Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 1, 2026

UNCERTAINTY-AWARE AI FOR RELIABLE AND CLINICALLY SAFE SEPSIS PREDICTION: CURRENT EVIDENCE AND RESEARCH GAPS

Shawal Khan[1], Didar Hussain[1,2], Hashim Ali[2*], Abdullah Sadiq[1], Nauman Khan[1], Danyal Khan[1], Fayaz Khan[1], Shahzad Alam[2]

[1]*Department of Computer Science, Bacha Khan University Charsadda*
[2]*Department of Computer Science, Abdul Wali Khan University Mardan*
**Corresponding Author Email: hashimali@awkum.edu.pk*

DOI: https://doi.org/10.5281/zenodo.23730191

Keywords:
Sepsis prediction, artificial intelligence, machine learning, uncertainty quantification, calibration, external validation, clinical safety.

Article History
Received: 12 Dec, 2025
Accepted: 16 Jan, 2026
Published: 29 Jan, 2026

Copyright @Author
Corresponding Author
***Hashim Ali**
hashimali@awkum.edu.pk

Abstract

Sepsis is a potentially life-threatening clinical state which needs to be detected as early as possible to ensure timely intervention and improved outcomes. The use of artificial intelligence (AI) and machine learning (ML) for the early identification of sepsis, based on routinely collected clinical data, has shown promise. Clinically safe or trustworthy decision-making is not a given with predictive performance. Despite the remarkable progress of AI sepsis predictive algorithms, they still have several drawbacks, including poor calibration, poor generalizability, uncertainty in the algorithm's prediction and absence of external validation. This work offers a thorough assessment of the various AI approaches for sepsis prediction, including their performance, uncertainty quantification, calibration, and clinical safety. The review's compass are three research questions regarding the AI models used, the performance of the models used for predictions and calibrations, and the main research gaps and future research directions. The suggested review strategy includes a planned literature search procedure, eligibility criteria, a method to select literature for review and a qualitative synthesis of data. From the practical clinical applications, the first literature review highlights the importance of independent validation, model robustness, uncertainty-aware prediction and assessment. Additionally, it shows that clinical reliability can be attained without relying just on conventional performance metrics. The most important areas for future study are the use of calibrated prediction probabilities, explicit statements of uncertainty, prospective and multicenter validation, and clinically important assessment of safety. The purpose of this review is to inspire the development of more trustworthy AI systems to be used in clinical practice and for early sepsis detection in a responsible way.

1. Introduction

Sepsis is a serious and life-threatening clinical condition which requires prompt diagnosis and management. Sepsis is a complex and rapidly evolving condition that poses a major challenge in clinical practice when it comes to early diagnosis. This need for methods to identify high-risk patients before the onset of serious sepsis is recognised is because traditional clinical methods including known diagnostic criteria and scoring mechanisms are not always sufficiently sensitive to predict the onset of sepsis [1].

In recent years machine learning (ML) and artificial intelligence (AI) have emerged as viable tools to facilitate early sepsis prediction.

Large and complicated clinical datasets, such as vital signs, lab results, and other regularly gathered patient data, can be analyzed using machine learning algorithms to find trends linked to the onset of sepsis. In a comprehensive review and meta-analysis, Fleuren et al. [2] showed the growing role of machine learning in the prediction of sepsis and highlighted the potential of machine learning to improve decision-making for initial diagnostic evaluation of sepsis. Similarly, in the critical care setting, Moor et al. described the use of machine learning in predicting early sepsis, and highlighted the increasing interest in data-driven approaches to early sepsis detection [3].

For the purpose of early sepsis identification, a number of studies have created and assessed machine learning models. Nemati et al. [5] developed an interpretable machine learning approach to predicting sepsis in the intensive care unit (ICU) setting and Calvert et al. devised a computational technique to detect the onset of sepsis from clinical data [4]. There have been other studies on the use of deep learning methods and the potential impact on clinical outcomes. In fact, beyond merely creating the model, Boussina et al. [6] evaluated the clinical effect of a deep learning sepsis prediction model, to understand the impact on survival and quality of treatment.

Moreover, the progress of sepsis prediction systems to clinical use has been further demonstrated by FDA-approved AI/ML technologies [7].

