Review Article | Data Sciences

The role of artificial intelligence in sepsis in the Emergency Department: a narrative review

Mui Teng Chua[1,2]^, Yuru Boon[1,2]^, Zi Yao Lee[1,2]^, Jian Hao Jaryl Kok[1,2]^, Clement Kee Woon Lim[1]^, Nicole Mun Teng Cheung[1,2]^, Lorraine Pei Xian Yong[1,2]^, Win Sen Kuan[1]^

[1]Emergency Medicine Department, National University Hospital, National University Health System, Singapore, Singapore; [2]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Contributions: (I) Conception and design: MT Chua, WS Kuan; (II) Administrative support: MT Chua, WS Kuan; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Mui Teng Chua, MBBS (Singapore), MMed (EM) (Singapore), MPH. Emergency Medicine Department, National University Hospital, National University Health System, 9 Lower Kent Ridge Road, Level 4, Singapore 119085, Singapore; Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. Email: mui_teng_chua@nuhs.edu.sg.

Background and Objective: Early recognition and treatment of sepsis in the emergency department (ED) is important. Traditional predictive analytics and clinical decision rules lack accuracy in identifying patients with sepsis. Artificial intelligence (AI) is increasingly prevalent in healthcare and offers application potential in the care of patients with sepsis. This review examines the evidence of AI in diagnosing, managing and prognosticating sepsis in the ED.

Methods: We performed literature search in PubMed, Embase, Google Scholar and Scopus databases for studies published between 1 January 2010 and 30 June 2024 that evaluated the use of AI in adult patients with sepsis in ED, using the following search terms: ("artificial intelligence" OR "machine learning" OR "neural networks, computer" OR "deep learning" OR "natural language processing"), AND ("sepsis" OR "septic shock", AND "emergency services" OR "emergency department"). Independent searches were conducted in duplicate with discrepancies adjudicated by a third member.

Key Content and Findings: Incorporating multiple variables such as vital signs, free text input, laboratory tests and electrocardiogram was possible with AI compared to traditional models leading to improvement in diagnostic performance. Machine learning (ML) models outperformed traditional scoring tools in both diagnosis and prognosis of sepsis. ML models were able to analyze trends over time and showed utility in predicting mortality, severe sepsis and septic shock. Additionally, real-time ML-assisted alert systems are effective in improving time-to-antibiotic administration and ML algorithms can differentiate sepsis patients into distinct phenotypes to tailor management (especially fluid therapy and critical care interventions), potentially improving outcomes. Existing AI tools for sepsis currently lack generalizability and user acceptance. This is risk of automation bias with loss of clinicians' skills if over-reliance develops.

Conclusions: Overall, AI holds great promise in revolutionizing management of patients with sepsis in the ED as a clinical support tool. However, its application is currently still constrained by inherent limitations. Balanced integration of AI technology with clinician input is essential to harness its full potential and ensure optimal patient outcomes.

Keywords: Artificial intelligence (AI); machine learning (ML); sepsis; emergency medical services; prognosis

^ ORCID: Mui Teng Chua, 0000-0002-6326-4914; Yuru Boon, 0000-0001-5463-9713; Zi Yao Lee, 0009-0000-4946-9457; Jian Hao Jaryl Kok, 0000-0001-5012-2127; Clement Kee Woon Lim, 0009-0004-9514-0766; Nicole Mun Teng Cheung, 0000-0003-1572-7328; Lorraine Pei Xian Yong, 0009-0001-1885-6830; Win Sen Kuan, 0000-0002-2134-7842.

Submitted Aug 10, 2024. Accepted for publication Dec 16, 2024. Published online Feb 25, 2025.
doi: 10.21037/atm-24-150
View this article at: https://dx.doi.org/10.21037/atm-24-150

Introduction

Sepsis is a life-threatening condition caused by dysregulated host response to infection leading to acute organ dysfunction (1,2). In 2017, this condition led to 11 million deaths, which is almost 20% of all global mortality annually (3). In the emergency department (ED), about one-fifth of adult visits occur due to serious infections with risk of progression to sepsis with organ failure (4), representing a significant burden of care. Given its high prevalence, the early recognition and prompt treatment of sepsis in the ED is important to improve patient outcomes (5,6). However, the diagnosis, prognostication and management of patients with sepsis remains a challenge in the ED due to the heterogeneity of clinical symptoms and limited definitive and rapid diagnostic tests available (7).

Traditionally, predictive analytics in emergency care has relied on clinical decision rules, such as heuristics and scoring systems, to guide the decision-making process. These models typically utilize a limited set of clinically relevant variables and straightforward calculations (8). Studies have reported that early warning scores do not possess high performance in predicting sepsis mortality in the ED (9) and poorly predicts in-hospital mortality in high-risk patients with infections (10,11). As the population ages and comorbidities increase, EDs are likely to experience an increasing number of patients presenting with early symptoms and signs of sepsis. More reliable, rapid and affordable methods are needed to identify, risk-stratify and manage septic patients in the ED.

Artificial intelligence (AI) in healthcare

AI has become increasingly prevalent in healthcare, offering valuable decision-making support across various domains, including outcome prediction and disease diagnosis. AI enables machines to imitate human cognitive functions such as problem-solving and learning. Machine learning (ML) is a branch of AI focused on leveraging data to develop computer systems that are able to learn and improve from experience without being explicitly programmed (12). Statistical methods and algorithms play a role in recognizing patterns and discerning relationships from data to construct models capable of making informed predictions or decisions (13). Various types of ML have been proposed (14), including random forest (RF), gradient tree boosting (such as Extreme Gradient Boosting) and deep learning or deep neural networks (DNNs) (13,15-19) (*Table 1*). AI models offer advantages because they can be designed to harness ever-

Table 1 Summary of machine learning models

Machine learning (ML) models	Description
Decision trees	A supervised learning algorithm structured like a flowchart that can be used for both classification and regression tasks to predict the target variable of future instances based on a set of decision rules (13)
Deep learning or deep neural network (DNN)	A subfield of ML that employs Artificial Neural Networks with multiple processing layers. Several input neurons, which are information from the initial dataset, feed into hidden layers before passing to an output final layer (15,16). It has gained popularity in healthcare due to its success on a variety of complex classification tasks (17)
Random forest	An ensemble learning method used for classification, regression and feature selection. It builds and aggregates predictions from multiple decision trees to improve accuracy and is useful when large number of variables must be considered (18)
Gradient tree boosting	An ensemble ML technique that combines a set of weak models, to increase reliability by minimizing the biases and variances produced by the model itself (19). Extreme Gradient Boosting (XGBoost) is a powerful ML algorithm that has high speed, able to train on large datasets and improves model generalization

growing amounts of data found in the electronic health records (EHRs) (20), improve from experience to develop more accurate predictions, has the ability to consider large number of variables and can combine weak models to increase reliability by reducing biases and variances (18,19).

Despite the exciting potential of AI in clinical applications, there are significant barriers to safe implementation including potential bias in datasets, the proprietorship of systems and regulation (13). This review aims to examine the latest applications of AI in diagnosing, managing, and prognosticating sepsis in the ED, spanning from the development of novel ML models for early sepsis diagnosis to real-time ML-assisted sepsis alerts (MLASAs) to expedite triage-to-antibiotic time, and the use of predictive ML models for sepsis prognostication. We present this article in accordance with the Narrative Review reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-24-150/rc).

Methods

We performed a literature search of published studies on the use of AI in the diagnosis, management and prognosis of sepsis in the ED in PubMed, Embase, Google Scholar and Scopus databases using the following search terms: ("artificial intelligence" OR "machine learning" OR "neural networks, computer" OR "deep learning" OR "natural language processing"), AND ("sepsis" OR "septic shock", AND "emergency services" OR "emergency department"). Study team members conducted independent searches of articles and any discrepancy between two members was resolved by a third independent co-investigator. Our inclusion criteria were studies that evaluated the use of AI in the diagnosis, management, and prognosis of sepsis among adult patients aged 18 years and above, in the EDs of hospitals. All English language and peer-reviewed articles that were published from 1 January 2010 to 30 June 2024 were eligible for inclusion. The inclusion period from January 2010 was selected due to increased availability and advancement in ML in 2010s (21). In addition, references of narrative reviews, scoping reviews, systematic reviews and meta-analyses were searched to include the original articles. Articles that were not published in English, studies with statistical modelling involving only logistic regression, studies that included animals, pediatric patients or conducted in pre-hospital settings, conference proceedings, editorials, letters to editors and abstracts were excluded. The search strategy is summarized in *Table 2*.

