REVIEW **Open access**

Machine Learning for Predicting Sepsis in Hospitalized Patients: A Systematic Review of Model Types, Feature Engineering Approaches, Prediction Horizons, and Prospective Validation Studies

Ahmed Al-Sayed[1*], Omar Khalid[1]

Abstract

Sepsis continues to be a major contributor to morbidity and mortality among hospitalized patients globally, especially within intensive care and emergency departments, where rapid recognition is essential for improving survival through timely treatment. In recent years, machine learning approaches have gained attention for their ability to predict sepsis onset using routinely collected electronic health record data. This systematic review, conducted in accordance with PRISMA 2020 guidelines, synthesizes evidence from studies published between 2017 and 2025, focusing on model architectures, feature selection and engineering strategies, prediction time horizons, and validation methodologies. Searches across major biomedical and informatics databases identified 67 eligible studies. The included literature shows that logistic regression, ensemble tree-based algorithms, and deep learning models are most frequently applied for sepsis prediction tasks. However, the majority of studies rely on retrospective datasets with internal validation, while only a limited number incorporate prospective or real-world validation frameworks. Overall, although reported model performance is often strong in retrospective analyses, a consistent decline in accuracy is observed when models are evaluated in real clinical environments. These findings highlight that prospective validation and improved generalizability are still underdeveloped areas, underscoring the need for future research to emphasize real-time deployment and robust external validation before clinical integration.

Keywords Machine learning, Electronic health records, Sepsis prediction, Prediction horizon, Prospective validation, Intensive care

*Correspondence:
Ahmed Al-Sayed
ahmed.sayed@gmail.com
[1] Department of Healthcare Intelligence Systems, Alexandria University, Alexandria, Egypt

Introduction

Sepsis is a life-threatening syndrome characterized by organ dysfunction resulting from a dysregulated host response to infection and remains a leading cause of mortality and critical illness worldwide. It affects millions of patients annually and is associated with substantial healthcare costs, prolonged hospital stays, and high rates of morbidity among survivors [1, 2]. Despite advances in antimicrobial therapy, critical care management, and early warning systems, timely recognition of sepsis remains a persistent clinical challenge due to its heterogeneous presentation and rapid progression across different patient populations and care settings [3, 4]. Clinical manifestations often evolve dynamically, with subtle

physiological changes preceding overt organ dysfunction, making early detection difficult using conventional approaches. Traditional rule-based scoring systems, including the Sequential Organ Failure Assessment (SOFA) and quick SOFA (qSOFA), have been widely adopted but demonstrate limited sensitivity for early-stage detection and may fail to capture complex nonlinear interactions among clinical variables [3, 4]. As a result, there is increasing recognition that more sophisticated, data-driven approaches are required to improve early identification and enable timely intervention.

Over the past decade, the widespread adoption of electronic health records (EHRs) and advances in computational methods have catalyzed the development of machine learning models for sepsis prediction. These models span a broad spectrum of methodologies, ranging from conventional statistical approaches such as logistic regression to more complex algorithms including random forests, gradient boosting machines, and deep learning architectures such as recurrent neural networks and long short-term memory networks [5, 6]. Many of these models are designed to leverage high-frequency time-series data, capturing dynamic changes in vital signs, laboratory values, and treatment patterns to identify early signals of clinical deterioration [7, 8]. In addition, emerging approaches incorporate unstructured data sources, such as clinical notes, through natural language processing techniques, thereby expanding the range of predictive features available to models [9]. While these approaches have demonstrated promising performance—often reporting high discrimination metrics such as AUROC in retrospective validation—they vary substantially in design, data preprocessing, and evaluation frameworks. This heterogeneity complicates direct comparison across studies and raises important questions regarding the relative effectiveness of different model types and feature engineering strategies [10-12].