Recent advances in developing reliable sepsis prediction systems using artificial intelligence (AI) remain challenging. Research has recorded significant differences in the datasets, the definition of sepsis, the prediction methods and the evaluation techniques [8]. The performance of a prediction model may vary when applied to another clinical setting, or to another population of patients. Therefore, external validation studies are of particular importance. For instance, studies on the use of proprietary sepsis prediction algorithms have observed that there is a need to evaluate the algorithm performance in different hospital settings to not rely solely on a single setting.

In the case of proprietary sepsis prediction algorithms, studies that evaluate the performance of these algorithms have recognized the need to test the model in independent hospital settings [9, 10] in addition to results from the development-dataset. Recent multicenter prospective validation [11] has again emphasized the importance of evaluating prediction models in the actual clinical context.

There is also a lot of uncertainty over AI predictions. Whenever a model is used in clinical decision making, it should be able to tell you when it might not be accurate in predicting the situation.

Shashikumar et al. [12] emphasized the importance of uncertainty-aware prediction in clinical AI systems with their sepsis prediction algorithm that was designed to articulate uncertainty. This is particularly important in healthcare settings, where hidden uncertainty can taint trust in the machine learning system and negatively impact informed healthcare choices. Research on robustness and uncertainty has brought to attention the importance of measuring model reliability beyond the conventional metrics of model performance.

In addition, prediction models can be affected by the different definitions of sepsis. Diagnostic criteria, when evaluating and comparing AI-based systems, is an important consideration. In recent years, there has been a specific focus on the performance of sepsis prediction models provided different sepsis definitions are used [13]. Systematic analysis of clinical narrative to understand how these can be used for early diagnosis and identification of sepsis has also emerged as an area of great interest, as has clinical text as a source of information for machine learning-based sepsis prediction [14]. They represent the shift from traditional statistical approaches to more complex, data-driven and clinically relevant AI systems for sepsis prediction. As a result, even though AI and machine learning have shown a great deal of promise for early sepsis prediction,

there are still significant research gaps in the areas of predictive reliability, uncertainty quantification, calibration, external validation, generalizability, and clinical safety. There have been many studies focused on the prediction accuracy of the model, but little on how well the model will be predictive of patient response in other patient populations and other clinical contexts.

Due to this, there is a need for a systematic study of AI models and assessments of their reliability and uncertainty. Thus, this study aims at analysing existing machine learning and AI techniques for sepsis prediction, including the most frequently used models and their predictive and calibration capabilities, the most prominent research gaps, and the possible future directions for developing clinically safe and trustworthy uncertainty-aware AI systems.

The achievements of this work are:

1. A systematic literature review on current AI and ML approaches to sepsis prediction and their clinical safety and reliability
2. Considers the possibility of using uncertainty quantification to uncover possibly incorrect predictions and to facilitate safer therapeutic decisions
3. Emphasizes that while accuracy is important, it is also important for the performance and calibration, Generalizability, external validation and robustness.
4. Identifies important gaps in the body of current research, such as the absence of uncertainty assessment, uneven evaluation techniques, and scant clinical effect data.
5. The advocates of the most important next steps in the development of reliable sepsis prediction models, such as prospective clinical evaluation, external validation, standardization of assessment, and estimation of uncertainty.

2. Literature Review

The reviewed literature highlights the increasing awareness of the need for precise, clinically safe and well-calibrated predictions, as well as the growing importance of machine-learning (ML) and artificial intelligence (AI) in sepsis prediction. Numerous techniques and predictive models have been explored, as well as how to identify possible false predictions. Important limitations in terms of uncertainty estimation, external validation, generalizability, and clinical assessment are also highlighted by the literature collected for this study.

These findings suggest that, besides the standard performance measures of AI-based sepsis prediction, the assessment of such prediction should consider the reliability of such predictions in various clinical environments and whether it can be safely used in clinical decision making.

The study found that deep learning and machine learning are potential to predict sepsis at an early stage. While Boussina et al. [6] evaluated the effects of a deep learning sepsis prediction model on survival and quality of treatment, studies by Calvert et al. [4] and Nemati et al. [5] looked into computational and interpretable machine learning methodologies.

In particular, Shashikumar et al. [12] looked at an AI sepsis prediction system that was intended to identify when it was unable to give a trustworthy response. This work is particularly relevant to uncertainty-aware prediction as it emphasizes that when a model's reliability is limited, it does not always give a confident prediction. However, there is not enough comparable information in the present literature table to identify the optimal uncertainty-estimation technique.

The literature searched for the research and themes related to the main is arranged in the following table.