Discussion

Early detection of sepsis

Early and accurate sepsis detection is challenging due to the complex pathophysiology and heterogeneity of the host response to infection (22). Clinicians differ in their knowledge and application of sepsis definitions (23), with under-recognition among patients with vague symptoms and delays in initiating the sepsis care plans (24,25). One of the main difficulties is that there is no universal gold standard for the diagnosis of sepsis, with changing definitions over time (*Table 3*) (1,26,27).

The use of quick Sequential Organ Failure Assessment (qSOFA) score as a quick bedside assessment tool outside of the intensive care unit (ICU) to predict mortality in sepsis has poor sensitivity (28,29). Other available tools such as the Modified Early Warning Score (MEWS) and Acute Physiology and Chronic Health Evaluation (APACHE) II score are more appropriate for prognostic rather than diagnostic purposes in the ED (30). Hence, there is a dearth in sepsis diagnostic scores in the current state. AI-based tools for sepsis have shown to be promising in the ICU setting due to the availability of frequent physiologic monitoring as well as plentiful laboratory and imaging data sets but there are not as many studies on AI-based tools involving ED patients (31).

The timing of sepsis diagnosis is crucial as early identification and treatment lead to improvement in outcomes (32). ML models can be trained to identify patients with sepsis within one hour of ED arrival, using data available such as mean arterial blood pressure, temperature, age, heart rate and white blood cell count. These models have a clinically acceptable sensitivity of close to 70% and above, allowing clinicians to identify septic patients early, with aim of translation to early treatment within one hour of ED attendance (29,33).

Recently, ML models were shown to consistently outperform traditional rule-based screening tools in detecting risk of developing sepsis. A study by Kijpaisalratana *et al.* demonstrated that ML models for the diagnosis of sepsis in the ED outperformed traditional tools like qSOFA [area under the receiver operating characteristic curve (AUROC), 0.635], systemic inflammatory response syndrome (SIRS) (AUROC, 0.814) and MEWS (AUROC, 0.688), with a high discriminatory power for RF (AUROC, 0.931) (34), which demonstrated reproducible results in the same center in a cluster randomized control trial (AUROC, 0.93) (35). Other studies similarly showed that ML models outperformed

Table 2 The search strategy summary

Items	Specification
Date of search	1 July 2024
Databases and other sources searched	PubMed, Embase, Google Scholar, Scopus
Search terms used	("Artificial intelligence" OR "machine learning" OR "neural networks, computer" OR "deep learning" OR "natural language processing"), AND ("Sepsis" OR "septic shock"), AND ("Emergency services" OR "emergency department")
Timeframe	1 January 2010 to 30 June 2024
Inclusion and exclusion criteria	Inclusion criteria • Original peer-reviewed research—retrospective, prospective, cross-sectional, case-control and randomized controlled trials • Study setting in emergency departments • English language papers • Focused on use of artificial intelligence on diagnosis, management and prognosis of sepsis Exclusion criteria • Studies involving animals, pediatric population less than 18 years of age, conducted in pre-hospital settings, only logistic regression used for statistical modelling • Editorials, letters to editors, abstracts, conference proceedings • Non-English language papers
Selection process	Two independent study team members searched the databases, and any discrepancy was resolved by a third team member
Additional considerations	Narrative reviews, scoping reviews, systematic reviews and meta-analyses searched for their primary references, and primary studies included if fulfil inclusion criteria

Table 3 Definitions of sepsis

Sepsis severity	Sepsis-3 definition [2016] (1)	Traditional definition [1992] (2)
Sepsis	Suspicious/known infection and rise in SOFA score ≥2 from baseline	Suspicious/known infection and ≥2 of 4 SIRS criteria
Severe sepsis	Not a category	Sepsis and organ dysfunction, hypoperfusion or hypotension
Septic shock	Sepsis and vasopressors required to maintain MAP >65 mmHg and lactate >2 mmol/L despite adequate fluid resuscitation	Sepsis and hypotension despite adequate fluid resuscitation

MAP, mean arterial pressure; SIRS, systemic inflammatory response syndrome; SOFA, sequential organ failure assessment.

various risk-stratifying scores such as SIRS, qSOFA, SOFA, MEWS and National Early Warning Score (NEWS) (*Table 4*) (22,26,29,34,35,37-39). McCoy *et al.* utilized ML to assist in sepsis prediction in a single center quality improvement initiative. Prior to implementation, the center was utilizing SIRS criteria as an indication for initiation of a sepsis bundle. Post-implementation, there was a reduction in sepsis-related in-hospital mortality by 60.24%, with reduction in length of stay by 9.55% and sepsis-related 30-day readmission rate reduction by 50.14% (38).

Different combinations of variables have been incorporated into ML models to diagnose sepsis. The most common variables include vital signs (e.g., temperature, blood pressure, respiratory rate, heart rate), demographic data (e.g., age and gender), and laboratory markers (e.g., full blood count, C-reactive protein, procalcitonin, lactate).

Table 4 Studies using machine learning models to diagnose sepsis in the emergency department

Study	Machine learning method(s)	Variable(s) used	Diagnostic criteria	Diagnostic performance	
				AUROC of ML methods	AUROC of comparator risk scores
Studies using data available at triage					
Mao 2018 (22)	GB	Vital signs, change in vital signs	SIRS criteria for sepsis and severe sepsis	Sepsis: 0.92	Sepsis: MEWS, 0.76; SIRS, 0.75; SOFA, 0.63
				Severe sepsis: 0.87	Severe sepsis: MEWS, 0.77; SIRS, 0.72; SOFA, 0.65
Brann 2024 (31)	XGBoost	Natural language processing of nursing triage notes, demographics, vital signs	Health system sepsis committee criteria for sepsis	Comprehensive model: 0.97 Time of triage: 0.94	–
Kijpaisalratana 2022 (34)	RF, LR, NN, GB	Vital signs, demographics, emergency severity index, mode of arrival, free text chief complaint	ICD-10-CM coding for sepsis	RF, 0.931; LR, 0.930; NN, 0.926; GB, 0.919	SIRS, 0.814; MEWS, 0.688; qSOFA, 0.635
Kijpaisalratana 2024 (35)	RF	Vital signs, demographics, emergency severity index, mode of arrival, free text chief complaint	Sepsis-3 criteria for sepsis	0.93	MEWS, 0.86; SIRS, 0.84; qSOFA, 0.73
Prasad 2023 (36)	L2-regularized LR	Vital signs, demographics, past medical conditions, symptoms, history of present illness	Clinical criteria for sepsis by Rhee *et al.*, 2017 (23)	Full model, training: 0.86	Full model, training: qSOFA, 0.66
Studies using a combination of demographics, vital signs, laboratory tests or medications					
Lin 2021 (26)	XGBoost	Vital signs, standard laboratory tests	Sepsis-3 criteria for sepsis	0.75	qSOFA, 0.66; SIRS, 0.57
Aygun 2024 (27)	XGBoost, LightGBM, AdaBoost	Demographics, vital signs, standard laboratory tests	Sepsis-3 criteria for sepsis	XGBoost, 0.940; LightGBM, 0.931; AdaBoost, 0.917	–
Delahanty 2019 (29)	GB	Shock index, vital signs, standard laboratory tests	Clinical criteria for sepsis by Rhee *et al.*, 2017 (23)	0.97	SOFA, 0.90; NEWS, 0.84; qSOFA, 0.80; MEWS, 0.78; SIRS, 0.77
Brown 2016 (33)	Naïve Bayes	Age, vital signs, white blood cell count	SIRS criteria for sepsis	0.953	SIRS, 0.606
Bedoya 2020 (37)	Multi-output Gaussian process and recurrent NN	Demographics, comorbidities, vital signs, standard laboratory tests, medications	≥2 SIRS criteria, blood cultures ordered, ≥1 end-organ failure	0.882	SIRS, 0.756; NEWS, 0.619

Table 4 (*continued*)

Table 4 (*continued*)

