



A scoping review on pediatric sepsis prediction technologies in healthcare



Ryan Tennant¹✉, Jennifer Graham², Juliet Kern¹, Kate Mercer^{1,3}, J. Mark Ansermino^{4,5} & Catherine M. Burns¹

This scoping review evaluates recent advancements in data-driven technologies for predicting non-neonatal pediatric sepsis, including artificial intelligence, machine learning, and other methodologies. Of the 27 included studies, 23 (85%) were single-center investigations, and 16 (59%) used logistic regression. Notably, 20 (74%) studies used datasets with a low prevalence of sepsis-related outcomes, with area under the receiver operating characteristic scores ranging from 0.56 to 0.99. Prediction time points varied widely, and development characteristics, performance metrics, implementation outcomes, and considerations for human factors—especially workflow integration and clinical judgment—were inconsistently reported. The variations in endpoint definitions highlight the potential significance of the 2024 consensus criteria in future development. Future research should strengthen the involvement of clinical users to enhance the understanding and integration of human factors in designing and evaluating these technologies, ultimately aiming for safe and effective integration in pediatric healthcare.

Sepsis is typically characterized by a dysregulated immune system response to bodily pathogens^{1,2}. While it can impact individuals of all ages, sepsis poses a particularly grave threat to children³. In 2017, approximately 20.3 million new cases of sepsis and 2.9 million deaths occurred in children under five years of age, where the burden of sepsis is greater in low-resource settings, as well as rural and remote communities^{4–6}. Prevention, early recognition, and treatment are critical, as sepsis can significantly damage various organ systems⁷. Following a diagnosis, one-third of children may experience a new disability or fail to fully recover their previous health-related quality of life after one to three months^{8–10}. The World Health Organization has positioned sepsis prevention, recognition, and treatment as a global priority, and there remain significant opportunities to improve pediatric sepsis prediction through the appropriate design and implementation of digitalized clinical tools^{3,11}, including data-driven algorithms, machine learning (ML), and artificial intelligence (AI).

Compared with adults and neonates, the design of technologies for predicting pediatric sepsis—the focus of this review—is uniquely challenging, with emergency departments potentially missing 7.1% of cases in children aged 1–19 years old¹². The sources and development of infection, maturation-based differences, age-based clinical differences, commonalities with other conditions and syndromes, and diagnostic

modalities contribute to these complexities. For example, compared to adults, the absence of standardized diagnostic criteria and age-adjusted cardiovascular and respiratory vital signs make it challenging to distinguish symptom presentation from heterogeneous clinical baselines in children with comorbidities, common febrile infections, and other syndromes such as Kawasaki disease and multisystem inflammatory syndrome in children¹³. Recognizing early signs of sepsis in children is also complicated by their limited ability to articulate their symptoms and their capacity to maintain physiological function until reaching an advanced stage of septic shock compared to adults¹⁴.

Among children, sepsis etiology significantly differs by age and varies globally¹⁵. In neonates, different bacterial and viral pathogens are more commonly acquired in-hospital within high-income countries and through vertical transmission from the vaginal tract, or community-acquired in low- and middle-income countries¹⁵. In older children, the most common infection sources are through the respiratory tract and bloodstream, and viral sepsis accounts for almost one-third of childhood sepsis cases⁵. The neonatal adaptive and innate immune response is also distinctly altered compared to the developing immune system of older children; it is further complicated by differences in clinical signs and biomarkers^{16–18}. For diagnosis, many clinicians prioritize microbiological findings over nonspecific

¹Department of Systems Design Engineering, Faculty of Engineering, University of Waterloo, 200 University Avenue West, Waterloo, N2L3G1 Ontario, Canada.

²Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo, N2L3G1 Ontario, Canada. ³Library, University of Waterloo, 200 University Avenue West, Waterloo, N2L3G1 Ontario, Canada. ⁴Centre for International Child Health, British Columbia Children's Hospital, 305-4088 Cambie Street, Vancouver, V5Z2X8 British Columbia, Canada. ⁵Department of Anesthesiology, The University of British Columbia, 950 West 28th Avenue, Vancouver, V5Z4H4 British Columbia, Canada. ✉e-mail: drtennan@uwaterloo.ca

Table 1 | Characteristics of the included studies

	Studies (n = 27), n (%) ^a
Publication year	
2017–2020	6 (22.2)
2021–2024	21 (77.8)
Quality Assessment (scale: levels 1 to 5)	
Level 2: well-designed controlled trial without randomization; prospective comparative cohort trial	9 (33.3)
Level 3: case-control studies; retrospective cohort study	21 (77.8)
Expected automation design	
Analysis automation	9 (33.3)
Decision automation	18 (66.7)
Expected data flow	
Provider entered data	25 (92.6)
Provider entered & acquisition automation	2 (7.4)
Outcome(s)	
Sepsis	20 (74.0)
Severe Sepsis	4 (14.8)
Septic Shock	10 (37.0)
Healthcare setting	
Emergency department ^b	17 (63.0)
Intensive care unit ^c	15 (55.6)
Inpatient unit ^d	6 (22.2)
Implemented	
Yes ^e	5 (18.5)
No	22 (81.5)
Included study sites	
Single-site study	23 (85.2)
Multi-site study	5 (18.5)
Dataset sources	
United States of America	16 (59.3)
Unspecified State	5 (18.5)
Colorado	3 (11.1)
California	2 (7.4)
Maryland	2 (7.4)
Virginia	2 (7.4)
Illinois	1 (3.7)
Massachusetts	1 (3.7)
Pennsylvania	1 (3.7)
Tennessee	1 (3.7)
Washington	1 (3.7)
China	5 (18.5)
Zhejiang	3 (11.1)
Anhui	1 (3.7)
Shanghai	1 (3.7)
Spain	2 (7.4)
Australia	1 (3.7)
Dhaka, Bangladesh	1 (3.7)
Medellin, Colombia	1 (3.7)
Germany	1 (3.7)
Nairobi/Kilifi, Kenya	1 (3.7)
Mexico City, Mexico	1 (3.7)

Table 1 (continued) | Characteristics of the included studies

	Studies (n = 27), n (%) ^a
Jinja, Uganda	1 (3.7)
New Zealand	1 (3.7)

^aPercentages may sum to more than 100 due to overlapping categories.
^bIncludes pediatric emergency departments (n = 10), emergency care sites (n = 1), urgent care (n = 1), general emergency departments (n = 3), dual-track emergency department/urgent care (n = 2).
^cIncludes pediatric intensive care units (n = 11), surgical intensive care units (n = 1), mixed intensive care units (n = 1), and general intensive care units (n = 2).
^dIncludes pediatric inpatient units (n = 3), general inpatient units (n = 1), hematology inpatients (n = 1), and general wards (n = 1).
^eIncludes implemented electronically (n = 3), nonelectronically with the potential to be digitalized (n = 1) and silent implementation (n = 1). Note that analysis automation is a second level of automation where algorithms or calculations are applied to incoming data to enable a prediction. Decision automation is a third level of automation where options or additional information to support decision-making is provided. Acquisition automation is where information is collected without human intervention, such as through continuous physiological sensors or monitors.

organ dysfunction criteria in neonates^{19–21}, whereas multiorgan dysfunction is prioritized in pediatric and adult populations^{1,22}.

The current deployment of data-driven technologies focuses mainly on adult and neonatal sepsis prediction^{23,24}, neglecting the non-neonatal pediatric age group. This age group has also been neglected in previous reviews^{25–36}. Indeed, existing systematic and scoping reviews often exclude young infants and adolescents^{27–31,33,37,38} while also excluding the engineering disciplines from their searches^{25,30,31,39,40} and limiting their scope to specific ML tools⁴¹. Reviews that have not excluded young infants have not been comprehensive, reviewing only one²⁵ or three^{36,42} non-neonatal pediatric articles, thereby lacking the specificity needed to inform technology development for this intermediate pediatric age range. However, prior reviews highlight the potential for non-invasive predictive measures³⁰ and the possibility of enhancing early detection of sepsis by integrating prediction models with clinical judgment³¹. They also identify barriers to implementation, including the risk of bias, overfitting, and the absence of standardized validation protocols³². Yet, beyond the absence of a comprehensive pediatric-focused review, it is also important to recognize that most of these prior reviews primarily report on quantitative performance metrics. These reviews do not consider how these systems are designed for the clinical environment or the factors that may facilitate or hinder effective clinician interaction, such as situational awareness, interface design, and the timing of an alert^{25,27–36}. With the anticipated widespread use of AI in healthcare in the coming decade, a thorough analysis of the existing literature on data-driven technologies for predicting pediatric sepsis is essential for directing research efforts to ensure appropriate and effective integration for this age group.

The literature on pediatric sepsis prediction technologies is highly varied, with differences in study design, development approaches, implementation stages, and evaluation methods. This variability highlights the importance of assessing the potential scope and depth of the available pediatric sepsis prediction research. Importantly, in this scoping review, we address the limitations of prior reviews by exploring both the data-driven development aspects and the human factors considerations of sepsis, severe sepsis, and septic shock prediction systems in the pediatric population. Focusing on young infants, children, and adolescents, the objective of this review is two-fold: (1) to explore current design strategies and performance metrics of developed technologies across healthcare settings, thereby providing valuable insights for future research and development, and (2) to critically analyze human factors aspects within the literature to inform current and future integration needs and research directions that support safe and effective clinical use. Our research questions are specifically designed to address the previously identified gaps and objectives and inform future clinical applications of this technology: (1) What are the

current design and performance characteristics of data-driven pediatric sepsis technologies? (2) How have developed systems considered human factors in their design for clinical use? (3) What specific challenges and research gaps need to be addressed to inform the future clinical implementation of these technologies?

Results

There were 36,757 records identified. After removing 9,586 duplicates and screening titles and abstracts, 163 records were assessed for inclusion, resulting in 27 articles in the final review (Supplementary Fig. 1). Most articles (23 [85%]) involved single-center datasets (Table 1). Five (19%) were multi-site studies: one combined datasets from multiple continents, including low-middle-income countries⁴³, and four (15%) used more than one healthcare site within North America or Australia^{44–46}. There were 17 (63%) articles using development datasets from the continent of North America, six (22%) from Asia, three (11%) from Europe, two (7%) from Africa, and two (7%) from Oceania. Included healthcare settings were mostly emergency departments (17 [63%]) and intensive care units (15 [56%]). Six (22%) articles included inpatient units. The completed extraction template is included in Supplementary Data 1, with the categorization and organization of the model features in Supplementary Data 2.

Endpoint definitions

More than half (20 [74%]) of the articles focused on predicting sepsis, compared to severe sepsis and septic shock. International consensus definitions and guidelines were used to define sepsis endpoints in most articles (16 [59%] of 27), including Goldstein et al.⁴⁷, Rhodes et al.⁴⁸, Singer et al.²², and Weiss et al.¹⁴, as well as criteria from Eisenberg et al.⁴⁹, Scott et al.⁵⁰, and Sepanski et al.⁵¹, with modifications (Table 2)^{52,53}. Other definitions included treatment or diagnostic codes (8 [30%] of 27). Less frequent approaches were mortality or extracorporeal membrane oxygenation use, Delphi processes, and outcomes from another screening algorithm^{43,54–56}. One article led to the most recent international consensus criteria for pediatric sepsis and septic shock: the Phoenix Sepsis Score⁴³.

We noted variability in how clinicians defined sepsis^{46,57,58}. Changes in treatment practices compared between when the data was collected and when the model was developed were also noted, as were expected changes in diagnostic definitions requiring model re-validation^{59,60}. One article mentioned the risk of including overtreated patients through the use of intention and treatment-based definitions⁶¹.