Table 1. ***Summary of related studies, AI approaches, reported performance, and key research gaps in healthcare AI***

Study	Models	Performance	Gaps
[15]	Healthcare digital twins; physiological, patient, care-delivery, hospital, and health-system levels. No specific ML/DL uncertainty-quantification method reported.	No quantitative AI performance metrics reported.	External validation was uncommon; prospective evaluation was rare; limited evidence of patient benefit, operational effectiveness, and system-level impact. Interoperability, governance, organisational, and

			equity challenges remain.
[16]	Random Forest, XGBoost, Parallel Lightweight Diagnosis Models, Deep Neural Networks, Extremely Randomized Trees (ERT)	ERT demonstrated the best/optimal performance. SHAP was the dominant XAI technique for interpretability.	Limited generalizability, data limitations, difficulty translating results into policy/practice, limited use of LIME, lack of explainability in ERT, and need for causal-XAI-ML approaches.
[17]	Deep Learning was the most common AI approach. Applications included AI for screening, diagnosis, treatment decision support, and rehabilitation.	9 studies found AI to be cost-effective; 6 were cost-neutral or somewhat cost-effective; 11 were not cost-effective. Overall, AI showed promising cost-effectiveness, particularly for screening and early diagnosis.	Evidence strength was moderate and highly context-dependent. Results were sensitive to methodological assumptions, implementation costs, and healthcare-system characteristics. Considerable heterogeneity existed in model structure, perspective, and time horizon. Methodological quality ranged from low to high.
[18]	Active-learning software (ASReview) for literature screening; Large Language Model (QWQ-32B) for data extraction; Bayesian Model Averaging (BMA) for pooled effect estimation.	AI-assisted screening and data extraction achieved ≥94% accuracy and $\kappa = 0.91$ inter-rater reliability. Bayesian analysis strongly supported early EN (≤24 hours), showing reduced mortality, infectious/GI complications, and ICU length of stay.	High methodological heterogeneity persisted. AI workflows still require expert oversight to address inherent limitations. Future work should focus on automated continuous evidence updating while maintaining human supervision.
[19]	Digital Twins, integrated with Artificial Intelligence (AI), Machine Learning (ML), and Internet of Medical Things (IoMT) technologies. Key applications include predictive/preventive digital twins and digital twin-based treatment planning and patient-response simulation.	The research shows a clear shift toward personalized, predictive, and simulation-oriented healthcare. Digital twins are emerging as central frameworks for integrating AI, ML, and IoMT.	A significant gap exists between technological development and routine clinical implementation. Further research is needed to translate digital twin technologies into practical healthcare applications and clinical settings.
[20]	Machine Learning (ML) models used in healthcare/medical decision-support systems. The review particularly emphasizes external validation and uncertainty quantification for evaluating these models.	Fewer than 4% of studies in high-impact medical informatics journals over the past 13 years validated their systems using data from a setting different from the training-data setting. This indicates limited evidence of performance across diverse real-world healthcare environments.	Poor external validation and reproducibility across healthcare settings. There is a need for robustness assessment, external validation, and uncertainty quantification. Hidden uncertainty can lead to unreliable ML models and potentially risky clinical decisions.
[21]	SHAP (54.0%), LIME (18.3%),	Diagnosis was the main	Strong method/domain

	and Grad-CAM (17.8%) were the dominant XAI methods. SHAP was mainly used for structured data (60.1%), while Grad-CAM was dominant in imaging studies (85%).	application (49.9%). Evaluation was mainly technical: 63.4% used quantitative metrics, 8.4% included human assessment, and 12.3% performed external validation. The three dominant methods accounted for 90.6% of studies.	concentration and limited evidence of clinical utility. 91.6% of studies lacked human assessment. There is a need for comparative evaluation of XAI methods, broader coverage of underrepresented clinical domains, multimodal approaches, and clinical validation protocols that assess whether explanations are actually useful to clinicians.
[22]	Event-log, Petri-net, graph, object-centric, neural, uncertainty-aware, sequential, bandit, reinforcement learning, and audit architecture approaches. These are organized into four layers: process representation, learning, inspection allocation, and governance.	The review covered 89 publication families. Substantial research exists for individual layers, but cross-layer evaluation remains fragmented, with heterogeneous datasets, objectives, and validation protocols.	Need for audit-ready benchmarks, explicit inspection-budget protocols, calibrated uncertainty, transfer across organizational contexts, and reproducible governance interfaces. A major gap is the lack of integrated evaluation across all four layers under real-world constraints such as limited human review, computation, latency, and documentation capacity.