Study	Machine learning method(s)	Variable(s) used	Diagnostic criteria	Diagnostic performance	
				AUROC of ML methods	AUROC of comparator risk scores
McCoy 2017 (38)	Machine learning algodiagnostic	Vital signs, standard laboratory tests if available	Sepsis-3 criteria for sepsis and severe sepsis	Sepsis: 0.91 Severe sepsis: 0.96	– Severe sepsis: SOFA, 0.77; SIRS, 0.76; MEWS, 0.55; qSOFA, 0.55
Knack 2024 (39)	LASSO	Demographics, vital signs, standard laboratory tests, skin findings	ICD-10-CM codes for sepsis	At 60 minutes: 0.87	At 60 minutes: physicians' gestalt, 0.93; qSOFA, 0.71; SIRS, 0.74
Upadhyaya 2024 (40)	RF, SVM, DNN, LR	Vital signs, complete blood count elements, monocyte distribution width	Sepsis-3 criteria for sepsis	RF, 0.90; SVM, 0.87; DNN, 0.87; LR, 0.83	–
Shashikumar 2021 (41)	COMPOSER	Vital signs, demographics, standard laboratory tests	Sepsis-3 criteria for sepsis	0.938	–
Taneja 2021 (42)	RF	Demographics, Glasgow Coma Scale, vital signs, standard laboratory tests, procalcitonin, interleukin-6, C-reactive protein	Sepsis-3 criteria for sepsis	0.83	–
Studies using specialized tests					
Niemantsverdriet 2022 (30)	L1 regularization, RF, LR	Demographics, vital signs, laboratory tests, advanced hematology variables from Abbott CELL-DYN Sapphire Analyzer	Sepsis-3 criteria for sepsis	L1, 0.85; RF, 0.84; LR, 0.73	–
Velly 2021 (43)	GB	18 plasma biomarkers, 12 biomarkers on monocytes, neutrophils, B- and T-lymphocytes, and 1 bacterial biomarker	SIRS and Sepsis-3 criteria for sepsis	SIRS: 0.880 Sepsis-3: 0.923	–
Study using ECG					
Kwon 2021 (44)	Convolutional NN	12-, 6- and 1-lead ECGs	Sepsis-3 criteria for sepsis	Sepsis: 12-lead ECG, 0.863; 6-lead ECG, 0.856; 1-lead ECG, 0.845	–

AdaBoost, adaptive boosting; AUROC, area under the receiver operating characteristic curve; COMPOSER, COnformal Multidimensional Prediction Of SEpsis Risk; DNN, deep neural network; ECG, electrocardiogram; GB, gradient boosting; ICD-10-CM, International Classification of Diseases, 10th Revision, Clinical Modification; ICD-9-CM, International Classification of Diseases, 9th Revision, Clinical Modification; LASSO, least absolute shrinkage and selection operator; LightGBM, Light Gradient Boosting Machine; LR, logistic regression; MEWS, Modified Early Warning Score; ML, machine learning; NEWS, National Early Warning Score; NN, neural network; qSOFA, quick Sequential Organ Failure Assessment; RF, random forest; SIRS, systemic inflammatory response syndrome; SOFA, Sequential Organ Failure Assessment; SVM, support vector machine; XGBoost, extreme gradient boosting.

Vital signs are easily available physiological clinical data that can be collected early during the patient's initial presentation at triage. These studies generated modest to excellent discriminatory power (AUROCs from 0.67 to 0.92) and their findings are summarized in *Table 4* (22,36,45). Apart from vital signs, the patients' chief complaints and nursing assessment can be useful in the diagnosis of sepsis. Horng *et al.* showed an improvement in the AUROC for diagnosis of sepsis from suboptimal (0.67) to excellent when chief complaint (0.83) or nursing free text (0.86) was supplemented into the variables (45). Brann *et al.* used a natural language processing-based ML model to interpret the triage nursing notes and combined this with clinical variables such as patient's demographics and vital signs available at triage. Sepsis was detected in 76% of cases where sepsis was not considered and in 97.6% of suspected cases at triage (31). This may be helpful for patients with vague symptoms where diagnostic suspicion is low, and clinicians may fail to initiate timely testing (36).

Non-invasive bedside investigations such as electrocardiogram (ECG) have been explored in AI studies for the diagnosis of sepsis. Previous studies have shown that 50% of patients with sepsis demonstrated signs of myocardial dysfunction and their ECGs may show prolonged QRS duration with decreased amplitude (46,47). On this basis, Kwon *et al.* trained a convolutional neural network to detect sepsis using single-, 6- and 12-lead ECGs, with all achieving excellent discriminatory power (AUROC of 0.8 to 0.9) (44).

No single biomarker is considered diagnostic for sepsis (48). Clinicians often use white blood cell count and its differentials to indicate presence or absence of infection. Incorporating blood counts into ML algorithms with vital signs, discriminatory power of these algorithms could further enhance accurate identification of sepsis (40,49). However, some markers are still not readily available in the clinical setting and may only be obtainable under experimental conditions. For instance, Velly *et al.* used a gradient tree boosting approach and found the best combination in human leukocyte antigen DR isotype (HLA-DR) on monocytes, myeloid-epithelial-reproductive tyrosine kinase (MerTk) on neutrophils and plasma matrix metaloproteinase-8 (MMP8) for the diagnosis of bacterial infection (43). This would require further validation in larger clinical cohorts.

Clinical decision support (CDS) in sepsis management

Sepsis poses a significant burden and challenge on the healthcare system (23,50). Timely empirical antibiotic administration, early adequate fluid resuscitation and appropriate vasopressor use can lead to better clinical outcomes (51). Protocolized sepsis bundles ensure standardized and timely treatment, improving patient outcomes (6,52,53); every one-hour delay in bundle completion increases in-hospital mortality by 4% (53). Although the benefit of expeditious appropriate treatment is evident (54), a large proportion of clinicians still fail to meet these time targets (55,56), resulting in suboptimal sepsis care. The use of AI and its integration into the EHR CDS tools have the potential to effect significant benefit in the timely management of sepsis (57-59), but widespread adoption would ultimately depend on effective implementation and user acceptance (60).

A cluster-randomized trial by Kijpaisalratana *et al.* implemented a real-time MLASA system integrated into the EHR that resulted in more patients receiving antibiotics in a timely fashion (8.3% and 5.5% increase in those receiving antibiotics in 1 and 3 hours, respectively) (35). The ML algorithm in this study utilized triage vital signs, age, gender, Emergency Severity Index (ESI) (61), mode of arrival and chief complaints through an RF model to generate a prediction probability of sepsis within minutes of triage. Patients with a sepsis probability above the diagnostic threshold were marked with a red banner. Treatment decisions thereafter were left to the judgment of the clinical teams. The earlier sepsis detection by ML appears to prompt clinicians to consider antibiotics sooner. Similar results in improvement of antibiotic timings were replicated elsewhere (41), showing potential in eventually modifying practice.

Previous research has recognized distinct subclasses and clinical phenotypes of patients in sepsis, linking them to different clinical outcomes and treatment effects (62,63). Specifically, 4 clinical phenotypes were not distinguished by severity of illness or site of infection alone (63). Briefly, the phenotypes were designated α, β, γ, and δ—α phenotype had patients with the lowest administration of a vasopressor due to fewer abnormal laboratory values and less organ dysfunction; β phenotype patients were older and had more chronic illness and renal dysfunction; γ phenotype patients had more inflammation and pulmonary dysfunction; and δ phenotype patients had more liver dysfunction and septic shock (63). Classification of each patient could enable precision medicine, tailoring individualized treatment strategies in fluid, vasopressor, corticosteroid and antibiotic use to improve patient care. ML algorithms can assist in

this. Boussina *et al.* input commonly used clinical variables from the EHR into a program trained for sepsis prediction. Using spectral clustering on the representations from this network (64), they were able to identify 4 consistent phenotypes of patients in sepsis. They found that the administration of fluids over 30 mL/kg was associated with poorer outcomes in particular patient phenotypes (64). This is particularly valuable in guiding fluid replacement therapy among heterogenous septic ED patients.

Apart from fluid therapy, the pillars of sepsis management commonly include many other interventions. Chang *et al.* created an extreme gradient boosting prediction model utilizing information immediately available on patient arrival in the ED (age, gender, triage vital signs, triage acuity score, mode of arrival, and mental status) to predict the need for six early critical care interventions, namely arterial line insertion, oxygen therapy, high flow nasal cannula use, intubation, massive transfusion and inotropes or vasopressor administration, all with excellent discriminatory power (65). They integrated the prediction model into the EHR system, with pop-up alerts to clinical teams to recommend critical interventions, guiding physician treatment decisions in the early stages of management.

Diagnosing severity of sepsis

Prognostic tools in sepsis, such as SIRS, Sequential Organ Failure Assessment (SOFA) and NEWS may help determine the mortality risk of a septic patient in the ED, assist in decisions regarding ICU admission and predict the length of hospital stay (66-69). However, each comes with their own set of advantages and limitations. The qSOFA score simplifies the assessment by relying on just three bedside clinical criteria-altered mental status, systolic blood pressure no more than 100 mmHg, and a respiratory rate more than 21 breaths per minute. A qSOFA score of 2 or more was associated with a higher probability of mortality (1). While most studies reveal a better performance of qSOFA over SIRS in terms of specificity (70,71), its sensitivity has been shown to be relatively lower (72).

