration with EHRs has been shown to reduce prediction errors and decrease clinicians' workload.[40] Intuitive SHAP visualisations also greatly assist clinicians in understanding the logic behind model predictions, thereby increasing their trust in decision support systems.[21,43]

# Discussion

Intensive care units (ICUs) constitute high-stakes clinical ecosystems in which interdisciplinary teams must continuously integrate multidimensional patient data to inform timely, high-stakes decisions amid the dynamic and often unpredictable trajectory of a patient's physiological status.[5,32,46] The sheer volume of data generated by continuous monitoring often contributes to cognitive fatigue and administrative burden among clinical staff, underscoring the need for data-driven decision support systems that can process this information effectively.[3] However, the greatest barrier to bedside adoption is not predictive accuracy alone but model transparency: many clinicians remain reluctant to trust AI systems they perceive as an impenetrable "black box".[4] The findings of this review suggest that ICU-based AI development is beginning to reflect this concern, with a growing, though still partial, emphasis on decision-support tools that are more transparent, auditable, and oriented toward clinicians' bedside needs, rather than on predictive performance alone.[19] However, as the following sections illustrate, this shift remains uneven: within the studies captured by this review, the corpus is still overwhelmingly dominated by black-box architectures rendered interpretable only after the fact, rather than by models that are transparent by design.

Findings from this corpus reinforce that tree-based algorithms remain the dominant and pragmatic choice for modelling structured ICU data, accounting for 24 of the 29 included studies (82.8%), including XGBoost,[22,29,36] CatBoost,[30,35] and LightGBM[20,43] compared with only four studies employing deep learning architectures.[10,19,42,45] Tree-based ensembles robustly handle complex variable interactions, even when adult ICU data are messy, high-dimensional, and riddled with missing values.[49] In such a high-pressure environment, clinicians tend to favour transparent, auditable models over black-box alternatives that offer only marginal accuracy improvements.[50] Bohlen et al (2025)[51] reported that the performance gap between transparent and state-of-the-art black-box models is typically as small as 0.2–0.9%, a margin that is arguably too narrow to justify trading interpretability for accuracy alone. Criticism of post hoc explanation techniques such as SHAP and LIME has grown in parallel, as these methods offer only approximate attributions and risk creating an "illusion of clarity" that can mislead rather than support clinical judgement.[52] Even within this review's own corpus, SHAP outputs were frequently visualised through dense summaries and beeswarm plots that, without a sufficiently intuitive interface, risk adding to the cognitive burden of already fatigued ICU clinicians rather than relieving it.[43] Against this backdrop, the single glass-box example identified in this review—the Explainable

Boosting Machine developed by Magunia et al (2021)[44]—illustrates an alternative pathway, allowing clinicians to inspect and, in principle, refine model logic before deployment rather than reconstructing it after the fact.[53] This distinction matters: the future utility of AI in hospital settings will likely depend less on marginal accuracy gains and more on a system's capacity to offer clinical reasoning that is honest, verifiable, and genuinely usable at the bedside.[54]

Reliance on public datasets from high-income countries, particularly MIMIC and eICU-CRD, has been widely noted to entrench geographic inequities in the global development of clinical AI models.[55–57] Models trained on relatively homogeneous populations often generalise poorly when deployed in hospitals or countries with different genetic backgrounds, clinical workflows, and resource availability.[58,59] This mismatch can produce substantial data drift and a corresponding drop in accuracy when applied to patients from a different institution altogether, posing a risk to patient safety wherever models are deployed without rigorous local validation.[56,60] This pattern is clearly visible within the present corpus: 22 of the 29 included studies (75.9%) originated from China, and 23 (79.3%) relied at least partly on MIMIC-III/IV, with a further 13 (44.8%) also using the eICU-CRD (Table 2). Thus, the studies reviewed here, although treated as 29 independent data points, in practice reflect a narrow methodological base, with predominantly Chinese research teams applying SHAP-explained tree ensembles to the same handful of public registries. Against this backdrop, the small number of studies departing from this pattern stand out precisely because they are exceptions rather than the norm: Magunia et al (2021)[44] validated their model through a nationwide German registry rather than a US public database, and Sjoding et al (2021)[10] applied an imaging-based approach in a US cohort distinct from the MIMIC/eICU database that otherwise dominates the field. Beyond raising concerns about generalisability, this geographic and methodological concentration carries an equity dimension, as the dominance of data from a narrow set of settings risks marginalising populations and healthcare contexts that are absent from the training data.[57,61] Federated learning has been proposed as a promising avenue for building more inclusive models across institutions without compromising patient data privacy,[62–64] although its application to ICU-based XAI remains largely unexplored in the studies reviewed.