It has been proven that ML and AI can be potentially useful in the early detection of sepsis. The current models, however, tend to focus on forecast accuracy, rather than uncertainty quantification, probability calibration, and external validation. These limitations make it difficult to evaluate the reliability of the forecast, the generalizability of the forecast, and the safety of the forecast for clinical use. To build clinically safe sepsis prediction systems, current research needs to be reviewed to explore uncertainty-aware AI approaches, analyze the reported performances and reliability, and recognize existing research gaps.

This review will focus on the state-of-the-art AI and ML methods for sepsis prediction, particularly predictive performance, UQ, calibration and clinical safety, to identify key research gaps and future directions.

The following goals assisted us in achieving this goal:

i. Identify the AI, ML and uncertainty quantification techniques used in predicting sepsis.
ii. Discuss the accuracy of existing sepsis prediction tools and their predictive performance.
iii. Discuss issues related to clinical safety, generalizability, external validation, and uncertainty.
iv. Identify areas of uncertainty and recommend future directions for the accurate prediction of sepsis in a clinically safe manner.

3. Methodology

The reliability and clinical safety of artificial intelligence (AI) and machine learning (ML) models for sepsis prediction are investigated in this study using a methodical literature review methodology. The methodology aims at providing a methodical process to identify, select, evaluate, and integrate relevant studies that are focused on predictive performance, uncertainty quantification, calibration, external validation, and clinical safety. The review follows a structured process that encompasses literature searches, selection of literature for the review, extraction of data, quality appraisal, and synthesis of the evidence to ensure transparency and repeatability. The suggested methodology is shown in Figure 1.

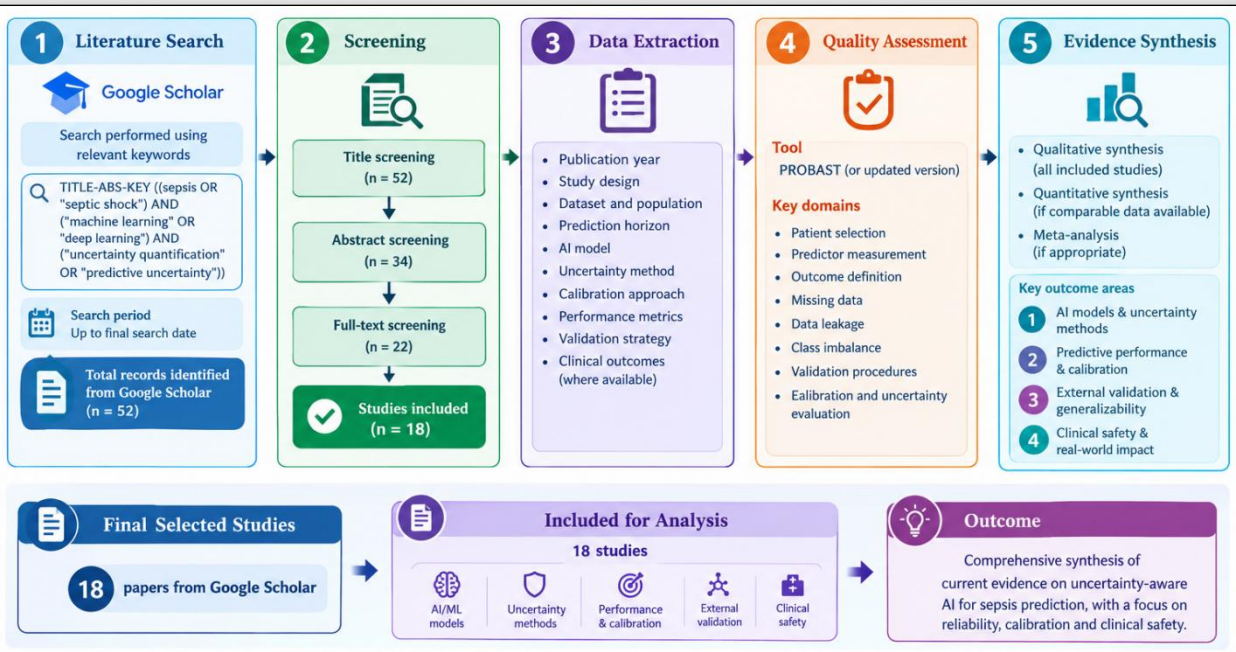


Figure 1. Proposed Methodology

3.1 Research Design

A literature review approach was employed to examine existing sepsis predictor technologies based on AI which focused on clinical safety and predictive accuracy. The main points discussed in the reviews were the methodologies used in the selected studies, the results reported, and the limitations of the studies regarding uncertainty estimation, calibration, generalizability, and clinical implementation. The results were synthesized to identify common issues in the research and study areas.