Model development approaches

More than half of the articles (16 [59%]) used logistic regression (Table 3)^{45,46,53–58,60–68}. Others used gradient boosting (6 [22%])^{52,57,69–72} and random forest (4 [15%])^{55,59,60,62}. Less common approaches were support vector machines or maximization (2 [7%])^{62,72}, stacked or classification and regression tree modelling (2 [7%])^{43,62}, tree augmented or Naïve Bayes (3 [11%])^{58,62,72}, elastic net regularization (1 [4%])⁴⁴, neural networks (1 [4%])⁵⁵, and empirical derivations for an age and temperature-adjusted index score (1 [4%])⁷³.

Patient demographics

We extracted the sample size based on the number of patients from most articles (19 [70%]), which had a median of 1238 (IQR 289–4403). Some reported the number of visits (3 [11%]), encounters (5 [19%]), or admissions (2 [7%]), which had a combined median of 35330 (IQR 2464–141510). Many articles (16 [59%]) reported median age statistics, ranging from 13.1 months to 10 years. One article specifically tested their models on different age groups, such as one-month to one-year-olds and those greater than one year old⁵⁸. 22 articles (81%) reported patient sex, and 10 (37%) included race or ethnicity data. Ten articles (37%) referenced the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines^{43–46,54,58,59,61,64,65}.

Most articles (20 [74%]) had a class imbalance in their datasets, with a low prevalence of sepsis-related outcomes (0.05% to 19.38%). Five articles (19%) used datasets with more than half of the patients classified with a sepsis-related outcome, with a prevalence ranging from 50.80% to 74.50% and sample sizes ranging from 65 to 667^{64–66,68,72}. One article normalized a pediatric sepsis dataset to an adult sepsis dataset because of the lack of readily available data⁷¹.

Other patient risk factor characteristics were included in 16 (59%) articles. For example, one study focused on septic shock outcomes in a post-chemotherapy population⁶³, and another included Kawasaki disease as a differentiating outcome from sepsis⁶⁸. Other health and socio-demographic risk factors that were characterized across studies were those related to immunodeficiency or immunosuppression^{56,61}, blastomas, bone marrow transplantation, aplastic anemia, organ transplantation^{57,67}, healthcare utilization and diagnoses history^{44,45,69}, hospital category (e.g., quaternary, dedicated pediatric, mixed) or resource setting^{43,46}, surgical status⁵⁹, complex and chronic conditions and comorbidities^{43–45,56,70}, history of prematurity^{56,58}, insurance status^{44,45,61,62}, weight or malnutrition^{43,46,54}, and medical technology needs^{43–45,59,61}.

Predictive features

The minimum number of features used in any article was three; the maximum was 107. The average number of features used across all models was 17.3 (SD = 19.6). Of the 27 articles in this review, we extracted ranked feature information from 29 individual models among 17 (63%) articles. For models defining a sepsis endpoint, the top six higher-ranking predictive features were age, platelet count, temperature, immunocompromised status, heart rate, and serum lactate (Supplementary Fig. 2). For severe sepsis endpoints, the top six higher-ranking predictive features were mean heart rate, maximum diastolic blood pressure, age, maximum heart rate, minimum systolic blood pressure, and heart rate (Supplementary Fig. 3). For septic shock endpoints, the top six higher-ranking predictive features were serum lactate, platelet count, maximum heart rate, maximum mean arterial pressure, Glasgow Coma Scale score, and respiratory sequential organ failure assessment score (Supplementary Fig. 4).

When categorized broadly (Fig. 1), blood pressure had the highest category ranking among each sepsis-related endpoint, followed by heart rate features for severe sepsis and septic shock, and thermoregulation for sepsis (e.g., temperature, fever, and hypothermia). Hematologic laboratory values (e.g., hematocrit, white blood cell count, platelet count, and basophils), immunological risk factors (e.g., up-to-date immunization, oncological comorbidity, cancer relapse, and chemotherapy), oxygen requirement (e.g., oxygen saturation and fraction of inspired oxygen), and demographics (e.g., age and home zip code) were also among the higher-ranking feature groups across the sepsis-related endpoints. Laboratory values were used in sepsis and septic shock models but not in the severe sepsis models included in this review.

Most articles without ranked features had less than 20 total features except one⁵⁸. One article was ranked for Kawasaki disease in differentiating from sepsis⁶⁸. When examining the frequency of the remaining top 20 feature groups from all developed models, blood pressure, heart rate, and thermoregulation were the most frequently occurring among each sepsis-related endpoint (Fig. 2). Hematologic laboratory values were the highest occurring laboratory feature group for sepsis and septic shock, followed by lactate. Electrolyte laboratory values, perfusion (e.g., capillary refill), parent and caregiver factors, and anthropometric and nutritional factors only appeared in sepsis modelling approaches. The renal function indicator feature (i.e., urine output) only appeared in septic shock modelling approaches.

Articles mentioned feature limitations such as having routine or non-routine access to laboratory or diagnostic information^{44–46,53,59,69,70,74}, differences in handling or excluding missing features or documentation^{45,57,58,62,65}, and concerns about human error^{43,44,69,73}. We also noted slightly different language for similar features (e.g., “perfusion” and “capillary refill”), feature exclusion due to reliability concerns⁶³, including only the most concerning

Table 2 | Summary of endpoint definitions for pediatric sepsis, severe sepsis, and septic shock

Endpoint definitions	Studies (n = 27), n (%) ^a
Sepsis	
<i>International definitions</i>	9 (33.3)
Schlapbach et al. ¹ International Consensus Criteria: Phoenix Sepsis Criteria. Potential life-threatening organ dysfunction of the respiratory, cardiovascular, coagulation, and/or neurological systems in children with suspected or confirmed infection—2 or more points: <ul style="list-style-type: none"> Respiratory: PaO₂:FIO₂, mmHg (score 0: ≥ 400 to 3: <100 and IMV) or SpO₂:FIO₂, % (score 0: ≥ 292 to 3: <148 and IMV). Cardiovascular: Vasoactive medication (score 0: none to 2: ≥ 2). Cardiovascular: Lactate, mmol/L (score 0: < 5, 1: 5–10.9, 2: ≥ 11). Age based: MAP, mmHg (score 0 to 2, by age in months and years). Coagulation: Platelets, x10³/μl (score 0: ≥ 100 to 1: < 100). Coagulation: International normalized ratio (score 0: ≤ 1.3 to 1: > 1.3). Coagulation: D-dimer, mg/L FEU (score 0: ≤ 2 to 1: > 2). Coagulation: Fibrinogen, mg/dL (score 0: ≥ 100 to 1: < 100). Neurological: Glasgow Coma Scale (score 0: >10; pupils reactive to 2: fixed pupils bilaterally). 	..
Goldstein et al. ⁴⁷ International Pediatric Sepsis Consensus Conference. Suspected or present source of infection plus at least two SIRS criteria, one of which must be abnormal temperature or leukocyte count: <ul style="list-style-type: none"> Core Temperature: > 38.5 °C or < 36 °C. Tachycardia or bradycardia (if < 1 year). Tachypnea or mechanical ventilation. Abnormal leukocyte count or > 10% bands. 	7 (25.9)
Rhodes et al. ⁴⁸ Surviving Sepsis Campaign Guidelines: Life threatening organ dysfunction caused by a dysregulated host response to infection. The authors using this definition also specify as follows: <ul style="list-style-type: none"> Signs and symptoms of inflammation. Signs and symptoms of infection. Rectal Temperature: 38.5 °C or 35 °C. Tachycardia. Signs of altered function in at least one of the following organs: mental status, hypoxemia, increased serum lactate levels, or bounding pulses. 	1 (3.7)
Unspecified application of Weiss et al. ¹⁴ Surviving Sepsis Campaign International Guidelines and Singer et al. ²² Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)—SOFA ≥ 2: <ul style="list-style-type: none"> Respiration: PaO₂:FIO₂, mmHg (score 0: ≥ 400 to 4: < 100 with respiratory support). Coagulation: Platelets, x10³/μl (score 0: ≥ 150 to 4: < 20). Liver: Bilirubin, mg/dL (score 0: < 1.2 to > 4: 12.0). Cardiovascular: MAP, mmHg (score 0: ≥70 to 1: <70). Cardiovascular: Vasoactive medication and dose given (score: 2 to 4). Central nervous systems: GCS score (score 0: 15 to 4: < 6). Renal: Creatinine, μmol/L (score 0: <1.2 to 4: >5.0). Renal: Urine output, mL/d (score 3: < 500 to 4: < 200). 	1 (3.7)
<i>Treatment and diagnostic codes</i>	10 (37.0)
Intention-to-treat: decision of senior clinician of sepsis being likely. Treatment with intravenous antibiotics.	1 (3.7)
Scott et al. ⁵⁰ Improving Pediatric Sepsis Outcomes Collaborative Investigators. Intention-to-treat criteria. <ul style="list-style-type: none"> Intravenous antibiotics. ≥2 intravenous fluid boluses or ≥ 1 bolus and a vasoactive agent within 6 hours and a blood culture within 72 hours. 	2 (7.4)
Eisenberg et al. ⁴⁹ and Scott et al. ⁵⁰ : Receipt of intravenous anti-infective infusion and either an intravenous vasoactive infusion or one of ≥ 2 boluses of isotonic intravenous fluid, ≥ 35 mL/kg intravenous bolus fluid, or ≥ 2 L intravenous fluid. Diagnostic code for sepsis entered.	1 (3.7)
Infection and organ dysfunction with receipt of intravenous antibiotics and at least two 20 mL/kg boluses of isotonic fluid within 6 hours of presentation; based on the Improving Pediatric Sepsis Outcomes Collaborative expert panel.	1 (3.7)
Disseminated infection with systemic inflammatory response syndrome within 24 hours of emergency department arrival resulting in major or extreme severity of illness: assigned by the 3 M Corporations All Patient Refined—Diagnosis-related Groups algorithm.	2 (7.4)
ICD-10 who had discharge diagnosis of sepsis: systemic inflammatory response syndrome and suspected infection—microbiological culture and antibiotic administration in 72 hours, or antibiotic administration with cultures in 24 hours).	1 (3.7)
ICD-10 combined with intensive care unit transfer.	1 (3.7)
ICD-9-CM and ICD-10-CM R-Codes and explicit microbiological codes for sepsis.	1 (3.7)
<i>Mortality</i>	2 (7.4)
In-hospital mortality and early death (within 72 hours of hospital arrival) or requirement of extracorporeal membrane oxygenation support.	1 (3.7)
Mortality within 30 days of intensive care unit admission, calculated using the length of intensive care unit and hospital stay for those who died in either setting.	1 (3.7)
<i>Other</i>	3 (11.1)
A modified Delphi process by the Pediatric Sepsis Predictor Standardization working group and hospital admission.	1 (3.7)
Vital sign thresholds derived from a multi-centered database of approximately 1.2 million emergency department visits.	2 (7.4)
Severe sepsis	
<i>International definitions</i>	5 (18.5)
Goldstein et al. ⁴⁷ International Pediatric Sepsis Consensus Conference. Sepsis criteria plus one of: <ul style="list-style-type: none"> Cardiovascular organ dysfunction. Acute respiratory distress syndrome. Two or more other organ dysfunctions (i.e., Neurologic, Hematologic, Renal, Hepatic). 	5 (18.5)

Table 2 (continued) | Summary of endpoint definitions for pediatric sepsis, severe sepsis, and septic shock

