

Development and Internal Validation of a Safety-Gated Artificial Intelligence System for Predicting Serious Bacterial Infection among Children in Northern Nigeria

Olanrewaju Eniade

olanrewaju.eniade@ifain.org

International Foundation Against Infectious Diseases in Nigeria

Adewemimo Olaosebikan

International Foundation Against Infectious Diseases in Nigeria

Emmanuel Egbe

International Foundation Against Infectious Diseases in Nigeria

Bernard Ebruke

International Foundation Against Infectious Diseases in Nigeria

Babatunde Ogunbosi

University of Ibadan

Joshua Akinyemi

University of Ibadan

Ehimario Igumbor

International Foundation Against Infectious Diseases in Nigeria

Stephen Obaro

University of Alabama at Birmingham

Research Article

Keywords: serious bacterial infection, artificial intelligence, pediatric infection, clinical decision support, machine learning, low-resource settings, TRIPOD-AI

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10619493/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Introduction:

Serious bacterial infection (SBI), remain a major cause of morbidity and mortality among children under five, particularly in low- and middle-income countries where diagnostic capacity is limited. Despite increasing use of artificial intelligence (AI) in pediatric risk prediction, context-specific AI models for pediatric SBI in Nigeria remain limited. To develop and internally validate an AI-based clinical prediction framework for identifying SBI risk among children in Nigerian hospital settings, with additional safety-gating features to support clinically responsible use.

Methods

We conducted an AI-enabled retrospective prediction model development and internal validation using pediatric clinical records from Nigerian hospital settings. Ensemble prediction combined logistic regression, decision tree, and Extreme Gradient Boosting models implemented in R. Internal validation used a stratified 80:20 development-held-out test split, with tuning, class weighting, and threshold optimization performed within the development partition only. The system architecture incorporated a local large language model, by finetuning MedGemma, for schema-constrained extraction from free-text clinical-note inputs, and a Python-based deterministic gating engine for completeness, plausibility, consistency, and prediction-readiness checks. Metrics including AUROC, sensitivity, specificity, accuracy, predictive values, balanced accuracy, and simulated workflow-abstention rates.

Results

Among 11,466 labeled pediatric records, 936 (8.2%) were SBI-positive. Culture-confirmed SBI was associated with key demographic, anthropometric, clinical, physiologic, and comorbidity factors, while the strongest model predictors were mainly growth, age-related, and maternal factors, with selected clinical signs contributing additional predictive value. The finalized ensemble model achieved AUROC 0.76 (95% CI: 0.72–0.80), sensitivity 0.76 (95% CI: 0.69–0.82), specificity 0.63 (95% CI: 0.61–0.65), accuracy 0.64 (95% CI: 0.62–0.66), PPV 0.13 (95% CI: 0.11–0.16), NPV 0.97 (95% CI: 0.96–0.98), and balanced accuracy 0.70 (95% CI: 0.66–0.73), using an optimized threshold of 0.50. In simulated workflow testing, schema-conforming extraction succeeded in 95% of inputs, while prediction was withheld in 20% because of missing, implausible, or conflicting information.

Conclusion

This study presents a context-specific AI framework for pediatric SBI prediction using Nigerian clinical data. The findings support further external validation, prospective evaluation, and testing with real-world

clinical documentation before routine deployment.

Introduction

Serious bacterial infection (SBI), including bacteremia, sepsis, meningitis, severe pneumonia, and other invasive bacterial infections, remains a major contributor to preventable childhood morbidity and mortality worldwide. In 2024, an estimated 4.9 million children died before their fifth birthday globally, including approximately 2.3 million newborns, with an estimated global under-five mortality rate of 37.4 deaths per 1,000 live births (WHO, 2026; UN IGME, 2026). These deaths remain highly concentrated in low- and middle-income countries, particularly in sub-Saharan Africa, which accounted for 58% of all global under-five deaths in 2024. Infectious diseases remain a major contributor to child mortality in the region, with the nine leading infectious diseases accounting for 54% of under-five deaths in sub-Saharan Africa in 2024 (UNICEF, 2026). Nigeria is among the highest-burden countries; in 2024, WHO reported that only Niger, Nigeria, and Somalia had under-five mortality rates above 100 deaths per 1,000 live births. These patterns reflect persistent weaknesses across the pediatric infection-care pathway, including delayed care-seeking, under-recognition of severe illness, limited diagnostic capacity, variable antimicrobial access, and referral-system constraints. In northwest Nigeria, Bakare et al. found that only 4.9% of caregivers could identify both cough and fast or difficult breathing as pneumonia symptoms, and only 32.3% of children with pneumonia-specific symptoms were taken for care, highlighting important delays before children reach formal health services (Ukwaja et al., 2012; Bakare et al. (2023). Falade et al. reported in Lagos that health workers in primary health centres rarely diagnosed chest indrawing pneumonia under routine programme conditions, reinforcing concerns about frontline recognition of severe pediatric infection (Bakare et al., 2020; Falade et al. (2026). Other Nigerian implementation studies have shown that strengthening pulse oximetry, oxygen services, and health-worker training can improve recognition and management of hypoxaemic pneumonia, but these services are not uniformly available across routine care settings (Colbourn et al., 2024; Graham et al., 2022). At the treatment level, recent global analyses show that antimicrobial resistance and limited access to effective antibiotics remain major barriers to appropriate management of serious bacterial infections in LMICs (Naghavi et al., 2024; Mishra et al., 2025).

Early recognition of SBI is critical because delayed diagnosis and treatment can lead to rapid progression to sepsis, prolonged hospitalization, inappropriate antimicrobial use, preventable complications, and death. However, identifying SBI in young children remains challenging. Clinical presentations are often non-specific, with symptoms such as fever, irritability, respiratory symptoms, or poor feeding overlapping with viral and non-invasive bacterial illnesses. Furthermore, microbiological confirmation is frequently unavailable or delayed, particularly in resource-constrained settings. In many frontline pediatric facilities, clinicians must make rapid triage and treatment decisions based on incomplete histories, limited physiologic measurements, variable documentation quality, and restricted laboratory capacity. These challenges are especially pronounced in Nigerian hospital settings, where under-five mortality are among the highest globally and diagnostic uncertainty continues to shape pediatric infectious disease care.