The NEWS is widely used by the National Health Service in the UK. It assigns a score to patients based on physiological parameters such as respiratory rate, oxygen saturations, temperature, systolic blood pressure, pulse rate and level of consciousness (73). The improved NEWS2 aims to provide a better prediction of deterioration in patients with hypercapnic respiratory failure and has been praised in some studies for its accuracy in predicting early mortality (74,75). However, unlike the other prognostic tools, its utility in all patient subgroups is often debatable (76-78).

While these traditional tools rely on static snapshots of patient data, AI models can analyze trends over time, identifying subtle patterns that may indicate the onset or progression of sepsis. A review by Islam *et al.* evaluated ML models for early sepsis prediction using EHRs, showing that the use of continuous patient data like vital signs and trending of laboratory data improved the accuracy of sepsis predictions as compared to traditional static models (79). By harnessing the strengths of ML and deep learning in processing large datasets including serial longitudinal data inputs to learn and predict complex patterns, more precise prognostic tools can be developed. Studies evaluating the use of AI in sepsis prognostication are summarized in *Table 5*.

Predicting outcomes of sepsis

Approximately 12% to 22% of patients with sepsis progress to septic shock within 72 hours of hospital admission, which is associated with worse outcomes including higher mortality rates (99,100). ML models have shown a high prediction rate for latent shock using vital signs trend in the ED (80) and even at the point of triage using a combination of chief complaint and physiological parameters, outperform qSOFA and early warning scores in predicting septic shock (81,82).

In an observational cohort study done by Wardi *et al.* (83), the Artificial Intelligence Sepsis Expert algorithm was developed using an ML algorithm trained on 40 most measured input variables from EHRs. This algorithm was able to predict the development of delayed septic shock during its derivation phase and its validation study at a different site showed an improvement in the AUROC from 0.778 to 0.85 after applying transfer learning techniques. This highlights the benefits of using ML algorithms as transfer learning can be used to improve portability and accuracy in various population groups. AI tools like deep learning models have also shown ability in predicting septic shock by capturing subtle changes on ECG (44), again highlighting its ability to extract and develop a predictive algorithm based on various types of data.

Despite advances in medical care, mortality rates from sepsis range between 25% and 30% (101). Many studies have shown the ability of ML models in predicting mortality among septic patients in clinical settings. For example, Karlsson *et al.* found that ML utilizing the

Table 5 Summary of studies using machine learning models in prognosis of sepsis

Study	Machine learning method(s)	Variable(s) used	Outcome(s)	Prognostic performance	
				AUROC of ML methods	AUROC of comparator prognostic methods
Predicting septic shock					
Mao 2018 (22)	Gradient tree boosting	Vital signs, change in vital signs	Septic shock during admission	0.9992	MEWS, 0.94; SOFA, 0.86; SIRS, 0.82
Kwon 2021 (44)	CNN	12-, 6- and 1-lead ECGs	Septic shock during admission	12-lead ECG, 0.899; 6-lead ECG, 0.893; 1-lead ECG, 0.860	C-reactive protein, 0.724; body temperature, 0.680
Chang 2023 (80)	RF, MLP, LR, RNN	Vital signs	Latent shock within 24 hours	At 3 hours: RF, 0.852; MLP, 0.841; LR, 0.830; RNN, 0.822	Adjusted shock index, 0.732; shock index, 0.546
Kim 2020 (81)	GBM, RF, MARS, SVM, MLP, LASSO, RR	Demographics, vital signs, laboratory data, chief complaints	Septic shock within 24 hours of ED arrival	GBM, 0.923; RF, 0.920; MARS, 0.915; SVM, 0.914; MLP, 0.911; LASSO, 0.905; RR, 0.904	Adjusted qSOFA, 0.832; qSOFA, 0.813; adjusted MEWS, 0.813; MEWS, 0.790
Yun 2021 (82)	XGBoost, LR with linear classification boundary, ANN	NEWS, demographics, vital signs	Septic shock within 24 hours of ED arrival	XGBoost, 0.845; LR, 0.844; ANN, 0.835	NEWS 0.804
Wardi 2021 (83)	Weibull-Cox proportional hazards model modified with a 2-layer neural network	Vital signs, laboratory data, SOFA scores, comorbidities, length of stay, patient outcomes	Septic shock within 4 to 48 hours of ED triage	At 12 hours: 0.778 At 12 hours with transfer learning: 0.85	At 12 hours: SOFA, 0.792 At 12 hours with transfer learning: 0.838
Predicting mortality					
Kwon 2021 (44)	CNN	12-, 6- and 1-lead ECGs	In-hospital mortality	12-lead ECG, 0.817; 6-lead ECG, 0.815; 1-lead ECG, 0.802	SOFA ,0.817; NEWS, 0.808; lactate, 0.801; qSOFA, 0.797; MEWS, 0.778; WBC, 0.591; C-reactive protein, 0.541
Karlsson 2021 (84)	RF	Vital signs, symptoms, observations, comorbidities	7- and 30-day mortality	7-day mortality, 0.83; 30-day mortality, 0.80	–
Rodríguez 2021 (85)	C4.5 decision tree, RF, ANN, SVM	Variables related to initial clinical care, vital signs, comorbidities, demographics, laboratory data	In-hospital mortality	Clinical care variables: C4.5 decision tree, 0.59; RF, 0.61; ANN, 0.58; SVM, 0.58 Direct measurement variables: C4.5 decision tree, 0.53; RF, 0.65; ANN, 0.69; SVM, 0.68	–
Cheng 2022 (86)	CNN, RF, LSTM	Demographics, vital signs	48-hour mortality	CNN, 0.82; RF, 0.75; LSTM, 0.74	–

Table 5 (*continued*)

Table 5 (*continued*)

Study	Machine learning method(s)	Variable(s) used	Outcome(s)	Prognostic performance	
				AUROC of ML methods	AUROC of comparator prognostic methods
Greco 2023 (87)	Balanced LR, unbalanced LR, RF	Demographics, provenience, comorbidities, nutritional status, delay to ED presentation, site of infection, clinical data at ED presentation, laboratory data, clinical scores	In-hospital mortality	RF, 0.834; LR, 0.826; balanced LR, 0.818	SOFA, 0.712; qSOFA, 0.706; APACHE II, 0.664
Raven 2023 (88)	XGBoost	Demographics, time of presentation, ED location, presenting complaint, vital signs, disease severity and urgency variables, laboratory data	In-hospital mortality	XGBoost with clinical judgment, 0.79; XGBoost, 0.79	Clinical judgement, 0.61
van Doorn 2021 (89)	XGBoost	Laboratory data, clinical data, demographics, vital signs, physical characteristics	31-day mortality	0.852	SOFA 0.752; clinical judgment, 0.735; abbMEDS, 0.631; mREMS, 0.630
Perng 2019 (90)	KNN, SVM, SoftMax, RF	Demographics, vital signs, laboratory data	72-hour and 28-day mortality	72-hour mortality: KNN, 0.83; SVM, 0.93; SoftMax, 0.91; RF, 0.89	72-hour mortality: qSOFA, 0.74; SIRS, 0.67
				28-day mortality: KNN, 0.84; SVM, 0.90; SoftMax, 0.88; RF, 0.89	28-day mortality: qSOFA, 0.68; SIRS, 0.59
Wong 2023 (91)	ANN	Demographics, vital signs, clinical variables, complete blood count	30-day mortality	0.811	Neutrophil-to-lymphocyte ratio, 0.644; platelet-to-lymphocyte ratio, 0.606; monocyte-to-lymphocyte ratio, 0.555
Jeon 2023 (92)	LightGBM, MLP, SVM, XGBoost	Demographics, vital signs, comorbidities, infection source, laboratory data, treatment, source control	7-day, 14-day and 30-day mortality	7-day mortality: LightGBM, 0.89; MLP, 0.89; SVM, 0.84; XGBoost, 0.85	7-day mortality: SOFA, 0.68; NEWS, 0.63; MEWS, 0.59; qSOFA, 0.59
				14-day mortality: LightGBM, 0.89; MLP, 0.88; SVM, 0.85; XGBoost, 0.84	14-day mortality: SOFA, 0.65; NEWS, 0.63; MEWS, 0.59; qSOFA, 0.57
				30-day mortality: LightGBM, 0.87; MLP, 0.86; SVM, 0.85; XGBoost, 0.84	30-day mortality: SOFA, 0.66; NEWS, 0.63; MEWS, 0.57; qSOFA, 0.57
Park 2024 (93)	SVM, RF, XGBoost, LightGBM, CatBoost	Clinical variables and SOFA score components	In-hospital mortality	CatBoost, 0.800; XGB, 0.797; LightGBM, 0.795; SVM, 0.771; RF, 0.736	–