The current AI implementation in the ICU remains dominated by passive tools that contribute to alert fatigue and clinician reluctance to engage with them.[65,66] In a fast-paced ICU environment, the need for manual data entry adds further cognitive burden and disrupts workflows that should remain centred on the patient.[67,68] A concrete example of bridging research and practice within this corpus is Jia et al (2021),[45] where counterfactual explanations were used to translate a black-box prediction into a personalised, actionable recommendation on extubation readiness. Therefore, the successful translation of XAI in adult ICUs depends less on technical accuracy alone than on a system's ability to deliver insights that clinicians can act on instantly without disrupting an already demanding workflow.[19] Realising this potential will likely require automatic integration into the electronic health record, surfacing only when clinically necessary to avoid compounding alert fatigue.[67,69,70] Yet the reliability of any such integration remains contingent on the quality of bedside documentation performed by nursing staff, since AI-based decision support is only as trustworthy as the clinical data entered at the point of care.[71] Taken together, the evidence synthesised in this review suggests that predictive accuracy is no longer the principal barrier to clinical AI in the ICU; the more pressing challenge lies in achieving seamless workflow integration and demonstrating robust external validity across diverse clinical settings.

## Practical Implications

AI in the ICU is shifting from risk-prediction tools toward decision-support systems that guide the care team more directly,[72,73] a shift illustrated within this review's own corpus by McWilliams et al (2019),[46] whose dashboard "nudge" was purpose-built to support nurse-led discharge decisions. Yet this shift has clear limits: explainable AI works best as a cognitive aid that lightens the burden of interpreting complex data, not as a substitute for physicians' clinical judgement,[74] a boundary that matters most for high-risk or complex patients, who still need fuller human decision-making rather than automated flagging.[32,46]

The implementation of this principle in everyday practice depends on the coordinated efforts of the care team. Automatic EHR integration is a starting point, ensuring that information reaches the team without manual-entry delays,[72,75] but technology alone is not enough: physicians need training to interrogate rather than passively trust XAI outputs, given the risk that post-hoc visualisations such as SHAP can create an illusion of clarity; nurses are well placed to standardise the

bedside documentation these systems depend on; management must prioritise investment in local data registries; and IT teams need to embed such solutions directly into clinical workflow rather than leaving them as standalone tools.

This last point, investment in local data, addresses a deeper problem: most current systems remain trained on legacy data from foreign settings, risking mismatches with local patients.[76] McWilliams et al (2019)[46] showed that testing across genuinely distinct cohorts is key to true generalisability, and transfer learning offers one route to local adaptation without building new pipelines from scratch. Confirming real-world benefits will still require prospective testing in ICU settings[73,74] built on local registries rather than continued reliance on external datasets such as MIMIC.[76] However, none of this will matter unless such infrastructure is paired with institutionally governed systems that preserve clinician autonomy and accountability; only then will these tools remain safe and equitable for the populations they serve.[5,75,76]

## Strengths and Limitations

The principal strength of this scoping review was its systematic and transparent methodological approach. Study selection, data extraction, and synthesis followed the PRISMA-ScR guidelines, ensuring that each step was replicated. The use of the PCC framework, involvement of two independent reviewers, and analysis across four main domains are additional advantages, as they facilitate the structured mapping of the translational potential, explainability strategies, and clinical relevance of AI tools in adult ICUs.