Prediction-support tools have the potential to improve pediatric triage and clinical decision-making by integrating multiple routinely available clinical features into structured risk estimates. Traditional clinical prediction rules and biomarker-based approaches, such as C-reactive protein and procalcitonin-have been used to stratify febrile children according to their risk of serious infection (Van den Bruel et al., 2011). More recently, machine-learning and artificial intelligence (AI) methods have expanded the scope of pediatric risk prediction by enabling flexible modelling of non-linear relationships, complex interactions among predictors, and high-dimensional clinical data. Several studies have demonstrated the potential of machine-learning models to predict SBI or invasive bacterial infection among febrile infants and children, particularly in emergency department and hospital settings in high-income countries (Ramgopal et al., 2020; Chiu et al., 2021; Lee et al., 2022; Park et al., 2025).

Despite these advances, important gaps remain. First, existing machine-learning models for pediatric SBI or invasive bacterial infection have largely been developed in emergency department or tertiary hospital datasets from high-income or non-African settings. For example, Ramgopal et al. developed a supervised machine-learning model for SBI among febrile infants aged ≤ 60 days in a United States pediatric emergency-care context (ref); Chiu et al. developed models for invasive bacterial infection among febrile infants attending emergency departments across three medical centers in Taiwan (ref); Lee et al. developed and externally validated an SBI prediction model using febrile children from three tertiary hospital emergency departments in South Korea (ref) ; and Park et al (2025) reported a machine-learning model for SBI among febrile young infants in a Korean tertiary-care setting. Although these studies demonstrate the potential of machine-learning approaches for pediatric infection risk prediction, their development populations, diagnostic pathways, laboratory availability, and healthcare-use patterns differ substantially from routine pediatric care in Nigeria.

Second, although AI applications in infectious disease research are increasing across sub-Saharan Africa, evidence remains uneven by geography, disease focus, and implementation maturity (Lokonon et al., 2026; Makola et al., 2026). A recent review of 62 studies, mainly from South Africa, Ethiopia, Kenya, and Tanzania, focused largely on major infectious diseases such as malaria, HIV, COVID-19, and tuberculosis (Lokonon et al., 2026). In Nigeria, emerging AI studies such as a hybrid Neural Network model trained on 4,027 clinical records focused only on neonates from the southwest Nigeria and a few conditions like sepsis, birth asphyxia, jaundice, and prematurity rather than culture-confirmed SBI in under-five children (Odeyemi et al., 2025). These suggests that pediatric SBI in Nigeria is underrepresented in the regional AI literature. This is important because clinical prediction models require external validation, impact assessment, and context-specific evaluation before clinical use, particularly in resource-constrained settings (Arshi et al., 2025; Waldock et al., 2026). Third, many AI prediction systems assume that clinical inputs are complete, structured, and internally consistent, which is often unrealistic in real-world LMIC practice. Missing data, implausible physiologic measurements, inconsistent symptom documentation, and heterogeneous free-text clinical notes are common challenges (Narasimhan et al., 2021). Consequently, models that generate risk estimates from incomplete or contradictory information may produce misleading outputs, potentially resulting in false reassurance or inappropriate escalation of care. Therefore, AI systems intended for frontline clinical

decision support should not only estimate risk but also assess whether the available information is sufficient, plausible, and reliable for prediction. This necessitates an informatics architecture capable of structured data extraction, prediction-readiness assessment, uncertainty management, and clinician oversight.

Large language models offer potential for transforming unstructured clinical notes into structured variables suitable for downstream prediction tasks. However, unrestricted generative outputs potentially introduce safety concerns; hallucination of clinical information, misclassification of key variables, and overconfident interpretations (Singhal et al., 2023). This study uses an internally validated AI-based, context-specific, ensemble machine-learning model to identify SBI risk among children in Nigerian clinical settings. In addition, it utilises an implementation-oriented, safety-gated architecture that integrates schema-constrained extraction from simulated clinical-note inputs, deterministic checks for completeness and plausibility, consistency validation, and clinician authorization prior to model scoring.

Methods

Study Design

This was a retrospective prediction model development and internal validation study using pediatric clinical records obtained from the Community-Acquired Pneumonia and Invasive Bacterial Diseases in Young Nigerian Children (CAPIBD) surveillance study conducted in Kano State, Nigeria (Ref). The study was designed and reported in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis-Artificial Intelligence (TRIPOD-AI) guideline for artificial intelligence-based clinical prediction models (Collins et al., 2024).

Study Setting and Data Source

The development dataset consisted of pediatric clinical records obtained from the

CAPIBD study. The dataset included demographic, symptom-history, signs, and microbiologic variables collected from children presenting with suspected bacterial infections.

Participants

Eligible participants were children younger than five years presenting with suspected bacterial infections at Aminu Kano Teaching Hospital, Kano. Records with missing observation for the outcome variable were excluded from the model development. A total of 11,466 pediatric records were available for system development.

Outcome Definition

The primary outcome was SBI status, defined as culture-confirmed invasive bacterial infection, modeled as a binary classification outcome: Positive SBI: blood culture positive for a clinically significant bacterial pathogen (excluding contaminants such as coagulase-negative staphylococci (CONS) and Bacillus Species). Negative SBI: blood culture negative or culture interpreted as contaminant. Culture interpretation followed existing clinical microbiology procedures used within the source datasets as guided by the Clinical and Laboratory Standards Institute (CLSI). The operational definition was selected to support clinically relevant binary prediction modeling while acknowledging the limitations of culture-based reference standards in resource-constrained environments.

Candidate Predictors

Candidate predictor variables were identified through a structured review of recent literature on pediatric serious bacterial infection (SBI), invasive bacterial infection, bacteremia, and sepsis prediction models published between 2020 and 2025. Variables consistently associated with SBI risk across previous studies were prioritized, particularly those routinely available during early clinical assessment in resource-constrained settings (Chiu et al., 2021; Schinkel et al., 2022). A total of 32 variables were available for use in the dataset (Table 1).