Endpoint definitions	Studies (n = 27), n (%) ^a
<i>Treatment and Diagnostic codes</i>	2 (7.4)
ICD-9 codes; cardiovascular dysfunction definitions were modified because it was not possible to assign causal attribution for fluid-refractory hypotension, and exam components could not be determined retrospectively.	1 (3.7)
ICD-9-CM and ICD-10-CM R-Codes and explicit microbiological codes for sepsis.	1 (3.7)
Infection and organ dysfunction that is not responsive to intravenous antibiotics or at least two 20 mL/kg of fluid boluses, requiring vasopressors and ongoing resuscitation with more than two 20 mL/kg of fluid boluses with intensive care management; based on the Improving Pediatric Sepsis Outcomes Collaborative expert panel.	1 (3.7)
<i>Other</i>	4 (14.8)
Sepanski et al. ⁵¹ screening algorithm results. <ul style="list-style-type: none"> Modified SIRS: two abnormal measures, one of which must be heart rate (by age) or respiratory rate (by age), and the other one temperature (by neonate or non-neonate age), WBC (by age), or > 10% bands. And one major organ dysfunction (i.e., Respiratory, Cardiovascular). Or at least two minor organ dysfunctions (i.e., Respiratory without mechanical ventilation, Hematologic, Renal, Hepatic). 	1 (3.7)
Sepsis with organ dysfunction developing within 48 hours of emergency department arrival. Vital sign thresholds derived from a multi-centered database of ~1.2 million emergency department visits.	2 (7.4)
Septic shock	
<i>International definitions</i>	6 (22.2)
Schlapbach et al. ¹ International Consensus Criteria: Phoenix Sepsis Criteria. Potential life-threatening organ dysfunction of the respiratory, cardiovascular, coagulation, and/or neurological systems in children with suspected or confirmed infection – Sepsis and ≥ 1 cardiovascular point(s).	..
Goldstein et al. ⁴⁷ International Pediatric Sepsis Consensus Conference. Sepsis criteria plus cardiovascular organ dysfunction despite adequate fluid resuscitation.	4 (14.8)
Weiss et al. ¹⁴ Surviving Sepsis Campaign International Guidelines: a severe infection leading to cardiovascular dysfunction (including hypotension, need for treatment with a vasoactive medication, or impaired perfusion).	1 (3.7)
Singer et al. ²² Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3): SOFA ≥ 2 using PELOD-2 cut offs for MAP and creatinine. <ul style="list-style-type: none"> Respiration: PaO₂:FIO₂, mmHg (score 0: ≥ 400 to 4: < 100 with respiratory support) Coagulation: Platelets, x10⁹/μl (score 0: ≥ 150 to 4: < 20) Liver: Bilirubin, mg/dL (score 0: < 1.2 to > 12.0) <ul style="list-style-type: none"> Cardiovascular: MAP, mmHg (by age in months) Cardiovascular: Vasoactive medication and dose given (score: 2 to 4) Central nervous systems: GCS score (score 0: 15 to 4: < 6) Renal: Creatinine, μmol/L (by age in months) Renal: Urine output, mL/d (score 3: < 500 to 4: < 200) 2- or 6-point rise in PELOD-2: <ul style="list-style-type: none"> Neurologic: GCS score (score 0: ≥ 11, 1: 5–10, 4: 3–4) Pupillary reaction (score 0: both reactive, 5: both fixed) Cardiovascular: Lactatemia, mmol/L (score 0: < 5.0, 1: 5.0–10.9, 4: ≥ 11.0) Cardiovascular: MAP, mmHg (by age in months) Renal: Creatinine, μmol/L (by age in months) Respiratory: PaO₂:FIO₂, mmHg (score 0: ≥ 61, 2: ≤ 60) Respiratory: PaCO₂, mmHg (score 0: ≤ 58, 1: 59–94, 3: ≥ 95) Respiratory: Invasive ventilation (score 0: No, 3: Yes) Hematologic: WBC count, x10⁹/L (score 0: > 2, 2: ≤ 2) Hematologic: Platelets, x10⁹/L (score 0: ≥ 142, 1: 77–141, 2: ≤ 76) Adequately fluid resuscitated as per Weiss et al. ¹⁴ Surviving Sepsis Campaign International Guidelines: 40 mL/kg of fluids in a 3-hour window or resuscitation target met for MAP at the 5 th percentile or higher for age. Given vasopressors, serum lactate > 2 mmol/L.	1 (3.7)
<i>Treatment and diagnostic codes</i>	4 (14.8)
Within the first 24 hours after arrival: systolic hypotension and the administration of vasoactive medication or isotonic crystalloids.	2 (7.4)
Eisenberg et al. ⁴⁹ : Patients with sepsis who received an intravenous vasoactive medication within 24 hours of arrival. Diagnostic code for septic shock entered.	1 (3.7)
Infection and organ dysfunction that is not responsive to intravenous antibiotics or at least two 20 mL/kg of fluid boluses, requiring vasopressors and ongoing resuscitation with more than two 20 mL/kg of fluid boluses with intensive care management; based on the Improving Pediatric Sepsis Outcomes Collaborative expert panel.	1 (3.7)
<i>Mortality</i>	2 (7.4)
In-hospital mortality and early death (within 72 hours of hospital arrival) or requirement of extracorporeal membrane oxygenation support.	1 (3.7)
Mortality within 30 days of intensive care unit admission, calculated using the length of intensive care unit and hospital stay for those who died in either setting.	1 (3.7)

(..) not applicable, CM clinical modification, FEU fibrinogen equivalent units, FIO₂ fraction of inspired oxygen, GCS Glasgow Coma Scale, ICD international classification of disease, IMV invasive mechanical ventilation, MAP mean arterial pressure, PaCO₂ partial pressure of carbon dioxide, PaO₂ partial pressure of oxygen, PELOD-2 Pediatric Logistic Organ Dysfunction-2, R Codes residential design codes, SOFA sequential organ failure assessment, WBC white blood cell count.

^aPercentages may sum to more than 100 due to combinations of multiple definitions.

Table 3 | Summary of development and performance outcomes reported for each article by sepsis-related endpoint

	Sample Size	Participant age range	Sex (% male)	Setting	Time	Prevalence (%)	Validation	Features and ranking	Approach(es)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Sepsis												
Ehwerhemuepha et al. ⁶⁹	537837 (visits)	Median: 4; IQR: 8 years	53.60	ED	5–7 seconds from triage to predict	0.20	50% training, 15% validation, 35% testing, 10-fold cross-validation	47 (ranked)	Stochastic gradient boosting	0.976 (0.972–0.981) ^a	0.4414 (.)	0.999 (.)
Georgette et al. ⁷³	141510 (encounters)	Median: 4.5; IQR: (1.8–9.7) years	51.70	ED	3.2 hours before IV vasoactive infusion	3.40	66% training, 33% validation	4	Empirically derived shock index	0.69–0.78 (0.65–0.80)	0.835 (0.817–0.854)	0.428 (0.424–0.433)
Gilholm et al. ⁴⁶	3473 (patients)	Median: 2.1; IQR: (0.9–5.6) years	55.00	ED ^b	After clinical suspicion	15.10	10-fold cross-validation	16 (ranked)	Logistic regression (score)	0.80 (0.78–0.82)	0.90 (0.87–0.92)	0.51 (0.49–0.53)
Lamping et al. ⁵⁹	289 (patients)	Medians: 28–46; IQR: (4–120) months	53.30–62.50	ICU	After clinical suspicion	19.38	3-fold cross-validation	8 (ranked)	Random forest	0.78 (0.70–0.87)
Li et al. ⁶⁸	608 (patients)	28.45 ± 21.47 months	62.70–64.80	ICU	During patient encounter	50.80	70% training, 30% validation	7 (ranked for KD)	Logistic regression (score)	0.878 (.)	0.94 (.)	0.75 (.)
Marassi et al. ⁷¹	12749 (patients)	0 to 18 years	..	ICU	At onset	2.11	80% training, 20% validation	11	XGBoost
Mawji et al. ⁶⁴	1612 (patients)	Medians: 13.1–16.5; IQR: (21.8–32.9) months	51.50	ED	At triage	5.16	80% training, 20% validation; 10-fold cross-validation	9 (ranked)	Logistic regression	0.86 (.)	0.59–0.91 (0.54–0.94)	0.50–0.92 (0.45–0.94)
Mercurio et al. ⁶²	35074 (encounters)	97.7% between 0.11 and 18.0 years	53.00	ED	During patient encounter	0.54	80% training, 20% validation; Cross-validation	20 (ranked)	Random forest	0.81 (.) ^a	0.93 (.)	0.84 (.)
Nguyen et al. ⁶⁶	3014 (admissions)	Median: 1.13; IQR: (0.15–4.30) years	56.30	ICU, Inpatient	Within the first 24, 48 hours with clinician suspicion	4.40	10-fold cross-validation	107	Tree augmented naive bayes	0.866 (0.826–0.906), 0.867 (0.828–0.906)	0.463 (0.376–0.550), 0.469 (0.377–0.550)	0.949 (0.940–0.956), 0.953 (0.941–0.957)
Sanchez-Pinto et al. ⁴³	218839 (encounters)	Medians: 2.6–3.7; IQR: (0.6–9.4) years	47.00–51.50	ED, ICU	During patient encounter	1.20 (Mortality), 0.60–0.80 (Early mortality)	10-fold cross-validation	13	Stacked regression models (score)	0.71–0.92 (0.70–0.92), 0.79–0.96 (0.77–0.96) ^a	0.23–0.81 (0.20–0.85), 0.31–0.90 (0.26–0.93)	0.53–0.99 (0.52–1.00), 0.52–0.99 (0.52–0.99)
Schlappbach et al. ⁵⁶	4403 (patients)	Median: 2.1; IQR: (0.62–6.68) years	55.50	ICU	Within 1 hour of admission	5.20 (30-day mortality)	..	6 (ranked)	Logistic regression (score)	0.810–0.817 (0.779–0.855)
Sepanski et al. ⁵¹	35586 (encounters)	Mean: 5.7; SD: 5.4 years	..	ED ^b	Within 24–48 hours	0.20	Split-sample validation	12	Logistic regression	0.849–0.918 (0.782–0.985)	0.703–0.770 (0.599–0.866)	0.977–0.981 (0.976–0.982)
Solá-Ribalta et al. ⁶⁵	210 (patients)	Median: 42.7; IQR: (5.7–138.6) months	52.40	ED, Inpatient	After clinical suspicion	69.52	k-fold cross-validation	11 (ranked)	Logistic regression (score)	0.886 (0.845–0.927)	0.98 (.)	0.767 (.)
Solá-Ribalta et al. ⁶⁸	266 (patients)	Median: 37.2; IQR: (4.2–127.8) months	53.80	ED	After clinical suspicion	64.66	k-fold cross-validation	4 (ranked)	Logistic regression (score)	0.825 (0.772–0.878)	0.88–0.90 (.)	0.45–0.62 (.)