Table 5 (*continued*)

Table 5 (*continued*)

Study	Machine learning method(s)	Variable(s) used	Outcome(s)	Prognostic performance	
				AUROC of ML methods	AUROC of comparator prognostic methods
Taylor 2016 (94)	RF, CART	Demographics, previous health status, ED health status, ED services rendered, operational details, laboratory data, ED diagnosis	28-day mortality	RF, 0.860; CART, 0.693	CURB-65, 0.734; REMS, 0.717; MEDS, 0.705
Kwon 2020 (95)	qSOFA-based ML models using XGBoost, LightGBM, RF	Demographics, ED diagnoses, vital signs, laboratory data, length of stay, intensive care admission, mechanical ventilation	3-day and inpatient mortality	3-day mortality: 0.86	3-day mortality: qSOFA, 0.78; MEWS, 0.77; SIRS, 0.68
				Inpatient mortality: 0.75	Inpatient mortality: qSOFA, 0.71; SIRS, 0.66; MEWS, 0.65
Katz 2022 (96)	RF	Demographics, laboratory data, vital signs, clinical observations	30-day mortality for patients with necrotizing soft-tissue infections	0.91	SOFA, 0.77
Ko 2022 (97)	RF, XGBoost, LR	Demographics, 6-hour bundle therapy components, vital signs, laboratory data	28-day mortality in stage 4 cancer patients with septic shock	Balanced RF, 0.826; RF, 0.811; XGBoost, 0.779; LR, 0.763	Lactate, 0.683; SOFA, 0.672; APACHE II, 0.662
Chiew 2019[†] (98)	GBM, AdaBoost, SVM, RF, KNN	Demographics, vital signs, heart rate variability	30-day mortality	GBM AUPRC, 0.35; AdaBoost AUPRC, 0.31; SVM AUPRC, 0.29; RF AUPRC, 0.27; KNN AUPRC, 0.10	qSOFA (worst) AUPRC, 0.29; NEWS AUPRC, 0.28; MEWS AUPRC, 0.25; qSOFA (initial) AUPRC, 0.21

[†], AUPRC reported in this study instead of AUROC curve. abbMEDS, abbreviated Mortality in Emergency Department Sepsis; AdaBoost, adaptive boosting; ANN, artificial neural network; APACHE II, Acute Physiology and Chronic Health disease Classification System II; AUPRC, area under precision-recall curve; AUROC, area under the receiver operating characteristic curve; CART, classification and regression tree; CatBoost, categorical boosting; CNN, convolutional neural networks; CURB-65, confusion, urea, respiratory rate, blood pressure, ≥65 years; ECG, electrocardiogram; ED, emergency department; GBM, gradient-boosting machine; KNN, k-nearest neighbors; LASSO, least absolute shrinkage and selection operator; LightGBM, Light Gradient Boosting Machine; LR, logistic regression; LSTM, long short-term memory; MARS, multivariate adaptive regression splines; MEDS, Mortality in Emergency Department Sepsis; MEWS, Modified Early Warning Score; ML, machine learning; MLP, multilayer perceptron; mREMS, modified Rapid Emergency Medicine Score; NEWS, National Early Warning Score; qSOFA, quick Sequential Organ Failure Assessment; REMS, Rapid Emergency Medicine Score; RF, random forest; RNN, recurrent neural network; RR, ridge regression; SIRS, systemic inflammatory response syndrome; SOFA, Sequential Organ Failure Assessment; SVM, support vector machine; WBC, white blood cell; XGBoost, extreme gradient boosting.

Balanced Random Forest classifier was able to predict 7-day and 30-day mortality with an AUROC of 0.83 and 0.80, respectively using clinical variables available on presentation to the ED (84). Another study by Rodríguez *et.al*. utilized multiple supervised ML techniques such as RF, ANN and SVM models to predict in-hospital mortality found both SVM and ANN provided the best results, with an AUROC of 0.69 (85). Similarly, Cheng *et al.*, with the use of vital signs data, utilized and compared three different ML methods, all of which were able to predict in-hospital mortality in septic patients, with an accuracy ranging from 0.817 to 0.905 at a lead time of 6 hours, and an accuracy of 0.759 to 0.828 at a lead time of 48 hours (86).

Traditional scoring tools such as qSOFA and SOFA have been found to poorly predict in-hospital mortality in the setting of sepsis (11,102) and were outperformed by AI tools and ML models (*Table 5*). AI tools showed good predictive performance for mortality in ED septic patients, achieving an AUROC of 0.863 compared to previously published SOFA, qSOFA and APACHE II scores (AUROCs 0.712, 0.706 and 0.664, respectively) (87). Other studies were able to successfully predict mortality in both the short term (e.g., 3 days) and longer term (28 to 31 days), and have outperformed physicians' clinical judgment (88) as well as clinical risk scores [e.g., SOFA, SIRS, Mortality in Emergency Department Sepsis (MEDS)] (*Table 5*) (89-95).

ML-based prediction models for mortality were explored in subgroups with specific medical conditions. Models utilizing prospectively collected data inclusive of patient demographics, vital signs, clinical findings, laboratory investigations and medications were able to predict mortality better than traditional SOFA scores among patients with necrotizing soft-tissue infections and patients with stage 4 cancer (96,97). Apart from clinical data, AI models utilizing heart rate variability based ML have shown superiority in comparison to NEWS and MEWS in the prediction of 30-day in-hospital mortality (98). In addition, learning algorithms have shown utility in predicting other sepsis related outcomes like hospital length of stay, ICU admission and re-admission rates (42,103).

ML techniques have also been used on existing statistical models to improve their accuracy. A study by Zhao *et al.* demonstrated that utilizing a gradient boosting machine model together with their constructed nomogram to predict 28-day mortality in septic patients admitted to the ED helped to improve the AUROC from 0.826 to 0.867 (104). When compared to logistic regression, neural networks outperformed logistic regression in predicting 28-day mortality among ED patients with sepsis, with AUROCs of 0.878 and 0.752, respectively (105). This highlights the potential of AI methods both as a standalone tool and as a valuable complement to traditional models to refine risk stratification which can lead to more targeted interventions.

Current limitations and future directions

Although AI has made significant strides in advancing healthcare in recent years and offers great promise, several limitations need to be considered as we embrace and integrate AI as a support tool. Much has been published about the general limitations of AI in healthcare (106,107) and many of these issues are also pertinent to the ED population.

From a technical perspective, the use of AI in diagnosis of sepsis has several challenges about its test characteristics and derivation. With a focus on high sensitivity for detecting sepsis, there is a risk for a high false positive rate resulting in alarm fatigue in the end users coupled with unnecessary testing (41,108). Additionally, because of the specificity of the model to the training dataset, there is lack of generalizability and external validity when applied to other patient populations or EDs with differing operating characteristics. Performance of ML models tend to decrease when used in an external validation cohort that is different from the derivation set, hence the need for recalibration and re-validation (26). This limits the usefulness of the AI tool outside of its original derivation and validation cohorts. Even within the same operating environment that the original AI tool was derived, disease patterns and prevalence may have sudden shifts such as in pandemic viral illnesses (108-110).

Not all studies have shown superiority of AI learning models over other prediction models for sepsis. Wang *et al.* utilized logistic regression to create a nomogram which showed comparative predictive power to AI models utilizing RF and stacking methods in predicting 30-day mortality in patients with sepsis (111). This would suggest that traditional analytic methods may still be advantageous in settings where advanced computational resources are less readily available (111). Additionally, in a single-center study by Knack *et al.*, physician gestalt was found to outperform ML learning models in identifying ED patients with sepsis (39).

Likewise, a study done by Kijpaisalratana *et al.* showed that despite the real-time MLASA system's success in expediting antibiotic administration, it did not significantly affect the length of stay or 30-day mortality rates in sepsis

patients (35). This highlights the complexity of sepsis management and implies that while AI tools may be able to successfully alert clinicians to sepsis and lead to faster initial responses, whether their integration into clinical workflows or decision-making processes have overall improvement in patient outcomes is still unknown.