Table 1
List of Variables available for modelling

List of Variables		
Age group	Reported fever	Local infection signs
Birth weight	Lethargy	Immunocompromised status
Place of delivery	Seizure	MUAC
Maternal age	Cough	Previous healthcare visit for current illness
Maternal education	Vomiting	Sickle cell disease
Weight	Diarrhea	Chronic illness present
Height	Rash	
Temperature	Fever duration	
Heart rate	Prior antibiotics	
Respiratory rate	HIV exposure or known HIV	
Systolic blood pressure	Ill appearance	
Diastolic blood pressure	AVPU	
Sex	Respiratory distress signs	

Missing Data Handling

Missing data was handled differently during model development and live clinical workflow operation. During model development, missing numeric and categorical variables were imputed using Multiple Imputation by Chained Equations (MICE). For selected categorical variables where missingness was considered potentially informative in routine clinical practice, an explicit “unknown” category was retained and evaluated. During live workflow execution, prediction was blocked if required predictors were missing. Follow-up information was requested before scoring and cases with unresolved missingness remained in abstention mode. This distinction was intentional because statistical imputation may support predictive model development, whereas real-time clinical deployment requires more conservative safety safeguards.

To improve prediction safety and reduce inappropriate model inference, the application incorporated deterministic validation rules and age-specific physiologic plausibility ranges for edge-case handling. Validation thresholds were informed by pediatric clinical references, sepsis prediction studies, and implementation guidance from trustworthy AI frameworks (Moor et al., 2023; Shashikumar, et al., 2021). Core variables including temperature, heart rate, respiratory rate, oxygen saturation, blood pressure, anthropometric measures, and laboratory parameters were evaluated against predefined biologically plausible ranges before model scoring. The rule engine additionally assessed cross-variable consistency, such as fever-temperature mismatches, implausible age-vital sign combinations, and contradictory respiratory findings. Inputs violating hard physiologic or logical constraints triggered prediction abstention or rejection, while clinically abnormal but plausible values were retained and flagged as high-risk indicators. Missing mandatory variables, implausible physiologic measurements, and out-of-distribution patterns were similarly handled through deterministic abstention logic rather than unrestricted prediction generation.

System Architecture

The system architecture consisted of three interconnected layers: extraction, deterministic validation, and prediction (Fig. 1).

Structured Extraction Layer

Free-text clinical notes were processed using a locally deployed MedGemma large language model. The extraction workflow constrained outputs to a predefined JSON schema to reduce hallucination risk and improve extraction consistency. The extraction layer was restricted to information structuring only and was not permitted to generate treatment recommendations or independent clinical decisions.

AI Development Workflow and Reproducibility

The AI system was developed as a modular workflow comprising structured extraction, deterministic safety validation, ensemble prediction, clinician-facing display, and validation testing. Each component was implemented using open-source or locally deployable tools where possible, with separation between information extraction, prediction-readiness checks, model inference, and output presentation. The major development phases, tools, and reproducibility functions are summarized in Table 2.

Table 2
AI Development Workflow

Phase	How it was built	Tools used	Reproducibility value
LLM extraction	Converted simulated free-text clinical notes into structured clinical variables using schema-constrained outputs	MedGemma, Python, JSON schema, retrieval-augmented generation, ChromaDB, SentenceTransformer	Makes unstructured or semi-structured clinical narratives usable for downstream prediction
Safety gate (Edge-Cases)	Checked extracted variables for missingness, physiologic implausibility, internal inconsistency, and prediction readiness	Python rule engine, predefined pediatric plausibility ranges	Prevents unsafe model scoring when inputs are incomplete, implausible, or contradictory
ML prediction	Built three component models and combined their predicted probabilities using a probability-averaging ensemble	R, logistic regression, decision tree, XGBoost, custom R ensemble script/function, held-out test evaluation	Generates reproducible SBI risk estimates from structured predictors and documents how component model outputs were combined
Clinician interface	Displayed extracted variables, risk estimates, warnings, uncertainty indicators, and authorization controls	Flask, HTML, CSS, JavaScript	Supports clinician review and prevents hidden or autonomous model output
Validation workbench	Tested frozen models on held-out datasets and summarized uncertainty around performance estimates	Python, R, bootstrap reporting	Enables reproducible validation and transparent reporting of performance metrics

Deterministic Safety Gating Layer

A Python-based deterministic gating engine validated prediction readiness before model inference. The gating system evaluated completeness of required variables, plausibility of numeric ranges, internal consistency between variables, out-of-distribution input patterns and explicit clinician authorization before scoring (Table 3).

Table 3
Deterministic Safety Controls in the Prediction Workflow

Safety Control	Purpose	Action
Completeness check	Ensures required predictors are present	Requests follow-up information
Plausibility check	Rejects impossible or out-of-range values	Blocks scoring pending correction
Consistency check	Detects contradictory information	Flags for human review
Out-of-distribution detection	Identifies unfamiliar patterns	Allows abstention or caution labeling
Permission gate	Preserves clinician control	Requires authorization before scoring
Structured output formatting	Reduces unsafe free-text interpretation	Returns fixed clinician-facing summaries

Prediction Layer

Prediction modeling was implemented in R (version 4.5.2) using an ensemble machine learning approach combining logistic regression, decision tree, and XGBoost. The final ensemble probability was calculated as the mean predicted probability across component models.

Model Specification

The logistic-regression component estimated SBI probability using the generalized linear form:

$$\log \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

where:

- p represents the predicted probability of SBI,
- β_0 is the intercept,
- and $X_1 \dots X_n$ represent predictor variables.

Decision Tree Prediction

A Decision Tree (DT) classifier was developed using recursive partitioning to predict serious bacterial infection. The model partitions the predictor space into mutually exclusive regions R_m and assigns a class label to each terminal node. Mathematically, the decision tree can be expressed as

$T(X) = \sum_{m=1}^M c_m I(X \in R_m)$ where M denotes the number of terminal nodes, R_m represents the partitioned regions of the predictor space, and c_m denotes the class prediction associated with each terminal node.

XGBoost prediction

For SBI prediction, XGBoost typically uses the logistic objective

$$\hat{y}_i = \sigma \left(\sum_{k=1}^K f_k(x_i) \right)$$

Where:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

This converts the boosted tree output into a predicted probability between 0 and 1.

Train-Test Splitting Strategy

The harmonized labeled dataset was randomly divided into development and held-out test partitions using an 80:20 ratio. Stratified splitting was applied to preserve the SBI-positive and SBI-negative class distribution across partitions. The development partition (80%) was used for preprocessing, feature prioritization, hyperparameter tuning, threshold optimization, and model fitting, whereas the held-out test partition (20%) was reserved exclusively for final internal evaluation. To reduce information leakage, train-test splitting was performed before preprocessing, imputation, scaling, feature prioritization, or imbalance handling. All preprocessing parameters, including imputation values and categorical-level mappings, were estimated using the development partition only and then applied unchanged to the held-out test set.