This study had several limitations. This review did not conduct a formal critical appraisal of the methodological quality or risk of bias in the included studies; therefore, the strength of the evidence for each study could not be compared. Synthesis was performed narratively because of substantial heterogeneity in clinical tasks, data sources, model architectures, explainability strategies, and translational indicators, which precluded a meta-analysis. Although two reviewers conducted screening and extraction independently, a formal inter-reviewer agreement statistic (eg Cohen's kappa) was not calculated; consistency was instead achieved through discussion and consensus. Because eligibility required studies to demonstrate at least one indicator of translational potential, this approach involved sampling dependent variables. The reported prevalence figures characterise the translation-oriented slice of ICU-XAI literature rather than the field as a whole and may understate the true proportion of early stage, non-deployed AI research in this domain. Additionally, the search strategy did not include explicit terms for inherently interpretable (glass-box) modelling approaches (eg "interpretable model", "glass-box", "Explainable Boosting Machine"), meaning studies employing such approaches may have been under-ascertained relative to those using post-hoc explainability methods, which were directly targeted by the search terms. Consequently, any observed predominance of post-hoc over glass-box approaches should be interpreted with caution, as it may partly reflect the search strategy design rather than the true distribution of explainability approaches in the ICU AI literature.

Most studies have employed retrospective designs and relied heavily on public datasets from North America. The included corpus also exhibited a degree of methodological homogeneity, with a substantial proportion of studies relying on similar public datasets and tree-based ensemble architectures explained via SHAP; this concentration may limit the independence of the evidence reviewed. Therefore, the generalisation of the findings to local contexts or other populations should be interpreted with caution.

## Conclusion

This scoping review of 29 studies indicates that predictive accuracy is no longer the principal barrier to AI-based decision support in adult ICUs; the more consequential challenge lies in explainability that the interdisciplinary care team can trust and in translating retrospective models into tools embedded within real clinical workflows. Tree-based ensembles explained through SHAP dominate the field, while genuinely interpretable, glass-box alternatives remain rare, illustrated by the single Explainable Boosting Machine identified in this corpus; this near-total reliance on post-hoc explanation should temper any narrative of a field already transitioning toward transparency by design.

Translational activity is similarly uneven: web-based prototypes and calculators are common, but direct integration into clinical workflows remains the exception, and the evidence base itself is concentrated in a narrow set of institutions,

datasets, and modelling choices that limit the generalisability of these findings. Closing this gap will depend less on further gains in discrimination and more on the interdisciplinary and institutional commitments outlined above. Future research should prioritise external validation in genuinely independent settings, prospective evaluation of clinical impact, and explicit comparison between post-hoc and inherently interpretable approaches, rather than continued optimisation of retrospective discrimination alone.

## Declaration of Generative AI

The authors used Google Gemini 3 to improve the clarity and organisation of the writing during the manuscript preparation. All content was subsequently reviewed and revised by the authors, who took full responsibility for the accuracy and integrity of the final manuscript.

## Acknowledgments

This publication charge is funded by Unpad through the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science, and Technology, and managed under the EQUITY Program (Contract No. 4303/B3/DT.03.08/2025 and 3927/UN6. RKT/HK.07.00/2025).

## Disclosure

The authors declare that there are no conflicts of interest related to the research, authorship, or publication of this study.