3.2 Research Questions

On the premise of the three research questions above, the following questions were formulated in order to assess the unmet research needs in this field and to identify the most important findings from the analysis of the 18 selected papers.

Q1. What are the prediction methods for sepsis with AI models and uncertainty techniques?

Q2. What are the outcomes of the calibration and prediction performance?

Q3. What are some possible directions for future study and what are the gaps in the state of the art for uncertainty-aware AI in sepsis prediction?

3.3 Search Strategy

Relevant research papers were obtained from Google Scholar. The primary focus for the search was sepsis prediction, artificial intelligence, machine learning, deep learning, uncertainty quantification, predictive performances and clinical validations. By looking at the titles, abstracts, and substance of publications in connection to the goals of the study, pertinent papers were found. Eighteen publications in all were selected and summarized for the review.

The following is a description of inclusion/exclusion criteria.

Studies that focused on machine learning or deep learning techniques, uncertainty quantification, model assessment, external validation, interpretability, or other challenges in healthcare AI applications were relevant to the review; only publications that were relevant to and contributed to the goals of the study were included in the synthesis. They were selected because they could be applied to clinical safety, were predictive and could be usefully applied based on AI-based sepsis prediction.

The publications selected contained studies that focused on sepsis predictions, as well as studies that were more broadly related to AI in healthcare relevant to the research topics outlined in this review.

3.4 Study Selection and Data Extraction

We have found 18 papers by using Google Scholar and reviewed and collated them. Each paper underwent a review of relevant information such as suggested methodologies, AI/ML models, performance/result(s) mentioned, and restrictions identified. The retrieved data has been used to create a comparative literature table that includes references to the literature studied, methodology, findings and research gaps. This process was very easy to conduct a comparison of the selected studies and helped to highlight ongoing clinical safety, external validation, calibration, uncertainty quantification, and generalizability problems.

This is an important assessment of the selected research conducted.

Selected papers were rigorously analysed in terms of methodology, conclusions and limitations. There was special attention to the assessment of prediction performance, uncertainty management, and interpretability of AI models, as well as external validation and evidence of clinical usefulness. The analysis also considered the limitations of the original studies, such as data limitations, methodological differences, limited generalizability, and challenges implementing the research findings into clinical practice. This comprehensive assessment was used to identify gaps in research.

3.5 Evidence Synthesis

Qualitative synthesis of the results of the 18 papers that were selected were performed before grouping them by the main themes of the review. Clinical safety, external validation and generalizability, uncertainty quantification, predictive performance and calibration, and sepsis prediction using AI and ML were all included in these themes. The results and limitations stated were compared to find common themes and unaddressed issues. Due to the diversity of studies used, results were presented in the form of the qualitative synthesis rather than totaling the numbers.

The resulting synthesis guided the identification of research gaps and suggestions for the creation of dependable and clinically safe AI-based sepsis prediction systems in the future.

4. Results

The review of literature highlights the increasing recognition of the importance of accurate, clinically safe, and well-calibrated prediction and the growing importance of machine learning (ML) and artificial intelligence (AI) in this space. Many interpretable learning techniques and predictive models have been investigated, along with ways to spot potentially false predictions. Literature gathered for this study also underscores some significant uncertainties around uncertainty estimation, external validity, generalizability and clinical evaluation.

These findings suggest that, besides the standard performance measures of AI-based sepsis prediction, the assessment of such prediction should consider the reliability of such predictions in various clinical environments and whether it can be safely used in clinical decision making.

The study found that deep learning and machine learning are potential to predict sepsis at an early stage. While Boussina et al. [6] evaluated the effects of a deep learning sepsis prediction model on survival and quality of treatment, studies by Calvert et al. [4] and Nemati et al. [5] looked into computational and interpretable machine learning methodologies.

