SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 9416-9426
Open Access. Online & Print

DOI: 10.5281/zenodo.12426923

FROM ALGORITHM TO BEDSIDE: A NARRATIVE REVIEW OF MACHINE LEARNING AND DEEP LEARNING IN EARLY SEPSIS PREDICTION AND A ROADMAP FOR IMPLEMENTATION IN DEVELOPING COUNTRIES

Komal Kaushal[1], V S Randhawa[2*], Ram Murti Sharma[3], Supriya Mahajan[4], Ajay Kumar[5], Ankit[6], Kuldeep Raj[7]

[1]PG Student, Department of Microbiology, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India. Email: kaushalkomal8@gmail.com
[2]Professor, Department of Microbiology, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India. Email: vsrandh@gmail.com
[3]Medical Superintendent, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India. Email: sharmarammurti@gmail.com
[4]Associate Professor, Department of Microbiology, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India. Email: drsupriyamahajan@gmail.com
[5]Senior Consultant, ESI Medical College, Basai Darapur, India. Email: ajayk5@yahoo.com
[6]Consultant, Department of anaesthesia, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India. Email: ankit@sharda.ac.in
[6]General Surgeon, Department of surgery, Sharda School of Medical Sciences and Research and associated hospitals, Greater Noida, India. Email: drkuldeepraj14@gmail.com

Received: 02/11/2025
Accepted: 27/03/2026

Corresponding Author: V S Randhawa (vsrandh@gmail.com)

ABSTRACT

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection, responsible for approximately 11 million deaths annually. Low- and middle-income countries (LMICs) carry approximately 85% of global sepsis cases and over 90% of related mortality. Current bedside scoring toolsincluding quick Sequential Organ Failure Assessment (qSOFA) and Systemic Inflammatory Response Syndrome (SIRS) criteriamiss up to 71% of true sepsis cases and are ill-suited for early intervention in resource-constrained environments. We searched PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore for studies published January 2020 to March 2025, using pre-specified terms related to sepsis prediction, machine learning, deep learning, and resource-limited settings. After removing duplicates (167 to 144 records) and title/abstract screening, 82 studies met inclusion criteria (64 adult, 18 paediatric), all reporting at least one validated performance metric. Machine learning and deep learning algorithms substantially outperform conventional scoring systems. Classical methods (Random Forest, XGBoost) achieve area under the receiver operating characteristic curve (AUROC) values of 0.86–0.90, while deep learning models reach 0.88–0.93 with four to twelve hours of lead time. A recent meta-analysis demonstrated a pooled AUROC of 0.80 for artificial intelligence versus 0.69 for traditional scoring ($P = 0.008$; $I^2 = 42\%$). Prospective validation in developing

countries is virtually absent; the single available LMIC validation study reported specificity declining from 88% to 71% during tropical infection seasons. The Epic Sepsis Model achieved a positive predictive value of only 10% in real-world deployment. Algorithms intended for developing countries must account for local microbiology, antimicrobial resistance patterns, tropical disease epidemiology, and infrastructure constraints. Federated learning, mandatory explainability frameworksincluding SHapley Additive exPlanations and Local Interpretable Model-agnostic Explanationsand prospective multi-site validation are prerequisites for safe deployment. We propose a tiered, phased implementation roadmap for 2025–2030 spanning tertiary, secondary, and district-level facilities.

KEYWORDS: Sepsis; Machine learning; Deep learning; Intensive care unit; Developing countries; Paediatric sepsis; Clinical decision support systems; Explainable artificial intelligence; Global health; Health equity.

INTRODUCTION

Sepsis remains one of the most challenging conditions in critical care medicine globally. The World Health Organization's Global Sepsis Report documented roughly 49 million cases and 11 million deaths annually approximately one in every five deaths worldwide. LMICs shoulder about 85% of cases and over 90% of deaths. Solutions developed in Boston or Bern cannot be assumed to work in Bangalore or Bamako without rigorous local validation.

Consider a four-year-old boy brought to a district hospital in rural India at 2 AM with three days of fever. He is drowsy and not feeding. His vital signs are: heart rate 152 beats per minute, respiratory rate 36 breaths per minute, and systolic blood pressure 78 mmHg. The nurse recognizes he is critically unwell, but without adequate laboratory support, she faces a difficult judgment: is this early sepsis or a self-limiting viral illness? She makes the wrong call, and by morning the child is in refractory septic shock. This scenario is repeated thousands of times daily across LMIC settings with limited laboratory and imaging support.

Patients in developing countries often arrive late, after substantial physiological deterioration.[1] Blood cultures frequently yield multidrug-resistant organisms; meropenem resistance in *Acinetobacter baumannii* exceeds 60% in multiple published reports.[2] Mortality ranges from 27–34% in sepsis without shock to 45–55% in septic shockfar higher than the 15–25% reported from high-income registries.[3,4] Gram-negative organisms predominate across LMIC ICUs. Any prediction algorithm that ignores these epidemiological realities risks not only underperformance but clinically misleading outputs.