## References

1. Berkhout WEM, van Wijngaarden JJ, Workum JD, et al. Operationalization of artificial intelligence applications in the intensive care unit a systematic review. *JAMA Netw Open*. *Am Med Assoc*. 2025;8(7). doi:10.1001/jamanetworkopen.2025.22866
2. Valentin A, Ferdinande P. Recommendations on basic requirements for intensive care units: structural and organizational aspects. *Intensive Care Med*. *Springer Verlag*. 2011;37(10):1569–1571. doi:10.1007/s00134-011-2332-z
3. Vernic C, Tamas TP, Petre I, Ursoniu S. Transforming critical care: the digital revolution's impact on intensive care units. *Front Digit Health*. 2025;7(November):1–15. doi:10.3389/fdgth.2025.1664382
4. Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. *Healthcare*. 2025;13(17):1–28. doi:10.3390/healthcare13172154
5. Żerdziński K, Janiec J, Jóźwik K, Łajczak P, Krzych ŁJ. Artificial intelligence in intensive care: an overview of systematic reviews with clinical maturity and readiness mapping. *J Clin Med*. 2026;15(1):1–22. doi:10.3390/jcm15010185
6. Bignami EG, Berdini M, Panizzi M, et al. Artificial Intelligence in Sepsis Management: an overview for Clinicians. *J Clin Med*. 2025;14(1):1–20. doi:10.3390/jcm14010286
7. Gottlieb ER, Samuel M, Bonventre JV, Celi LA, Mattie H. Machine learning for acute kidney injury prediction in the intensive care unit. *Adv Chronic Kidney Dis*. 2022;29(5):431–438. doi:10.1053/j.ackd.2022.06.005
8. Huang J, Liu X, Jin W. Clinical decision support systems for 3-month mortality in elderly patients admitted to ICU with ischemic stroke using interpretable machine learning. *Digit Health*. 2024;10. doi:10.1177/20552076241280126
9. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus*. 2023;15(10). doi:10.7759/cureus.46454
10. Sjoding MW, Taylor D, Motyka J, et al. Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. *Lancet Digit Health*. 2021;3(6):e340–e348. doi:10.1016/S2589-7500(21)00056-X
11. Gupta J, Majumder AK, Sengupta D, Sultana M, Bhattacharya S. Investigating computational models for diagnosis and prognosis of sepsis based on clinical parameters: opportunities, challenges, and future research directions. *J Intensive Me*. 2024;4(4):468–477. doi:10.1016/j.jointm.2024.04.006
12. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell Nat Res*. 2019;1(5):206–215. doi:10.1038/s42256-019-0048-x
13. Marcinkevičs R, Vogt JE. Interpretable and explainable machine learning: a methods-centric overview with concrete examples. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2023;13(3). doi:10.1002/widm.1493
14. 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;26(1):64. doi:10.1186/s12911-026-03344-0
15. Chromik M, Eiband M, Buchner F, Krüger A, Butz A. I think i get your point, AI! the illusion of explanatory depth in explainable AI. In: *International Conference on Intelligent User Interfaces, Proceedings IUI*. Association for Computing Machinery; 2021:307–317. doi:10.1145/3397481.3450644.
16. Hu D. An examination of the justification for post hoc explanations of artificial intelligence. *AI Soc*. 2025. doi:10.1007/s00146-025-02742-8
17. Yang M, Shi N, Chen H, et al. Multimodal artificial intelligence for precision critical care: a scoping review. *Health Data Sci*. 2026;6. doi:10.34133/hds.0356
18. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–473. doi:10.7326/M18-0850