In particular, Shashikumar et al. [12] looked at an AI sepsis prediction system that was intended to identify when it was unable to give a trustworthy response. This work is particularly relevant to uncertainty-aware prediction as it emphasizes that when a model's reliability is limited, it does not always give a confident prediction. However, there is not enough comparable information in the present literature table to identify the optimal uncertainty-estimation technique.

The literature searched for the research and themes related to the main is arranged in the following table.

Table 2. ***Key sepsis prediction studies and their relevance to model interpretability, uncertainty, external validation, and clinical reliability***

Study	Approach or focus	Finding relevant to this review
Calvert et al. [4]	Computational approach to early sepsis detection	Demonstrates the application of clinical data for early sepsis prediction.
Nemati et al. [5]	Interpretable ML model	Highlights the use of interpretable prediction for ICU sepsis detection.

Boussina et al. [6]	Deep learning sepsis prediction	Examines quality of care and survival, extending evaluation beyond model performance alone.
Wong et al. [9]	External validation of a proprietary sepsis model	Highlights the importance of assessing performance in independent healthcare settings.
Shashikumar et al. [12]	Uncertainty-aware sepsis prediction	Investigates the ability of an AI algorithm to recognize when it cannot provide a reliable prediction.
Dutta et al. [13]	Evaluation across different sepsis definitions	Highlights the importance of diagnostic definitions when interpreting and comparing model performance.

How these models' estimated dangers are assessed also affects how reliable they are. Predicted probabilities cannot be used to determine whether these probabilities are correct, but they can be used to determine whether a model discriminates between patients at higher risk and lower risk. Thus, prediction scores are important when making clinical judgments, and calibration is important. Although the literature abounds with the notion of dependability and uncertainty, a uniform set of calibration measures is not available across the research. So, it is not possible to compare the performance of the different calibrations quantitatively with the evidence currently available.

Another important finding is the generalizability. The performance of a model developed with a patient population or healthcare organization may differ from what it is expected to do with another population or organization because of differences in patient characteristics, data collection, clinical procedures or criteria for sepsis. The multi-centre prospective validation study cited by Wong et al. [11] and the external-validation research which was highlighted by Wong et al. [9] both highlight the importance of testing models in environments other than where they were developed. Dutta et al. [13] further stress the significance of some of the definitions used for sepsis in evaluating the model.

However, the amount of information extracted from the evaluation provided is not enough to assess the percentage of research that has been externally validated and to evaluate the performance gains as a result of these external validations.

The main constraints and their consequences for uncertainty-aware sepsis prediction are outlined in the following table.

Table 3. ***Research gaps and recommended evaluations for improving clinical reliability and safety***

Identified issue	Implication for clinical use	Recommended evaluation
Unquantified predictive uncertainty	Clinicians may not recognize unreliable predictions.	Evaluate uncertainty estimation and selective prediction.
Inadequate calibration reporting	Predicted probabilities may not represent actual clinical risk accurately.	Report calibration plots, Brier scores, or other appropriate calibration measures.
Limited external validation	Performance may not generalize to new hospitals or patient populations.	Conduct independent and multicenter validation.
Differences in sepsis definitions	Performance comparisons may be misleading across studies.	Document outcome definitions and evaluate robustness across definitions where feasible.
Limited prospective clinical assessment	Technical performance may not translate into improved care.	Evaluate clinical impact, workflow integration, and patient outcomes.

These constraints also highlight the need to consider clinical safety and prediction accuracy separately. However, a good discriminator can still give too many false alarms, misdiagnose on a new patient or give false probabilities. Uncertainty-aware systems may be beneficial to recognise situations that require further assessment, but only empirical tests can determine the safety benefits of these systems. This evaluation is based on research in the areas of clinical impact and uncertainty-aware prediction, but additional prospective data is needed to assess the value of these approaches for improving clinical decisions and patient outcomes.

A conceptual description of the findings' relevance to the creation of trustworthy and clinically safe sepsis prediction systems is shown in Figure 1.