Current bedside scoring systems are blunt instruments. SIRS flags nearly every febrile patient (83–96% sensitivity, approximately 26% specificity). qSOFA misses 40–71% of true sepsis cases (sensitivity 29–60%).[5] Full SOFA requires six laboratory parameters often unavailable in district hospitals and was designed to quantify organ dysfunction after the fact, not to detect sepsis early. These tools describe what has already happened, not what is about to occur.

Machine learning (ML)algorithms that learn patterns from data without explicit programming for each decision ruleand deep learning (DL), a subset using layered neural networks designed to remember patterns over time and detect complex relationships in continuous data streams, offer something qualitatively different: prediction four to twelve hours before visible clinical deterioration. That window permits blood culture collection before antibiotics, preserving diagnostic yield and enabling targeted therapy.[6] Classical ML methods (Logistic Regression, Random Forest, XGBoost, LightGBM) operate on structured tabular data and run on standard hospital computers. DL architectures (Long Short-Term Memory [LSTM] networks, Transformers) detect subtle physiological patterns preceding deterioration but are harder to interpret without additional tools.[7]

The area under the receiver operating characteristic curve (AUROC) summarizes how well a model distinguishes sepsis from non-sepsis across all possible classification thresholds; values closer to 1.0 indicate better discrimination. Current evidence confirms that AI-driven ICU tools achieve AUROCs between 0.83 and 0.93 for sepsis prediction.[8] The Surviving Sepsis Campaign 2021 guidelines identify timely recognition as a central management pillar.[6] The PhysioNet/Computing in Cardiology Challenge 2019, testing 104 algorithms on over 40,000 patients, established ML superiority over conventional scoring.[9] However, none of those algorithms were developed or validated on patients from developing countries. That epidemiological mismatch motivated this review.

Box 1. Technical Primer Algorithm Architectures for Sepsis Prediction

Classical Machine Learning (Logistic Regression, Random Forest, XGBoost, LightGBM): These models operate on discrete variablesvital signs, laboratory results, and demographics. They are relatively transparent, especially when paired with SHapley Additive exPlanations (SHAP)a method that quantifies each feature's contribution to an individual predictionand they run on standard hospital infrastructure. Performance metrics are summarized in Table 1.

Deep Learning (LSTM Networks, Transformers): These models ingest temporal data streams and recognize subtle physiological divergence before overt clinical deterioration. They achieve superior discrimination but require integration with explainability frameworks such as SHAP or Local Interpretable Model-agnostic Explanations (LIME)a technique that approximates complex model behaviour with a simpler, interpretable surrogate locally around each prediction. Performance metrics appear in Table 1.

Federated Learning: A collaborative training approach enabling multiple institutions to build models without sharing patient-level data. Each site retains data locally and contributes only model parameter updates. This approach is particularly recommended for LMIC networks where cross-border data sharing is legally or logistically constrained. Rather than viewing this as merely a research endeavour, institutions should recognize federated learning as strategic infrastructure investment for long-term collaborative intelligence.

MATERIALS AND METHODS

This narrative review followed the PRISMA 2020 statement as an organizing framework but did not constitute a formal systematic review or meta-analysis.[10] We did not conduct quantitative pooling due to substantial heterogeneity in prediction windows, sepsis definitions, and performance metrics across included studies.

Definitions

"Developing countries" refers to nations classified as low-income or middle-income economies by the World Bank Atlas method (fiscal year 2025), corresponding broadly to WHO African, South-East Asian, and Eastern Mediterranean regions with per-capita healthcare spending below USD 500 annually. This classification aligns closely with healthcare spending levels and resource availability in critical-care settings. "Developing countries" and "LMICs" are used interchangeably throughout this review.

Search Strategy

We searched PubMed/MEDLINE, Scopus, Web of Science, and IEEE Xplore for literature published January 2020 to March 2025. Search terms included: "sepsis," "septic shock," "machine learning," "deep learning," "artificial intelligence," "prediction," "early warning," "intensive care unit," "emergency department," "clinical decision support," "developing countries," "low and middle income countries," and "resource-limited settings." Reference lists of included studies were hand-searched for additional publications.

Inclusion and Exclusion Criteria

Studies were included if they evaluated an ML or DL algorithm for sepsis prediction in adult or paediatric ICU or emergency department populations and reported at least one validated performance metric (AUROC, sensitivity, specificity, positive predictive value [PPV], or alert rate). We excluded studies focusing exclusively on mortality prediction rather than sepsis onset, those with only internal validation, case reports, editorials, commentaries, and studies without accessible full text.

