



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE

AI-DRIVEN EARLY WARNING SYSTEMS FOR SEPSIS DETECTION
IN ICU SETTINGS: A STRUCTURED REVIEW OF PREDICTIVE
PERFORMANCE, CLINICAL INTEGRATION AND EQUITY
OUTCOMES

DOI

[https://doi.org/10.31435/ijitss.4\(52\).2026.6372](https://doi.org/10.31435/ijitss.4(52).2026.6372)

RECEIVED

16 July 2026

ACCEPTED

30 September 2026

PUBLISHED

05 October 2026

LICENSE



The article is licensed under a **Creative Commons Attribution 4.0 International License**.

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

AI-DRIVEN EARLY WARNING SYSTEMS FOR SEPSIS DETECTION IN ICU SETTINGS: A STRUCTURED REVIEW OF PREDICTIVE PERFORMANCE, CLINICAL INTEGRATION AND EQUITY OUTCOMES

Zuzanna Julia Komin (Corresponding Author, Email: zuza.k.1414@gmail.com)

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0001-6284-6410

Eliza Zofia Kabat

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-6531-967X

Michalina Justyna Kozłowska

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-9561-2957

Krzysztof Konopka

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-4055-0074

Maciej Kamil Majchrzak

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0007-4739-0092

Faustyna Magdalena Kołodziej

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0002-8778-7100

Alicja Kwiatkowska

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0003-1027-2114

Jakub Juskowiak

Poznań University of Medical Sciences, Poznań, Poland

ORCID ID: 0009-0001-0372-400X

Mateusz Lis

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0006-8441-8296

Maryan Liasota

Pomeranian Medical University, Szczecin, Poland

ORCID ID: 0009-0006-0266-4849

ABSTRACT

Background: Sepsis remains a primary cause of mortality and organ dysfunction in Intensive Care Units (ICUs) globally. Although AI models show considerable promise for the early identification of sepsis, moving these tools from retrospective performance to bedside utility is hindered by socio-technical barriers, workflow friction, and algorithmic equity issues.

Objective: This structured review evaluates the state of AI-driven sepsis Early Warning Systems (EWS), identifying systemic barriers to clinical integration and assessing algorithmic fairness across demographic cohorts.

Methodology: A structured review of peer-reviewed literature published in PubMed/MEDLINE and Google Scholar between January 2018 and June 2026 was conducted, focusing on predictive performance, clinical implementation, and algorithmic equity.

Results: Retrospective assessments of deep sequential architectures demonstrate high diagnostic accuracy (AUROC>0.95) (Boussina et al., 2024; Siddiqui, 2026). Prospective trials of landmark systems such as TREWS (Adams et al., 2022), COMPOSER (Shashikumar et al., 2025), and CONCERN EWS (Rossetti et al., 2025), confirm clinical utility, yielding mortality reductions of 17% to 35.6%. However, real-world effectiveness is frequently hindered by severe alert fatigue (Wong et al., 2021). Furthermore, because models rely on historical health data, they often inherit systemic biases, risking the exacerbation of existing disparities related to race, gender and socioeconomic status (Chen et al., 2023; Obermeyer et al., 2019; Seyyed-Kalantari et al., 2021).

Conclusion: Maximizing the clinical impact of AI-driven EWS requires a paradigm shift from purely performance-centric model development to rigorous, equity-oriented clinical governance guided by international reporting frameworks like TRIPOD-AI (Collins et al., 2024) and DECIDE-AI (Vasey et al., 2022). To operationalize this, we propose the integration of the Clinical AI Safety Officer (CAISO) to maintain real-time algorithmic safety, audit demographic equity, and manage the clinical accountability loop.

KEYWORDS

Sepsis, Artificial Intelligence, Early Warning Systems, Intensive Care Units, Health Equity, Clinical Decision Support

CITATION

Komin, Z. J., Kabat, E. Z., Kozłowska, M. J., Konopka, K., Majchrzak, M. K., Kołodziej, F. M., Kwiatkowska, A., Juskowiak, J., Lis, M., & Liasota, M. (2026). AI-driven early warning systems for sepsis detection in ICU settings: A structured review of predictive performance, clinical integration and equity outcomes. *International Journal of Innovative Technologies in Social Science*, 4(52). [https://doi.org/10.31435/ijitss.4\(52\).2026.6372](https://doi.org/10.31435/ijitss.4(52).2026.6372)

COPYRIGHT

© The author(s) 2026. This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

1.1 Background and Clinical Context

Sepsis, explained under the Sepsis-3 definition as a life-threatening organ dysfunction caused by a dysregulated host response to infection, remains a primary factor in global critical care mortality and it accounts for up to 35% of all in-hospital deaths in high-income countries. A rapid cascade of cellular injury, microvascular leakage, and endothelial damage drives the transition from localized infection to systemic inflammatory response syndrome (SIRS), eventually progressing to distributive septic shock. This state induces systemic hypotension and tissue hypoperfusion, which, without immediate therapeutic intervention, inevitably trigger multi-organ failure.

Clinical consensus underscores that early recognition is the primary determinant of patient survival. Every hour of delay in administering broad-spectrum intravenous antibiotics and initiating aggressive fluid resuscitation is associated with a linear, cumulative increase in absolute mortality. However, early detection still remains a challenge. Sepsis presents with extreme heterogeneity, and its initial indicators- such as mild tachypnea, low-grade pyrexia, subtle tachycardia, and marginal blood pressure fluctuations- are highly non-specific, often confounded by other common critical care conditions. For years, clinical practice has been dominated by decision tools and scales like SIRS, qSOFA, and the National Early Warning Score (NEWS). While these tools are widely implemented, they present severe trade-offs: SIRS demonstrates high sensitivity but poor specificity, leading to widespread false alarms, while qSOFA often lacks sufficient sensitivity to flag patients until organ dysfunction is already established.

1.2 The Rise of AI-Driven Early Warning Systems for Sepsis Detection in ICU Settings

To overcome the limitations of traditional scoring tools, the combination of artificial intelligence (AI) and machine learning (ML) has emerged as a promising approach. Most modern ICUs accumulate electronic health record (EHR) data. AI-driven early warning systems (EWS) are designed to ingest these complex, multi-modal data streams in real-time, leveraging sophisticated algorithms to identify subtle, non-linear correlations and temporal trajectories that precede clinical recognition.