Duplicate and Cohort-Overlap Checks

The training workflow included exact row-signature checks for duplicate records and exact overlap between the development and held-out internal test partitions. In the verification runs used to audit the updated pipeline, no exact row overlap was identified between training and test partitions.

Leakage Prevention and Data-Separation Safeguards

To reduce information leakage risk, train-test splitting preceded preprocessing, imputation parameters were learned exclusively on training data, feature prioritization avoided direct use of the held-out test set, duplicate-row checks were performed, and exact train-test overlap detection was implemented. Internal evaluation emphasized held-out test-set performance from the finalized leakage-safer workflow. Performance measures included the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, balanced accuracy, accuracy, and calibration slope. Also, calibration intercept, Brier score, and confusion-matrix statistics were assessed. Five-fold cross-validation remained available as an internal stability procedure but was not used as the primary final performance estimate. Following stratified 80:20 splitting, approximately 80% of labeled records were allocated to model development and 20% reserved for held-out internal evaluation.

Hyperparameter Tuning and Class-Imbalance Handling

Hyperparameter tuning was conducted within the 80% development partition using internal train-validation resampling. The 20% held-out test set was reserved exclusively for final evaluation and was not used for feature selection, preprocessing, tuning, or threshold selection. Logistic regression, decision tree, and XGBoost models were tuned using model-specific parameters, including regularization strength, tree complexity, learning rate, subsampling, and boosting rounds. Because SBI-positive cases were uncommon, class weighting was applied to reduce majority-class dominance, and threshold optimization was performed within the development data to prioritize balanced accuracy while maintaining clinically acceptable sensitivity. Random oversampling, undersampling, and SMOTE were available as sensitivity options but were not used as default approaches. Final performance was assessed once on the held-out test set, with Wilson 95% confidence intervals for sensitivity, specificity, and accuracy, and bootstrap percentile 95% confidence intervals for AUROC and balanced accuracy.

Deterministic Validation and Edge-Case Handling

To improve prediction safety and reduce inappropriate model inference, the application incorporated deterministic validation rules and age-specific physiologic plausibility ranges for edge-case handling. Validation thresholds were informed by pediatric clinical references, sepsis prediction studies, and implementation guidance from trustworthy AI frameworks (Moor et al., 2023; Shashikumar, et al., 2021). Variables including temperature, heart rate, respiratory rate, oxygen saturation, blood pressure, anthropometric measures, and laboratory parameters were evaluated against predefined biologically plausible ranges before model scoring. The rule engine additionally assessed cross-variable consistency, such as fever-temperature mismatches, implausible age-vital sign combinations, and contradictory respiratory findings. Inputs violating hard physiologic or logical constraints triggered prediction abstention or rejection, while clinically abnormal but plausible values were retained and flagged as high-risk indicators. Missing mandatory variables, implausible physiologic measurements, and out-of-distribution patterns were similarly handled through deterministic abstention logic rather than unrestricted prediction generation.

Operational Workflow Testing

The end-to-end workflow was tested using 30 simulated, de-identified clinical-note inputs derived from the study variables and designed to reflect plausible pediatric presentations in the source dataset. Testing was conducted by the study data-management/model-development team, with clinical review of flagged outputs. Each test case was processed through the full pipeline: schema-constrained extraction, completeness checks, physiologic plausibility checks, consistency checks, clinician-authorization simulation, model scoring where permitted, and structured output generation. Successful extraction was defined as production of a valid schema-conforming output. Prediction blocking was defined as any case in which model scoring was prevented because of missing required predictors, implausible physiologic values, internally inconsistent information, or absence of authorization. Blocked cases were reviewed to classify the reason for abstention as missing data or implausible/conflicting input.

Ethical Considerations

Ethical approval

for the Community-Acquired Pneumonia and Invasive Bacterial Diseases in Young Nigerian Children (CAPIBD) study was obtained from the National Health Research Ethics Committee of Nigeria (NHREC) (Approval No: NHREC/21/08/2008/AKTH/1633) and the Ethics Review Board of Aminu Kano Teaching Hospital (Approval No: AKTH/MAC/SUB/12A/P-3/VI/1733). Written informed consent to participate in the original CAPIBD study was obtained from the parent or legal guardian of every enrolled child before enrolment. Authorization to access and use the dataset for this study was granted by the International Foundation Against Infectious Diseases in Nigeria (IFAIN). All data used for analysis were retrospectively collected and de-identified prior to model development and analysis to protect participant confidentiality and privacy. The AI framework was designed as a clinician-support tool and not as an autonomous diagnostic system.

Results

Dataset Composition

A total of 11,466 pediatric records were available for system development, including 936 culture-confirmed SBI-positive cases, corresponding to an event prevalence of 8.2% (Table 4).

Table 4
Dataset Composition

Dataset Component	Count	Proportion
Total labeled records	11,466	
SBI-positive cases	936	8.2%
SBI-negative cases	10,530	91.8%

Participant Characteristics

The cohort consisted primarily of infants and young children, with a mean age of 605 days (IQR: 284–1300). The mean height was 65 cm (SD 15), while the median weight was 8.1 kg (IQR 4.5–12.0). The median temperature at presentation was 38.2°C (IQR 37.7–38.8). The most common symptom was fever seen in 92.5% of participants, other symptoms, and prior antibiotic exposure – 33.3%, are as shown in Table 5. of cases (Table 5).

Table 5
Descriptive Characteristics of Selected Variables

Variable	Summary
Age- Median (IQR)	605 days (IQR: 284–1300)
Mean height (cm)	65 (SD 15)
Weight (kg)- Median (IQR)	8.1 kg (IQR 4.5–12.0)
Temperature (°C) - Median (IQR)	38.2°C (IQR 37.7–38.8)
Reported fever present	92.5%
Poor feeding present	30%
Cough present	40.7%
Vomiting present	34.2%
Diarrhea present	25.8%
Rash present	5.6%
Prior antibiotic exposure	33.3%

Association Between Demographic Characteristics and Culture-Confirmed SBI

Factors associated with a higher proportion of culture-confirmed SBI included; neonates and older children ($p < 0.001$), birthweight less than 2.5 kg ($p = 0.013$), home-delivery ($p = 0.001$), mothers with formal education ($p = 0.012$), and lower weight and extreme height categories $p < 0.001$). However, the findings for birth weight, place of delivery, maternal age, and maternal education should be interpreted cautiously because of the high proportion of missing observations for these variables (Table 6).