Table 3 (continued) | Summary of development and performance outcomes reported for each article by sepsis-related endpoint

Sample Size	Participant age range	Sex (% male)	Setting	Time	Prevalence (%)	Validation	Features and ranking	Approach(es)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Spaeder et al. ⁶⁰	1521 (patients) Medians: 1.4–3.2; IQR: (0.3–7.4) years	50.3–53.7	ICU	Within 24 hours	10.20 (of admissions)	Cross-validation	37 (ranked)	Random forest	0.762 (0.728–0.789)	0.102 (.)	0.16 (.)
Stephen et al. ⁶¹	1425 (patients) 79.8–84.2% between 1 and > 18 years	53.70	ICU ^b	Within 24 hours, Anytime	1.40 (of encounters)	84.6% training, 15.4% validation; 40-fold cross-validation	13 (ranked)	Logistic regression (score)	.. ^a	0.49 (0.37–0.61), 0.72 (0.60–0.81)	0.98 (0.98–0.98), 0.96 (0.96–0.96)
Yang et al. ⁶⁴	65 (patients) Between 8.8–9.3 ± 4.6–4.9 years	42.9–56.7	ICU	After clinical suspicion	53.80	10-fold cross-validation	4	Logistic regression (score)	0.826–0.917 (.)
Ying et al. ⁷²	667 (patients) 0 to 10 years	..	ICU	After clinical suspicion	74.50	67% training, 33% validation; 10-fold cross-validation	18	Gradient boosting machine	0.943 (0.900–0.978) ^a	0.941 (.)	0.787 (.)
Severe Sepsis											
Enverhemuepha et al. ⁶⁹	537837 (visits) Median: 4; IQR: (8) years	53.60	ED	5–7 seconds from triage to predict	0.05	50% training, 15% validation, 35% testing; 10-fold cross-validation	47 (ranked)	Stochastic gradient boosting	0.99 (0.985–0.995) ^a	0.845 (.)	0.99 (.)
Kamaleswaran et al. ⁶⁵	493 (patients) 6 to 18 years	..	ICU	2, 8 hours earlier than existing alert	4.06	5-fold cross-validation	3 (ranked), 6 (ranked), 14 (ranked)	Logistic regression Random forest	0.77 (0.63–0.91), 0.56 (0.39–0.76)	0.55 (.), 0.80 (.), 0.611 (.)	0.874 (.), 0.771 (.), 0.796 (.), 0.823 (.)
Le et al. ⁶²	9486 (patients) Median: 10; IQR: (5–14) years	50.39	Inpatient	4 hours before onset, at onset	1.06	4-fold cross-validation	7 (ranked), 6 (ranked)	Boosted ensembles of decision trees	0.718 (.), 0.916 (.)	0.75 (.), 0.75 (.)	0.70 (.), 0.94 (.)
Mercurio et al. ⁶²	35074 (encounters) 97.7% between 0.11 and 18.0 years	53.00	ED	During patient encounter	0.54	80% training, 20% validation; Cross-validation	20 (ranked)	Random forest Classification and regression tree	0.81 (.) ^a	0.93 (.)	0.84 (.)
Sepanski et al. ⁵¹	35586 (encounters) Mean: 5.7; SD: 5.4 years	..	ED ^b	Within 24–48 hours	0.20	Split-sample validation	12	Logistic regression	0.849–0.918 (0.782–0.985)	0.703–0.770 (0.599–0.866)	0.977–0.981 (0.976–0.982)
Septic Shock											
Aviles-Robles et al. ⁶³	404 (febrile neutropenia episodes) Median: 7.7; IQR (4.4–11.7) years	48.80	ED	During patient encounter	16.1 (of febrile neutropenia episodes)	80% training, 20% validation	9 (ranked)	Logistic regression	0.66 (0.56–0.76)	0.96 (.)	0.33 (.)
Georgette et al. ⁷³	141510 (encounters) Median: 4.5; IQR: (1.8–9.7) years	51.70	ED	3.2 hours before IV vasoactive infusion	0.36	66% training, 33% validation	4	Empirically derived shock index	0.82–0.95 (0.65–0.97)	0.813 (0.752–0.874)	0.835 (0.832–0.838)

Table 3 (continued) | Summary of development and performance outcomes reported for each article by sepsis-related endpoint

Sample Size	Participant age range	Sex (% male)	Setting	Time	Prevalence (%)	Validation	Features and ranking	Approach(es)	AUROC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Liu et al. ⁷⁰	6161 (patients) Mean: 6.04; SD: 5.75 years	55.70	ICU	Median early warning: 8.9 hours Median early warning: 12.0 hours	17.58	70% training, 30% validation	26 (top 10 ranked)	XGBoost Generalized linear model	0.90 (.)	0.84 (.)	0.82 (.)
Mercurio et al. ⁶²	35074 (encounters) 97.7% between 0.11 and 18.0 years	53.00	ED	During patient encounter	0.54	80% training, 20% validation; Cross-validation	20 (ranked)	Random forest Classification and regression tree	0.81 (.) ^a 0.77 (.) ^a	0.93 (.)	0.84 (.) 0.70 (.)
Sanchez-Pinto et al. ⁴³	218839 (encounters) Medians: 2.6–3.7; IQR: (0.6–9.4) years	47.00–51.50	ED, ICU	During patient encounter	1.20 (Mortality), 0.60–0.80 (Early mortality)	10-fold cross-validation	13	Stacked regression models (score)	0.71–0.92 (0.70–0.92), 0.79–0.96 (0.77–0.96) ^a	0.23–0.81 (0.20–0.85), 0.31–0.90 (0.26–0.93)	0.53–0.99 (0.52–1.00), 0.52–0.99 (0.52–0.99)
Schlapbach et al. ⁵⁶	4403 (patients) Median: 2.1; IQR: (0.62–6.68) years	55.50	ICU	Within 1 hour of admission	5.20 (30-day mortality)	..	6 (ranked)	Logistic regression (score)	0.810–0.817 (0.779–0.855)
Scott et al. ⁴⁴	2464 (visits) Medians: 5.6–6.0; IQR: (2.1–13.4) years	49.00–55.00	ED	At hospital arrival: temporal, geographic	11.40	10-fold cross validation, temporal and geographic validation	9 (ranked)	Elastic net regularization	0.75 (0.69–0.81), 0.87 (0.73–1.00)	0.82 (0.72–0.90), 0.90 (0.55–1.00)	0.48 (0.44–0.52), 0.32 (0.21–0.46)
Scott et al. ⁴⁵	2318 (visits) Medians: 4.9–5.9; IQR: (1.0) years	52.00–55.00	ED, ICU	Within 2 hours of hospital arrival: temporal, geographic	8.50	Temporal and geographic validation	20 (ranked)	Logistic regression	0.83 (0.78–0.89), 0.83 (0.60–1.00)	0.84 (0.71–0.92), 0.80 (0.28–0.99)	0.65 (0.61–0.69), 0.40 (0.27–0.54)
Xiang et al. ⁵⁷	1238 (patients) Median: 58.3; IQR: (25.8–111.0) months	59.60	Inpatient	4, 8, 12, 24 hours before onset	2.84 (of observation periods)	78% training, 22% validation	23 (ranked)	XGBoost	0.925 (.), 0.925 (.), 0.900 (.), 0.930 (.)
Ying et al. ⁷²	667 (patients) 0 to 10 years	..	ICU	After clinical suspicion	74.50	67% training, 33% validation; 10-fold cross-validation	18	Gradient boosting machine	0.943 (0.900–0.978) ^a	0.941 (.)	0.787 (.)

(.) not available or reported, AUROC area under the receiver operating curve, CI confidence interval, ED emergency department, ICU intensive care unit, IQR interquartile range, KD Kawasaki Disease, SD standard deviation, XGBoost extreme gradient boosting. ^aAlso reports the area under the precision-recall curve (see Supplementary Data 1).

^bImplemented. Note that some articles appear more than once if they include multiple sepsis-related endpoints. See Supplementary Data 1 for Martinez et al.⁷⁴ and Stephen et al.⁶⁷.

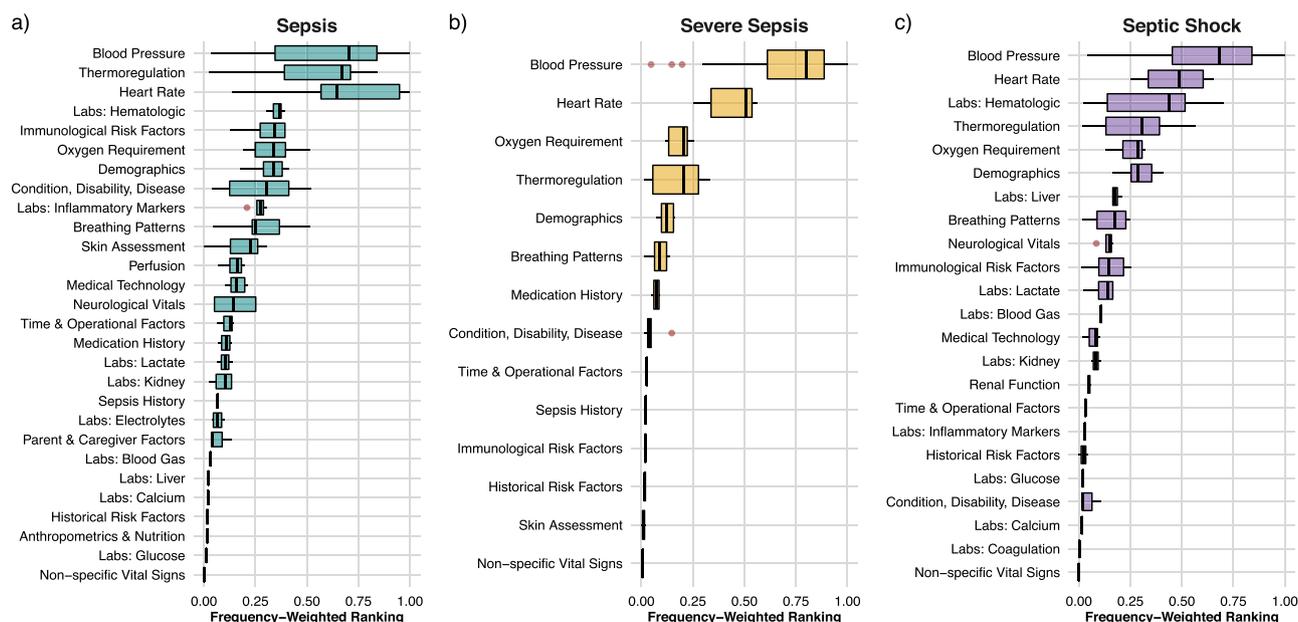


Fig. 1 | Frequency-weighted feature group rankings based on the top 20 extracted features. The normalized rankings of combined feature groups used in models for pediatric **a** sepsis, **b** severe sepsis, and **c** septic shock, weighted by how often each feature group appeared within the three endpoints and listed from highest (top) to lowest (bottom) median value. The individual features included within the

developed categorizations are available in Supplementary Data 2. The data contained in this graph is from 17 articles and 29 models reporting ranked features for pediatric sepsis, severe sepsis, and septic shock endpoints, with some models including more than one sepsis-related endpoint.

feature values within a time range⁵⁸, using qualitative variables⁶⁵, and not knowing a patient’s treatment time⁷⁴.

Performance metrics

The average area under the receiver operating curve (AUROC) for sepsis endpoint logistic regression-based models developed into a score-based tool was 0.84. For the linear and tree-based sepsis models, the average AUROCs were 0.81 and 0.85, respectively. For the severe sepsis models, the average AUROCs for the linear and tree-based approaches were 0.74 and 0.84, respectively. For septic shock models, the average AUROC for logistic regression-based models developed into a score-based tool, linear models, and tree-based models was 0.81, 0.83, and 0.89, respectively, with the latter score heavily weighted by two articles with AUROCs above 0.90 at multiple prediction time points^{57,72}.

Few articles (5 [18%]) reported the area under the precision-recall curve (AUPRC)^{43,61,62,69,72}. Three articles (11%) reported AUPRCs less than 0.50 for sepsis, severe sepsis, and septic shock, using stochastic gradient boosting, regression-based scoring tools, or random forest methods^{43,62,69}. Two articles (7%) reported AUPRCs higher than 0.90 for a logistic regression-based scoring tool and gradient boosting model for sepsis and septic shock^{61,72}. Most articles reported sensitivity (20 [74%]) and specificity (19 [70%]). Some reported positive predictive value (16 [59%]), negative predictive value (12 [44%]), positive or negative likelihood ratios (4 [15%]), and F₁ scores (3 [11%]).