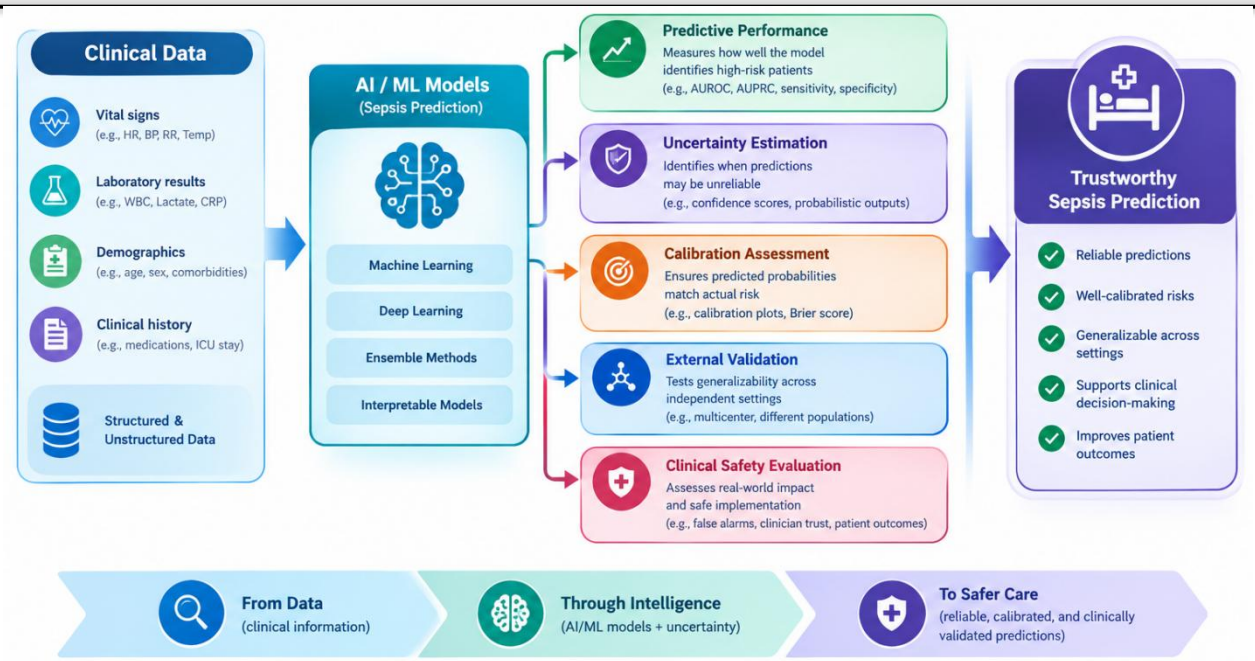


Figure 1. Conceptual framework for evaluating the reliability and clinical safety of AI-based sepsis prediction

In summary, the literature reviewed shows that there is a need for developing reliable AI models for sepsis prediction to include an integrated assessment of predictive performance, uncertainty estimation, calibration, external validation, and clinical safety. The evidence compiled in this publication does not yet provide conclusive evidence of the relative efficacy of different uncertainty-aware strategies, although previous work demonstrated progress in interpretable modeling and early prediction. Standardized evaluation procedures, independent and prospective validation, open reporting of calibration and uncertainty metrics, and evaluation of clinical impact should be the top priorities for future research.

These criteria can be used to develop AI systems that can not only predict sepsis but also communicate how accurate they are in their prediction, thus improving safe clinical decision-making.

5. Conclusion

Artificial intelligence (AI), and machine learning (ML), have demonstrated promising potential for early prediction of sepsis based on clinical data and identifying patterns associated with patient deterioration. The literature reviewed identifies several challenges, including limited uncertainty quantification, difficulty in assessing the calibration, no external validation, concerns about generalizability and clinical safety, which limit the reliability and clinical applicability of these systems. Previous work highlights the importance of interpretability, transparency of the model and evaluating the effectiveness of the model in real healthcare environments.

These findings imply that AI-based sepsis prediction systems' accuracy is insufficient to confirm their dependability. Instead, detailed evaluation of prediction performance, prediction uncertainty, calibration, and resilience and clinical impact are required for reliable clinical use. This systematic review highlights how crucial it is to move beyond traditional performance-based assessment and move towards AI systems that can report on the uncertainty of their predictions, facilitating more trusted clinical decisions.

5.1 Future Work

Further studies are needed to design and test uncertainty-aware AI models which accurately predict outcomes and determine when further clinical evaluation is needed. To make sure that the probabilities projected by the calibration are more in line with clinical risks, calibration should be used. The model needs to be validated at other institutions and patient populations to assess its robustness and generalizability. Standardized assessment methods, such as discrimination, calibration, uncertainty estimation, and clinically relevant safety measures, should be used by investigators.

Future research in the field is also needed to demonstrate that uncertainty-aware prediction algorithms can lead to better clinical decision making, fewer unnecessary alarms and better patient outcomes, as well as to show that these algorithms are better than traditional algorithms.