Table 6
Association Between Demographic Characteristics and Culture-Confirmed SBI

Predictor	Total n	SBI-negative n (%)	SBI-positive n (%)	p-value*
Age group				
Neonate (0–28 d)	227	202 (89.0)	25 (11.0)	< 0.001
Infant (29–364 d)	3,517	3,309 (94.1)	208 (5.9)	
Child (≥ 365 d)	7,722	7,019 (90.9)	703 (9.1)	
Weight				
< 2.5 kg	945	826 (87.4)	119 (12.6)	< 0.001
2.5–4.9 kg	1,706	1,576 (92.4)	130 (7.6)	
5.0–9.9 kg	3,491	3,305 (94.7)	186 (5.3)	
≥ 10 kg	3,800	3,512 (92.4)	288 (7.6)	
Height				
< 50 cm	1,171	1,072 (91.5)	99 (8.5)	< 0.001
50.0–74.9 cm	2,423	2,314 (95.5)	109 (4.5)	
75.0–109.9 cm	2,232	2,105 (94.3)	127 (5.7)	
≥ 110 cm	388	338 (87.1)	50 (12.9)	
Sex				
Female	5,069	4,662 (92.0)	407 (8.0)	0.641
Male	6,397	5,868 (91.7)	529 (8.3)	
MUAC				
< 11.5 cm	421	393 (93.3)	28 (6.7)	0.431
11.5–12.4 cm	377	359 (95.2)	18 (4.8)	
≥ 12.5 cm	3,826	3,624 (94.7)	202 (5.3)	
Birth weight				
< 2.5 kg	79	69 (87.3)	10 (12.7)	0.013
2.5–3.9 kg	214	205 (95.8)	9 (4.2)	
≥ 4.0 kg	20	20 (100.0)	0 (0.0)	
Place of delivery				

Predictor	Total n	SBI-negative n (%)	SBI-positive n (%)	p-value*
Age group				
Health facility	222	211 (95.0)	11 (5.0)	0.001
Home	49	39 (79.6)	10 (20.4)	
Maternal age				
< 20 years	10	8 (80.0)	2 (20.0)	0.346
20–34 years	371	343 (92.5)	28 (7.5)	
≥ 35 years	55	51 (92.7)	4 (7.3)	
Maternal education				0.012
None	36	28 (77.8)	8 (22.2)	
Primary	45	43 (95.6)	2 (4.4)	
Secondary	137	128 (93.4)	9 (6.6)	
Tertiary	224	207 (92.4)	17 (7.6)	

Percentages are row percentages and were calculated using available non-missing data for each predictor. Denominators vary across predictors because of missing observations.

Clinical Presentation and Physiologic Factors Associated with Culture-Confirmed SBI

Also, children presenting with hypothermia ($< 36.5^{\circ}\text{C}$) or high-grade fever ($\geq 38.5^{\circ}\text{C}$) demonstrating ($p < 0.001$). Similarly, for tachypneic children (≥ 60 breaths/minute - $p < 0.001$), fever duration of seven or more days ($p < 0.001$), absence of cough ($p < 0.001$), and unresponsive children using AVPU neurologic status assessment ($p < 0.001$) all had higher proportions of SBI-positive cases. For heart rate, seizure, vomiting, diarrhea, ill appearance, and respiratory distress signs, SBI-positive proportions were broadly comparable across categories and should therefore be interpreted cautiously (Table 7).

Table 7
Clinical Presentation and Physiologic Factors Associated with Culture-Confirmed SBI

Predictor	Total n	SBI-negative n (%)	SBI-positive n (%)	p-value*
Temperature				
< 36.5°C	770	696 (90.4)	74 (9.6)	< 0.001
36.5–37.4°C	1,497	1,395 (93.2)	102 (6.8)	
37.5–38.4°C	4,092	3,888 (95.0)	204 (5.0)	
≥ 38.5°C	4,222	3,893 (92.2)	329 (7.8)	
Heart rate				
< 120 bpm	2,198	2,055 (93.5)	143 (6.5)	0.617
120–159 bpm	6,516	6,068 (93.1)	448 (6.9)	
≥ 160 bpm	1,880	1,743 (92.7)	137 (7.3)	
Respiratory rate				
< 40 bpm	5,032	4,738 (94.2)	294 (5.8)	< 0.001
40–59 bpm	3,547	3,301 (93.1)	246 (6.9)	
≥ 60 bpm	2,166	1,962 (90.6)	204 (9.4)	
Systolic BP				
< 70 mmHg	11	11 (100.0)	0 (0.0)	0.300
70–89 mmHg	93	84 (90.3)	9 (9.7)	
≥ 90 mmHg	419	364 (86.9)	55 (13.1)	
Diastolic BP				
< 40 mmHg	6	4 (66.7)	2 (33.3)	0.305
40–59 mmHg	86	75 (87.2)	11 (12.8)	
≥ 60 mmHg	183	161 (88.0)	22 (12.0)	
Lethargy				
Yes	403	368 (91.3)	35 (8.7)	0.151
No	10,866	10,123 (93.2)	743 (6.8)	
Seizure				
Yes	804	752 (93.5)	52 (6.5)	0.733

Predictor	Total n	SBI-negative n (%)	SBI-positive n (%)	p-value*
Temperature				
No	9,480	8,837 (93.2)	643 (6.8)	
Cough				
Yes	4,182	3,967 (94.9)	215 (5.1)	< 0.001
No	6,102	5,622 (92.1)	480 (7.9)	
Vomiting				
Yes	3,518	3,298 (93.7)	220 (6.3)	0.142
No	6,766	6,291 (93.0)	475 (7.0)	
Diarrhea				
Yes	2,651	2,468 (93.1)	183 (6.9)	0.730
No	7,633	7,121 (93.3)	512 (6.7)	
Rash				
Yes	573	526 (91.8)	47 (8.2)	0.156
No	9,711	9,063 (93.3)	648 (6.7)	
Fever duration				
0–2 days	3,528	3,337 (94.6)	191 (5.4)	< 0.001
3–6 days	3,323	3,117 (93.8)	206 (6.2)	
≥ 7 days	1,364	1,235 (90.5)	129 (9.5)	
Ill appearance				
Yes	2,019	1,868 (92.5)	151 (7.5)	0.150
No	8,265	7,721 (93.4)	544 (6.6)	
AVPU				
Alert	10,707	9,991 (93.3)	716 (6.7)	< 0.001
Unresponsive	159	132 (83.0)	27 (17.0)	
Voice	403	368 (91.3)	35 (8.7)	
Respiratory distress signs				
Yes	1,512	1,404 (92.9)	108 (7.1)	0.519
No	8,772	8,185 (93.3)	587 (6.7)	

Percentages are row percentages and were calculated using available non-missing data for each predictor. Denominators vary across predictors because of missing observations.