19. Zhu L, Chen Z, Zhang H, et al. Explainable AI unravels sepsis heterogeneity via coagulation-inflammation profiles for prognosis and stratification. *Nat Commun*. 2025;16(1). doi:10.1038/s41467-025-65365-z
20. Ge X, Chen W, Shi J, et al. Prediction of moderate-to-severe sepsis-associated acute kidney injury using a dual-timepoint machine learning model: development, multiregional validation, and clinical deployment study. *J Med Internet Res*. 2025;27(1):1–26. doi:10.2196/73840
21. Chen M, Li P, Xu Y, et al. Development and validation of interpretable machine learning models to predict intensive care unit outcomes in patients on hemodialysis: a multicenter study. *BMC Med Inform Decis Mak*. 2026;26(1). doi:10.1186/s12911-025-03301-3
22. Ma M, Liu J, Li C, et al. Thirty-day mortality risk prediction for geriatric patients undergoing non-cardiac surgery in the surgical intensive care unit. *Eur J Med Res*. 2025;30(1):1–15. doi:10.1186/s40001-025-02543-1
23. Ge Y, Wang G, Liu T, Ji W, Sun J, Zhang Y. Predicting 30-day in-hospital mortality in ICU asthma patients: a retrospective machine learning study with external validation. *BMC Pulm Med*. 2025;25(1). doi:10.1186/s12890-025-03881-w
24. Yang R, Huang T, Yao R, et al. Risk factors and an interpretability tool of in-hospital mortality in critically ill patients with acute myocardial infarction. *Clin Med J Royal College Phys London*. 2025;25(3):100299. doi:10.1016/j.clinme.2025.100299
25. Hu W, Jin T, Pan Z, et al. An interpretable ensemble learning model facilitates early risk stratification of ischemic stroke in intensive care unit: development and external validation of ICU-ISPM. *Comput Biol Med*. 2023;166(August):107577. doi:10.1016/j.compbiomed.2023.107577
26. Wang SQ, Qiu K, Zheng QR, et al. Development and validation of web-based, interpretable predictive models for sepsis and mortality in extensive burns. *Front Cell Infect Microbiol*. 2025;15(August):1–15. doi:10.3389/fcimb.2025.1586087
27. Gu K, Lu S. Machine learning-based predictive tools and nomogram for in-hospital mortality in critically ill cancer patients: development and external validation using retrospective cohorts. *BMC Med Inform Decis Mak*. 2025;25(1). doi:10.1186/s12911-025-03054-z
28. Wang ZJ, Li FY, Cai JJ, Xue ZT, Zhou Y, Wang Z. Construction and validation of a machine learning-based prediction model for short-term mortality in critically ill patients with liver cirrhosis. *Clin Res Hepatol Gastroenterol*. 2025;49(1):102507. doi:10.1016/j.clinre.2024.102507
29. Liu Y, Xu Y, Guo L, et al. Development and external validation of machine learning models for the early prediction of malnutrition in critically ill patients: a prospective observational study. *BMC Med Inform Decis Mak*. 2025;25(1). doi:10.1186/s12911-025-03082-9
30. Cheng Y, Guo Y, Zhao Y, et al. Development and validation of a machine learning model to predict 30-day mortality in ischemic stroke patients with consciousness impairment: insights from MIMIC-IV database and multicenter ICU data in China. *Int J Med Inform*. 2026;207 (November 2025):106203. doi:10.1016/j.ijmedinf.2025.106203
31. Wang Y, Li W, Cui J, Wang Z, Li Y. Development and multicenter validation of a machine learning model for postoperative sepsis risk in critically Ill traumatic spinal injury patients. *Injury*. 2026;57(2):112949. doi:10.1016/j.injury.2025.112949
32. Bai X, Huang S, Huang S, et al. Development and validation of interpretable machine learning models for dynamic prediction of prognosis in acute pancreatitis complicated by acute kidney injury: a multicenter study. *Int J Med Inform*. 2026;209(December 2025):106260. doi:10.1016/j.ijmedinf.2025.106260