Automation type

We categorized nine articles (33%) as “analysis automation,” which is the automation of algorithms or other calculations to inform clinicians about certain sepsis criteria being met^{53,55,57,62,69–71,73,74}. We categorized 18 (67%) articles as “decision automation,” in which the article described more details about how the technology provides additional information to support sepsis suspicions, stratify risk or triage levels, identify workflow actions, or achieve greater situational awareness^{43–46,52,54,56,58–61,63–68,72}. All articles included manual provider-entered information within an electronic health record. Two articles (7%) involved “acquisition automation” through collecting data without human interaction⁷⁵, such as continuous physiological monitoring^{55,60}.

Prediction timings

The prediction timings ranged from the onset of the sepsis-related condition^{52,71}, within five to seven seconds after initial triage⁶⁹, and between two and 12 hours earlier^{52,55,70,73}. Other prediction timings were within 24 hours^{60,61,67} and up to 24 hours before sepsis-related outcome onset⁵⁷. For eight (30%) articles, clinical suspicion was first required to initiate the model and provide a result^{44–46,58,59,64–66,72}. The four (11%) remaining articles were identified as being used during a patient encounter^{43,62,63,68}.

The prediction timings also ranged with respect to other design and performance elements (Table 3). For example, one XGBoost model that could predict the onset of septic shock 24 hours in advance across over 1200 patients with hematological malignancies in the inpatient unit, with a median age of 58 months using 24 features, had a reported AUROC of 0.93⁵⁷. One decision tree ensemble model that could predict the onset of severe sepsis four hours in advance across over 9400 patients in the inpatient unit, with a median age of 10 years using seven features, had a reported AUROC of 0.718⁵². Comparatively, the Phoenix Sepsis Criteria, a regression-model informed integer-based scoring tool using up to 13 variables to identify sepsis and septic shock across over 200000 patient encounters in high- and low-income healthcare settings, with a median age of 2.6–3.7 years, had a reported AUROC between 0.71–0.96 and an AUPRC between 0.14–0.48⁴³.

Interface and interaction design

Few articles described specific details related to how clinicians would interact or interface with the developed model or alerting systems. Two articles (7%) described a two-tiered notification system: an “alert” tier, with a score greater than or equal to 15, to prompt a team-based bedside evaluation, or an “aware” tier, with a score greater than or equal to 30, to increase situational awareness, including highlighting the score in a bright colour upon opening the patient’s electronic health record chart^{61,67}. The purpose was to reduce false alerts and unnecessary and unplanned team huddles while ensuring children who met the “aware” threshold were discussed during rounds, handoffs, and planned huddles^{61,67}. The screening scores were visualized on patient lists and

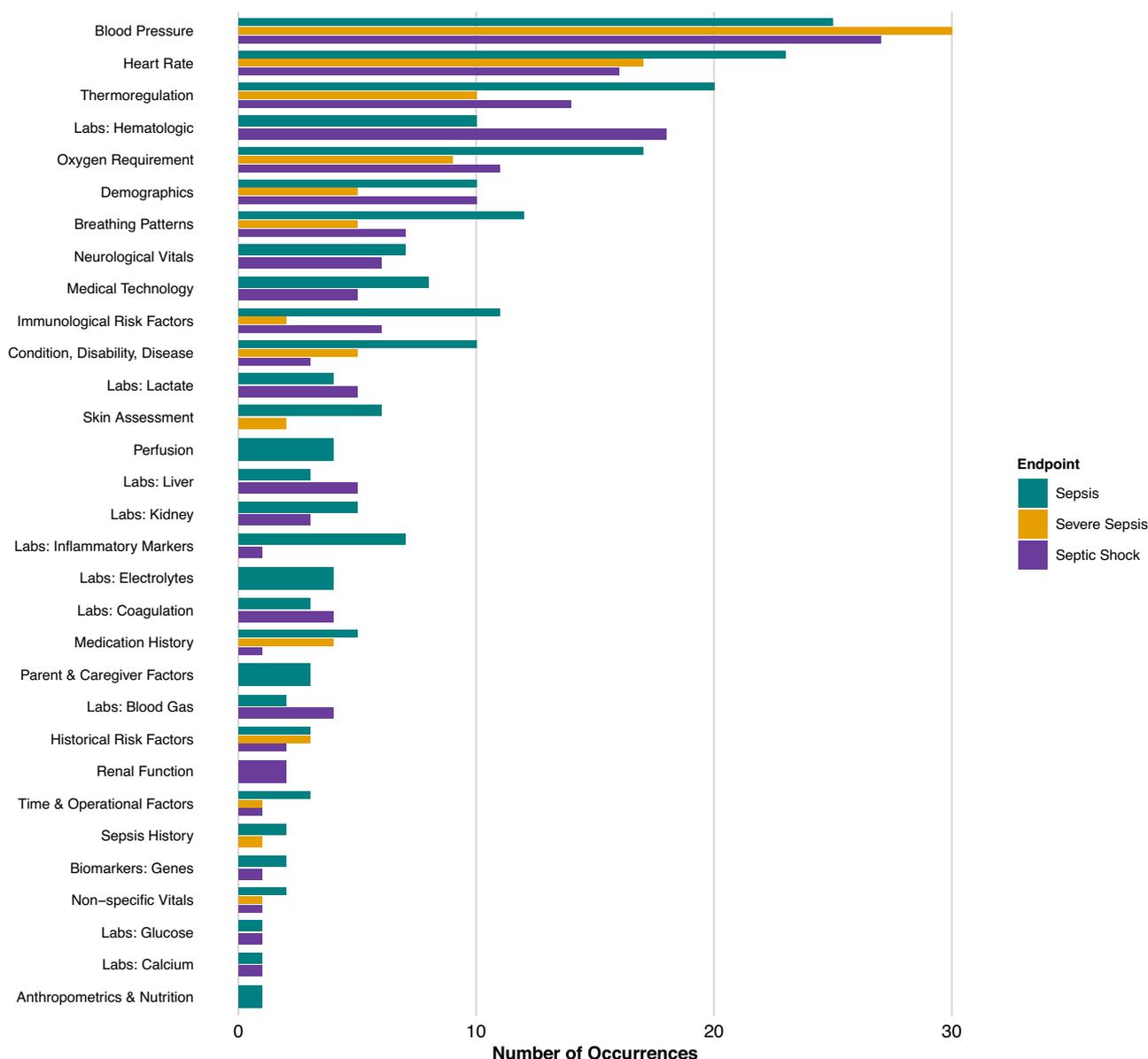


Fig. 2 | Overall frequency of categorized features for pediatric sepsis, severe sepsis, and septic shock endpoints. The frequency of each feature category is listed in descending order from highest to lowest occurrence. The individual features included within the developed categorizations are available in Supplementary Data 2. All articles with 20 or fewer features are included. For articles with more than 20

features, only the top 20 are included. Articles with a ranking for Kawasaki Disease vs Sepsis ($n = 1$) and articles with more than 20 non-ranked features ($n = 1$) are excluded from this figure. The “Biomarkers: Genes” category was only used in two articles, accounting for 100% of the model features when used, and so is only counted once for septic shock and twice for sepsis in this figure.

on-screen banners, updated in real-time to prompt a huddle at the bedside⁶⁷. The clinical variables and their scoring values contributing to the overall score were also displayed, including provider-specific steps when the “alert” tier was activated to call a sepsis huddle. Clinician feedback from an urban quaternary children’s hospital indicated higher compliance with the sepsis huddle process and minimal alert fatigue, but this was partly attributed to pre-implementation and pilot testing⁶⁷.

Other articles mentioned a tier-like system design to identify risk levels or support triage of the most at-risk children^{54,56}, support risk stratification^{43,46,52,56,60}, identify low-risk post-chemotherapy children⁶³, and identify children who are at a higher risk of hypotensive septic shock⁴⁵. One article combined their data-driven sepsis decompensation risk score with a red or yellow spotlight colour to indicate patient status⁵³, available to all emergency department staff and clinicians⁷⁴. The red or yellow spotlight colour specified how often a bedside huddle was needed for the patient,

which was every 30 or 60 min, respectively⁷⁴. A tracking board was used to alert all clinicians with a checkerboard flag displayed if the sepsis score indicated a child at risk, suggesting the need for a bedside assessment, with the most recent risk score shown next to the patient on the tracking board for increased situational awareness⁷⁴.

Clinical implementation and decision-making

Three (11%) technologies were implemented electronically^{53,67,74}, one (4%) non-electronically⁴⁶, and one (4%) in silent mode⁶¹. One-third (9 [33%]) of the articles mentioned specific human factors considerations regarding implementation^{44–46,58,59,61,67,72,74}. This included descriptions of implementation within sepsis workflows^{46,61,67}, the impact of the technology on team-based activities, including huddles and team situational awareness^{61,67,74}, and considerations for minimizing alert fatigue^{61,67}. Two articles found positive impacts on huddle compliance^{67,74}, and after electronic implementation, one

found a significant improvement in team-based assessments, situational awareness, patient huddles, and communication, as well as increased fluid bolus and antibiotics administration by three hours post-activation⁷⁴. However, clinicians were not always blinded to the tool, which may have biased their decisions to diagnose more or fewer patients, or the study was not powered to determine the impact on patient outcomes⁴⁶. Additionally, one study mentioned that early treatment or transfers may inadvertently perturb the measured outcome or lead to overusing healthcare resources⁶⁷.

For decision support, two articles differentiated sepsis from other conditions, such as non-infectious systemic inflammatory response syndrome or Kawasaki disease^{59,68}. A few articles (9 [33%]) mentioned that users would interact with their system after clinician-initiated suspicion^{44–46,58,59,64–66,72}. The predictive features would be entered into a smartphone-based calculator application^{54,66}, a web calculator^{59,72}, or an electronic health record^{43–45} once a provider thought a child might develop sepsis or septic shock. The system would provide a probability risk score, or other information would be returned to support or refute their judgment.

While at least one of the models was not ultimately designed to be clinically implemented⁴⁶, none of the articles using complex data-driven modelling explicitly described the explainable aspects of their technologies if they were to be implemented. Notably, one article explained that “black box” models do not support direct inferences about how specific features impact outcome predictions⁵⁹. Given the diversity of patient characteristics and demographics, features, prediction timings, healthcare settings, outcome definitions, and performance outcomes, we did not identify detailed considerations for providing prediction uncertainty information or evaluating clinician interpretations of data-driven predictions beyond limited measurements of huddle and treatment outcomes.

Potential biases

We noted risks of bias potentially impacting testing and validation, which may introduce uncertainties for appropriate clinical use. For example, selection bias was reported in one screening tool’s performance⁶⁸, and feature data may be collected at different times between different technologies⁷⁰. As some models were only applied to patients for whom clinicians suspected sepsis, this may introduce bias in performance outcomes^{44,46}. Some patient groups, such as children with oncological diseases⁶³, were more heavily represented in some datasets, and the model may perform better or worse if used in a different patient group^{52,54–56,60,61}. One model was developed and tested on the same patient cohort⁶⁵. Some models were intentionally biased for maximum sensitivity to alert clinicians earlier than existing technologies⁵⁵, and datasets may not have been specifically collected for data-driven algorithmic or AI prediction purposes^{52,56,57,59}.

Discussion

This scoping review is the first to comprehensively identify and summarize the current design and performance of data-driven technologies for predicting pediatric sepsis-related endpoints in healthcare settings, including aspects of human factors related to clinical interaction and implementation. Our findings encompass 27 studies exploring endpoint definitions, prediction approaches, datasets, feature combinations, performance metrics, automation types, prediction timings, interface elements and implementation outcomes. There are crucial research directions to consider for the future development and integration of data-driven technologies for pediatric sepsis prediction, particularly in terms of feature use, data inclusivity with standardized reporting, and human factors. Research focusing on these aspects is essential to facilitate effective comparisons among technologies, support performance outcomes across heterogeneous patient cohorts, and enable safe clinical integration that augments—rather than replaces—clinical decision-making.