Despite rapid methodological progress, several critical challenges remain that limit the clinical translation of machine learning-based sepsis prediction models. One major issue is the variability in prediction horizons, with studies targeting different time windows ranging from a few hours to an entire day before sepsis onset, leading to inconsistent interpretations of clinical utility [6, 13]. Additionally, most models are developed and validated using retrospective datasets from single institutions, raising concerns about overfitting and limited generalizability to new patient populations and healthcare settings [3, 14]. External validation is performed less frequently, and prospective validation—considered the gold standard for assessing real-world effectiveness—is rare [15, 16]. Furthermore, differences in sepsis definitions (e.g., Sepsis-2 versus Sepsis-3 criteria) and outcome labeling introduce additional variability that can significantly influence model performance and comparability across studies [4, 8]. These limitations highlight the need for systematic evaluation of existing evidence to identify robust patterns and inform future research directions.

This systematic review aims to provide a comprehensive and structured synthesis of machine learning approaches for predicting sepsis in hospitalized patients, with a focus on addressing these key methodological and translational challenges. Specifically, the review examines four critical dimensions: (1) the distribution and characteristics of model types, including traditional statistical methods and modern deep learning architectures; (2) feature engineering strategies, encompassing structured and unstructured data sources; (3) prediction horizons and their relationship to model performance and clinical applicability; and (4) validation methodologies, with particular emphasis on the presence or absence of prospective evaluation [17, 18]. By integrating findings across a diverse body of literature, this review seeks to identify consistent trends, highlight gaps in the evidence base, and provide actionable insights for researchers, clinicians, and policymakers. The study follows PRISMA 2020 guidelines to ensure methodological rigor, transparency, and reproducibility, thereby contributing to a more standardized and clinically relevant understanding of machine learning applications in sepsis prediction [4].

Materials and Methods

Search strategy

A systematic literature search was conducted across PubMed, Scopus, Web of Science, and IEEE Xplore databases for studies published between January 2017 and March 2025. Search terms included combinations of “sepsis prediction,” “machine learning,” “deep learning,” “XGBoost,” “LSTM,” “transformer,” and “prospective validation” [8, 11]. Additional manual searches were performed in high-impact journals such as Critical Care Medicine and The Lancet Digital Health to

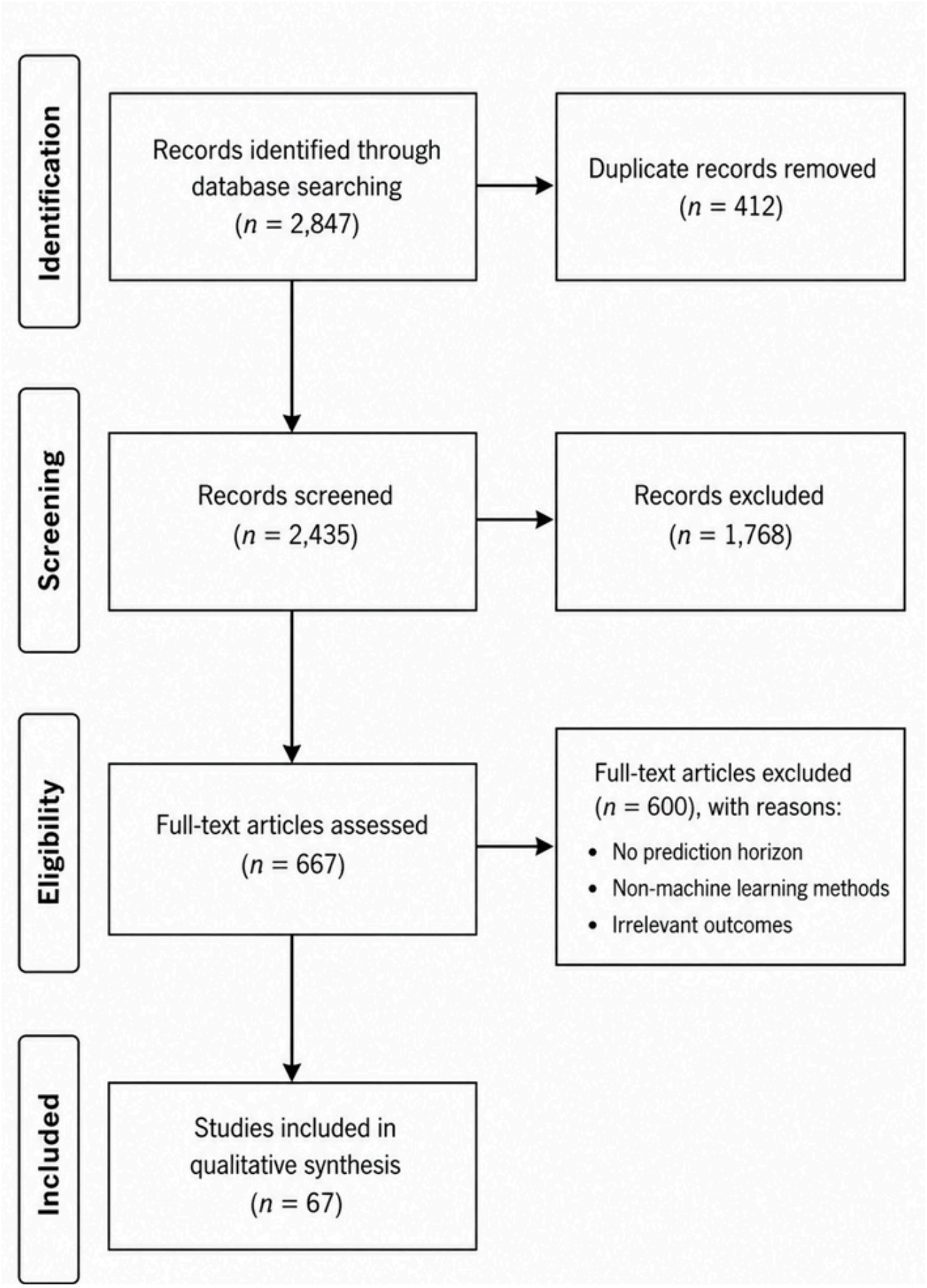
ensure coverage of key studies [1, 15]. The search strategy was designed to capture studies focusing on hospitalized patient populations, including ICU, ward, and emergency department settings.