33. Yang C, Ma H, Zeng X, et al. CatBoost Machine Learning Model for Thrombosis Risk Prediction in Critically Ill Cancer Patients: a MIMIC-IV Database Study. *Clin App Thrombosis/Hemostasis*. 2026;2026:32. doi:10.1177/10760296251408357
34. Liang X, Zhao W, Liufu W, et al. An explainable machine learning model for comorbidity risk stratification in patients with fractures admitted to the intensive care unit: a multicenter study. *Arch Gerontol Geriatr*. 2026;141(May 2025):106082. doi:10.1016/j.archger.2025.106082
35. Wei S, Dong H, Yao W, et al. Machine learning models for predicting in-hospital mortality from acute pancreatitis in intensive care unit. *BMC Med Inform Decis Mak*. 2025;25(1). doi:10.1186/s12911-025-03033-4
36. Ding R, Deng M, Wei H, et al. Machine learning-based prediction of clinical outcomes after traumatic brain injury: hidden information of early physiological time series. *CNS Neurosci Ther*. 2024;30(7):1–12. doi:10.1111/cns.14848
37. Kim YK, Koo JH, Lee SJ, Song HS, Lee M. Explainable Artificial Intelligence Warning Model Using an Ensemble Approach for In-Hospital Cardiac Arrest Prediction: retrospective Cohort Study. *J Med Internet Res*. 2023;25(1):1–17. doi:10.2196/48244
38. Chen J, Hou D, Song Y. Development and multi-database validation of interpretable machine learning models for predicting In-Hospital mortality in pneumonia patients: a comprehensive analysis across four healthcare systems. *Respir Res*. 2025;26(1). doi:10.1186/s12931-025-03348-w
39. Liu Y, Lyu X, Yang B, et al. Early Triage of Critically Ill Adult Patients With Mushroom Poisoning: machine Learning Approach. *JMIR Form Res*. 2023;7:e44666. doi:10.2196/44666
40. Yang M, Zhuang J, Hu W, et al. Enhancing Patient Selection in Sepsis Clinical Trials Design Through an AI Enrichment Strategy: algorithm Development and Validation. *J Med Internet Res*. 2024;26:1–18. doi:10.2196/54621
41. Wang S, Liu X, Yuan S, Bian Y, Wu H, Ye Q. Artificial intelligence based multispecialty mortality prediction models for septic shock in a multicenter retrospective study. *NPJ Digit Med*. 2025;8(1):1–10. doi:10.1038/s41746-025-01643-w
42. Rong R, Gu Z, Lai H, et al. A deep learning model for clinical outcome prediction using longitudinal inpatient electronic health records. *JAMIA Open*. 2025;8(2). doi:10.1093/jamiaopen/ooaf026
43. Lin TH, Chung HY, Jian MJ, et al. AI-Driven Innovations for Early Sepsis Detection by Combining Predictive Accuracy With Blood Count Analysis in an Emergency Setting: retrospective Study. *J Med Internet Res*. 2025;27(1):1–13. doi:10.2196/56155
44. Magunia H, Lederer S, Verbuecheln R, et al. Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. *Crit Care*. 2021;25(1):1–14. doi:10.1186/s13054-021-03720-4
45. Jia Y, Kaul C, Lawton T, Murray-Smith R, Habli I. Prediction of weaning from mechanical ventilation using Convolutional Neural Networks. *Artif Intell Med*. 2021;117(May):102087. doi:10.1016/j.artmed.2021.102087
46. McWilliams CJ, Lawson DJ, Santos-Rodriguez R, et al. Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. *BMJ Open*. 2019;9(3):1–8. doi:10.1136/bmjopen-2018-025925
47. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021:372. doi:10.1136/bmj.n71
48. Zhu H, Bai J, Li N, et al. FedWeight: mitigating covariate shift of federated learning on electronic health records data through patients re-weighting. *NPJ Digit Med*. 2025;8(1). doi:10.1038/s41746-025-01661-8
49. Dhami A, Onyeukwu KA, Sattar S, et al. The Prognostic Performance of Artificial Intelligence and Machine Learning Models for Mortality Prediction in Intensive Care Units: a Systematic Review. *Cureus*. 2025;17(Ml). doi:10.7759/cureus.90465
50. Pasupuleti MK. Building Interpretable AI Models for Healthcare Decision Support. *Int J Acad Ind Res Innov*. 2025;05(05):549–560. doi:10.62311/nesx/rphcr16