Medical History and Comorbidity Factors Associated with Culture-Confirmed SBI

Children with a previous healthcare visit for the current illness also had significantly higher SBI positivity compared with those without prior healthcare visits ($p = 0.001$). The presence of chronic illness was also significantly associated with SBI occurrence ($p = 0.034$) [Table 8].

Table 8
Medical History and Comorbidity Factors Associated with Culture-Confirmed SBI

Predictor	Total n	SBI-negative n (%)	SBI-positive n (%)	p-value*
Prior antibiotics				
Yes	3,086	2,883 (93.4)	203 (6.6)	0.809
No	6,169	5,755 (93.3)	414 (6.7)	
HIV exposure/known HIV				
Yes	37	33 (89.2)	4 (10.8)	0.094
No	1,651	1,572 (95.2)	79 (4.8)	
Previous healthcare visit				
Yes	3,235	2,975 (92.0)	260 (8.0)	0.001
No	7,793	7,298 (93.6)	495 (6.4)	
Sickle cell disease				
Yes	84	82 (97.6)	2 (2.4)	0.264
No	1,629	1,546 (94.9)	83 (5.1)	
Chronic illness present				
Yes	363	335 (92.3)	28 (7.7)	0.034
No	2,487	2,362 (95.0)	125 (5.0)	

Percentages are row percentages and were calculated using available non-missing data for each predictor. Denominators vary across predictors because of missing observations.

Internal Validation Performance

The finalized ensemble model demonstrated moderate discrimination on the held-out internal test set, with an AUROC of 0.76 (95%CI: 0.72–0.80). Sensitivity and specificity were 0.76 (95%CI: 0.69–0.82) and

0.63 (95%CI: 0.61–0.65), respectively, indicating reasonable detection of SBI-positive cases with moderate exclusion of SBI-negative cases. Overall accuracy was 0.64 (0.62–0.66), while the balanced accuracy of 0.70 (95%CI: 0.66–0.73) suggested moderate combined classification performance despite class imbalance. Binary classification was performed using an optimized probability threshold of 0.50, selected within the development data (Fig. 2).

Feature importance ranking for the SBI prediction model

Feature-importance analysis showed that anthropometric and early-life variables contributed most strongly to the SBI prediction model. Birth weight was the highest-ranked predictor, followed by current weight and gestational age. Age, maternal age, and MUAC also showed substantial contribution, suggesting that developmental, maternal, and nutritional factors were important in model discrimination. Clinical presentation variables, including respiratory rate, heart rate, cough, fever duration, appetite status, lethargy, reported fever, ill appearance, and respiratory distress, contributed moderately but ranked below the anthropometric and early-life predictors. Overall, the model appeared to draw most strongly on growth, age-related, and perinatal characteristics, while acute clinical signs provided additional discriminatory information (Fig. 3).

Operational Workflow Performance

Among 30 simulated clinical-note inputs evaluated, the structured extraction layer generated schema-conforming outputs in 95% of cases. Overall, prediction was blocked in 20% of cases by the deterministic safety-gating layer, including 12% because of missing required data and 8% because of implausible or conflicting inputs. Blocked cases did not proceed to model scoring until the missing, implausible, or contradictory information was reviewed or corrected. This demonstrated the intended safety behavior of the workflow, prioritizing prediction readiness and controlled inference over unrestricted model output generation (Table 9).

Table 9
Operational Workflow Outcomes

Workflow Measure	Result
Successful structured extraction	95%
Prediction blocked overall	20%
Blocked because of missing data	12%
Blocked because of implausible/conflicting inputs	8%

Clinical Output Structure

The system returned clinician-facing summaries organized into five key sections. Clinical impression provided an overview of the patient presentation and initial assessment. Supporting clinical features detailed the relevant findings that informed this impression. Model-based assessment presented the algorithmic predictions and their rationale. Safety and uncertainty indicators highlighted areas of confidence and potential limitations in the model's output. Finally, immediate clinical considerations outlined actionable next steps and critical alerts. This structured reporting strategy was designed to improve interpretability and reduce unsafe reliance on raw technical outputs.

Discussion

This study describes the development and validation of a safety-gated artificial intelligence (AI) framework for predicting culture-confirmed SBI among children in resource-constrained healthcare settings. The proposed framework integrates large language model-based structured extraction, deterministic workflow governance, and ensemble machine-learning prediction within a unified clinician-facing architecture. Unlike many conventional healthcare AI systems that proceed directly from data ingestion to prediction generation, the present framework incorporated explicit validation and safety controls before model inference. This design was intended to reduce unsafe prediction under conditions of incomplete, implausible, or contradictory clinical information. This design aligns with the need for clinically responsible AI systems that are adapted to local data realities and capable of supporting decision-making under diagnostic uncertainty. The model demonstrated moderate discriminatory performance during internal validation. Sensitivity was intentionally prioritized over specificity to reduce missed SBI-positive cases, reflecting the clinical realities of pediatric infectious disease triage where delayed identification of bacterial infection may contribute to progression to sepsis, prolonged hospitalization, and increased mortality. In many frontline pediatric settings, especially within LMIC environments, prediction systems designed for triage support often require greater emphasis on sensitivity and negative predictive value than on maximizing specificity alone. The high negative predictive value observed in internal validation suggests that the framework may have utility in identifying lower-risk presentations while maintaining cautious identification of potentially severe bacterial infections. However, the relatively low positive predictive value reflects the low-prevalence nature of SBI within the evaluated cohorts and reinforces the importance of interpreting the model as a supportive triage aid rather than a confirmatory diagnostic system.