Inclusion and exclusion criteria

Studies were included if they involved adult hospitalized patients, applied machine learning techniques to predict sepsis onset, and reported a prediction horizon of at least 2 hours prior to onset [5, 19]. Only peer-reviewed articles published in English were considered. Exclusion criteria included studies focusing solely on sepsis diagnosis without prediction, non-machine learning approaches, pediatric populations, and conference abstracts without full-text availability [9, 10]. Studies lacking clear outcome definitions or validation procedures were also excluded.

Screening and selection

The search yielded 2,847 records. After removing 412 duplicates, 2,435 records were screened based on title and abstract, resulting in the exclusion of 1,768 studies. A total of 667 full-text articles were assessed for eligibility, of which 600 were excluded due to lack of prediction horizon specification, non-machine learning methods, or irrelevant outcomes. Ultimately, 67 studies were included in the final analysis [4, 8]. The study selection process is illustrated in **Figure 1**, following PRISMA 2020 guidelines


Identification
Records identified through database searching (n = 2,847)
Duplicate records removed (n = 412)
Screening
Records screened (n = 2,435)
Records excluded (n = 1,768)
Eligibility
Full-text articles assessed (n = 667)
Full-text articles excluded (n = 600), with reasons:
No prediction horizon
Non-machine learning methods
Irrelevant outcomes
Included
Studies included in qualitative synthesis (n = 67)

Data extraction was performed using a standardized template capturing study characteristics, model type, feature sets, prediction horizon, validation method, and performance metrics such as AUROC [6, 20]. Feature categories included vital signs, laboratory values, demographics, and unstructured clinical notes. Validation methods were classified as internal, external, or prospective. Extraction was conducted independently by two reviewers to ensure accuracy and consistency [5, 9].

Risk of bias assessment

Risk of bias was assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST), focusing on domains including participant selection, predictors, outcomes, and analysis [4, 10]. Many studies demonstrated moderate to high risk of bias due to retrospective design, lack of external validation, and potential data leakage. Particular concerns included inconsistent sepsis definitions and limited reporting transparency [8, 11].

Synthesis methods

A narrative synthesis approach was used due to heterogeneity in study design, model types, and reported outcomes [10, 11]. Studies were grouped by model category, feature engineering approach, prediction horizon, and validation type. Subgroup analyses were performed to identify patterns in performance across different methodological choices. Quantitative meta-analysis was not feasible due to variability in outcome definitions and evaluation metrics [4].