Over the last decade, predictive modeling has evolved from relying on static statistical rules toward the data-driven frameworks. There has been a clear transition from classical tree-based ensembles (e.g., XGBoost) to advanced architectures like LSTMs and RNNs, which offer a more robust way to model the sequential nature of clinical deterioration over time (Boussina et al., 2024; Siddiqui, 2026). More recently, the research focus has evolved toward Transformer-based deep learning. By employing self-attention mechanisms, these models can capture long-range temporal dependencies and complex multi-modal relationships, offering superior predictive performance and earlier warning horizons compared to traditional sequence models (Shashikumar et al., 2025; Vasey et al., 2022).

1.3 From Technical Performance to Equity and Engagement

While retrospective evaluations of AI-driven sepsis models frequently report outstanding diagnostic accuracy- with AUROC metrics often exceeding 0.90- these technical benchmarks rarely translate into real-world clinical utility (Adams et al., 2022). This discrepancy is best understood as the "implementation paradox": the observation that models with superior retrospective predictive power often exhibit lower clinical impact when deployed at the bedside. This disconnect highlights a critical reality: AI tools are not independent factors; they operate within real-world clinical environments where their success depends as much on institutional culture and human-factor design as on the algorithm itself. These models are built on historical EHR data and they often inherit and inadvertently amplify the systemic biases embedded in past clinical practice. Crucially, this is not merely a "data problem" that can be solved by cleaning datasets. As Hooker (2021) argues, algorithmic bias is deeply rooted in the design choices, model architectures, and incentive structures that prioritize aggregate performance metrics over individual and equitable patient outcomes. Real-world healthcare data is shaped by decades of structural inequality, uneven geographic distribution, and significant disparities in access to care (Shashikumar et al., 2025). If a model relies on cost-based proxies for illness severity or is trained on datasets where marginalized groups are underrepresented, the resulting predictions may exhibit disparate performance across race, gender, and socioeconomic status (Chen et al., 2023; Gichoya et al., 2022). By underdiagnosing or delaying care for already vulnerable groups, these systems may widen the very health inequities they were meant to solve (Obermeyer et al., 2019; Seyyed-Kalantari et al., 2021).

1.4 Research Gap and Rationale

Although the computer science and medical informatics literature is predominantly characterized by studies detailing retrospective model development and optimization (Boussina et al., 2023, 2024), there is a lack of structured syntheses evaluating the socio-technical reality of deployed AI sepsis systems. Specifically, existing reviews frequently analyze algorithmic performance in isolation, separate from clinical workflow barriers, implementation science, and demographic equity audits. Furthermore, there is insufficient exploration of how contemporary AI clinical validation studies align with recently established reporting standards, such as the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis - Artificial Intelligence (TRIPOD-AI) (Collins et al., 2024) and the Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (DECIDE-AI) guidelines (Vasey et al., 2022). Understanding this alignment is crucial, as it is the only way to ensure that AI-driven interventions move beyond academic performance to become genuinely safe, equitable, and reliable tools in clinical practice.

1.5 Aim of the Review

This structured review synthesizes the contemporary literature (2018–2026) regarding AI-driven early warning systems for sepsis detection in ICU and critical care settings. Specifically, this review evaluates the prevailing machine learning architectures and data modalities utilized for early sepsis prediction, with a particular focus on how their predictive metrics evolve as the lead-time horizon increases (Siddiqui, 2026). It also examines the underlying technical and socio-technical hurdles that stall these tools at the bedside; most notably, the persistent challenge of alert fatigue often undermines their potential for meaningful clinical

integration (Adams et al., 2022; Wong et al., 2021). In addition to examining performance and implementation, this review analyzes the extent to which existing research evaluates algorithmic bias and the mechanisms through which these models may inadvertently reinforce health inequities across diverse demographic cohorts (Chen et al., 2023; Seyyed-Kalantari et al., 2021). Finally, the study assesses how the current evidence base for AI-driven sepsis detection aligns with established clinical reporting standards, specifically the TRIPOD-AI and DECIDE-AI guidelines, to ensure that future interventions are clinically safe, fair, and replicable (Collins et al., 2024; Vasey et al., 2022).

2. Methodology

2.1 Research Design

To investigate the multi-dimensional challenges of AI-driven sepsis detection, this study utilized a structured review design. Meta-analyses are valuable for evaluating statistical performance, but they often provide limited insight into practical function of AI tools. A structured review addresses this issue by bringing together the mathematical, clinical, and ethical considerations inherent in these interventions. This is essential because sepsis EWS operate within complex clinical environments dictated by human workflows, institutional goals, and the persistent biases embedded in historical records. Consequently, the analytical framework of this review was structured to evaluate these systems as socio-technical clinical interventions, analyzing their performance, workflow integration, and equity implications.

2.2 Data Source and Search Strategy

A literature search was conducted using the PubMed/MEDLINE database selected due to its comprehensive indexing of peer-reviewed biomedical, public health, and interdisciplinary health policy research. Targeted follow-up searches across Google Scholar, SciSpace, and Semantic Scholar were performed to capture emerging prospective trials and established clinical guidelines. The search window was restricted to publications from January 2018 to June 2026 to capture the modern era of deep learning in clinical medicine, the evolution of real-world implementation trials, and the emergence of up-to-date AI reporting guidelines.

Our search strategy employed a combination of MeSH terms and Boolean operators to identify studies at the intersection of sepsis, artificial intelligence, critical care workflows, and socio-ethical outcomes. The primary search string was:

("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Predictive Model") AND ("Sepsis" OR "Septic Shock" OR "Systemic Inflammatory Response Syndrome") AND ("ICU" OR "Intensive Care" OR "Critical Care") AND ("Early Warning System" OR "EWS" OR "Clinical Decision Support") AND ("Clinical Integration" OR "Implementation" OR "Alert Fatigue" OR "Equity" OR "Algorithmic Bias")

This query was designed to deliberately bypass studies focused solely on computational model tuning, instead isolating papers that actively engage with the clinical realities, implementation hurdles, and ethical dimensions of deployed predictive tools. All of the included studies were published in English and peer-reviewed, ranging from randomized controlled trials to prospective pragmatic studies, multi-center cohort studies and policy analyses. This strategy allowed to synthesize both clinical data and conceptual analysis, ensuring a comprehensive view.