A major contribution of this study lies in the explicit separation between generative extraction, deterministic validation, and model scoring. In this architecture, the language model performed structured information extraction, while the rule-based safety layer assessed whether extracted data met predefined criteria for completeness, physiologic plausibility, internal consistency, and clinician authorization before prediction was permitted. This separation is particularly important in clinical decision-support settings where AI tools are intended to augment, rather than replace, clinical judgment. Importantly, the framework could abstain from prediction when critical information was missing or implausible. This “safe abstention” philosophy aligns with emerging principles of trustworthy and human-centered clinical AI, which emphasize uncertainty awareness, controlled inference, and explicit

safeguards against unsafe autonomous prediction (World Health Organization, 2021; Shashikumar et al., 2021).

The workflow findings further support the relevance of deterministic safety governance in constrained healthcare environments. Approximately one-fifth of evaluated cases triggered prediction blocking because of missing or implausible information. In many conventional machine-learning systems, these records might still generate predictions after silent imputation or unconstrained inference, potentially producing misleading outputs. By explicitly preventing prediction under unsafe conditions, the framework prioritized transparency and clinician oversight over unrestricted automation. This distinction is especially important in LMIC healthcare systems where inconsistent documentation and incomplete clinical observations are common operational realities rather than isolated exceptions.

This study also contributes to the growing literature on AI implementation in low-resource healthcare settings. Many published prediction models are developed using highly curated datasets from high-income countries with mature electronic health record systems and extensive laboratory infrastructure. Such models may not generalize effectively to settings characterized by paper-based documentation, fragmented workflows, and variable diagnostic availability. The present framework was intentionally designed around these operational realities by supporting free-text clinical documentation, incorporating local deployment workflows, and prioritizing routinely obtainable clinical predictors. The use of schema-constrained extraction and deterministic validation may therefore represent a more feasible implementation pathway for AI-assisted pediatric triage support in constrained environments. Several findings from the bivariate analysis were consistent with existing pediatric SBI and sepsis literature. Younger age, abnormal temperature, tachypnea, poor nutritional indicators, and several symptom-history variables demonstrated significant associations with SBI occurrence (Ref). Previous healthcare contact and chronic illness showed associations with SBI positivity, while other comorbidity indicators showed limited separation in the available-case analysis. These findings support the clinical plausibility of the selected predictor set and align with previous studies evaluating machine-learning approaches for invasive bacterial infection and sepsis prediction in pediatric populations (Chiu et al., 2021; Schinkel et al., 2022).

The study has some limitations, the retrospective design introduces the possibility of selection bias, documentation inconsistency, and residual confounding. Despite implementation of leakage-reduction safeguards, retrospective harmonized datasets may still contain latent correlations capable of inflating apparent model performance. Additionally, the development cohort exhibited class imbalance, which may have influenced threshold-dependent performance metrics despite the use of class weighting, and threshold optimization. Operational workflow testing used simulated clinical-note inputs rather than prospectively collected routine clinical notes; therefore, future implementation studies should evaluate extraction performance, clinician interaction, and safety-gating behavior using real-world clinical documentation. Similarly, the present study did not formally evaluate clinician usability, workflow integration, fairness across demographic subgroups, or downstream effects on patient outcomes. Consequently, the findings should be interpreted primarily as evidence of developmental feasibility and

governance-oriented system design rather than proof of deployment readiness. Future studies will be conducted to evaluate clinician usability, workflow integration, fairness across demographic subgroups, or downstream effects on patient outcomes

Nevertheless, this study demonstrates how deterministic governance mechanisms can be integrated with machine-learning prediction systems to support safer AI-assisted pediatric triage in constrained healthcare environments. The framework moves beyond conventional predictive modeling by embedding explicit safety controls, abstention logic, and clinician authorization directly into the inference pipeline. In this context, the principal contribution of the system may not solely be predictive discrimination, but rather the demonstration of a governance-oriented AI architecture that prioritizes transparency, controlled inference, and operational safety within real-world pediatric care environments.

Conclusion

We developed a gated AI system for pediatric serious bacterial infection prediction that combines local large language model-based structured extraction with deterministic workflow safeguards and ensemble machine learning prediction. The architecture was specifically designed for resource-constrained clinical environments where incomplete documentation and variable data quality may make unrestricted prediction unsafe. Internal validation demonstrated promising discrimination and operational feasibility. This work demonstrates the feasibility of integrating deterministic safety governance into AI-assisted pediatric triage systems and provides a potentially scalable framework for trustworthy clinical AI deployment in resource-constrained healthcare environments.

Declarations

Consent for publication:

Not applicable. This manuscript does not contain any individual person's data, images, or identifiable clinical information.

Competing interests:

The authors declare that there are no competing interests.

Authors' contributions

O.E. conceptualised the study, designed the safety-gated AI architecture, developed the machine-learning pipeline, performed the statistical and computational analyses, interpreted the results, and drafted the manuscript. A.O. supported the statistical and computational analyses, results interpretation, and the manuscript draft. E.E., B.E., and B.O. contributed to the clinical, microbiological, and pediatric infectious

disease interpretation of the study and reviewed the manuscript for intellectual content. J.A. contributed to methodological review, statistical interpretation. E.I. provided implementation science, public health, and institutional guidance and critically reviewed the manuscript. S.O. provided senior scientific leadership, pediatric infectious disease expertise. All authors read and approved the final manuscript.

Funding:

No specific funding was secured for the development and internal validation of this model.

Author Contribution

O.E. conceptualised the study, designed the safety-gated AI architecture, developed the machine-learning pipeline, performed the statistical and computational analyses, interpreted the results, and drafted the manuscript. A.O. supported the statistical and computational analyses, results interpretation, and the manuscript draft. E.E., B.E., and B.O. contributed to the clinical, microbiological, and pediatric infectious disease interpretation of the study and reviewed the manuscript for intellectual content. J.A. contributed to methodological review, statistical interpretation. E.I. provided implementation science, public health, and institutional guidance and critically reviewed the manuscript. S.O. provided senior scientific leadership, pediatric infectious disease expertise. All authors read and approved the final manuscript.

Acknowledgement

The authors acknowledge the International Foundation Against Infectious Diseases in Nigeria (IFAIN), the CAPIBD study team, participating health facilities, clinical and laboratory personnel, and data teams whose work contributed to the generation and curation of the data used for this study. The authors also acknowledge the children and families whose clinical records contributed to this research.

Data Availability

The source code and project materials supporting the model development are publicly available on GitHub at: <https://github.com/enicrowndav/Predicting-SBI/tree/master> and are distributed under the GNU General Public License v3.0 (GPL-3.0). The clinical dataset used for model development is not publicly available because it contains sensitive patient-level health information and is subject to institutional data governance and ethical restrictions. Reasonable requests for access to de-identified data may be directed to the corresponding author and will be considered subject to approval by the relevant data custodian and ethics/governance bodies.