With widespread use of AI tools in diagnosis and prompting management strategies, there is also a potential for automation bias (112), loss of clinician skill and gestalt in future generations if clinicians become over-reliant on AI and default to AI suggestions without rigorous clinical assessment and judgment (113). Conversely, there is also uncertainty from significant sections of practicing clinicians with regard to AI and lack of user acceptance (113,114). Many clinicians are wary of the "black-box" nature of AI prediction as the end-user does not understand how the algorithm makes predictions (115). This may eventually lead to poor uptake and utilization by unconvinced clinicians or those who are less savvy with information technology (60). A qualitative study was done by Sandhu *et al.* to investigate the factors affecting the adoption of an ML sepsis early warning system by frontline clinicians into regular clinical practice (116). They found that a barrier to the integration of ML models into clinical workflows was the clinicians' unfamiliarity with such systems. This negatively influenced their perceived accuracy of ML-based CDS tools, resulting in a lack of trust in the system and hence, hesitancy in adoption. Addressing such gaps in their knowledge and creating follow-up feedback loops on patient encounters to demonstrate the utility of the ML models could address this barrier. Facilitators of adoption included ease-of-use of the application and having a human intermediary (dedicated nurses) to review and discuss the ML model recommendations with the physicians.

Lee *et al.* described the implementation of a sepsis monitoring platform that alerted ED physicians to incomplete components of the sepsis bundle (117). They sought to minimize alert fatigue while still maintaining the clinical relevance of the alerts. More than one-third of patients had at least one alert sent, supporting the clinical need for such a system. However, of the missing sepsis bundle components for which an alert was sent, only 38.2% were successfully completed on time. Further studies would need to be done on the integration of such systems into clinical practice, and their impact on actual patient outcomes. Moreover, the practice of medicine is as much an art as a science with empathetic human physicians caring for patients and recommending investigations and treatments in their best interest. One of the pillars of evidence-based healthcare in considering patients' values and preferences are usually not considered building AI models. It may be challenging for patients to accept treatment suggestions by ML algorithms without clinicians to ensure their safety and to preserve their autonomy (118).

Additionally, since ML algorithms require learning from large datasets, it may also potentiate underlying misrepresentation bias from datasets that might have precluded certain patient groups and ethnic minorities (119). Algorithmic bias can also be present in ML, when higher disease severity is attributed to lower socioeconomic classes who may, in fact, be receiving less treatment due to costs and healthcare accessibility issues (119). The datasets for ML learning may also have unequal representation from all countries, leading to under-representation of certain nationalities and healthcare systems (120). On the other hand, if ML algorithms can be trained purposefully, it has the ability to help identify social disparities and patients at-risk of under-treatment, allowing resources allocation and social support (121,122).

From our review of the literature, many of the studies looking at AI in sepsis are observational studies, of which many are retrospective in nature. The use of retrospective datasets for ML learning and testing would mean that training sets have inherent information bias due to erroneous diagnosis coding and may therefore affect ML models' performance. Most were focused on test performance characteristics though there were some that explored prognostication and investigated patient-centered outcomes such as mortality or length of hospitalization. Additionally, all but one of the studies evaluated AUROC, which can remain high even when the model performs poorly on the minority class, as in septic cases; the calculation of precision-recall area-under-curve could have better reflected the performance of ML models in such imbalanced datasets (123). There is also a lack of high quality prospective randomized controlled trials with only two such studies (Kijpaisalratana and Tarabichi) (35,124) in our review of the literature. As such, while the data is promising and has shown improvements in areas such as time to antibiotic administration or completion of sepsis care bundles, there is still inconclusive evidence that this translates to improved patient outcomes (35). A systematic review or meta-analysis in this field may also add further scientific information to the current body of knowledge but the heterogeneity of studies at the current stage may limit pooling of results.

Finally, there are several other factors to consider before any real-world implementation of AI technologies, especially in medicine. Practical, ethical and regulatory issues such as data sharing, privacy, transparency of algorithms, data standardization, interoperability across multiple platforms, and patient safety need to be addressed (125). Regional variability and current lack of consensus may hinder the adoption of any new AI technologies in the medical field (126). Therefore, establishing validated frameworks and guidelines that cover transparency, reproducibility, ethics, effectiveness, and engagement will be required before application in the clinical realm (127).

Conclusions

AI holds considerable potential in revolutionizing management of septic patients in the ED. It offers great promise as a clinical support tool and has shown significant improvements in measured performance markers. However, its application is currently still constrained by inherent limitations. The complexity of sepsis diagnosis, risk of over-reliance on AI, dynamic nature of sepsis, and ethical and legal considerations underscore the indispensable role of human expertise and clinical judgment in complementing AI-driven approaches. Moving forward, a balanced integration of AI technologies with clinician input and oversight is essential to harness the full potential of AI while ensuring optimal patient outcomes in sepsis management.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-24-150/rc

Peer Review File: Available at https://atm.amegroups.com/article/view/10.21037/atm-24-150/prf

Funding: None.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-24-150/coif). M.T.C. serves as an unpaid editorial board member of *Annals of Translational Medicine* from November 2024 to October 2026. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315:801-10.
2. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. Chest 1992;101:1644-55.
3. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet 2020;395:200-11.
4. Wang HE, Jones AR, Donnelly JP. Revised National Estimates of Emergency Department Visits for Sepsis in the United States. Crit Care Med 2017;45:1443-9.
5. PRISM Investigators, Rowan KM, Angus DC, et al. Early, Goal-Directed Therapy for Septic Shock - A Patient-Level Meta-Analysis. N Engl J Med 2017;376:2223-34.
6. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021;47:1181-247.
7. Vincent JL. The Clinical Challenge of Sepsis Identification and Monitoring. PLoS Med 2016;13:e1002022.
8. Green SM, Schriger DL, Yealy DM. Methodologic standards for interpreting clinical decision rules in emergency medicine: 2014 update. Ann Emerg Med

2014;64:286-91.
9. Hamilton F, Arnold D, Baird A, et al. Early Warning Scores do not accurately predict mortality in sepsis: A meta-analysis and systematic review of the literature. J Infect 2018;76:241-8.
10. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315:762-74.
11. Song JU, Sin CK, Park HK, et al. Performance of the quick Sequential (sepsis-related) Organ Failure Assessment score as a prognostic tool in infected patients outside the intensive care unit: a systematic review and meta-analysis. Crit Care 2018;22:28.
12. Helm JM, Swiergosz AM, Haeberle HS, et al. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Curr Rev Musculoskelet Med 2020;13:69-76.
13. Mueller B, Kinoshita T, Peebles A, et al. Artificial intelligence and machine learning in emergency medicine: a narrative review. Acute Med Surg 2022;9:e740.
14. Baştanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol 2014;1107:105-28.
15. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19:64.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
17. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25:24-9.
18. Medina-Merino R, Ñique-Chacón C. Random Forests as an extension of the classification trees with the R and Python programs. Interfases Internet 2017;10:165-89.
19. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21.
20. Reyna MA, Josef CS, Jeter R, et al. Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. Crit Care Med 2020;48:210-7.
21. Sánchez Fernández I, Peters JM. Machine learning and deep learning in medicine and neuroimaging. Ann Child Neurol Soc 2023;1:102-22.
22. Mao Q, Jay M, Hoffman JL, 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:e017833.
23. Rhee C, Dantes R, Epstein L, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. JAMA 2017;318:1241-9.
24. Filbin MR, Lynch J, Gillingham TD, et al. Presenting Symptoms Independently Predict Mortality in Septic Shock: Importance of a Previously Unmeasured Confounder. Crit Care Med 2018;46:1592-9.
25. Filbin MR, Thorsen JE, Zachary TM, et al. Antibiotic Delays and Feasibility of a 1-Hour-From-Triage Antibiotic Requirement: Analysis of an Emergency Department Sepsis Quality Improvement Database. Ann Emerg Med 2020;75:93-9.
26. Lin PC, Chen KT, Chen HC, et al. Machine Learning Model to Identify Sepsis Patients in the Emergency Department: Algorithm Development and Validation. J Pers Med 2021;11:1055.
27. Aygun U, Yagin FH, Yagin B, et al. Assessment of Sepsis Risk at Admission to the Emergency Department: Clinical Interpretable Prediction Model. Diagnostics (Basel) 2024;14:457.
28. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the Emergency Department. Am J Emerg Med 2019;37:1490-7.
29. Delahanty RJ, Alvarez J, Flynn LM, et al. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. Ann Emerg Med 2019;73:334-44.
30. Niemantsverdriet MSA, de Hond TAP, Hoefer IE, et al. A machine learning approach using endpoint adjudication committee labels for the identification of sepsis predictors at the emergency department. BMC Emerg Med 2022;22:208.
31. Brann F, Sterling NW, Frisch SO, et al. Sepsis Prediction at Emergency Department Triage Using Natural Language Processing: Retrospective Cohort Study. JMIR AI 2024;3:e49784.
32. Levy MM, Rhodes A, Phillips GS, et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. Intensive Care Med 2014;40:1623-33.
33. Brown SM, Jones J, Kuttler KG, et al. Prospective evaluation of an automated method to identify patients with severe sepsis or septic shock in the emergency department. BMC Emerg Med 2016;16:31.
34. Kijpaisalratana N, Sanglertsinlapachai D, Techaratsami S, et al. Machine learning algorithms for early sepsis detection in the emergency department: A retrospective study. Int J Med Inform 2022;160:104689.
35. Kijpaisalratana N, Saoraya J, Nhuboonkaew P, et al.