2.3 Inclusion and Exclusion Criteria

To ensure methodological rigor and thematic coherence, all retrieved literature was evaluated against strict, predefined inclusion and exclusion criteria, summarized in Table 1.

Table 1. Inclusion and Exclusion Criteria.

Criterion	Inclusion Criteria	Exclusion Criteria
Study Type	RCTs, prospective pragmatic trials, multi-center cohort studies, implementation evaluations, and clinical guidelines.	Single-center retrospective performance tuning studies; non-peer-reviewed preprints, editorials, or conference abstracts.
Focus	AI/ML-driven EWS for sepsis prediction; evaluation of clinical integration, workflow utility, alert fatigue, or demographic equity.	Non-sepsis deterioration, general diagnostic imaging, or treatment optimization; models without prospective evaluation.
Database	PubMed/MEDLINE, Google Scholar, Semantic Scholar, and SciSpace.	Databases not subject to standard academic indexing.
Language	Articles published in English.	Articles in languages other than English.
Timeline	January 2018 – June 2026.	Studies published before 2018.

2.4 Study Selection Process

The study selection process followed an initial structured screening based on titles and abstracts to exclude papers focusing on non-clinical machine learning optimization or unrelated diagnostic tools. Articles that met the initial criteria underwent full-text evaluation to confirm that they addressed at least one of the core domains: predictive metrics over time, workflow implementation barriers (such as alert fatigue), or algorithmic fairness and demographic bias.

2.5 Data Extraction

A structured data extraction form was utilized to capture both qualitative and quantitative parameters from the final study corpus. This form recorded computational characteristics (algorithms, input variables, sequential architectures, and prediction lead times); prospective clinical outcomes (mortality, length of stay, bundle compliance); implementation science metrics (alert PPV, adoption, override rates); and fairness parameters (demographic representation, bias mitigation). Two authors independently extracted the data, and any disagreements were resolved through discussion and consensus.

2.6 Thematic Synthesis

The evidence gathered in this review was examined through a qualitative thematic synthesis. Instead of attempting a quantitative meta-analysis of performance metrics, the objective was to examine the core implementation hurdles and systemic challenges associated with deploying sepsis early warning systems in clinical practice. The analysis focused on recurring patterns covering effectiveness, workflow dynamics, and the socio-ethical implications of integrating these models into practice. We employed an inductive process to explore the literature. Through repeated review of the literature, both overlapping and conflicting perspectives were identified and transformed into main analytical categories. A socio-technical perspective served as our primary interpretive framework, allowing us to evaluate predictive algorithms not as static software, but as dynamic components embedded within the complex, high-stakes environment of the ICU. The findings were consolidated into three distinct thematic domains:

1. Architectural Performance and Temporal Utility: Assessing the tension between advanced machine learning architectures (such as Transformers and LSTMs) and the practical requirement for clinically meaningful warning lead-times.

2. Workflow Integration and Human-Factor Engineering: Investigating how technical systems interact with clinical habits, specifically focusing on the causes of alert fatigue and the conditions required to foster meaningful clinician-AI collaboration.

3. Systemic Bias and Algorithmic Equity: Deconstructing how inherent data limitations and historical disparities are encoded into predictive models, and analyzing the resulting risk of automated diagnostic or treatment gaps across diverse patient groups.

This approach allowed us to move beyond metrics of model accuracy to uncover the structural influences at play. By highlighting the interplay between data-driven prediction and existing clinical governance, this synthesis illustrates how AI integration is a two-way process: these tools are not only changing the delivery of critical care, but are also being continuously redefined by the clinical workflows and institutional biases they encounter.

3. Results

3.1 Structural Overview of the Evidence

Trends in clinical AI research in sepsis detection significantly changed between 2018 and 2026. From 2018 to 2021 the dominant tendencies were retrospective, single-center evaluations, with researchers leveraging datasets such as MIMIC-IV to push the boundaries of predictive model performance (Boussina et al., 2024). These foundational studies were crucial for demonstrating that machine learning architectures could identify complex, non-linear correlations in physiological time-series data signals that traditional scoring systems frequently overlooked. This phase also included the early development of scalable infrastructure needed to operationalize these models, shifting the focus from isolated algorithms to integrated, real-time healthcare predictive analytics platforms capable of handling high-velocity EHR data (Boussina et al., 2023). However, this early period was marked by an over-emphasis on technical benchmarks, often leading to models that functioned exceptionally well in "dry-lab" environments but exhibited poor generalization when confronted with the data-entry delays and heterogeneous charting practices of real-world hospitals. Since 2022, the field has transitioned toward a new era characterized by prospective, multi-site implementation trials. This shift reflects a growing understanding that clinical AI is a dynamic, socio-technical intervention. Current research prioritizes "clinical utility" over "algorithmic accuracy". Landmark systems, including the Targeted Real-time Early Warning System (TREWS) (Adams et al., 2022), the Conformal Multidimensional Prediction Of Sepsis Risk (COMPOSER) system (Shashikumar et al., 2025), and the Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) EWS (Rossetti et al., 2025), now serve as the primary evidence base for how AI models function when subjected to the operational constraints of real-world ICU workflows. The contemporary evidence base suggests that the impact of these technologies is deeply context-dependent; while some institutions report substantial life-saving benefits, others demonstrate that poorly integrated or inadequately calibrated algorithms fail to improve clinical outcomes due to systemic workflow friction, low user adoption, and, most critically, persistent alert fatigue.

3.2 Predictive Performance and Clinical Outcomes: Benchmarks and Mortality Impact

Deep sequential architectures, such as LSTMs and Transformer-based networks, have emerged as the new standard for retrospective predictive performance (Boussina et al., 2024; Siddiqui, 2026). These architectures excel by processing high-dimensional, longitudinal data, effectively capturing the temporal interplay between evolving physiological markers, such as the nuanced relationship between a rising serum lactate, falling mean arterial pressure, and subtle heart rate variability, that static clinical rules completely miss. The literature reports that Transformer-based and hybrid CNN-Transformer models can achieve AUROC metrics as high as 0.99 in retrospective cohorts (Boussina et al., 2024; Siddiqui, 2026). However, the literature highlights a consistent, mathematically documented "precision decay" as the prediction lead-time increases. When moving from a 1-hour to a 12-hour prediction horizon, even the most sophisticated models face a decline in positive predictive value (PPV). This decay is not merely a technical failure, it is an inherent clinical challenge. Predicting sepsis prior to organ dysfunction is inherently probabilistic. Despite this, prospective trials have confirmed that AI systems can have measurable clinical value if they are paired with appropriate, time-sensitive interventions.