Higher-ranking individual predictive features for pediatric sepsis, severe sepsis, and septic shock generally included cardiovascular vital signs (e.g., heart rate, mean arterial pressure, and capillary refill), laboratory values (e.g., lactate and platelet counts), respiratory vital signs (e.g., oxygen

saturation and PaO₂:FIO₂), and neurological vital signs (e.g., Glasgow Coma Scale score). These outcomes align with the latest pediatric sepsis and septic shock international consensus criteria, which can guide future model development¹. However, it is also important to recognize that the identified lower-ranking or less frequently occurring features in this review should not necessarily be seen as unimportant for pediatric sepsis outcomes—this may simply reflect the overrepresentation of high-income healthcare settings in which models were more commonly developed. Lower-ranked features may also have resulted from fewer articles that reported ranked features and the lack of investigation into potentially underappreciated features across models, such as parental or caregiver concerns⁷⁶, reducing their frequency of occurrence.

Among the sepsis endpoint models, age was a notably high-ranking feature. Indeed, identifying abnormal vital signs often includes age-based thresholds in pediatric sepsis consensus definitions, such as for mean arterial pressure and creatinine^{1,48}. In this review, age was an individual predictor for sepsis-related endpoints, with included models identifying older children (e.g., five-year-olds) or increasing age as being at higher risk of sepsis^{60,62,69} and higher risks for one-year-olds than the very youngest patients⁶⁰. These findings could be due to challenges with underreporting or under-recognition of vital sign abnormalities among older infants and preschoolers⁶² and to how abnormal vital sign thresholds change with age⁶⁹. There is a potential for future research to consider age as a predictive feature in data-driven models, with further investigation into its importance.

In the few models with a severe sepsis endpoint, none included laboratory values as predictive features. While severe sepsis is often referred to as organ dysfunction (i.e., cardiovascular or respiratory or two other organ system dysfunctions) in addition to a sepsis diagnosis¹, laboratory values for sepsis may already be confirmed before these models are intended for use. The utility of a severe sepsis model may be for identifying deterioration and increasingly abnormal organ dysfunction beyond an initial sepsis diagnosis. However, the term “severe sepsis” has been recently noted as redundant to sepsis and septic shock in the latest international consensus criteria¹, suggesting that future model development should focus on the newest sepsis and septic shock definitions.

In contrast, laboratory values were the top two predictive features for septic shock models. Serum lactate emerged as the highest-ranking individual feature on the basis of median values, followed by platelet count. Lactate is a clinical indicator for hemodynamic instability in children and, if infected, indicates a greater risk of multiorgan dysfunction syndrome and mortality⁷⁷. Lactate is also included in the latest pediatric septic shock criteria¹. Yet, controversy surrounds this measurement as it may not always be available in lower-resource settings¹, and children with normal lactate levels are not necessarily excluded from a sepsis diagnosis^{1,78}. Low platelet counts are most clearly understood in the pediatric sepsis literature as a “consumptive coagulopathy” leading to simultaneously increased bleeding and clotting⁷⁹. Other research suggests that low platelet counts may also be a good predictor of 28-day mortality from septic shock children⁸⁰. However, similar to lactate, acquiring platelet counts in all settings may not be possible. Scoring tools with built-in redundancies if laboratory values are unavailable—or models that do not rely on laboratory values—may have greater utility and practicality in lower-resource environments^{43,44,62,73}, and this consideration for resiliency in response to resource availability should be emphasized in future model development.

Patient datasets varied widely across studies with respect to the volume and balance of sepsis and non-sepsis outcomes, as well as patient demographics. This is a crucial challenge to ensuring the appropriate representativeness of patient groups in model development and for classification-based approaches. Current trends in research predominantly originate from North America and Asia, focusing on single-site studies, with the United States of America contributing the majority of pediatric datasets⁴³. Notably, only four studies have aggregated data from multiple healthcare sites, one of which included low-to-middle-income settings^{13–46}. This lack of inclusion highlights substantial limitations with respect to model development and technology access in remote and rural communities, Indigenous

communities, and low-to-middle-income countries, which are disproportionately affected by sepsis^{1–6}. These areas are frequently overlooked in the design, validation, and implementation of predictive technologies, resulting in a scarcity of studies representing these underrepresented environments^{43–45,54}.

Furthermore, there is a risk of over-representing the same patient populations across the various studies, contributing to the prevalent issue of class imbalance within the available datasets. In some instances, multiple approaches have been explored using identical datasets from single-site studies, and both multi-site and single-site studies have used the same datasets, including those from the Chinese pediatric intensive care dataset⁸¹. Future research should address the issue of class imbalance by employing appropriate model evaluation methods tailored for imbalanced datasets, implementing strategies to balance the datasets, and refining algorithmic approaches to manage class disparities better. Ultimately, there is an urgent need for more international collaboration and diverse data collection practices to increase patient representativeness, alongside a focus on innovations that democratize access to pediatric sepsis prediction technologies in underserved regions. Without these advancements, relying on individual healthcare sites to independently develop and validate their own systems risks perpetuating disparities in under-resourced settings.

Data-driven pediatric sepsis prediction tools should also work effectively with children of all ages. Current model development shows wide variations in the age ranges of patients in the studies reviewed. Additionally, some articles specifically included or excluded certain pediatric age ranges^{55,56,68,72} and reported small sample sizes of children in different age groups^{55,58,62,64,72}. This is significant because pediatric sepsis involves age-based differences in symptom manifestation²⁶. Excluding different pediatric age groups runs the risk of model overfitting, leading to technologies that may not equitably apply to all children, complicating performance comparisons among articles and warranting caution for widespread clinical implementation. Future research should prioritize collaboration and develop robust dataset-sharing methods while exploring opportunities to ensure the optimal performance of these data-driven prediction models for all ages⁸².

In addition to concerns about pediatric age ranges, reporting and including patient sex, race, and ethnicity, especially in studies with relatively large sample sizes, is inconsistent. Only a few articles referenced the TRI-POD reporting guidelines, and some articles reported neither patient sex nor race or ethnicity. Limited transparency about this aspect of developing data-driven prediction tools may result in hidden discrimination, such as prioritizing treatment actions in children from different socio-demographic backgrounds. In fact, considerable concerns have been raised about the risks of codifying racism into data-driven prediction modelling for pediatric sepsis and exacerbating inequalities in the medical care of children⁸³. In one article, race was included as a predictive feature in a top-performing model⁶², whereas another study that included Aboriginal descent as a prediction feature reported that their algorithm performed similarly to when this feature was removed in a smaller model⁴⁶. The potential impacts on diverse pediatric populations have yet to be fully understood, but ensuring that all predictive models perform equally among racialized and historically disadvantaged groups must be prioritized in future research, ultimately providing equitable healthcare outcomes once implemented. The comprehensiveness of reporting patient demographics in future research should follow the most updated best practices, such as the latest TRI-POD + AI guidelines⁸⁴, and ensure strong calibration to reduce the uncertainties of unequal and inequitable model performance among diverse pediatric patients.

As a data-driven, standardized definition of pediatric sepsis has only been recently published^{1,43}, the endpoint definitions among the reviewed models varied over time. This finding was unsurprising, given the variability in clinical definitions from international research surveys⁸⁵. Definition inconsistency was a substantial hindrance to comparing quantitative outcomes of model performance and potential outcomes, given modifications

and variable use of treatment and diagnostic codes, which risk clinician bias. Adopting the latest international consensus criteria holds significant promise for standardizing future development^{1,43}. It is crucial for future research to consolidate the use of pediatric sepsis definitions for data-driven prediction tools in ways that optimize equitable outcomes for all children, and to support comparative analyses among future prospective studies for this high-risk population.

Despite the development of some higher-performing models in this review, it is unlikely that future developments will achieve perfectly generalizable performance in predicting pediatric sepsis or septic shock across diverse healthcare environments and patient populations. Consequently, human factors will be crucial in supporting clinical users in understanding the functionality of these models—how they work or fail to work—and in guiding when and how to incorporate this information into decision-making processes alongside other indicators for determining pediatric sepsis prognoses and preventative measures. From the reviewed articles, several aspects of current model development and design will influence how clinicians engage with these technologies in a healthcare environment, contributing to a more holistic view of understanding overall system performance.

The type of automation implemented will influence workflows, and the timing of predictions could be crucial in whether a prediction model assists or complicates sepsis-related decision-making²⁴. Early predictions generated through analysis automation may enable a greater frequency of timely and effective interventions. As noted in this review, decision automation has the potential to prompt or guide clinical actions both before and following the onset of clinical suspicion of sepsis. Furthermore, automation in data acquisition may decrease the risk of human error in data entry. Additional research is necessary to identify the most effective automation strategies for pediatric sepsis technologies, including the possible integration of diverse approaches within sepsis workflows that enhance clinical judgment and situational awareness, ultimately leading to improved patient outcomes compared with current practices.

The variety of prediction timings highlights the lack of a standardized target timepoint for optimal use, which may differ depending on the healthcare context and the specific decisions involved. For example, predictions made several hours before the onset of sepsis may be more beneficial for determining the timing of patient transfers to specialized pediatric centres or better-equipped facilities. In contrast, shorter prediction windows may be beneficial for prompting earlier clinical interventions to prevent deterioration. This review indicates that prediction timings can range from during a patient encounter to as much as 24 h before the onset of septic shock. The timing of a prediction and its associated risk introduce an element of uncertainty regarding integration into decision-making processes, particularly when other vital signs or indicators may not reflect sepsis-related risks or decompensation. Consequently, future research must focus on how clinicians will navigate the uncertainties tied to predictions of a potential septic state in children, as well as how they might respond to and act upon different levels of risk prediction. This should encompass considerations for effectively displaying or communicating such early prediction information, along with its limitations, to support safer and more effective integration into practice.

Research has shown that model interpretability is important for the acceptance of data-driven technologies among clinicians⁸⁶. However, a recent systematic review and meta-analysis of 106 experiments on human-AI systems revealed no significant impacts on human-AI effectiveness by including explanations. Instead, human-AI performance was better when the human knew when to trust their judgment or the prediction, and vice-versa⁸⁷. In the context of pediatric sepsis, the variability in patient characteristics among training and testing datasets introduces uncertainties when clinicians make decisions for patients for whom a potential model *has not* encountered. Understanding whether the training and testing datasets include representative patient characteristics—including the healthcare environments and available features—can affect perceptions of fairness in model outputs. This highlights the ethical considerations that clinicians

must navigate to use model outputs responsibly^{88–90}. Future research should focus on practical strategies to assist clinicians and clinical teams in determining how and when to trust their judgment alongside prediction outputs. These strategies should mitigate cognitive overload while enhancing situational awareness when these technologies are used in clinical decision-making.

This review aimed to capture the implementation stage of data-driven pediatric sepsis prediction technologies. However, some articles did not focus on identifying significant differences in outcomes, such as mortality rates, and lacked detailed descriptions of implementation factors. Additionally, some studies primarily evaluated performance metrics and alert firing rates. Among the articles reporting on implementation outcomes, data-driven pediatric sepsis prediction technologies may benefit clinical workflows in emergency departments and inpatient settings in high-income countries. These technologies can improve situational awareness with shared displays that communicate model outputs as patient decompensation risk scores, employing colour coding to facilitate understanding⁷¹ while imposing minimal alert fatigue on clinicians⁶⁷. Furthermore, their implementation may reinforce compliance with workflow processes, such as bedside huddles and timely administration of antibiotics and fluids^{67,74}. However, more research involving collaboration among clinicians, data scientists, and human factors engineers is needed to explore the factors that either promote or hinder the effective implementation of these technologies in various clinical environments, including the digital formats in which they are presented. Additionally, future longitudinal studies are essential to assess the experiences and impacts of model design and implementation on clinical teams and patient stakeholders. Such research will help identify the potential influence on decision-making and clinician workflow, as well as the possibilities for improved patient outcomes and new risks for human error compared with current practices and screening tools.