Study Selection Flow

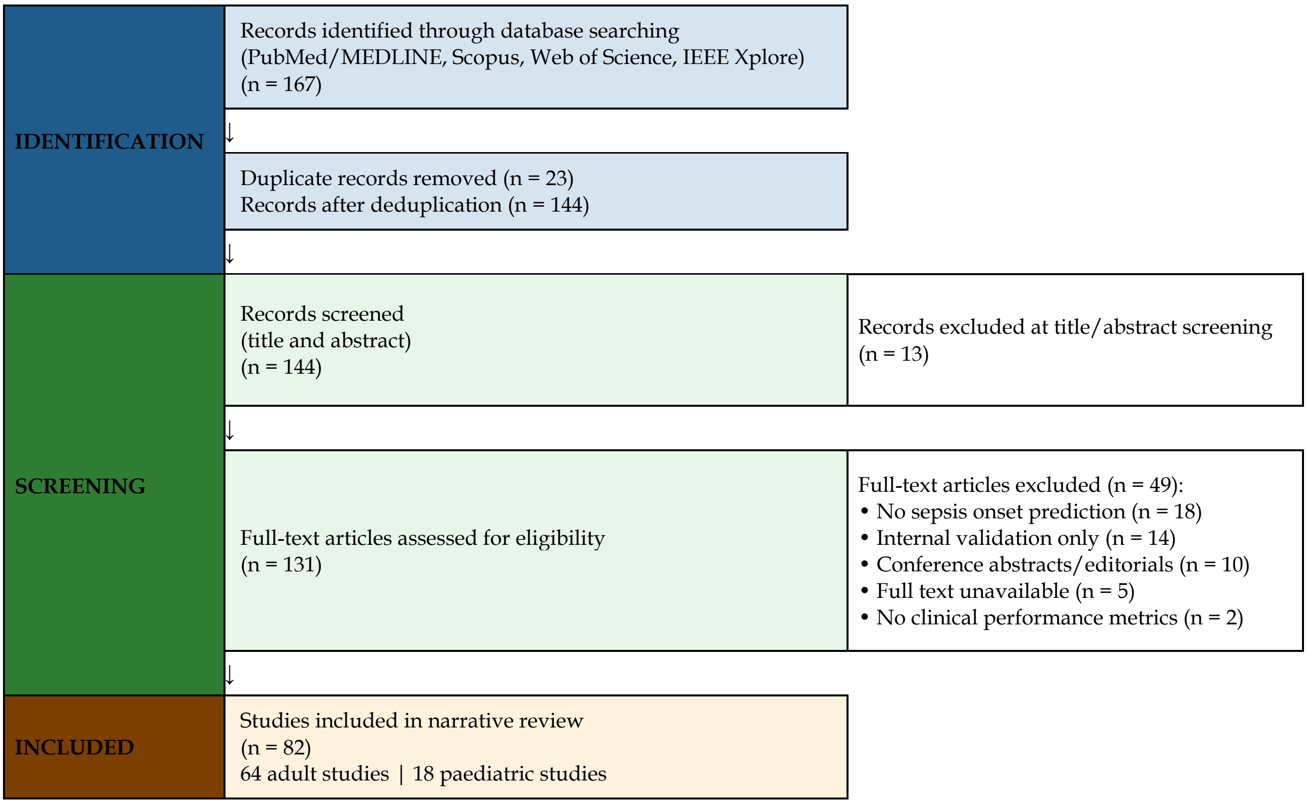

Figure 1. PRISMA 2020 Study Selection Flow Diagram. Records identified through database searching (n = 167); after deduplication (n = 144); title/abstract screened (n = 144); full-text assessed for eligibility (n = 131); studies included in narrative review (n = 82): 64 adult and 18 paediatric studies. A total of 49 full-text articles were excluded (no sepsis onset prediction, n = 18; internal validation only, n = 14; conference abstracts/editorials, n = 10; full text unavailable, n = 5; no clinical performance metrics, n = 2).

Study Selection and Data Extraction

One reviewer extracted data; a second reviewer independently verified a randomly selected 20% subset. We extracted the following variables: study design, sample size, algorithm architecture, prediction window, performance metrics (AUROC, sensitivity, specificity, PPV, alert rate), external validation status, and contextual applicability to LMIC settings.

Limitations

This narrative review carries susceptibility to selective citation bias inherent to the study design. The search was restricted to English-language publications. One reference predating the 2020 search window was retained because no subsequent LMIC publication addressed procalcitonin's diagnostic performance in neonatal sepsis with comparable methodological rigor.[11]

RESULTS

Algorithm Performance Metrics

ML and DL algorithms substantially outperform conventional scoring systems across all reviewed studies. Classical ML approaches (Random Forest, XGBoost) achieve AUROCs of 0.86–0.90 with four to eight hours of lead time before clinical deterioration.[12,13] DL models (LSTMs, Transformers) reach 0.88–0.93 across four to twelve-hour prediction windows.[7] For comparison, SIRS sits at approximately 0.64, qSOFA at 0.74–0.81, and SOFA at 0.74–0.79.[5]

A 2025 systematic review of 39 studies confirmed AI efficacy while flagging pronounced geographic bias toward high-income countries.[14] A recent meta-analysis of 13 studies (56,502 patients with septic shock) demonstrated pooled AUROC of 0.80 (95% CI 0.75–0.85) for AI versus 0.69 (95% CI 0.64–0.74) for traditional scoring ($P = 0.008$; $I^2 = 42\%$).[15] Recurrent neural network architectures reached AUROC of 0.91 within this meta-analysis.