1. TREWS (Adams et al., 2022): Implemented across five academic and community hospitals, the TREWS trial provides a granular look at how AI functions as a behavioral intervention. Unlike "best practice" alerts that simply notify a clinician of a risk score, TREWS was designed to prioritize the delivery of the Surviving Sepsis Campaign (SSC) bundles. When the system triggered an alert, it provided the clinician with a structured interface that prioritized the three "core" components of resuscitation: the timely acquisition of

blood cultures, the administration of broad-spectrum intravenous antibiotics, and the initiation of fluid resuscitation. The AI did not just "warn", it effectively acted as an electronic checklist that synchronized the team's response. The trial demonstrated that when a provider confirmed an alert within three hours, the system facilitated a 1.85-hour reduction in the time to first antibiotic administration. Crucially, the AI's ability to prioritize "antibiotic-first" logic- often delayed by the practical challenge of obtaining blood cultures- resulted in an 18.7% relative decrease in in-hospital mortality. By nudging clinicians to execute these bundle components in a high-priority sequence, TREWS demonstrated that the clinical impact of an AI system is not merely in its sensitivity, but in its ability to reduce the "latency of coordination" between the recognition of a threat and the delivery of life-saving therapy.

1. COMPOSER (Shashikumar et al., 2025): By surveilling over 150 clinical variables and utilizing conformal prediction to quantify uncertainty, the COMPOSER model achieved a 1.9% absolute reduction in sepsis-related mortality in Emergency Department settings. This model demonstrated that the integration of uncertainty quantification allows clinicians to better interpret when the algorithm is "confident" in its prediction, a crucial step for building trust.

2. CONCERN EWS (Rossetti et al., 2025): This study provides the most compelling evidence for the value of "nursing-centric" data. By modeling narrative documentation as a proxy for clinical intuition, the CONCERN EWS identified deterioration up to 42 hours earlier than physiological scores. Furthermore, the trial demonstrated that the system's ability to flag these risks in real-time directly translated into a 24.9% increase in proactive, unanticipated ICU transfers and a significant reduction in overall length of stay (Lee et al., 2024), proving that clinical concern is often captured in unstructured notes long before it is reflected in vital sign abnormalities.

3.3 Clinical Integration, Workflow, and Alert Fatigue Patterns

The most significant barrier to AI adoption in the ICU is the "implementation paradox": models with the highest retrospective predictive power often exhibit the lowest clinical utility in practice. This discordance is driven primarily by severe alert fatigue. The literature indicates that the technical design of the user interface is just as critical as the algorithm itself (Wong et al., 2021). Interruptive, modal dialog boxes - the standard for "Best Practice Advisories" (BPAs) in many legacy EHR systems - often force clinicians to perform extra clicks to close alerts, which disrupts the bedside workflow and generates significant cognitive friction. To minimize cognitive load, HFE research advocates for a shift from interruptive alerts to systems designed for attentive monitoring- a critical distinction for effective clinical support. HFE studies highlight that the visual syntax of an alert - its color, placement, and mode of notification - dictates its acceptance. For instance, the transition from red, high-urgency flashing alerts to subtle, color-coded status icons in the patient summary view represents a major shift in design philosophy. By moving to passive, context-aware notification, systems can reduce the "interruptive tax" on clinical staff. In contrast to systems that prioritize forced compliance, the TREWS system (Adams et al., 2022) integrated alerts as subtle, non-intrusive icons within the patient list, allowing clinicians to review information at their own discretion. This approach honors clinicians' autonomy and leans into the concept of "human-machine teaming". Evidence from the CONCERN trial further supports this, showing that clinicians' usage patterns are highly dependent on the system's ability to augment rather than interrupt their existing workflow (Lee et al., 2025). Recent tendencies toward combining deep learning with LLMs are helping to address the problem of alert fatigue. COMPOSER-LLM is an excellent example of this evolution (Shashikumar et al., 2025). By utilizing natural language processing to extract contextual concepts from free-text physician notes, the model can automatically filter out "sepsis mimics," such as patients experiencing acute anxiety or chronic heart failure exacerbations. By reducing the volume of clinically irrelevant false positives, these hybrid systems maintain clinician engagement, ensuring that when an alert is finally triggered, it is viewed as a signal of high clinical importance.

3.4 Equity of Outcomes: Algorithmic Bias and Demographic Disparities

A critical finding across the contemporary evidence base is that clinical AI models are highly vulnerable to learning and propagating systemic bias, which introduces the risk of automated disparities in care. The literature conceptualizes this not as a coding error, but as a systemic consequence of structural inequities encoded within three primary domains:

1. Historical Bias: Models often rely on proxy variables that reflect societal inequities. A landmark example is the use of cost-based proxies for illness severity (Obermeyer et al., 2019). Because Black patients have historically faced systemic barriers to accessing high-cost care, algorithms systematically underestimate their clinical needs. As a result, the model systematically underestimates the true severity of their clinical needs and prioritizes healthier, wealthier demographic groups for advanced resources.

2. Representation Bias: Models developed in elite, well-resourced academic centers frequently falter when deployed in different clinical settings, such as rural hospitals or public health systems that lack the same infrastructure (Seyyed-Kalantari et al., 2021). Without local retraining, these models may fail to identify the unique disease prevalence or physiological trajectories of patient populations that were not represented in the training data.

3. Measurement Bias: Research has shown that clinical monitoring isn't demographically neutral, for example, serum lactate levels are less frequently measured in female patients than in male patients with similar levels of physiological distress (Chen et al., 2023; Seyyed-Kalantari et al., 2021). When AI models are trained on these datasets, they frequently apply a "normal-value imputation" strategy, treating missing laboratory data as "healthy." This systematically obscures the illness severity of female patients, embedding gender-biased clinical practice directly into the model's predictive architecture.