While our search strategy was broad, this approach to a scoping review for pediatric sepsis prediction was necessary for multiple reasons: inconsistencies in data reporting, endpoint definitions, patient characteristics, and the range of development strategies. The significant heterogeneity among the included studies subsequently limited our ability to compare performance outcomes systematically. Notably, most studies were retrospective, and some included small sample sizes. However, the reviewed studies, including some of those that were excluded (Supplementary Table 1), can inform eligibility criteria in future systematic reviews and meta-analyses. Future research can also address the identified limitations for future reviews through standardized reporting and analysis with specific patient cohorts—including high-risk segments, social determinants of health, and specific age groups. Aligning future endpoint definitions with the 2024 consensus criteria will also increase the capacity for rigorous and systematic analyses in this critical area.

Overall, this comprehensive scoping review highlights the heterogeneity of data-driven approaches for predicting sepsis-related outcomes in children and the challenges in translating current research into clinical practice. While promising approaches for improving pediatric sepsis prediction exist, they must be pursued with an awareness of the nuanced disparities across healthcare settings and patient demographics. In working towards enhancing patient outcomes in this vulnerable population, the 2024 international consensus criteria for pediatric sepsis and septic shock¹ and standardized reporting practices can guide future research. Furthermore, there is a need to integrate clinical user perspectives and prioritize human factors in both the development and evaluation of predictive systems, specifically for prediction timings, approaches to automation, and supporting clinical judgment. The incorporation of this design component in future research will be critical to advancing the safe and effective implementation of data-driven pediatric sepsis prediction technologies.

Methods

Search strategy

The review adhered to the PRISMA-ScR guidelines⁹¹. Following the published protocol⁹², we searched the ACM Guide to Computing

Literature, CINAHL, Embase, IEEE, PubMed, SCOPUS, and Web of Science for peer-reviewed studies published until March 1st, 2024. The search strategy for each database is outlined in Supplementary Table 2. Our search used keywords related to sepsis, the pediatric population, and digital technologies. We also used snowball sampling to find additional articles. The overall search was a complex process due to heterogeneity among the articles regarding development approaches and objectives, sepsis-related outcome definitions, patient demographics, reported performance characteristics, and outcome measures for implemented systems. We worked with an Information Specialist at the University of Waterloo (KM) to ensure a thorough and robust search strategy despite the lack of standardized reporting and data.

Selection criteria

The inclusion criteria encompassed articles published in English about data-driven approaches to predict sepsis-related outcomes for children aged 90 days to 21 years old, with no restrictions to country of origin or being commercially available. Articles were only included if there was a digital aspect in their current or future use within a computerized system, such as an electronic health record, bedside monitor, or mobile or online application. Studies focusing solely on sepsis mortality risk were excluded, as this review's primary focus was on predicting sepsis, severe sepsis, and septic shock before mortality unless mortality outcomes were used as a defining criterion.

Screening, data extraction, and analysis

Two members (RT and JG) independently completed the abstract and title screening and the full-text review, with support from a third member (JK). A fourth member (KM) resolved disagreements. One member (RT) initially completed the extraction, verified by a second member (JG or JK). Study quality was evaluated by two members (RT and JK) using a simplified version of the Oxford Centre for Evidence-Based Medicine levels of evidence on the basis of the study design⁹³.

We extracted general article information, the prediction task and automation type⁷⁵, the prediction method, patient demographics, prediction features, validation measures, outcomes related to implementation, and human factors considerations. We extracted details about the best performers for articles outlining multiple approaches, as identified by the authors of each article when applicable. We attempted to contact all corresponding authors for missing information, but only ten responded to provide clarifications or confirm the availability of additional data. We summarized the data in tables using descriptive statistics and plotted the predictive features in graphical format using R (v 4.1.2). We calculated normalized frequency-weighted rankings to analyze the features when the articles included data on feature importance, coefficient values, or odds ratios. We visualized the weighted rankings for the top 20 most predictive individual features of the developed approaches—organized by sepsis, severe sepsis, and septic shock endpoints—by multiplying each feature's rank by occurrence across all studies with ranked features. Features that included multiple related aspects were separated and given the same ranking. For example, if “fever or hypothermia” was listed 10th (Rank 10), it was separated into “fever” and “hypothermia,” both with the same rank of 10. We also categorized features and visualized weighted rankings of these grouped features.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Code availability

The code used in the analysis of this study can be made available from the corresponding author upon reasonable request.

Received: 9 July 2024; Accepted: 25 November 2024;

Published online: 04 December 2024

References

- Schlapbach, L. J. et al. International Consensus Criteria for Pediatric Sepsis and Septic Shock. *JAMA* **331**, 665 (2024).
- Hall, M. Immune Modulation in Pediatric Sepsis. *J. Pediatr. Intensive Care* **08**, 042–050 (2019).
- World Health Organization. *Global Report on Epidemiology and Burden of Sepsis: Current Evidence, Identifying Gaps and Future Directions*. <https://apps.who.int/iris/handle/10665/334216> (2020).
- Rudd, K. E. et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet* **395**, 200–211 (2020).
- Palasantiran, P. & Bowen, A. C. The excess burden of severe sepsis in Indigenous Australian children: can anything be done? *Med. J. Aust.* **206**, 71–72 (2017).
- Ostrowski, J. A. et al. The burden of invasive infections in critically ill Indigenous children in Australia. *Med. J. Aust.* **206**, 78–84 (2017).
- Killien, E. Y., Farris, R. W. D., Watson, R. S., Dervan, L. A. & Zimmerman, J. J. Health-Related Quality of Life Among Survivors of Pediatric Sepsis*. *Pediatr. Crit. Care Med.* **20**, 501–509 (2019).
- Weiss, S. L. et al. Global Epidemiology of Pediatric Severe Sepsis: The Sepsis Prevalence, Outcomes, and Therapies Study. *Am. J. Respir. Crit. Care Med.* **191**, 1147–1157 (2015).
- Farris, R. W. D., Weiss, N. S. & Zimmerman, J. J. Functional Outcomes in Pediatric Severe Sepsis: Further Analysis of the Researching Severe Sepsis and Organ Dysfunction in Children. *Pediatr. Crit. Care Med.* **14**, 835–842 (2013).
- Ravikumar, N., Sankar, J. & Das, R. R. Functional Outcomes in Survivors of Pediatric Sepsis: A Scoping Review and Discussion of Implications for Low- and Middle-Income Countries. *Front. Pediatr.* **10**, 762179 (2022).
- Reinhart, K. et al. Recognizing Sepsis as a Global Health Priority — A WHO Resolution. *N. Engl. J. Med.* **377**, 414–417 (2017).
- Cifra, C. L. et al. An estimate of missed pediatric sepsis in the emergency department. *Diagnosis* **8**, 193–198 (2021).
- Cox, M. I. & Voss, H. Improving sepsis recognition and management. *Curr. Probl. Pediatr. Adolesc. Health Care* **51**, 101001 (2021).
- Weiss, S. L. et al. Surviving sepsis campaign international guidelines for the management of septic shock and sepsis-associated organ dysfunction in children. *Pediatric Critical Care Medicine* E52–E106 (2020) <https://doi.org/10.1097/PCC.0000000000002198>.
- Bracken, A., Lenihan, R., Khanijau, A. & Carrol, E. D. The Aetiology and Global Impact of Paediatric Sepsis. *Curr. Pediatr. Rep.* **11**, 204–213 (2023).
- Zonneveld, R. et al. Soluble adhesion molecules as markers for sepsis and the potential pathophysiological discrepancy in neonates, children and adults. *Crit. Care* **18**, 204 (2014).
- Wynn, J. L. et al. The Influence of Developmental Age on the Early Transcriptomic Response of Children with Septic Shock. *Mol. Med.* **17**, 1146–1156 (2011).
- Molloy, E. J. et al. Neonatal sepsis: need for consensus definition, collaboration and core outcomes. *Pediatr. Res* **88**, 2–4 (2020).
- Attia Hussein Mahmoud, H. et al. Insight Into Neonatal Sepsis: An Overview. *Cureus* (2023) <https://doi.org/10.7759/cureus.45530>.
- McGovern, M. et al. Challenges in developing a consensus definition of neonatal sepsis. *Pediatr. Res* **88**, 14–26 (2020).
- Hayes, R. et al. Neonatal sepsis definitions from randomised clinical trials. *Pediatr. Res* **93**, 1141–1148 (2023).
- Singer, M. et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **315**, 801 (2016).
- Meeus, M. et al. Clinical Decision Support for Improved Neonatal Care: The Development of a Machine Learning Model for the Prediction of Late-onset Sepsis and Necrotizing Enterocolitis. *J. Pediatr.* **266**, 113869 (2024).
- Kamran, F. et al. Evaluation of Sepsis Prediction Models before Onset of Treatment. *NEJM AI* **1**, (2024).
- Despins, L. A. Automated Detection of Sepsis Using Electronic Medical Record Data: A Systematic Review. *J. Healthc. Qual.* **39**, 322–333 (2017).
- Eisenberg, M. A. & Balamuth, F. Pediatric sepsis screening in US hospitals. *Pediatric Research* (2021) <https://doi.org/10.1038/s41390-021-01708-y>.
- Fleuren, L. M. et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* **46**, 383–400 (2020).
- Hassan, N. et al. Preventing sepsis; how can artificial intelligence inform the clinical decision-making process? A systematic review. *Int. J. Med. Inform.* **150**, 104457 (2021).
- Makam, A. N., Nguyen, O. K. & Auerbach, A. D. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: A systematic review. *J. Hosp. Med.* **10**, 396–402 (2015).
- Persad, E. et al. Neonatal sepsis prediction through clinical decision support algorithms: A systematic review. *Acta Paediatr.* **110**, 3201–3226 (2021).
- Sahu, P. et al. Prediction modelling in the early detection of neonatal sepsis. *World Journal of Pediatrics* <https://doi.org/10.1007/s12519-021-00505-1> (2022).
- Schinkel, M., Paranjape, K., Nannan Panday, R. S., Skyttberg, N. & Nanayakkara, P. W. B. Clinical applications of artificial intelligence in sepsis: A narrative review. *Comput. Biol. Med.* **115**, 103488 (2019).
- Smyth, M. A., Brace-McDonnell, S. J. & Perkins, G. D. Identification of adults with sepsis in the prehospital environment: a systematic review. *BMJ Open* **6**, e011218 (2016).
- Teng, A. K. & Wilcox, A. B. A Review of Predictive Analytics Solutions for Sepsis Patients. *Appl. Clin. Inform.* **11**, 387–398 (2020).
- Wu, M., Du, X., Gu, R. & Wei, J. Artificial Intelligence for Clinical Decision Support in Sepsis. *Front. Med.* **8**, (2021).
- Wulff, A., Montag, S., Marscholke, M. & Jack, T. Clinical Decision-Support Systems for Detection of Systemic Inflammatory Response Syndrome, Sepsis, and Septic Shock in Critically Ill Patients: A Systematic Review. *Methods Inf. Med.* **58**, e43–e57 (2019).
- Muralitharan, S. et al. Machine Learning–Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. *J. Med. Internet Res.* **23**, e25187 (2021).
- Fernandes, M. et al. Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review. *Artif. Intell. Med.* **102**, 101762 (2020).
- Li, L., Ackermann, K., Baker, J. & Westbrook, J. Use and Evaluation of Computerized Clinical Decision Support Systems for Early Detection of Sepsis in Hospitals: Protocol for a Scoping Review. *JMIR Res. Protoc.* **9**, e24899 (2020).
- Patton, L. & Young, V. Effectiveness of provider strategies for the early recognition of clinical deterioration due to sepsis in pediatric patients. *JBI Database Syst. Rev. Implement. Rep.* **15**, 76–85 (2017).
- Pepic, I. et al. Early detection of sepsis using artificial intelligence: a scoping review protocol. *Syst. Rev.* **10**, 28 (2021).
- Al-Shwaheen, T. I., Moghbel, M., Hau, Y. W. & Ooi, C. Y. Use of learning approaches to predict clinical deterioration in patients based on various variables: a review of the literature. *Artif. Intell. Rev.* **55**, 1055–1084 (2022).
- Sanchez-Pinto, L. N. et al. Development and Validation of the Phoenix Criteria for Pediatric Sepsis and Septic Shock. *JAMA* **331**, 675 (2024).
- Scott, H. F. et al. Development and Validation of a Predictive Model of the Risk of Pediatric Septic Shock Using Data Known at the Time of Hospital Arrival. *J. Pediatr.* **217**, 145–151.e6 (2020).
- Scott, H. F. et al. Development and Validation of a Model to Predict Pediatric Septic Shock Using Data Known 2 h After Hospital Arrival. *Pediatr. Crit. Care Med.* **22**, 16–26 (2021).
- Gilholm, P. et al. Validation of a paediatric sepsis screening tool to identify children with sepsis in the emergency department: a