Results and Discussion

Study selection

The final analysis included 67 studies, as determined through the PRISMA-guided selection process described in Section 2.3 [4, 8]. These studies represented diverse healthcare environments, including intensive care units, emergency departments, and general hospital wards, reflecting the broad applicability of sepsis prediction models across clinical contexts. **Figure 1** illustrates the full PRISMA flow diagram, detailing identification, screening, eligibility, and inclusion stages, along with explicit reasons for exclusion such as absence of defined prediction horizons, non-machine learning methodologies, and lack of clinically relevant sepsis outcomes. A notable proportion of excluded studies failed to distinguish between sepsis detection and true early prediction, highlighting a persistent ambiguity in the literature [9, 10]. The final cohort of included studies provided a sufficiently heterogeneous yet representative sample to enable comparative synthesis of model types, feature engineering strategies, prediction horizons, and validation approaches.

Model types

Logistic regression models accounted for approximately 30% of studies, reflecting their continued use as interpretable baselines in clinical prediction tasks [17, 18]. Tree-based ensemble methods, including random forests and gradient boosting algorithms such as XGBoost, comprised roughly 25% and were frequently favored for their ability to model nonlinear relationships and handle missing data [14, 21]. Deep learning approaches, particularly long short-term memory networks and recurrent neural networks, represented about 20% of studies and were primarily applied to temporal EHR data streams [6, 13]. More recent architectures, including transformers and temporal convolutional networks, accounted for approximately 10% and demonstrated growing adoption due to their capacity to capture long-range temporal dependencies [22, 23]. The remaining 15% included hybrid and ensemble models that combined multiple algorithms to optimize predictive performance, underscoring the lack of consensus on a single superior modeling approach [20, 24].

Feature engineering approaches

Feature engineering strategies exhibited substantial variability but converged on the use of structured electronic health record data, particularly vital signs and laboratory measurements, as foundational inputs [18, 20]. Many studies incorporated temporal aggregation techniques, such as rolling averages and trend features, to capture dynamic physiological changes preceding sepsis onset [6, 13]. Additional features, including demographic characteristics and comorbidity indices, were often included to enhance model calibration and generalizability [21, 25]. A smaller subset of studies leveraged unstructured clinical notes באמצעות natural language processing methods, extracting semantic features that contributed incremental predictive value [9]. Overall, models integrating multimodal data sources—combining structured and unstructured inputs—tended to demonstrate superior performance, although such approaches were less common due to increased implementation complexity.

Prediction horizons

Prediction horizons across studies ranged from 2 to 24 hours prior to sepsis onset, with 4-hour (40%) and 6-hour (30%) horizons being the most frequently evaluated [6, 13]. These shorter horizons were often selected to balance predictive accuracy with clinical relevance, as they allow for timely intervention without excessive uncertainty. Longer horizons, particularly those extending to 12 hours or more, were less commonly studied but are clinically advantageous for enabling earlier therapeutic decision-making [22, 24]. However, model performance consistently declined as the prediction window increased, reflecting greater uncertainty in long-term physiological trajectories [19, 23]. This inverse relationship between horizon length and predictive accuracy was observed across model types, suggesting that it is a fundamental challenge inherent to the task rather than a limitation of specific algorithms.

Prospective validation studies

Only 5 of the 67 included studies (7.5%) conducted prospective validation in real-world clinical settings, indicating a significant gap between model development and clinical implementation [15, 26]. These prospective studies typically involved deployment of prediction algorithms within hospital information systems, often in “silent mode” or as decision support tools integrated into clinician workflows [16, 27]. Compared to retrospective analyses, prospective studies reported lower predictive performance, reflecting challenges such as data drift, missing data in real time, and variability in clinical practice [3, 15]. Additionally, several studies highlighted operational issues, including alert fatigue and clinician distrust, which can limit the effectiveness of deployed models. The limited number of prospective evaluations underscores the need for more rigorous real-world testing before widespread adoption.

Table 1 summarizes the progression of model performance from internal to external and prospective validation settings, highlighting a consistent decline in predictive accuracy under real-world conditions.