Table 1. Comparative Algorithm Performance and Implementation Feasibility

Algorithm	AUROC	Sensitivity (%)	Specificity (%)	Prediction Window	Feasibility in LMICs
SIRS	0.64	83–96	~26	Bedside	Very high
qSOFA	0.74–0.81	29–60	80–84	0 h	Very high
SOFA	0.74–0.79	74–80	75–79	0–2 h	Moderate
Logistic Regression	0.72–0.80	72–78	75–82	1–3 h	High
Random Forest	0.84–0.88	80–85	80–90	3–6 h	Medium
XGBoost	0.86–0.90	80–87	82–91	4–8 h	Medium
LightGBM	0.85–0.89	79–86	81–90	4–8 h	Medium
LSTM Network	0.88–0.93	80–88	82–92	4–12 h	Low
Transformer	0.88–0.93	83–90	84–93	4–12 h	Very low
Federated LSTM	0.91	82–88	82–90	6–10 h	Strategic Infrastructure

AUROC: Area Under the Receiver Operating Characteristic Curve; SIRS: Systemic Inflammatory Response Syndrome; qSOFA: quick Sequential Organ Failure Assessment; SOFA: Sequential Organ Failure Assessment; XGBoost: Extreme Gradient Boosting; LightGBM: Light Gradient Boosting Machine; LSTM: Long Short-Term Memory. Feasibility graded based on: (i) computational infrastructure requirements; (ii) need for continuous electronic data streams; (iii) local technical expertise; and (iv) documented experience in resource-limited settings. Prediction windows reflect ranges reported across studies; individual implementations may target narrower horizons.

Operational Metrics and Clinical Utility

AUROC reveals nothing about PPV or operational alert rate in practice. External validation of the Epic Sepsis Model documented real-world PPV of approximately 10%.[16] In a hypothetical 50-bed ICU, this generates over 50 alerts daily, 45 of which are false positives. Clinical experience suggests that within four to six weeks of deployment, clinicians learn to disregard the systema phenomenon termed "alert fatigue."[17]

The Targeted Real-time Early Warning System (TREWS) trial demonstrated this outcome is not inevitable. With proper calibration, institutional configuration, and clinician training, PPV reached 30–40%, time to antibiotics decreased by 1.8 hours, and all-cause mortality relative risk was 0.84 (95% CI 0.72–0.98).[18]

Geographic and Epidemiological Gaps

Prospective validation from developing countries is virtually absent from the published literature. The single published LMIC validation study, conducted at a tertiary ICU in India, reported specificity falling from 88% to 71% during tropical infection seasons, nearly doubling the false-positive burden.[19] Every model reviewed was trained on high-income country data: MIMIC-IV from Boston,[20] the eICU Collaborative Research Database from 208 United States hospitals, and HiRID from Bern, Switzerland.[21]

These datasets contain virtually no cases of dengue haemorrhagic fever, malaria, or scrub typhusinfections producing physiological profiles nearly indistinguishable from bacterial sepsis.[22]

Table 2. Key Studies (2020–2025) and Their Implications for Developing Countries

Study (Year)	Algorithm	Sample Size	Key Finding	Relevance for LMICs
Reyna et al. 2020 [9]	104-algorithm benchmark	40,336	ML outperforms conventional scores across all sites	Site-specific validation mandatory
Wong et al. 2021 [16]	Epic Sepsis Model (ESM)	9,585	PPV 10% in external validation; high false-alarm burden	Primary cautionary example for LMIC deployment
Adams et al. 2022 [18]	TREWS	590 clinicians	Mortality RR 0.84; time to antibiotics reduced by 1.8 h	Best prospective evidence; U.S. setting only
Yang et al. 2022 [23]	XGBoost + SHAP	8,450	SHAP explanations increased alert acceptance by 31%	Explainability is prerequisite for clinician uptake
Li et al. 2024 [24]	Federated LSTM	98,432 (12 sites)	Cross-national training without sharing patient data	Template for LMIC multi-centre collaboration
Kumar et al. 2024 [19]	XGBoost	3,400	Specificity fell from 88% to 71% in tropical season	Only published LMIC prospective validation study
Systematic review 2025 [14]	Multiple ML/DL	39 studies	Confirms AI efficacy; significant geographic bias to HICs	Underscores urgent need for LMIC validation
Popoola et al. 2025 [25]	Explainable ML	LMIC hospital	SHAP/LIME deployable with resource constraints	Proof of concept for low-resource settings
Meta-analysis 2025 [15]	AI vs. traditional	56,502	Pooled AUROC: AI 0.80 vs. 0.69 ($P = 0.008$; $I^2 = 42\%$)	Confirms AI superiority; all studies from HICs

TREWS: Targeted Real-time Early Warning System; PPV: Positive Predictive Value; RR: Relative Risk; SHAP: SHapley Additive exPlanations; LSTM: Long Short-Term Memory; LIME: Local Interpretable Model-agnostic Explanations; HIC: High-Income Country.