3.5 The Role of Governance and Reporting Standards in Mitigation

To combat the disparities, the literature suggests that technical fixes are insufficient without a foundational shift in clinical governance. The adoption of frameworks like TRIPOD-AI (Collins et al., 2024) and DECIDE-AI (Vasey et al., 2022) is becoming a mandatory requirement for transparency. TRIPOD-AI emphasizes the necessity of reporting calibration statistics, which are often overlooked in favor of AUROC metrics but are vital for ensuring that a model's risk scores are reliable across different demographic groups. Similarly, DECIDE-AI focuses on the "early-stage clinical evaluation," mandating that developers track human-factor metrics such as the time required for clinicians to achieve proficiency and the overall usability of the interface before a model is scaled across an entire healthcare system. The distinction between the TRIPOD-AI and DECIDE-AI frameworks is critical for institutional safety, as they address fundamentally different phases of the clinical AI lifecycle.

Table 2. Comparison of TRIPOD-AI vs. DECIDE-AI

Feature	TRIPOD-AI	DECIDE-AI
Primary Focus	Performance Reporting	Clinical Safety & Usability
Development Stage	Model Development & Validation	Early Clinical Deployment
Key Metrics	Calibration, Discrimination (AUROC)	Alert Fatigue, Workflow Integration
Stakeholder Goal	Algorithmic Transparency	Patient Safety & Process Improvement
Core Question	"Does the model predict accurately?"	"Does the model improve care?"

While TRIPOD-AI provides the mathematical rigor needed to verify that a model's risk scores are reliable and statistically sound, it is inherently limited to the static data. In contrast, DECIDE-AI is designed for the "live" world of the ICU, mandating the evaluation of how clinicians interact with the system under stress. Relying solely on TRIPOD-AI would leave a hospital blind to the operational risks - such as high override rates or workflow disruption - that characterize the implementation paradox. Conversely, using only DECIDE-AI would ignore the foundational need to verify that the model is mathematically calibrated for the institution's specific patient population. Integrating these frameworks into a mandatory quality-assurance loop is the only way for hospitals to ensure their AI systems remain technically sound, clinically safe, and fair across their entire deployment lifecycle. Unfortunately, recent bibliometric studies reveal that adherence to these standards remains inconsistent (Collins et al., 2024). Many papers continue to prioritize "model performance reporting" over "utility reporting". This creates a transparency vacuum where hospitals may deploy models without knowing if they are calibrated for their specific patient mix or if they are prone to demographic failure. To address this, current research mandates that all AI deployments include a "clinical accountability loop": a process where performance, bias, and alert-fatigue metrics are monitored in real-time, with clear thresholds for taking a model "offline" if it demonstrates drift or inequity. This level of rigorous lifecycle governance is the only way to transform sepsis EWS from black-box software into safe, equitable, and sustainable components of the critical care tools.

4. Discussion

4.1 AI Sepsis Detection as a Socio-Technical Intervention

As demonstrated by Boussina et al. (2024), the efficacy of these systems is less a product of mathematical sophistication and more a result of how well the system aligns with the organizational culture and workflows of the host institution. Unlike the early "dry-lab" research conducted on databases like MIMIC-IV, which treated models as isolated artifacts, current prospective trials (Adams et al., 2022; Rossetti et al., 2025) show that AI interacts directly with pre-existing institutional power structures, nursing documentation habits, and multidisciplinary team dynamics. A significant question surrounding AI in the ICU is whether it risks 'de-skilling' junior clinicians. We must consider if becoming overly reliant on automated alerts undermines the development of their own diagnostic intuition. Fortunately, as highlighted by recent human-factor research, these systems may actually "up-skill" the bedside team by providing them with a high-fidelity "second opinion" that synthesizes data points across tens of thousands of past patient trajectories (Rossetti et al., 2025). We must therefore treat these tools as active agents that both shape and are shaped by the ICU environment, where the ultimate goal is not replacing the clinician's judgment, but enhancing their capacity for complex decision-making (Seyyed-Kalantari et al., 2021; Wong et al., 2021).

4.2 The Implementation Paradox: Predictive Power vs. Clinical Utility

This review identifies a clear "implementation paradox": high retrospective predictive performance frequently fails to translate into improved outcomes (Wong et al., 2021). While Transformer-based architectures have pushed AUROC metrics toward 0.99 (Siddiqui, 2026), the clinical reality often involves alert fatigue and system avoidance (Adams et al., 2022; Wong et al., 2021). As highlighted by Wong (2021), interruptive, modal alerts often disrupt life-saving bedside care, leading to high override rates. In contrast, the success of the TREWS (Adams et al., 2022) and CONCERN (Rossetti et al., 2025) systems underlines that the most effective AI minimizes cognitive load. These systems move away from forced compliance, instead leveraging passive, context-aware notification that respects provider autonomy. As shown in the COMPOSER trials (Shashikumar et al., 2025), the future lies in the human-factor engineering - prioritizing systems that provide actionable insights at the precise moment of clinical decision-making, rather than relying on high-frequency, non-specific alerts.

4.3 The Psychology of Trust-Calibration in Critical Care

A critical yet under-theorized dimension of the "implementation paradox" is the cycle of trust-calibration that clinicians undergo when interacting with AI. Trust in AI is not a binary state but a dynamic process. When a clinician first interacts with an EWS, they often exhibit one of two extremes: automation bias (the tendency to overly rely on the algorithm, even when the patient's clinical status contradicts the AI) or algorithm aversion (the tendency to ignore the AI after one or two false positives). The literature suggests that the transition to a sustainable clinical partnership requires "calibrated trust," where the clinician views the AI as a fallible peer rather than a deterministic oracle (Shashikumar et al., 2025). To achieve this, the clinical interface must transition from providing simple "risk scores" to providing evidence-based explanations. Moreover, as Shashikumar et al. (2021) demonstrate, AI systems should be designed to quantify their own uncertainty and "learn to say I don't know", rather than forcing a deterministic prediction in high-ambiguity scenarios. As deep learning architectures evolve, the "black box" is being replaced by XAI (Explainable Artificial Intelligence) techniques that highlight the specific variables - such as a trend in creatinine levels or a specific pattern in heart rate variability - that triggered the alert. This transparency is vital; when the algorithm can show how it works, it empowers the clinician to perform a sanity check, thereby aligning the AI's computational speed with the clinician's diagnostic intuition. This process of trust-calibration effectively transforms the AI from a source of cognitive friction into a high-fidelity "second opinion", fostering a more nuanced, collaborative environment where the clinician's experience is complemented, not bypassed, by the algorithm (Rossetti et al., 2025; Wong et al., 2021).