Real-time machine learning-assisted sepsis alert enhances the timeliness of antibiotic administration and diagnostic accuracy in emergency department patients with sepsis: a cluster-randomized trial. Intern Emerg Med 2024;19:1415-24.
36. Prasad V, Aydemir B, Kehoe IE, et al. Diagnostic suspicion bias and machine learning: Breaking the awareness deadlock for sepsis detection. PLOS Digit Health 2023;2:e0000365.
37. Bedoya AD, Futoma J, Clement ME, et al. Machine learning for early detection of sepsis: an internal and temporal validation study. JAMIA Open 2020;3:252-60.
38. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 2017;6:e000158.
39. Knack SKS, Scott N, Driver BE, et al. Early Physician Gestalt Versus Usual Screening Tools for the Prediction of Sepsis in Critically Ill Emergency Patients. Ann Emerg Med 2024;84:246-58.
40. Upadhyaya DP, Tarabichi Y, Prantzalos K, et al. Machine Learning Interpretability Methods to Characterize the Importance of Hematologic Biomarkers in Prognosticating Patients with Suspected Infection. Computers in Biology and Medicine 2024;183:109251.
41. Shashikumar SP, Wardi G, Malhotra A, et al. Artificial intelligence sepsis prediction algorithm learns to say "I don't know". npj Digital Medicine volume 2021;4:134.
42. Taneja I, Damhorst GL, Lopez-Espina C, et al. Diagnostic and prognostic capabilities of a biomarker and EMR-based machine learning algorithm for sepsis. Clin Transl Sci 2021;14:1578-89.
43. Velly L, Volant S, Fitting C, et al. Optimal combination of early biomarkers for infection and sepsis diagnosis in the emergency department: The BIPS study. J Infect 2021;82:11-21.
44. Kwon JM, Lee YR, Jung MS, et al. Deep-learning model for screening sepsis using electrocardiography. Scand J Trauma Resusc Emerg Med 2021;29:145.
45. Horng S, Sontag DA, Halpern Y, et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One 2017;12:e0174708.
46. Court O, Kumar A, Parrillo JE, et al. Clinical review: Myocardial depression in sepsis and septic shock. Crit Care 2002;6:500-8.
47. Rich MM, McGarvey ML, Teener JW, et al. ECG changes during septic shock. Cardiology 2002;97:187-96.
48. Pierrakos C, Vincent JL. Sepsis biomarkers: a review. Crit Care 2010;14:R15.
49. Aguirre U, Urrechaga E. Diagnostic performance of machine learning models using cell population data for the detection of sepsis: a comparative study. Clin Chem Lab Med 2023;61:356-65.
50. Buchman TG, Simpson SQ, Sciarretta KL, et al. Sepsis Among Medicare Beneficiaries: 1. The Burdens of Sepsis, 2012-2018. Crit Care Med 2020;48:276-88.
51. Lat I, Coopersmith CM, De Backer D, et al. The surviving sepsis campaign: fluid resuscitation and vasopressor therapy research priorities in adult patients. Intensive Care Med Exp 2021;9:10.
52. Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. Intensive Care Med 2018;44:925-8.
53. Seymour CW, Gesten F, Prescott HC, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med 2017;376:2235-44.
54. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med 2017;43:304-77.
55. Rhee C, Filbin MR, Massaro AF, et al. Compliance With the National SEP-1 Quality Measure and Association With Sepsis Outcomes: A Multicenter Retrospective Cohort Study. Crit Care Med 2018;46:1585-91.
56. Mah JW, Bingham K, Dobkin ED, et al. Mannequin simulation identifies common surgical intensive care unit teamwork errors long after introduction of sepsis guidelines. Simul Healthc 2009;4:193-9.
57. Manaktala S, Claypool SR. Evaluating the impact of a computerized surveillance algorithm and decision support system on sepsis mortality. J Am Med Inform Assoc 2017;24:88-95.
58. Arabi YM, Al-Dorzi HM, Alamry A, et al. The impact of a multifaceted intervention including sepsis electronic alert system and sepsis response team on the outcomes of patients with sepsis and septic shock. Ann Intensive Care 2017;7:57.
59. Islam MM, Nasrin T, Walther BA, et al. Prediction of sepsis patients using machine learning approach: A meta-analysis. Comput Methods Programs Biomed 2019;170:1-9.
60. Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17.

61. Gilboy N, Tanabe P, Travers D, et al. Emergency severity index, version 4: implementation handbook. In: Rockville, MD: Agency for Healthcare Research and Quality; 2005.
62. DeMerle KM, Angus DC, Baillie JK, et al. Sepsis Subclasses: A Framework for Development and Interpretation. Crit Care Med 2021;49:748-59.
63. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA 2019;321:2003-17.
64. Boussina A, Wardi G, Shashikumar SP, et al. Representation Learning and Spectral Clustering for the Development and External Validation of Dynamic Sepsis Phenotypes: Observational Cohort Study. J Med Internet Res 2023;25:e45614.
65. Chang H, Yu JY, Yoon S, et al. Machine learning-based suggestion for critical interventions in the management of potentially severe conditioned patients in emergency department triage. Sci Rep 2022;12:10537.
66. Rangel-Frausto MS, Pittet D, Costigan M, et al. The natural history of the systemic inflammatory response syndrome (SIRS). A prospective study. JAMA 1995;273:117-23.
67. Taniguchi LU, Pires EMC, Vieira JM Jr, et al. Systemic inflammatory response syndrome criteria and the prediction of hospital mortality in critically ill patients: a retrospective cohort study. Rev Bras Ter Intensiva 2017;29:317-24.
68. Dulhunty JM, Lipman J, Finfer S, et al. Does severe non-infectious SIRS differ from severe sepsis? Results from a multi-centre Australian and New Zealand intensive care unit study. Intensive Care Med 2008;34:1654-61.
69. Wang C, Xu R, Zeng Y, et al. A comparison of qSOFA, SIRS and NEWS in predicting the accuracy of mortality in patients with suspected sepsis: A meta-analysis. PLoS One 2022;17:e0266755.
70. Freund Y, Lemachatti N, Krastinova E, et al. Prognostic Accuracy of Sepsis-3 Criteria for In-Hospital Mortality Among Patients With Suspected Infection Presenting to the Emergency Department. JAMA 2017;317:301-8.
71. Jiang J, Yang J, Mei J, et al. Head-to-head comparison of qSOFA and SIRS criteria in predicting the mortality of infected patients in the emergency department: a meta-analysis. Scand J Trauma Resusc Emerg Med 2018;26:56.
72. Ruan H, Ke D, Liao D. Prognostic Accuracy of qSOFA and SIRS for Mortality in the Emergency Department: A Meta-Analysis and Systematic Review of Prospective Studies. Emerg Med Int 2022;2022:1802707.
73. Smith GB, Prytherch DR, Meredith P, et al. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation 2013;84:465-70.
74. Wei S, Xiong D, Wang J, et al. The accuracy of the National Early Warning Score 2 in predicting early death in prehospital and emergency department settings: a systematic review and meta-analysis. Ann Transl Med 2023;11:95.
75. Hsieh MS, Chiu KC, Chattopadhyay A, et al. Utilizing the National Early Warning Score 2 (NEWS2) to confirm the impact of emergency department management in sepsis patients: a cohort study from taiwan 1998-2020. Int J Emerg Med 2024;17:42.
76. Sardidi H, Bawazeer D, Alhafi M, et al. The Use of the Initial National Early Warning Score 2 at the Emergency Department as a Predictive Tool of In-Hospital Mortality in Hemodialysis Patients. Cureus 2023;15:e39678.
77. Almutary A, Althunayyan S, Alenazi K, et al. National Early Warning Score (NEWS) as Prognostic Triage Tool for Septic Patients. Infect Drug Resist 2020;13:3843-51.
78. Welch J, Dean J, Hartin J. Using NEWS2: an essential component of reliable clinical assessment. Clin Med (Lond) 2022;22:509-13.
79. Islam KR, Prithula J, Kumar J, et al. Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review. J Clin Med 2023;12:5658.
80. Chang H, Jung W, Ha J, et al. Early prediction of unexpected latent shock in the emergency department using vital signs. Shock 2023;60:373-8.
81. Kim J, Chang H, Kim D, et al. Machine learning for prediction of septic shock at initial triage in emergency department. J Crit Care 2020;55:163-70.
82. Yun H, Park JH, Choi DH, et al. Enhancement in Performance of Septic Shock Prediction Using National Early Warning Score, Initial Triage Information, and Machine Learning Analysis. J Emerg Med 2021;61:1-11.
83. Wardi G, Carlile M, Holder A, et al. Predicting Progression to Septic Shock in the Emergency Department Using an Externally Generalizable Machine-Learning Algorithm. Ann Emerg Med 2021;77:395-406.
84. Karlsson A, Stassen W, Loutfi A, et al. Predicting mortality among septic patients presenting to the emergency department-a cross sectional analysis using machine learning. BMC Emerg Med 2021;21:84.
85. Rodríguez A, Mendoza D, Ascuntar J, et al. Supervised classification techniques for prediction of mortality in adult patients with sepsis. Am J Emerg Med