Explainability and Clinician Trust

Explainability is both an ethical and practical prerequisite for clinical adoption.[26]TreeSHAP computes exact per-patient feature contributions, enabling real-time bedside explanations that directly attribute each prediction to specific physiological derangements.[27] In a multi-site study, providing SHAP explanations increased clinician alert acceptance by 31%.[23] A 2025 Nigerian case study demonstrated that SHAP and LIME methods remain deployable even with unreliable electronic records and intermittent electricity, providing a proof of concept for low-resource settings.[25]

DISCUSSION

Three central findings emerge from this review: (i) the technical superiority of ML and DL over conventional sepsis scoring is well-supported across diverse study populations; (ii) prospective validation in LMIC populations is virtually nonexistent; and (iii) operational metricsparticularly PPV and alert rateultimately determine clinical utility in practice. Deploying sepsis prediction algorithms in developing countries therefore requires matching algorithmic complexity to local clinical, technical, and infrastructural capacity.

The Local Lens: Bridging Global Evidence and Indian Reality

These findings underscore a significant implementation chasm. While the global data confirms algorithmic superiority (Finding 1), the near-total absence of LMIC validation (Finding 2) means we cannot simply import a model trained in Boston to our wards in Bangalore. The difference in specificity during monsoon season (71% versus 88%) is not a statistical nuance; it is the difference between rational antibiotic use and fueling our antimicrobial resistance crisis. Therefore, the remainder of this discussion pivots from "if AI works" to "how AI can work for us, specifically in an Indian medical college ecosystem."

Consider a 35-year-old woman admitted to a tertiary ICU in South India during monsoon season with fever, tachycardia, and hypotension. The ML model flags her as high sepsis risk. Her dengue NS1 antigen returns positive, and her platelet count is falling precipitously. The experienced clinician appropriately overrides the alert, withholds broad-spectrum antibiotics, and initiates targeted fluid resuscitation. She recovers fully without antibiotic exposure. This case illustrates how these tools should function: as intelligent decision support calibrated to local epidemiology, not decision replacement systems that override clinical judgment.

Implementation: A Tiered Framework

Effective implementation requires a tiered approach matching algorithmic complexity to institutional digital maturity. Table 3 provides detailed AUROC ranges and prediction windows for each tier.

Tier 1: Tertiary Academic Centres. Institutions with functional electronic health records (EHRs) should secure ethics approval, ensure data protection compliance, and conduct prospective stepped-wedge trials before clinical rollout.[24,28] Federated learning enables multi-institutional collaboration while maintaining patient data at source.[24] Generalizability from U.S. trial environments to settings with nurse-to-patient ratios exceeding 1:8 or intermittent EHR access remains entirely untested.

Tier 2: Secondary and Regional Hospitals. Simplified models using six to eight vital signs plus three to four laboratory values with manual data entry are appropriate and validated.[29] A prediction window of three to six hours is realistic given the clinical workflow constraints at this tier. ***Implementation Caution:*** Manual data entry remains a significant barrier to model fidelity in Tier 2 settings. Without dedicated data entry personnel or voice-to-text integration, the "prediction window" is eroded by the "data entry lag." Pilot programs should audit the time from vital signs measurement to EHR logging before algorithm deployment.

Tier 3: District Hospitals and Primary Care Settings. A simple four-point clinical referral rulerespiratory rate ≥24 breaths per minute sustained after ten minutes of rest; systolic blood pressure <90 mmHg after 500 mL fluid challenge; lactate ≥2.5 mmol/L if point-of-care (POC) testing is available; or clinical impression of critical illness based on gestalt assessmentcaptures approximately 30% more sepsis cases than qSOFA with a one to three-hour prediction window.[30]

Table 3. Tiered Implementation Framework for Developing Countries

Tier	Digital Maturity	Recommended Approach	Prediction Window	Expected Benefit
Tier 1	Tertiary Functional EHR	XGBoost or LSTM with integrated SHAP explanations; prospective stepped-wedge trial; federated learning across institutions	4–12 h	AUROC 0.88–0.93
Tier 2	Secondary Partial EHR	6–8 vital signs + 3–4 laboratory values with manual data entry; classical ML (Random Forest or XGBoost)	3–6 h	AUROC 0.83–0.90
Tier 3	District Paper-based	Four-point rule: RR ≥24/min; SBP <90 mmHg post-fluid challenge; lactate ≥2.5 mmol/L if POC available; or clinical impression of critical illness	1–3 h	~30% more sepsis detected vs. qSOFA

EHR: Electronic Health Record; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-Term Memory; SHAP: SHapley Additive exPlanations; qSOFA: quick Sequential Organ Failure Assessment; RR: Respiratory Rate; SBP: Systolic Blood Pressure; POC: Point-of-Care Testing.

Paediatric Considerations Within the Tiered Framework

Paediatric sepsis is not adult sepsis in a smaller body. Vital sign reference ranges vary dramatically across developmental age groups, requiring age-normalized z-scores rather than fixed thresholds. Blood pressure is a late and unreliable marker of paediatric sepsis; tachycardia and delayed capillary refill are earlier and more sensitive signals. Children can deteriorate with alarming rapidity and minimal clinical warning.[40]

A 2024 scoping review of 27 paediatric sepsis prediction studies found AUROCs ranging from 0.56 to 0.99; only five studies (18.5%) involved multi-site validation, and most employed logistic regression rather than contemporary ML methods.[41] Class imbalance between sepsis and non-sepsis cases requires specialized algorithmic handling. Age-specific models trained prospectively on LMIC paediatric populations are mandatory, not optional.