4.4 The Equity Gap: Demographic Bias and the Risk of Automated Disparities

The potential for AI to serve as an "equalizer" in critical care is currently threatened by structural bias. Our results highlight that clinical AI is a mirror reflecting systemic inequities (Obermeyer et al., 2019). First, we must acknowledge the technical nature of the problem: modeling studies have demonstrated that AI systems can inadvertently learn to identify protected demographic characteristics- such as patient race or sex- directly from raw medical data in ways that remain imperceptible to human observers (Gichoya et al., 2022; Glocker et al., 2023). This is compounded by "measurement bias," where differential monitoring practices (such as lactate testing frequency) across demographic lines are learned by AI as "normal" indicators of health status (Chen et al., 2023; Seyyed-Kalantari et al., 2021). As a result, marginalized populations- who have historically faced significant barriers to diagnostic testing- are systematically overlooked or misclassified by predictive algorithms (Obermeyer et al., 2019). Second, we must recognize the clinical and ethical consequences: the risk is not just misdiagnosis, but the widening of health disparities- a phenomenon well-documented in oncology literature where researchers have underscored that AI must be coupled with robust strategies to ensure it serves diverse patient populations (Dankwa-Mullan et al., 2025; Istasy et al., 2022; Luchini et al., 2022). Furthermore, as Kelkar et al. (2024) note, the opaque nature of AI decision-making poses a fundamental threat to patient dignity when algorithmic recommendations are implemented without meaningful human oversight. Finally, to address these systemic risks, the field is reaching a consensus on the necessary governance: we must move away from static mathematical metrics of fairness toward a holistic, context-aware integration of social and clinical determinants (Konate et al., 2026; Srivastav et al., 2025; Zhang et al., 2025). However, as McCradden (2020) cautions, we must also recognize the ethical limitations of simplistic "algorithmic fairness" solutions; technical patches cannot fully resolve deep-seated societal injustices and may provide a false sense of security if not paired with comprehensive clinical oversight. As argued by leading voices (Seyyed-Kalantari et al., 2021), we must transition toward active, equity-oriented governance that mandates stratified auditing to prevent these tools from automating and accelerating existing disparities in sepsis care.

4.5 Regulatory, Governance, and Ethical Implications

Frameworks such as TRIPOD-AI (Collins et al., 2024) and DECIDE-AI (Vasey et al., 2022) provide a necessary roadmap for transparency, yet our review suggests that adherence to these standards remains inconsistent. A critical ethical issue is the "black box" nature of deep learning; when a clinician acts on a prediction, they must trust a system that is often non-auditable in real-time (Shashikumar et al., 2025).

A central, often overlooked, governance challenge is the difference between AI models. The literature reveals that they are not static, their performance degrades as clinical environments change. A stark example of this was observed during the COVID-19 pandemic, where models trained on pre-2020 patient populations showed dramatic performance decay when faced with the unique physiological and documentation patterns of COVID-19-positive sepsis patients (Shashikumar et al., 2025; Siddiqui, 2026). Clinical protocols evolve, patient populations shift, and documentation patterns change. Consequently, governance cannot be a one-time approval but must instead function as an ongoing "clinical accountability loop" (Vasey et al., 2022). We argue that regulatory bodies should mandate continuous auditing, requiring hospitals to establish clear thresholds for model decommissioning if the system demonstrates performance drift or demographic inequity.

4.6 The Emergence of the Clinical AI Safety Officer

Clinical performance inevitably degrades due to temporal data drift- as shifts in demographics, protocols, or documentation practices render original training parameters obsolete- and institutional factors like local nursing workflows (Shashikumar et al., 2025; Siddiqui, 2026). We propose the creation of a Clinical AI Safety Officer (CAISO) role by expanding the existing Clinical Safety Officer position. Unlike traditional hospital IT or informatics staff, whose primary focus is often system uptime and EHR functionality, the CAISO's primary mandate is the longitudinal safety, equity, and reliability of deployed AI models. This role acts as the human anchor for the "clinical accountability loop" identified in our synthesis (Vasey et al., 2022). The CAISO ensures that AI performance is evaluated through the lens of patient outcomes and workflow usability, adhering to the safety principles outlined in the DECIDE-AI framework.

The CAISO's first responsibility is active surveillance: monitoring how the model performs daily across diverse ICU patient populations. This includes tracking "precision decay"- the systematic drop in predictive accuracy- and alerting clinical leadership when a model's performance falls outside of predefined, safety-assured thresholds (Shashikumar et al., 2025; Wong et al., 2021). Second, the CAISO manages "Demographic Auditing." They are tasked with conducting regular, stratified performance reviews to ensure that the AI is not

exhibiting disparities based on race, gender, or socioeconomic status, as well as investigating the "invisibility" of measurement biases that may lead to the under-diagnosis of vulnerable patient populations (Chen et al., 2023; Obermeyer et al., 2019).

Finally, the CAISO serves as the ultimate adjudicator in "disagreement protocols." In scenarios where a high-performing clinician disagrees with the AI's recommendation, the CAISO's role is to retrospectively analyze that disagreement, using it as a diagnostic tool to identify where the model's logic deviates from current clinical expertise. Through careful documentation, the CAISO ensures that clinical overrides are not simply dismissed as anomalies, but captured as critical data points to inform future model updates. In this capacity, the CAISO is the ultimate safeguard against the "black box" risks of deep learning (Rossetti et al., 2025), ensuring that human clinical judgment remains the primary driver of critical care while leveraging the predictive power of machine learning. The establishment of this role is not merely an administrative addition, it is a structural necessity for the safe, ethical, and equitable deployment of sepsis EWS, ensuring that clinical AI remains a partner in care rather than a source of hidden clinical risk (Seyyed-Kalantari et al., 2021; Vasey et al., 2022).