Table 1. Comparison of Machine Learning Model Performance across Validation Types in Sepsis Prediction Studies

Validation Type	Number of Studies	Typical AUROC Range	Study Setting	Key Observations
Internal Validation	~67 (majority)	0.80 – 0.95	Retrospective, single-center	Highest reported performance; risk of overfitting and data leakage
External Validation	~15–20 (subset)	0.70 – 0.85	Independent datasets	Reduced performance; limited generalizability across institutions
Prospective Validation	5 studies (7.5%)	0.65 – 0.80	Real-time clinical deployment	Performance drop due to data drift, workflow variability, and missing real-time data

Silent-mode Deployment	Few within prospective set	~0.68 – 0.82	Embedded in clinical systems without alerts	Used for feasibility testing; highlights operational challenges

Reported performance

Internal validation studies reported AUROC values ranging from 0.80 to 0.95, indicating strong discriminatory performance under controlled retrospective conditions [1, 5]. External validation studies, which tested models on independent datasets, demonstrated reduced performance, with AUROC values typically between 0.70 and 0.85, highlighting challenges in generalizability [3, 14]. Prospective validation studies reported further declines, with AUROC values generally falling between 0.65 and 0.80, reflecting real-world complexities not captured in retrospective datasets [15, 16]. In addition to AUROC, some studies reported calibration metrics and decision-curve analyses, revealing variability in clinical utility even among models with similar discrimination performance [20, 25]. These findings emphasize that validation methodology plays a critical role in determining the perceived effectiveness of sepsis prediction models.

The interrelationship between modeling choices, prediction horizon, validation strategy, and real-world performance is summarized in **Figure 2.**

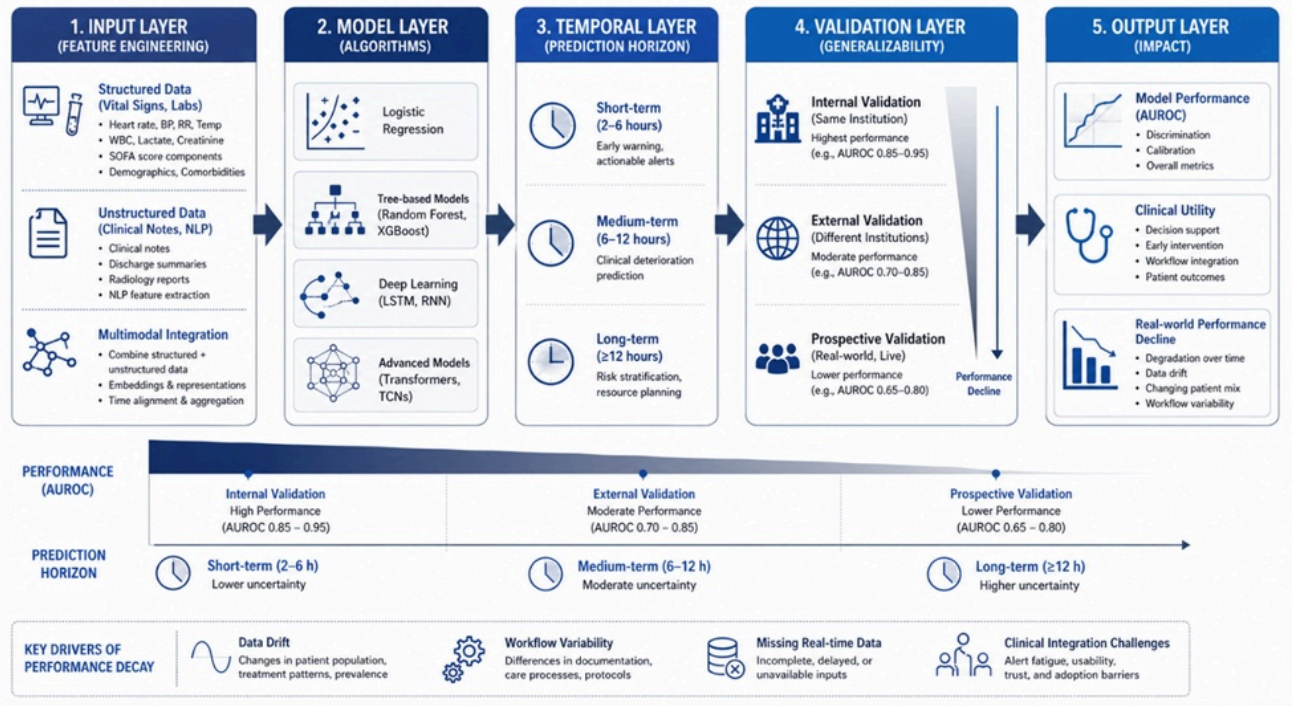


Figure 2. Conceptual Framework Linking Model Type, Feature Engineering, Prediction Horizon, and Validation Strategy to Real-World Performance