Leveraging Existing Indian Research Infrastructure

In India, integration of electronic ICU data from multicentre Indian Network of ICUs (INICUs) studies, antimicrobial resistance surveillance from the Indian Council of Medical Research (ICMR) network, and regional sepsis registries can form the backbone of locally tuned prediction models.[2,3] Federated learning architectures, aligned with the Digital Personal Data Protection Act 2023, allow tertiary and district hospitals to contribute data without compromising patient data sovereignty.[28] The existing research infrastructureincluding point-prevalence studies from the Indian Society of Critical Care Medicine and longitudinal antimicrobial resistance (AMR) surveillanceprovides a foundation more developed than in many other LMICs.

Leveraging these assets for prospective model development and validation should constitute a national priority.

Technical Considerations for LMIC Deployment

Input Data Requirements. Vital sign time-series are foundational inputs; the rate of physiological change matters more than single absolute values.[31,32] Lactate is the most powerful individual laboratory predictor of sepsis.[33] Importantly, delayed sample transport can artificially elevate lactate concentrations by 0.5–1.0 mmol/L per hour of pre-analytic delay, potentially triggering clinically misleading alerts.[34] Procalcitonin (PCT) and C-reactive protein should be incorporated as dynamic features; serial PCT guides antibiotic duration and facilitates evidence-based de-escalation.[35,36] Severe malnutritionprevalent in 35–40% of hospitalized children in LMICsblunts inflammatory responses and may cause standard algorithms to systematically underestimate sepsis risk.[37] Diabetes and tuberculosis similarly confound conventional inflammatory markers.[38,39]

Tropical Infections. Dengue haemorrhagic fever, malaria, and scrub typhus mimic bacterial sepsis with remarkable physiological fidelity.[22] When specificity fell from 88% to 71% during tropical infection season in the only published LMIC validation study, the false-positive burden nearly doubled, substantially increasing antibiotic overuse risk.[19] Practical institutional responses include: mandatory clinical override when tropical infection serological testing returns positive; algorithmic suppression of antibiotic prompts during confirmed peak endemic seasons; prospective tracking of weekly alert rates against laboratory-confirmed sepsis cases; and systematic threshold recalibration during peak tropical disease seasons.

Explainability in Practice. Clinicians need to know which specific physiological derangements drove an alert before they will modify prescribing behavior. An experienced intensivist will not change antibiotic prescribing based on an unexplained probability score from an opaque algorithm. TreeSHAP enables force plots that communicate at the bedside: "Sepsis probability 74%. Primary contributors: lactate rising from 1.9 to 3.6 mmol/L over four hours; respiratory rate 29 breaths per minute and climbing; mean arterial pressure falling from 84 to 66 mmHg. Reassuring features: temperature 37.3°C; no new antibiotics in past 24 hours."[27] Deploying ML alerts without explanatory output constitutes a substantial clinical and ethical risk.

Governance Considerations

Algorithmic bias arising from non-representative training data is particularly concerning for LMICs. Urban-rural differences in disease trajectory, socioeconomic factors affecting laboratory ordering patterns, affordability-driven missing data, and regional pathogen variation can all introduce systematic prediction bias that disproportionately harms already-marginalized populations.[42] Stratified performance auditing should be mandatory, with monitoring disaggregated by sex, age group, site, socioeconomic stratum, geographic origin, nutritional status, and language.

Institutions deploying sepsis AI tools must secure ethics committee approval with explicit data management provisions, establish compliant informed consent pathways aligned with local regulations, and define data retention and deletion schedules per national legislation.[28] The reputational and legal risks of deploying inadequately validated algorithms in vulnerable populations far outweigh any efficiency gains from premature implementation.[43]

Synthesis and Limitations

The recent meta-analysis confirms AI superiority for mortality risk stratification (pooled AUROC 0.80 vs. 0.69, P = 0.008).[15] The TREWS trial's mortality benefit (RR 0.84) supports prospective implementationbut generalizability to LMIC settings with fundamentally different nurse staffing, laboratory capacity, and pathogen profiles remains entirely untested.[18] The single LMIC validation study showed specificity dropping to 71% during tropical disease season.[19] The microbiological data layerAMR patterns, local pathogen distribution, and serial culture resultsis conspicuously absent from nearly all published prediction models.[2,44]

Implementation Roadmap: 2025–2030

Technical validation is only the first step in successful implementation. Training clinical and data science staff, securing institutional leadership buy-in, establishing sustainable operational funding, and building local model stewardship capacity are equally important prerequisites.[45]