4.7 Future Horizons: Multimodal Fusion and Context-Aware Prediction

Looking toward 2030, the next frontier in sepsis detection is the transition from static EHR-based predictive analytics to multimodal fusion. Current models rely on structured physiological data- heart rate, blood pressure, and laboratory tests- yet these metrics often fail to capture the molecular heterogeneity of sepsis. The next generation of AI-driven systems is poised to integrate high-dimensional, unstructured data, including genomics, proteomics, and metabolomics - to drive precision medicine (Gao et al., 2025). By utilizing AI to analyze these complex molecular biomarkers, researchers have demonstrated that AI can identify immune-phenotypes that correlate with specific sepsis trajectories, allowing for a more nuanced understanding of patient responsiveness to resuscitation. However, this expansion introduces new governance hurdles. Multimodal systems are prone to data drift, as the integration of such high-volume, diverse biological data necessitates significantly higher computational costs and more rigorous, standardized data-cleaning protocols (Gao et al., 2025). Furthermore, the inherent complexity of these architectures renders the "clinical accountability loop" and the CAISO's oversight role even more essential, as these models will become increasingly opaque without significant advances in explainable AI (XAI) and intuitive user-interface design. The success of future sepsis AI will ultimately depend on our ability to balance this increased data complexity with the requirement for actionable, explainable, and equity-focused decision support that remains grounded in the bedside needs of the ICU clinician.

4.8 Limitations of This Review

Several limitations must be acknowledged. First, the variety of sepsis definitions and study designs made a formal meta-analysis unfeasible, requiring a qualitative synthesis instead. Second, by limiting our search to English-language, peer-reviewed literature, this analysis likely favors data from large academic medical centers, potentially missing the specific implementation difficulties found in rural or resource-limited hospitals. Third, the rapid pace of AI development means that some of the most cutting-edge, LLM-driven deployments are still in the early pilot or pre-print stage, which may not have been fully represented here. Finally, as noted by various studies (Chen et al., 2023; Collins et al., 2024; Vasey et al., 2022), many publications still fail to report crucial data on demographic breakdowns and clinician override rates, limiting our ability to fully synthesize the equity and usability impacts of these systems.

5. Conclusions

As this review has synthesized, the field of sepsis management is undergoing a transition from the era of retrospective, "dry-lab" model optimization to a more mature, prospective phase focused on clinical utility and real-world safety. The evidence underscores that high-performing algorithms are necessary but insufficient for improving patient outcomes. The persistent challenges of alert fatigue, clinician resistance, and the "implementation paradox" highlight that technical success is inextricably linked to human-factor engineering and organizational culture. Furthermore, the risk of automated demographic disparities remains a critical ethical barrier. Without active, equity-oriented governance and rigorous, stratified performance auditing, there is a substantial risk that these systems will reinforce the very inequities they are designed to mitigate.

Moving forward, the policy implications are clear: the deployment of clinical AI must be governed by a framework of continuous accountability. We must move beyond "one-time" regulatory approvals and toward a lifecycle-based model of oversight that accounts for temporal performance drift, evolving documentation practices, and the shifting physiological patterns of diverse patient populations. Transparency frameworks like TRIPOD-AI and DECIDE-AI provide a vital starting point, yet they must be integrated into the mandatory quality-assurance loops of every health system.

Ultimately, the goal of AI-Driven Early Warning Systems for Sepsis Detection is to transcend the "black-box" model of computation and move toward a paradigm of clinical partnership. If we prioritize clinical judgment, algorithmic transparency and equity, predictive analytics can finally become a tool that supports patient-centered decision-making in the ICU. As the field advances toward 2030, the success of these technologies will be measured not by their AUC metrics, but by their demonstrated capacity to reduce the burden of sepsis and foster a more resilient, inclusive healthcare environment for all patients.

Author Contributions:

Conceptualization: Alicja Kwiatkowska, Jakub Juskowiak

Methodology: Mateusz Lis, Michalina Justyna Kozłowska

Software: Maryan Liasota

Validation: Eliza Zofia Kabat, Michalina Justyna Kozłowska

Formal Analysis: Krzysztof Konopka, Maciej Kamil Majchrzak

Investigation: Alicja Kwiatkowska, Faustyna Magdalena Kołodziej

Resources: Jakub Juskowiak

Data Curation: Zuzanna Julia Komin, Maryan Liasota

Writing - Original Draft: Zuzanna Julia Komin, Eliza Zofia Kabat,

Writing - Review & Editing: Krzysztof Konopka, Maciej Kamil Majchrzak, Faustyna Magdalena Kołodziej

Visualization: Alicja Kwiatkowska, Jakub Juskowiak

Supervision: Mateusz Lis, Zuzanna Julia Komin

Project Administration: Maryan Liasota, Eliza Zofia Kabat

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Declaration of the Use of Generative AI and AI-Assisted Technologies in the Writing Process:: In preparing this work, the authors used Gemini-AI for the purpose of improving language clarity and structural organization. After using this tool, the authors critically reviewed and edited the content and take full responsibility for the final version of the manuscript.

REFERENCES

1. Adams, R., Henry, K. E., Sridharan, A., Soleimani, H., Zhan, A., Rawat, N., Johnson, L., Hager, D. N., Cosgrove, S. E., Markowski, A., Klein, E. Y., Chen, E. S., Saheed, M. O., Henley, M., Miranda, S., Houston, K., Linton, R. C., Ahluwalia, A. R., Wu, A. W., & Saria, S. (2022). Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nature Medicine*, 28(7), 1455–1460. <https://doi.org/10.1038/s41591-022-01894-0>
2. Boussina, A., Shashikumar, S., Amrollahi, F., Pour, H., Hogarth, M., & Nemati, S. (2023). Development & deployment of a real-time healthcare predictive analytics platform. *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–4. <https://doi.org/10.1109/EMBC40787.2023.10340351>
3. Boussina, A., Shashikumar, S. P., Malhotra, A., Owens, R. L., El-Kareh, R., Longhurst, C. A., Quintero, K., Donahue, A., Chan, T. C., Nemati, S., & Wardi, G. (2024). Impact of a deep learning sepsis prediction model on quality of care and survival. *Npj Digital Medicine*, 7(1), 14. <https://doi.org/10.1038/s41746-023-00986-6>

4. Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S., & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, 7(6), 719–742. <https://doi.org/10.1038/s41551-023-01056-8>
5. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., ... Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385. <https://doi.org/10.1136/bmj-2023-078378>
6. Dankwa-Mullan, I., Ndoh, K., Akogo, D., Rocha, H. A. L., & Juaçaba, S. F. (2025). Artificial intelligence and cancer health equity: Bridging the divide or widening the gap. *Current Oncology Reports*, 27(2), 95–111. <https://doi.org/10.1007/s11912-024-01627-1>
7. Gao, Y., Chen, H., Wu, R., & Zhou, Z. (2025). AI-driven multi-omics profiling of sepsis immunity in the digestive system. *Frontiers in Immunology*, Volume 16-2025. <https://doi.org/10.3389/fimmu.2025.1590526>
8. Gichoya, J. W., Banerjee, I., Bhimireddy, A. R., Burns, J. L., Celi, L. A., Chen, L.-C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S.-C., Kuo, P.-C., Lungren, M. P., Palmer, L. J., Price, B. J., Purkayastha, S., Pyrros, A. T., Oakden-Rayner, L., Okechukwu, C., Seyyed-Kalantari, L., ... Zhang, H. (2022). AI recognition of patient race in medical imaging: A modelling study. *The Lancet Digital Health*, 4(6), e406–e414. [https://doi.org/10.1016/S2589-7500\(22\)00063-2](https://doi.org/10.1016/S2589-7500(22)00063-2)
9. Glocker, B., Jones, C., Bernhardt, M., & Winzeck, S. (2023). Algorithmic encoding of protected characteristics in chest X-ray disease detection models. *eBioMedicine*, 89. <https://doi.org/10.1016/j.ebiom.2023.104467>
10. Hooker, S. (2021). Moving beyond “algorithmic bias is a data problem.” *Patterns*, 2(4). <https://doi.org/10.1016/j.patter.2021.100241>
11. Istasy, P., Lee, W. S., Iansavichene, A., Upshur, R., Gyawali, B., Burkell, J., Sadikovic, B., Lazo-Langner, A., & Chin-Yee, B. (2022). The impact of artificial intelligence on health equity in oncology: Scoping review. *J Med Internet Res*, 24(11), e39748. <https://doi.org/10.2196/39748>
12. Kelkar, A. H., Hantel, A., Koranteng, E., Cutler, C. S., Hammer, M. J., & Abel, G. A. (2024). Digital health to Patient-Facing artificial intelligence: Ethical implications and threats to dignity for patients with cancer. *JCO Oncology Practice*, 20(3), 314–317. <https://doi.org/10.1200/OP.23.00412>
13. Konate, S., Gallifant, J., Senteio, C., Celi, L. A., & Seyyed-Kalantari, L. (2026). Rethinking fairness in AI to improve current practice in oncology. *Trends in Cancer*, 12(1), 1–4. <https://doi.org/10.1016/j.trecan.2025.11.005>
14. Lee, R. Y., Cato, K. D., Dykes, P. C., Lowenthal, G., Tuteja, S., & Rossetti, S. C. (2025). Understanding clinicians’ usage patterns of the concern early warning system: Insights from a Multi-Site pragmatic cluster randomized controlled trial. In M. S. Househ, Z. U. A. Tariq, M. Al-Zubaidi, U. Shah, & E. Huesing (Eds.), *Studies in Health Technology and Informatics*. IOS Press. <https://doi.org/10.3233/SHTI251040>
15. Lee, R. Y., Cato, K. D., Dykes, P. C., Lowenthal, G., Withall, J. B., Cho, S., Jia, H., & Rossetti, S. C. (2024). Influence of the concern early warning system on unanticipated icu transfers, In-Hospital mortality, and length of stay: Results from a multi-site pragmatic randomized controlled clinical trial. *AMIA ... Annual Symposium Proceedings. AMIA Symposium, 2024*, 655–663.
16. Luchini, C., Pea, A., & Scarpa, A. (2022). Artificial intelligence in oncology: Current applications and future perspectives. *British Journal of Cancer*, 126(1), 4–9. <https://doi.org/10.1038/s41416-021-01633-1>
17. McCradden, M. D., Joshi, S., Mazwi, M., & Anderson, J. A. (2020). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health*, 2(5), e221–e223. [https://doi.org/10.1016/S2589-7500\(20\)30065-0](https://doi.org/10.1016/S2589-7500(20)30065-0)
18. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>

19. Rossetti, S. C., Dykes, P. C., Knaplund, C., Cho, S., Withall, J., Lowenthal, G., Albers, D., Lee, R. Y., Jia, H., Bakken, S., Kang, M.-J., Chang, F. Y., Zhou, L., Bates, D. W., Daramola, T., Liu, F., Schwartz-Dillard, J., Tran, M., Bokhari, S. M. A., ... Cato, K. D. (2025). Real-time surveillance system for patient deterioration: A pragmatic cluster-randomized controlled trial. *Nature Medicine*, 31(6), 1895–1902. <https://doi.org/10.1038/s41591-025-03609-7>
20. Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, 27(12), 2176–2182. <https://doi.org/10.1038/s41591-021-01595-0>
21. Shashikumar, S. P., Mohammadi, S., Krishnamoorthy, R., Patel, A., Wardi, G., Ahn, J. C., Singh, K., Aronoff-Spencer, E., & Nemati, S. (2025). Development and prospective implementation of a large language model based system for early sepsis prediction. *Npj Digital Medicine*, 8(1), 290. <https://doi.org/10.1038/s41746-025-01689-w>
22. Shashikumar, S. P., Wardi, G., Malhotra, A., & Nemati, S. (2021). Artificial intelligence sepsis prediction algorithm learns to say “I don’t know.” *Npj Digital Medicine*, 4(1), 134. <https://doi.org/10.1038/s41746-021-00504-6>
23. Siddiqui, M. M. (2026). Advanced AI-Based early warning framework for icu patient monitoring using temporal clinical data. *Biomedical & Pharmacology Journal*, 1(19), 70. <https://doi.org/10.13005/bpj/3338>
24. Srivastav, A. K., Singh, A., Singh, S., Rivers, B., Lillard, J. W., & Singh, R. (2025). Revolutionizing oncology through AI: Addressing cancer disparities by improving screening, treatment, and survival outcomes via integration of social determinants of health. *Cancers*, 17(17), 2866. <https://doi.org/10.3390/cancers17172866>
25. Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., ... McCulloch, P. (2022). Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*, 377, e070904. <https://doi.org/10.1136/bmj-2022-070904>
26. Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestruie, J., Phillips, M., Konye, J., Penzoza, C., Ghous, M., & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8), 1065. <https://doi.org/10.1001/jamainternmed.2021.2626>
27. Zhang, Y., Peng, D., & Zhang, A. (2025). Artificial intelligence in cancer diagnostics: Governance for equity in low- and middle-income countries. *African Journal of Reproductive Health*, 29(12s), 93–103. <https://doi.org/10.29063/ajrh2025/v29i12s.10>