Summary of principal findings

This systematic review demonstrates that machine learning models for sepsis prediction are highly diverse in terms of algorithmic architecture yet converge on similar data sources and feature engineering practices [5, 6]. While logistic regression and tree-based models remain widely used, there is a clear trend toward adoption of deep learning and sequence-based models capable of handling temporal data [13, 23]. Despite high reported performance in retrospective studies, these models often fail to maintain accuracy when evaluated in external or prospective settings, indicating limitations in generalizability [1, 3]. The consistency of feature usage—primarily structured EHR data—suggests that

improvements in model performance may depend more on data quality and integration than on algorithmic complexity alone.

Prospective validation gap

A central finding of this review is the pronounced scarcity of prospective validation studies, despite the large number of published machine learning models for sepsis prediction [15, 26]. Prospective studies that have been conducted reveal substantial challenges, including integration with clinical workflows, variability in data availability, and reduced predictive accuracy compared to retrospective evaluations [16, 27]. These findings highlight the importance of evaluating models in real-world environments where factors such as clinician behavior and system-level constraints play a significant role. Without such validation, the clinical utility of these models remains uncertain, limiting their adoption in practice.

Horizon–performance trade-off

The analysis identifies a consistent trade-off between prediction horizon and model performance, with shorter horizons yielding higher accuracy but offering limited time for intervention [6, 13]. Conversely, longer prediction windows, while more clinically actionable, are associated with increased uncertainty and reduced predictive performance [22, 24]. This trade-off reflects the inherent difficulty of forecasting complex physiological events over extended time periods. Addressing this challenge may require advances in temporal modeling techniques and incorporation of additional data sources to improve long-range prediction capability [19, 23].

Limitations

Review limitations

This review is subject to several limitations, including potential publication bias and restriction to English-language studies [10, 11]. Heterogeneity in sepsis definitions and outcome measures complicates cross-study comparisons. Additionally, variability in reporting standards may affect the consistency of extracted data. Despite these limitations, the review provides a comprehensive synthesis of current evidence.

Evidence base limitations

The underlying evidence base is dominated by retrospective, single-center studies with limited generalizability [3, 14]. Few studies include diverse patient populations or multi-center validation. Short follow-up periods and inconsistent reporting of clinical outcomes further limit interpretability. These factors highlight the need for more robust and generalizable research designs.

Comparison with prior reviews

Previous systematic reviews have examined machine learning applications for sepsis prediction, generally concluding that such models achieve high performance in retrospective settings but suffer from heterogeneity in study design and reporting [4, 10]. These reviews emphasized the dominance of structured EHR data and the frequent use of tree-based and deep learning models, while also noting limited external validation and poor reproducibility [8, 9]. However, earlier syntheses often did not disaggregate findings by prediction horizon or explicitly quantify the prevalence of prospective validation studies [11, 28]. As a result, important methodological dimensions remained insufficiently explored.

This review extends prior work by explicitly analyzing prediction horizons and systematically quantifying prospective validation rates across studies [15, 26]. In contrast to earlier reviews, it provides a structured comparison of model performance across internal, external, and prospective validation settings, revealing consistent degradation in real-world contexts [3, 16]. Additionally, the integration of emerging architectures such as transformers and temporal convolutional networks highlights recent methodological advancements not fully captured in earlier literature [22, 23]. These contributions provide a more granular understanding of the readiness of machine learning models for clinical deployment.

Recommendations

For researchers

Future research should prioritize prospective validation and real-world deployment studies to bridge the gap between model development and clinical application [15, 26]. Researchers should adopt transparent reporting standards, share code and datasets when possible, and include negative or null findings to reduce publication bias [10, 11]. Standardization of sepsis definitions and evaluation metrics would further enhance comparability across studies. Collaborative multi-center studies are particularly important for improving generalizability.