Table 4. Implementation Roadmap Timeline: 2025–2030

Phase	Timeline	Key Activities	What Success Looks Like
Phase 1: Foundation	Years 1–2	Establish coordinating centre; multi-centre data collection with microbiological linkage; train junior faculty in AI methods	Ethics approvals secured; ≥5 sites enrolled; baseline dataset >5,000 admissions
Phase 2: Development	Years 3–4	Train federated LSTM model on ~20,000 ICU admissions; build SHAP explanation interface; prepare regulatory submissions	AUROC ≥0.85 on local validation; clinician satisfaction ≥80%; regulatory dossier submitted
Phase 3: Validation	Year 5	Prospective stepped-wedge trial; primary outcome: time from alert to pre-antibiotic blood culture; secondary outcome: 30-day mortality	Reduced time to antibiotics; favourable mortality signal; no increase in adverse events
Phase 4: Scale	Year 5+	Disseminate to Tier 2–3 facilities; continuous monitoring and seasonal recalibration; cost-effectiveness analysis	Demonstrated cost-effectiveness; sustained performance across all tiers; national policy endorsement

ICU: Intensive Care Unit; LSTM: Long Short-Term Memory; SHAP: SHapley Additive exPlanations; AUROC: Area Under the Receiver Operating Characteristic Curve.

CONCLUSION

The technical case for ML and DL-based sepsis prediction is solid and growing stronger with each successive meta-analysis. For developing countries, the question is no longer whether these tools work in principle, but whether we can adapt, validate, and deploy them in ways that are operationally feasible, clinically trustworthy, and genuinely beneficial within the constraints of real-world resource-limited environments.

Four imperatives consolidate the core themes of this review and point toward actionable next steps:

1. Prospective, Multi-Centre LMIC Validation with Mandatory Microbiological Linkage. Without local prospective validation, models will fail when encountering local pathogens and seasonal disease patterns that were absent from training data.[19]

2. Mandatory Explainability Integrated into Every Deployed System. Clear, actionable, patient-specific reasons for alerts improve clinician acceptance, reduce alert fatigue, and constitute an ethical obligation to patients and healthcare systems.[23,26,27]

3. Context-Specific Calibration and Seasonal Recalibration. Alert thresholds must explicitly account for local nurse-to-patient ratios, laboratory turnaround times, and tropical disease epidemiology that fluctuates predictably by season.[17,18,19]

4. Dedicated Paediatric Sepsis Prediction Research. Age-specific prediction models trained on LMIC paediatric populations with appropriate handling of malnutrition and class imbalance are urgently needed.[41]

The scientific and institutional capacity to develop context-appropriate sepsis prediction tools already exists within developing countries. What is needed now is concerted institutional and policy support to build solutions calibrated to local pathogens, local resistance patterns, local infrastructure constraints, and local patientsfrom the neonatal ICU in a tertiary teaching hospital to the district hospital referral chain in a remote rural setting. Regional consortia can build, validate, and continuously maintain these models with local stewardship. The children and adults who arrive at district hospitals in the early hours, drowsy and deteriorating, deserve tools that were built for themnot tools borrowed from settings that bear little resemblance to their own.

DECLARATIONS

Conflicts of interest: There is no any conflict of interest associated with this study.

Consent to participate: We have consent to participate.

Consent for publication: We have consent for the publication of this paper.

Authors' contributions: All the authors equally contributed the work.

REFERENCES

1. Nasa P, Juneja D, Singh O. Severe sepsis and septic shock in the intensive care unit: incidence, risk factors and outcome. J Intensive Care Med. 2021;36(10):1199–1207.

2. Indian Council of Medical Research. Annual Report: Antimicrobial Resistance Research and Surveillance Network 2023. New Delhi: ICMR; 2023.

3. Indian Society of Critical Care Medicine INICUs Study Group. Epidemiology and outcomes of sepsis in intensive care units: the INICUs point-prevalence study 2023. Indian J Crit Care Med. 2023;27(11):795–804.

4. Chatterjee S, Bhattacharya M, Bhattacharya PK, et al. Outcomes of sepsis and septic shock in critically ill patients: analysis of a prospective cohort 2020–2023. J Assoc Physicians India. 2023;71(9):11–18.