51. Bohlen L, Rosenberger J, Zschech P, Kraus M. Leveraging interpretable machine learning in intensive care. *Ann Oper Res*. 2025;347 (2):1093–1132. doi:10.1007/s10479-024-06226-8
52. Abgrall G, Holder AL, Chelly Dagdia Z, Zeitouni K, Monnet X. Should AI models be explainable to clinicians? *Crit Care*. 2024;28(1):1–8. doi:10.1186/s13054-024-05005-y
53. Hegselmann S, Ertmer C, Volkert T, Gottschalk A, Dugas M, Varghese J. Development and validation of an interpretable 3 day intensive care unit readmission prediction model using explainable boosting machines. *Front Med*. 2022;9. doi:10.3389/fmed.2022.960296
54. Caterson J, Lewin A, Williamson E. The application of explainable artificial intelligence (XAI) in electronic health record research: a scoping review. *Digit Health*. 2024;10. doi:10.1177/20552076241272657
55. Tang R, Zhang S, Ding C, Zhu M, Gao Y. Artificial Intelligence in Intensive Care Medicine: bibliometric Analysis. *J Med Internet Res*. 2022;24 (11):1–15. doi:10.2196/42185
56. Rockenschaub P, Hilbert A, Kossen T, et al. The Impact of Multi-Institution Datasets on the Generalizability of Machine Learning Prediction Models in the ICU. *Crit Care Med*. 2024;52(11):1710–1721. doi:10.1097/CCM.0000000000006359
57. Celi LA, Cellini J, Charpignon ML, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health*. 2022;1(3 March):1–19. doi:10.1371/journal.pdig.0000022
58. Yang J, Clifton L, Dung NT, et al. Mitigating machine learning bias between high income and low–middle income countries for enhanced model fairness and generalizability. *Sci Rep*. 2024;14(1):1–12. doi:10.1038/s41598-024-64210-5
59. Lasko TA, Strobl EV, Stead WW. Why do probabilistic clinical models fail to transport between sites. *NPJ Digit Med*. 2024;7(1):1–8. doi:10.1038/s41746-024-01037-4
60. Montomoli J, Bitondo MM, Cascella M, et al. Algor-ethics: charting the ethical path for AI in critical care. *J Clin Monit Comput*. 2024;38 (4):931–939. doi:10.1007/s10877-024-01157-y
61. Agnes K. Addressing Bias in AI Algorithms for Health Applications. *IAA J Biol Sci*. 2025;13(1):37–43. doi:10.59298/iaajb/2025/1313743
62. Teo ZL, Jin L, Li S, et al. Federated machine learning in healthcare: a systematic review on clinical applications and technical architecture. *Cell Rep Med*. 2024;5(2):101419. doi:10.1016/j.xcrm.2024.101419
63. Nguyen TV, Dakka MA, Diakiw SM, et al. A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. *Sci Rep*. 2022;12(1):1–12. doi:10.1038/s41598-022-12833-x
64. Ye H, Zhang X, Liu K, et al. A personalized federated learning approach to enhance joint modeling for heterogeneous medical institutions. *Digit Health*. 2025:11. doi:10.1177/20552076251360861
65. Peek N, Capurro D, Rozova V, van der Veer SN. Bridging the Gap: challenges and Strategies for the Implementation of Artificial Intelligence-based Clinical Decision Support Systems in Clinical Practice. *Yearb Med Inform*. 2025;33(1):103–114. doi:10.1055/s-0044-1800729
66. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int J Med Inform*. 2020;137(September 2019):104072. doi:10.1016/j.ijmedinf.2019.104072
67. Gonzalez FA, Santonocito C, Lamas T, et al. Is artificial intelligence prepared for the 24-h shifts in the ICU? *Anaesth Crit Care Pain Med*. 2024;43 (6). doi:10.1016/j.accpm.2024.101431
68. van de Sande D, Chung EFF, Oosterhoff J, van Bommel J, Gommers D, van Genderen ME. To warrant clinical adoption AI models require a multi-faceted implementation evaluation. *NPJ Digit Med*. 2024;7(1):1–5. doi:10.1038/s41746-024-01064-1
69. Horvat CM, Suresh S, Clark RSB. The stelar ICU: leveraging electronic health record data to foster research and optimize patient care. *Informatics*. 2018;5(3):5–9. doi:10.3390/informatics5030037
70. Contreras M, Silva B, Shickel B, et al. Real-time prediction of intensive care unit patient acuity and therapy requirements using state-space modelling. *Nat Commun*. 2025;16(1):1–15. doi:10.1038/s41467-025-62121-1
71. Yadav S. Embracing Artificial Intelligence: revolutionizing Nursing Documentation for a Better Future. *Cureus*. 2024;16(4):e57725. doi:10.7759/cureus.57725
72. Barea Mendoza JA, Valiente Fernandez M, Pardo Fernandez A, Gómez Álvarez J. Current perspectives on the use of artificial intelligence in critical patient safety. *Med Intensiva*. 2025;49(3):154–164. doi:10.1016/j.medine.2024.04.002
73. Hong N, Liu C, Gao J, et al. State of the Art of Machine Learning-Enabled Clinical Decision Support in Intensive Care Units: literature Review. *JMIR Med Inform*. 2022;10(3):1–15. doi:10.2196/28781
74. Abdelbaky AM, Elmasry WG, Awad AH, Khan S. Role of Artificial Intelligence in Critical Care Medicine: a Literature Review. *Cureus*. 2025;17 (Ml). doi:10.7759/cureus.90149
75. Naseer A, Kiran Q. Artificial intelligence – new horizons in intensive care. *Anaesth Pain Intensive Care*. 2025;29(5):254–256. doi:10.35975/apic.v29i5.2848
76. Moazemi S, Vahdati S, Li J, et al. Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: a systematic review. *Front Med*. 2023;10(March):1–18. doi:10.3389/fmed.2023.1109411



**Journal of Multidisciplinary Healthcare**

**Dovepress**
Taylor & Francis Group

**Publish your work in this journal**

The Journal of Multidisciplinary Healthcare is an international, peer-reviewed open-access journal that aims to represent and publish research in healthcare areas delivered by practitioners of different disciplines. This includes studies and reviews conducted by multidisciplinary teams as well as research which evaluates the results or conduct of such teams or healthcare processes in general. The journal covers a very wide range of areas and welcomes submissions from practitioners at all levels, from all over the world. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

**Submit your manuscript here:** https://www.dovepress.com/journal-of-multidisciplinary-healthcare-journal