Future work should explore the use of interpretable AI methods that balance the therapeutic benefits of the explanation with the prediction's accuracy. Finally, for the development of a sepsis prediction model to be used in clinical practice, openness in reporting, repeatable research processes, and collaboration among AI researchers and clinicians would be needed to be able to embed uncertainty in the results and facilitate the transition from the model to clinical use.

References

[1] Schertz, Adam R., et al. "Sepsis prediction model for determining sepsis vs SIRS, qSOFA, and SOFA." JAMA Network Open 6.8 (2023): e2329729.

[2] Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive care medicine, 46(3), 383-400.

[3] Moor, Michael, et al. "Early prediction of sepsis in the ICU using machine learning: a systematic review." Frontiers in medicine 8 (2021): 607952.

[4] Calvert, Jacob S., et al. "A computational approach to early sepsis detection." Computers in biology and medicine 74 (2016): 69-73.

[5] Nemati, Shamim, et al. "An interpretable machine learning model for accurate prediction of sepsis in the ICU." Critical care medicine 46.4 (2018): 547.

[6] Boussina, Aaron, et al. "Impact of a deep learning sepsis prediction model on quality of care and survival." NPJ digital medicine 7.1 (2024): 14.

[7] Bhargava, Akhil, et al. "FDA-authorized AI/ML tool for sepsis prediction: development and validation." Nejm Ai 1.12 (2024): AIoa2400867.

[8] Deng, Hong-Fei, et al. "Evaluating machine learning models for sepsis prediction: A systematic review of methodologies." Iscience 25.1 (2022).

[9] Wong, Andrew, et al. "External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients." JAMA internal medicine 181.8 (2021): 1065-1070.

[10] Wong, Andrew, et al. "External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients." JAMA internal medicine 181.8 (2021): 1065-1070.

[11] Wong, Andrew, et al. "Multicenter prospective validation of an updated proprietary sepsis prediction model." JAMA network open 9.2 (2026): e260181.

[12] Shashikumar, Supreeth P., et al. "Artificial intelligence sepsis prediction algorithm learns to say “I don’t know”." NPJ digital medicine 4.1 (2021): 134.

[13] Dutta, Sayon, et al. "Performance of a Sepsis Prediction Model Across Different Sepsis Definitions." JAMA Network Open 9.4 (2026): e265599.

[14] Yan, Melissa Y., Lise Tuset Gustad, and Øystein Nytrø. "Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review." Journal of the American Medical Informatics Association 29.3 (2022): 559-575.

[15] Azamfirei, Leonard, et al. "Healthcare Digital Twins Across Scales: A Narrative Review and Five-Level Conceptual Framework." Healthcare. Vol. 14. No. 16. MDPI, 2026.

[16] Edem, Agbehadji Israel, and Obagbuwa Ibidun Christiana. "Explainable artificial intelligence and machine learning for air pollution risk assessment and respiratory health outcomes: A systematic review." Atmosphere 16.10 (2025): 1154.

[17] Imani, Ali, and Elham Monaghesh. "Cost-effectiveness of Artificial Intelligence applications in healthcare: A systematic review." Health informatics journal 32.3 (2026): 14604582261460314.

[18] Mao, Yumei, Geying Yang, and Hongfei Xu. "Timing of enteral nutrition in intensive care: A machine-learning-assisted automated meta-analysis using robust Bayesian model averaging." Nutrition Clinique et Métabolisme 40.2 (2026): 103038.

[19] Lăzăroiu, George, et al. "Digital technologies supporting predictive healthcare: Review of research trends." Human Technology 21.3 (2025): 694-730.
[20] Marconi, Luca, and Federico Cabitza. "Show and tell: A critical review on robustness and uncertainty for a more responsible medical AI." International journal of medical informatics 202 (2025): 105970.
[21] Bello, Sadeeq Olalekan, et al. "Explainable AI in Medicine: A Systematic Literature Mapping Study." Medicine: A Systematic Literature Mapping Study (January 31, 2026) (2026).
[22] Becerril Villamil, Héctor R., et al. "Adaptive Process Mining and Selective Monitoring for Algorithmic Auditing: A Survey of Representations, Learning Policies, Decision Strategies, and Open Problems." Machine Learning and Knowledge Extraction 8.8 (2026): 234.

IEER
Institute for Excellence in Education & Research