- statewide prospective cohort study in Queensland, Australia. *BMJ open* **13**, e061431 (2023).
47. Goldstein, B., Giroir, B. & Randolph, A. International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics*. *Pediatr. Crit. Care Med.* **6**, 2–8 (2005).
 48. Rhodes, A. et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med* **43**, 304–377 (2017).
 49. Eisenberg, M., Madden, K., Christianson, J. R., Melendez, E. & Harper, M. B. Performance of an Automated Screening Algorithm for Early Detection of Pediatric Severe Sepsis. *Pediatr. Critical Care Med.* **1** (2019) <https://doi.org/10.1097/PCC.0000000000002101>.
 50. Scott, H. F. et al. Evaluating Pediatric Sepsis Definitions Designed for Electronic Health Record Extraction and Multicenter Quality Improvement*. *Crit. Care Med.* **48**, e916–e926 (2020).
 51. Sepanski, R. J. et al. Designing a Pediatric Severe Sepsis Screening Tool. *Front. Pediatr.* **2**, (2014).
 52. Le, S. et al. Pediatric Severe Sepsis Prediction Using Machine Learning. *Front. Pediatr.* **7**, (2019).
 53. Sepanski, R. J., Zaritsky, A. L. & Godambe, S. A. Identifying children at high risk for infection-related decompensation using a predictive emergency department-based electronic assessment tool. *Diagnosis* **8**, 458–468 (2021).
 54. Mawji, A. et al. Smart triage: Development of a rapid pediatric triage algorithm for use in low-and-middle income countries. *Front. Pediatr.* **10**, (2022).
 55. Kamaleswaran, R. et al. Applying Artificial Intelligence to Identify Physiometers Predicting Severe Sepsis in the PICU*. *Pediatr. Crit. Care Med.* **19**, e495–e503 (2018).
 56. Schlapbach, L. J. et al. Prediction of pediatric sepsis mortality within 1 h of intensive care admission. *Intensive Care Med* **43**, 1085–1096 (2017).
 57. Xiang, L. et al. Machine Learning for Early Warning of Septic Shock in Children With Hematological Malignancies Accompanied by Fever or Neutropenia: A Single Center Retrospective Study. *Front. Oncol.* **11**, (2021).
 58. Nguyen, T. M., Poh, K. L., Chong, S.-L. & Lee, J. H. Effective diagnosis of sepsis in critically ill children using probabilistic graphical model. *Transl. Pediatr.* **12**, 538–551 (2023).
 59. Lamping, F. et al. Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children - a data-driven approach using machine-learning algorithms. *BMC Pediatr.* **18**, 112 (2018).
 60. Spaeder, M. C. et al. Predictive analytics in the pediatric intensive care unit for early identification of sepsis: capturing the context of age. *Pediatr. Res.* **86**, 655–661 (2019).
 61. Stephen, R. J. et al. Sepsis Prediction in Hospitalized Children: Model Development and Validation. *Hospital Pediatrics* **13**, 760–767 (2023).
 62. Mercurio, L., Pou, S., Duffy, S. & Eickhoff, C. Risk Factors for Pediatric Sepsis in the Emergency Department: A Machine Learning Pilot Study. *Pediatr. Emer Care* **39**, e48–e56 (2023).
 63. Avilés-Robles, M. et al. Predictors of Septic Shock or Bacteremia in Children Experiencing Febrile Neutropenia Post-Chemotherapy. *J. Pediatr. Infect. Dis. Soc.* **11**, 498–503 (2022).
 64. Yang, Y. & Zhang, G. Lysosome-Related Diagnostic Biomarkers for Pediatric Sepsis Integrated by Machine Learning. *JIR* **ume 16**, 5575–5583 (2023).
 65. Solé-Ribalta, A. et al. New multivariable prediction model PEdiatric SEpsis recognition and stratification (PESERS score) shows excellent discriminatory capacity. *Acta Paediatr.* <https://doi.org/10.1111/apa.16321> (2022).
 66. Solé-Ribalta, A. et al. Quantification of a qualitative sepsis code: laying the foundations for the automation revolution. *Eur. J. Pediatr.* **182**, 2169–2172 (2023).
 67. Stephen, R. J. et al. Sepsis Prediction in Hospitalized Children: Clinical Decision Support Design and Deployment. *Hosp. Pediatr.* **13**, 751–759 (2023).
 68. Li, C. et al. A machine learning model for distinguishing Kawasaki disease from sepsis. *Sci. Rep.* **13**, 12553 (2023).
 69. Ehwerhemuepha, L. et al. Development and validation of an early warning tool for sepsis and decompensation in children during emergency department triage. *Sci. Rep.* **11**, 8578 (2021).
 70. Liu, R. et al. Prediction of Impending Septic Shock in Children With Sepsis. *Crit. Care Explorat.* **3**, e0442 (2021).
 71. Marassi, C., Socia, D., Larie, D., An, G. & Cockrell, R. C. Children are small adults (when properly normalized): Transferrable/generalizable sepsis prediction. *Surg. Open Sci.* **16**, 77–81 (2023).
 72. Ying, J., Wang, Q., Xu, T. & Lu, Z. Diagnostic potential of a gradient boosting-based model for detecting pediatric sepsis. *Genomics* **113**, 874–883 (2021).
 73. Georgette, N., Michelson, K., Monuteaux, M. & Eisenberg, M. A Temperature- and Age-Adjusted Shock Index for Emergency Department Identification of Pediatric Sepsis. *Ann. Emerg. Med.* **82**, 494–502 (2023).
 74. Martinez, E. M. et al. Optimizing Recognition and Management of Patients at Risk for Infection-Related Decompensation Through Team-Based Decision Making. *J. Health. Qual.* **45**, 59–68 (2023).
 75. Parasuraman, R., Sheridan, T. B. & Wickens, C. D. A model for types and levels of human interaction with automation. *IEEE Trans. Syst., Man, Cybern. A* **30**, 286–297 (2000).
 76. Harley, A., Latour, J. M. & Schlapbach, L. J. The Role of Parental Concerns in the Recognition of Sepsis in Children: A Literature Review. *Front. Pediatr.* **7**, 161 (2019).
 77. Miranda, M. & Nadel, S. Pediatric Sepsis: a Summary of Current Definitions and Management Recommendations. *Curr. Pediatr. Rep.* **11**, 29–39 (2023).
 78. Scott, H. F. et al. Lactate Clearance and Normalization and Prolonged Organ Dysfunction in Pediatric Sepsis. *J. Pediatr.* **170**, 149–155.e4 (2016).
 79. O'Reilly, D. et al. Platelets in pediatric and neonatal sepsis: novel mediators of the inflammatory cascade. *Pediatr. Res* **91**, 359–367 (2022).
 80. Zhang, R. et al. Relationship between thrombocytopenia and prognosis in children with septic shock: a retrospective cohort study. *Platelets* **35**, 2363242 (2024).
 81. Zeng, X. et al. PIC, a paediatric-specific intensive care database. *Sci. Data* **7**, 14 (2020).
 82. Rieke, N. et al. The future of digital health with federated learning. *npj Digit. Med.* **3**, 119 (2020).
 83. Sveen, W., Dewan, M. & Dexheimer, J. W. The Risk of Coding Racism into Pediatric Sepsis Care: The Necessity of Antiracism in Machine Learning. *J. Pediatr.* **247**, 129–132 (2022).
 84. Collins, G. S. et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **385**, e078378 (2024).
 85. Morin, L. et al. The Current and Future State of Pediatric Sepsis Definitions: An International Survey. *Pediatrics* **149**, e2021052565 (2022).
 86. Van Der Vegt, A. H. et al. Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework. *J. Am. Med. Inform. Assoc.* **30**, 1349–1361 (2023).
 87. Vaccaro, M., Almaatouq, A. & Malone, T. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nat. Human Behav.* <https://doi.org/10.1038/s41562-024-02024-1> (2024).
 88. Pushkarna, M., Zaldivar, A. & Kjartansson, O. Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. in *2022 ACM Conference on Fairness, Accountability, and Transparency*

- 1776–1826 (Association for Computing Machinery, New York, NY, USA, 2022). <https://doi.org/10.1145/3531146.3533231>.
89. Sendak, M. P., Gao, M., Brajer, N. & Balu, S. Presenting machine learning model information to clinical end users with model facts labels. *npj Digit. Med.* **3**, 41 (2020).
90. Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K. E. & Dugan, C. Explaining models: an empirical study of how explanations impact fairness judgment. in *Proceedings of the 24th International Conference on Intelligent User Interfaces 275–285* (ACM, Marina del Rey California, 2019). <https://doi.org/10.1145/3301275.3302310>.
91. Tricco, A. C. et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern Med* **169**, 467–473 (2018).
92. Tennant, R., Graham, J., Mercer, K., Ansermino, J. M. & Burns, C. M. Automated digital technologies for supporting sepsis prediction in children: a scoping review protocol. *BMJ Open* **12**, e065429 (2022).
93. Centre for Evidence-Based Medicine. Oxford Centre for Evidence-Based Medicine: levels of evidence (March 2009).

Acknowledgements

No funding was granted for this study. RT's doctoral research is supported by a National Sciences and Engineering Research Council of Canada Graduate Scholarship. We thank Brie McConnell, an information specialist at the University of Waterloo, for her feedback on our manuscript.

Author contributions

R.T., J.G., K.M., J.M.A., and C.M.B. were responsible for study conception and design. R.T. and J.G. developed the study searches with K.M. R.T., J.G., J.K., and K.M. reviewed the literature. R.T., J.G., and J.K. extracted and verified the data. R.T. and J.K. analysed and interpreted the data. R.T. drafted the original manuscript with support from J.K. and K.M. All co-authors reviewed the draft and provided feedback and editing. All authors contributed to and approved the final manuscript.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-024-01361-9>.

Correspondence and requests for materials should be addressed to Ryan Tennant.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024