2021;45:392-7.
86. Cheng CY, Kung CT, Chen FC, et al. Machine learning models for predicting in-hospital mortality in patient with sepsis: Analysis of vital sign dynamics. Front Med (Lausanne) 2022;9:964667.
87. Greco M, Caruso PF, Spano S, et al. Machine Learning for Early Outcome Prediction in Septic Patients in the Emergency Department. Algorithms 2023;16:76.
88. Raven W, de Hond A, Bouma LM, et al. Does machine learning combined with clinical judgment outperform clinical judgment alone in predicting in-hospital mortality in old and young suspected infection emergency department patients? Eur J Emerg Med 2023;30:205-6.
89. van Doorn WPTM, Stassen PM, Borggreve HF, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. PLoS One 2021;16:e0245157.
90. Perng JW, Kao IH, Kung CT, et al. Mortality Prediction of Septic Patients in the Emergency Department Based on Machine Learning. J Clin Med 2019;8:1906.
91. Wong BPK, Lam RPK, Ip CYT, et al. Applying artificial neural network in predicting sepsis mortality in the emergency department based on clinical features and complete blood count parameters. Sci Rep 2023;13:21463.
92. Jeon ET, Song J, Park DW, et al. Mortality prediction of patients with sepsis in the emergency department using machine learning models: a retrospective cohort study according to the Sepsis-3 definitions. Signa Vitae 2023;19;112-24.
93. Park SW, Yeo NY, Kang S, et al. Early Prediction of Mortality for Septic Patients Visiting Emergency Room Based on Explainable Machine Learning: A Real-World Multicenter Study. J Korean Med Sci 2024;39:e53.
94. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Jones A, editor. Acad Emerg Med 2016;23:269-78.
95. Kwon YS, Baek MS. Development and Validation of a Quick Sepsis-Related Organ Failure Assessment-Based Machine-Learning Model for Mortality Prediction in Patients with Suspected Infection in the Emergency Department. J Clin Med 2020;9:875.
96. Katz S, Suijker J, Hardt C, et al. Decision support system and outcome prediction in a cohort of patients with necrotizing soft-tissue infections. Int J Med Inform 2022;167:104878.
97. Ko BS, Jeon S, Son D, et al. Machine Learning Model Development and Validation for Predicting Outcome in Stage 4 Solid Cancer Patients with Septic Shock Visiting the Emergency Department: A Multi-Center, Prospective Cohort Study. J Clin Med 2022;11:7231.
98. Chiew CJ, Liu N, Tagami T, et al. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. Medicine (Baltimore) 2019;98:e14197.
99. Glickman SW, Cairns CB, Otero RM, et al. Disease progression in hemodynamically stable patients presenting to the emergency department with sepsis. Acad Emerg Med 2010;17:383-90.
100. Capp R, Horton CL, Takhar SS, et al. Predictors of patients who present to the emergency department with sepsis and progress to septic shock between 4 and 48 hours of emergency department arrival. Crit Care Med 2015;43:983-8.
101. Cohen J, Vincent JL, Adhikari NK, et al. Sepsis: a roadmap for future research. Lancet Infect Dis 2015;15:581-614.
102. Do SN, Dao CX, Nguyen TA, et al. Sequential Organ Failure Assessment (SOFA) Score for predicting mortality in patients with sepsis in Vietnamese intensive care units: a multicentre, cross-sectional study. BMJ Open 2023;13:e064870.
103. Tschoellitsch T, Seidl P, Böck C, et al. Using emergency department triage for machine learning-based admission and mortality prediction. Eur J Emerg Med 2023;30:408-16.
104. Zhao C, Wei Y, Chen D, et al. Prognostic value of an inflammatory biomarker-based clinical algorithm in septic patients in the emergency department: An observational study. Int Immunopharmacol 2020;80:106145.
105. Jaimes F, Farbiarz J, Alvarez D, et al. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. Crit Care 2005;9:R150-6.
106. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17:195.
107. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull 2021;139:4-15.
108. Serigstad S, Markussen DL, Ritz C, et al. The changing spectrum of microbial aetiology of respiratory tract infections in hospitalized patients before and during the COVID-19 pandemic. BMC Infect Dis 2022;22:763.
109. Ankert J, Hagel S, Schwarz C, et al. Streptococcus

pneumoniae re-emerges as a cause of community-acquired pneumonia, including frequent co-infection with SARS-CoV-2, in Germany, 2021. ERJ Open Res 2023;9:00703-2022.
110. Hyams C, Challen R, Begier E, et al. Incidence of community acquired lower respiratory tract disease in Bristol, UK during the COVID-19 pandemic: A prospective cohort study. Lancet Reg Health Eur 2022;21:100473.
111. Wang B, Chen J, Pan X, et al. A nomogram for predicting mortality risk within 30 days in sepsis patients admitted in the emergency department: A retrospective analysis. PLoS One 2024;19:e0296456.
112. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19:121-7.
113. Stewart J, Freeman S, Eroglu E, et al. Attitudes towards artificial intelligence in emergency medicine. Emerg Med Australas 2024;36:252-65.
114. Sivaraman V, Bukowski LA, Levin J, et al. Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 2023:1-18.
115. Wadden JJ. Defining the undefinable: the black box problem in healthcare artificial intelligence. J Med Ethics 2021. [Epub ahead of print]. doi: 10.1136/medethics-2021-107529.
116. Sandhu S, Lin AL, Brajer N, et al. Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study. J Med Internet Res 2020;22:e22421.
117. Lee AH, Aaronson E, Hibbert KA, et al. Design and Implementation of a Real-time Monitoring Platform for Optimal Sepsis Care in an Emergency Department: Observational Cohort Study. J Med Internet Res 2021;23:e26946.
118. Richardson JP, Smith C, Curtis S, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021;4:140.
119. Norori N, Hu Q, Aellen FM, et al. Addressing bias in big data and AI for health care: A call for open science. Patterns (N Y) 2021;2:100347.
120. Celi LA, Cellini J, Charpignon ML, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities-A global review. PLOS Digital Health 2022;1:e0000022.
121. Berdahl CT, Baker L, Mann S, et al. Strategies to Improve the Impact of Artificial Intelligence on Health Equity: Scoping Review. JMIR AI 2023;2:e42936.
122. Guevara M, Chen S, Thomas S, et al. Large language models to identify social determinants of health in electronic health records. NPJ Digit Med 2024;7:6.
123. Tharwat A. Classification assessment methods. Appl Comput Inform 2021;17:168-92.
124. Tarabichi Y, Cheng A, Bar-Shain D, et al. Improving Timeliness of Antibiotic Administration Using a Provider and Pharmacist Facing Sepsis Early Warning System in the Emergency Department Setting: A Randomized Controlled Quality Improvement Initiative. Crit Care Med 2022;50:418-27.
125. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25:30-6.
126. Schmidt J, Schutte NM, Buttigieg S, et al. Mapping the regulatory landscape for artificial intelligence in health within the European Union. NPJ Digit Med 2024;7:229.
127. Crossnohere NL, Elsaid M, Paskett J, et al. Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks. J Med Internet Res 2022;24:e36823.

Cite this article as: Chua MT, Boon Y, Lee ZY, Kok JHJ, Lim CKW, Cheung NMT, Yong LPX, Kuan WS. The role of artificial intelligence in sepsis in the Emergency Department: a narrative review. Ann Transl Med 2025;13(1):4. doi: 10.21037/atm-24-150