5. Gnanaraj JP, Bhatt M, Kanthimathinathan HK, Nadel S. qSOFA score in sepsis. Crit Care. 2021;25(1):388.
6. Evans L, Rhodes A, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med. 2021;49(11):e1063–e1143.
7. Moor M, Rieck B, Horn M, et al. Early prediction of sepsis in the ICU using machine learning: a systematic review. EClinicalMedicine. 2021;39:101064.
8. Bose S, Prakash A, Prusty A, et al. Artificial Intelligence (AI) supported decision-making in intensive care units: implications for nursing and medical practice. Cureus. 2025;18(2):e104266.
9. Reyna MA, Josef CS, Seyedi S, et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. Crit Care Med. 2020;48(2):210–217.
10. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
11. Bala Y, Randhawa VS, Kaur R, Saili A, Chitkara S. Procalcitonin as a diagnostic biomarker in neonatal sepsis: experience from a tertiary care centre. Int J Curr Microbiol Appl Sci. 2018;7(8):490–498.
12. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46(3):383–400.
13. Pettinati MJ, Chen G, Rajput KS, et al. Temporal trends in sepsis mortality: a systematic analysis of ICU cohorts using machine learning. Sci Rep. 2021;11(1):22456.
14. AbuHaweeleh MN, et al. Sepsis mortality prediction using machine learning and deep learning: a systematic review. BMC Med Inform Decis Mak. 2025;26:16.
15. Lian X, Liu Y, Liu X, Tao W, Cao B, Fu B, et al. Artificial intelligence for mortality risk stratification in septic shock: a systematic review and meta-analysis. Int J Med Inform. 2025;207:106197.
16. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181(8):1065–1070.
17. Ginestra JC, Giannini HM, Schweickert WD, et al. Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. Crit Care Med. 2020;48(11):1552–1559.
18. Adams R, Henry KE, Sridharan A, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nat Med. 2022;28(7):1455–1460.
19. Kumar R, Selvarajan S, George M, et al. Machine learning for sepsis prediction in a tertiary care ICU: XGBoost validation with seasonal performance analysis. J Intensive Care Med. 2024;39(4):312–321.
20. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10(1):1–9.
21. Hyland SL, Faltys M, Haner M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med. 2020;26(3):364–373.
22. Varghese GM, Trowbridge P, Janardhanan J, et al. Clinical profile and improving outcomes of scrub typhus: analysis of 1,165 cases from South India. Int J Infect Dis. 2020;23:39–43.
23. Yang M, Mehta S, Li Y, et al. Explainable artificial intelligence for sepsis prediction: a multi-site study. Crit Care Med. 2022;50(8):1191–1200.
24. Li F, Zheng X, Pan L, et al. FedSepsis: federated deep learning for sepsis prediction across multiple ICUs without sharing patient data. Nat Med. 2024;30(4):1102–1113.
25. Popoola T, John-Dewole T. Towards an explainable machine learning system for early detection of pediatric sepsis in low-resource hospital settings in Nigeria: challenges and applications. Int J Res Innov Appl Sci. 2025;10(10):1440–1445.
26. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.
27. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56–67.
28. Ministry of Electronics and Information Technology, Government of India. Digital Personal Data Protection Act, 2023. New Delhi: Government of India; 11 August 2023.
29. van Wyk F, Khojandi A, Mohammed A, et al. A minimal set of physiomarkers in continuous high-frequency data streams predict adult sepsis onset earlier. J Am Med Inform Assoc. 2022;29(2):347–354.
30. Ramanathan K, Murali A, Singh R. Clinical gestalt in sepsis triage: a prospective observational study from a high-volume emergency department. J Emerg Med. 2023;64(2):198–206.
31. Caicedo-Torres W, Gutierrez J. ISeeU: visually interpretable deep learning for mortality prediction inside

the ICU. Expert Syst Appl. 2021;169:114486.
32. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med. 2020;48(2):e113–e120.
33. Delahanty RJ, Alvarez J, Flynn LM, et al. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. J Am Med Inform Assoc. 2021;28(2):285–295.
34. Quintard H, Hubert S, Ichai C. Causes and consequences of hyperlactatemia. Ann Intensive Care. 2020;10(1):74.
35. Kumar A, Singh V, Bhargava A, Garg R. Procalcitonin-guided antibiotic therapy in sepsis: a randomised controlled trial. Indian J Med Res. 2022;155(3):412–421.
36. Gupta S, Kodan P, Soneja M, Biswas A, Wig N. Procalcitonin-guided de-escalation of carbapenem therapy in carbapenem-treated sepsis: a prospective observational study. Int J Infect Dis. 2023;128:192–199.
37. International Institute for Population Sciences. National Family Health Survey (NFHS-5), 2019–21: India Report. Mumbai: IIPS; 2021.
38. Pradeepa R, Anjana RM, Joshi SR, et al. Prevalence of generalized and abdominal obesity in urban and rural India: the ICMR-INDIAB Study. Indian J Med Res. 2021;154(3):388–398.
39. World Health Organization. Global Tuberculosis Report 2023. Geneva: WHO; 2023.
40. Weiss SL, Peters MJ, Alhazzani W, et al. Surviving Sepsis Campaign International Guidelines for the Management of Septic Shock and Sepsis-Associated Organ Dysfunction in Children. Pediatr Crit Care Med. 2020;21(2):e52–e106.
41. Tennant R, et al. A scoping review on pediatric sepsis prediction technologies in healthcare. NPJ Digit Med. 2024;7:245.
42. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in healthcare. Annu Rev Biomed Data Sci. 2021;4:123–144.
43. Char DS, Abramoff MD, Feudtner C. Identifying ethical considerations for machine learning healthcare applications. Am J Bioeth. 2020;20(11):7–17.
44. World Health Organization. Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2022. Geneva: WHO; 2022.
45. Shanmugam H, Airen L, Rawat S. Author Response: From Narrative to Navigation: a translational roadmap for AI-enabled early sepsis prediction. Indian J Crit Care Med. 2025;29(12):1052–1053.