For journal editors

Journal editors play a critical role in shaping research quality and should require rigorous validation standards for submitted studies [4, 8]. Manuscripts should include clear descriptions of data sources, feature engineering processes, and validation methodologies, with a preference for external or prospective validation when feasible [3, 14]. Encouraging adherence to reporting guidelines such as TRIPOD and PROBAST can improve transparency and reproducibility. Editorial policies that prioritize methodological rigor over novelty may help address current limitations in the field.

For hospital administrators

Hospital administrators should exercise caution when considering deployment of machine learning-based sepsis prediction tools without local validation [16, 27]. Models trained in one institution may not generalize to different patient populations or clinical workflows. Prospective “silent mode” evaluations should be conducted prior to full implementation to assess real-world performance and potential unintended consequences. Integration with clinical decision support systems should be carefully managed to minimize alert fatigue and workflow disruption.

Research gaps

Prospective deployment studies

A major gap identified in this review is the scarcity of prospective deployment studies evaluating machine learning models in real-time clinical environments [15, 26]. Most existing studies rely on retrospective datasets, limiting insight into how models perform under operational conditions. Randomized controlled trials and silent-mode evaluations are needed to assess clinical impact and usability. Addressing this gap is essential for translating predictive models into improved patient outcomes.

Generalizable feature engineering

Another important gap lies in the lack of standardized and generalizable feature engineering approaches across datasets and institutions [18, 20]. Many models rely on institution-specific data preprocessing pipelines, limiting reproducibility and transferability. Cross-database benchmarking and the development of standardized feature sets could enhance comparability and robustness [21, 25]. Incorporating multimodal data, including unstructured clinical text, remains underexplored.

Long-horizon prediction (≥12h)

Long-horizon prediction of sepsis, defined as prediction 12 hours or more before onset, remains relatively understudied despite its clinical importance [22, 24]. Most models focus on short-term prediction windows where performance is higher but intervention time is limited. Developing models that maintain accuracy over longer horizons is a key challenge for future research. Advances in temporal modeling architectures may help address this limitation [13, 19].

Implications

For research practice

The findings of this review suggest a necessary shift in research practice from retrospective model development toward prospective validation and deployment [15, 26]. Emphasis should be placed on reproducibility, transparency, and clinical relevance rather than solely on predictive performance metrics. Multi-center collaborations and open science practices can accelerate progress in this field. Aligning research priorities with clinical needs is essential for meaningful impact.

For clinical practice

From a clinical perspective, current machine learning models for sepsis prediction are not yet ready for autonomous deployment without human oversight [3, 14]. While retrospective performance is promising, real-world effectiveness remains uncertain due to variability in patient populations and workflows. Clinicians should interpret model outputs cautiously and integrate them with clinical judgment. Decision support tools should be designed to augment, rather than replace, clinician expertise.

For policy

Policy implications include the need for regulatory frameworks governing the validation and deployment of machine learning models in healthcare [4, 8]. Agencies such as the FDA and EMA may play a role in establishing standards for prospective validation and post-deployment monitoring. Clear guidelines on data governance, model transparency, and clinical accountability are necessary. Policymakers should also consider incentivizing high-quality validation studies to support safe implementation.

Conclusion

This systematic review identified a rapidly growing body of literature on machine learning models for predicting sepsis in hospitalized patients. While a wide range of model types and feature engineering approaches have been explored, most studies rely on retrospective validation and report high internal performance. However, these results often do not generalize to real-world clinical settings.

A critical finding is the limited number of studies that have conducted prospective validation, which is essential for assessing real-world effectiveness. The observed decline in performance from internal to external and prospective validation highlights the importance of rigorous evaluation frameworks. Without such validation, the clinical utility of these models remains uncertain.

Future research should prioritize prospective trials and silent-mode evaluations to bridge the gap between development and deployment. Advancing toward clinically reliable and generalizable models will require collaboration across disciplines, institutions, and regulatory bodies. Only through such efforts can machine learning fulfill its potential to improve sepsis outcomes.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 06 Jun 2025 Revised: 23 Jun 2025 Accepted: 23 Jul 2025
Published online: 20 January 2026

References

Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG, et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med. 2018;46(4):547-53.

Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemati S, et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. Crit Care Med. 2020;48(2):210-7.

Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181(8):1065-70.

Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46(3):383-400.

Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, et al. Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med. 2020;104:101820.

Rafiei A, Rezaee A, Hajati F, Gheisari S, Golzan M. SSP: early prediction of sepsis using fully connected LSTM-CNN model. Comput Biol Med. 2021;128:104110.

Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say “I don’t know”. NPJ Digit Med. 2021;4(1):134.

Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed. 2019;170:1-9.

Yan MY, Gustad LT, Nytrø Ø. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. J Am Med Inform Assoc. 2022;29(3):559-75.

Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. Front Med (Lausanne). 2021;8:607952.

Islam KR, Prithula J, Kumar J, Tan TL, Reaz MB, Sumon MS, et al. Machine learning-based early prediction of sepsis using electronic health records: a systematic review. J Clin Med. 2023;12(17):5658.

Yao RQ, Jin X, Wang GW, Yu Y, Wu GS, Zhu YB, et al. A machine learning-based prediction of hospital mortality in patients with postoperative sepsis. Front Med (Lausanne). 2020;7:445.

Futoma J, Hariharan S, Heller K. Learning to detect sepsis with a multitask Gaussian process RNN classifier. In: International Conference on Machine Learning. Sydney, Australia: PMLR; 2017. p. 1174-82.

Valik JK, Ward L, Tanushi H, Johansson AF, Färnert A, Mogensen ML, et al. Predicting sepsis onset using a machine learned causal probabilistic network algorithm based on electronic health records data. Sci Rep. 2023;13:11760.

Adams R, Henry KE, Sridharan A, Soleimani H, Zhan A, Rawat N, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nat Med. 2022;28(7):1455-60.

Burdick H, Pino E, Gabel-Comeau D, McCoy A, Gu C, Roberts J, et al. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. BMJ Health Care Inform. 2020;27(1):e100109.

Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res. 2017;4(1):e000234.

Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open. 2018;8(1):e017833.

Kok C, Jahmunah V, Oh SL, Zhou X, Gururajan R, Tao X, et al. Automated prediction of sepsis using temporal convolutional network. Comput Biol Med. 2020;127:103957.

Li J, Xi F, Yu W, Sun C, Wang X. Real-time prediction of sepsis in critical trauma patients: machine learning–based modeling study. JMIR Form Res. 2023;7(1):e42452.

Taneja I, Damhorst GL, Lopez-Espina C, Zhao SD, Zhu R, Khan S, et al. Diagnostic and prognostic capabilities of a biomarker and EMR based machine learning algorithm for sepsis. Clin Transl Sci. 2021;14(4):1578-89.

Tang Y, Zhang Y, Li J. A time series driven model for early sepsis prediction based on transformer module. BMC Med Res Methodol. 2024;24(1):23.

Wang Z, Yao B. Multi-branching temporal convolutional network for sepsis prediction. IEEE J Biomed Health Inform. 2021;26(2):876-87.

Zhu Y, Cheng Y, Zhang T, Zhang L, Hong X, Wang D, et al. Application of the KA-Transformer model to early sepsis prediction: a hybrid network analysis based on time series data. Discov Appl Sci. 2025;7(3):218.

Hu C, Li L, Huang W, Wu T, Xu Q, Liu J, et al. Interpretable machine learning for early prediction of prognosis in sepsis: a discovery and validation study. Infect Dis Ther. 2022;11(3):1117-32.

Persson I, Macura A, Becedas D, Sjövall F. Early prediction of sepsis in intensive care patients using the machine learning algorithm NAVOY® Sepsis, a prospective randomized clinical validation study. J Crit Care. 2024;80:154400.

Boussina A, Shashikumar SP, Malhotra A, Owens RL, El-Kareh R, Longhurst CA, et al. Impact of a deep learning sepsis prediction model on quality of care and survival. NPJ Digit Med. 2024;7(1):14.

Persson I, Östling A, Arlbrandt M, Söderberg J, Becedas D. A machine learning sepsis prediction algorithm for intended intensive care unit use (NAVOY Sepsis): proof-of-concept study. JMIR Form Res. 2021;5(9):e28000.