References

1. Arshi B, Cowley LE, Rijnhart E, Reeve K, Smits LJ, Wynants L. External validation, impact assessment and clinical utilization of clinical prediction models: a prospective cohort study. *J Clin Epidemiol*. 2025;186:111902. <https://doi.org/10.1016/j.jclinepi.2025.111902>.
2. Bakare AA, King C, Salako J, Bakare D, Uchendu OC, Burgess RA, Shittu F, Iuliano A, Isah A, Ahmed T, Ahmar S, Valentine P, Olowookere TF, McCollum ED, Colbourn T, Falade AG, Graham HR. Pneumonia knowledge and care seeking behavior for children under-five years in Jigawa, Northwest Nigeria: a cross-sectional study. *Front public health*. 2023;11:1198225. <https://doi.org/10.3389/fpubh.2023.1198225>.
3. Bakare AA, Graham H, Agwai IC, Shittu F, King C, Colbourn T, Iuliano A, Aranda Z, McCollum ED, Isah A, Bahiru S, Valentine P, Falade AG, Burgess RA. Community and caregivers' perceptions of pneumonia and care-seeking experiences in Nigeria: A qualitative study. *Pediatr Pulmonol*. 2020;55(Suppl 1S1):S104–12. <https://doi.org/10.1002/PPUL.24620>.
4. Chiu IM, Cheng CY, Zeng WH, Huang YH, Lin CR. Using Machine Learning to Predict Invasive Bacterial Infections in Young Febrile Infants Visiting the Emergency Department. *J Clin Med*. 2021;10(9):1875. <https://doi.org/10.3390/jcm10091875>.
5. Colbourn T, Falade AG, Graham HR, Olojede OE, Bakare AA, Beard J, et al. Pulse oximetry and oxygen services for under-five children with community-acquired pneumonia attending primary and secondary level health facilities in Lagos, Nigeria (INSPIRING-Lagos): a pre-implementation and post implementation study. *BMJ Public Health*. 2024;2:e001210. <https://doi.org/10.1136/bmjph-2024-001210>.
6. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, van Smeden M, Riley RD, Moons KG. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ open*. 2021;11(7):e048008. <https://doi.org/10.1136/bmjopen-2020-048008>.
7. Falade AG, Uchendu OC, Bakare AA, Colbourn T, Olojede OE, Abuo JK, Olasupo O, Graham HR, Burgess RA, Salako J, Bakare AA, Shittu F, Iuliano A, McCollum ED, Nisar YB, Qazi SA, King C. Outcome and management of 2–59-month-old Nigerian children with chest indrawing pneumonia at primary-level healthcare facilities: a prospective cohort study. *J Glob Health*. 2026;16:04004.
8. GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet (London England)*. 2024;404(10459):1199–226. [https://doi.org/10.1016/S0140-6736\(24\)01867-1](https://doi.org/10.1016/S0140-6736(24)01867-1).
9. GBD 2021 Causes of Death Collaborators. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet (London England)*. 2024;403(10440):2100–32. [https://doi.org/10.1016/S0140-6736\(24\)00367-2](https://doi.org/10.1016/S0140-6736(24)00367-2).
10. Graham HR, Bakare AA, Ayede AI, Eleyinmi J, Olatunde O, Bakare OR, et al. Cost-effectiveness and sustainability of improved hospital oxygen systems in Nigeria. *BMJ Global Health*. 2022;7:e009278.

<https://doi.org/10.1136/bmjgh-2022-009278>.

11. Lee B, Chung HJ, Kang HM, Kim DK, Kwak YH. Development and validation of machine learning-driven prediction model for serious bacterial infection among febrile children in emergency departments. *PLoS ONE*. 2022;17(3):e0265500. <https://doi.org/10.1371/journal.pone.0265500>.
12. Lokonon BE, Houetohossou SCA, Tchede BA, Yapi RB, Cailleau A, Haydon DT, Bonfoh B. The use of artificial intelligence based modelling techniques in One Health-related infectious disease studies in Sub-Saharan Africa: a review. *Front Artif Intell*. 2026;9:1778800. <https://doi.org/10.3389/frai.2026.1778800>.
13. Makola D, Dzinamarira T, Duah E, Morerwa S, Maluleke K. Application of artificial intelligence in infectious disease surveillance in sub-Saharan africa: A scoping review. *Smart Health*. 2026;41:100676. <https://doi.org/10.1016/J.SMHL.2026.100676>.
14. McMullan BJ, Andresen D, Blyth CC, Avent ML, Bowen AC, Britton PN, Clark JE, Cooper CM, Curtis N, Goeman E, Hazelton B, Haeusler GM, Khatami A, Newcombe JP, Osowicki J, Palasanthiran P, Starr M, Lai T, Nourse C, Francis JR. Antibiotic duration and timing of the switch from intravenous to oral route for bacterial infections in children: systematic review and guidelines. *Lancet Infect Dis*. 2016;16(8):e139–52. [https://doi.org/10.1016/S1473-3099\(16\)30024-X](https://doi.org/10.1016/S1473-3099(16)30024-X). ANZPID-ASAP group.
15. Mishra A, Dwivedi R, Faure K, Morgan DJ, Cohn J. Estimated undertreatment of carbapenem-resistant Gram-negative bacterial infections in eight low-income and middle-income countries: a modelling study. *Lancet Infect Dis*. 2025;25(9):1011–9. [https://doi.org/10.1016/S1473-3099\(25\)00108-2](https://doi.org/10.1016/S1473-3099(25)00108-2).
16. Moor M, Bennett N, Plečko D, Horn M, Rieck B, Meinshausen N, Bühlmann P, Borgwardt K. Predicting sepsis using deep learning across international sites: a retrospective development and validation study. *EClinicalMedicine*. 2023;62:102124. <https://doi.org/10.1016/j.eclinm.2023.102124>.
17. Mustafa ZU, Salman M, Khan AH, Harun SN, Meyer JC, Godman B. Antimicrobial Use Among Hospitalized Neonates and Children; Findings and Implications from a Comprehensive Point Prevalence Survey Among General Tertiary Hospitals in Pakistan. *Infect drug Resist*. 2024;17:5411–28. <https://doi.org/10.2147/IDR.S491454>.
18. Narasimhan P, Iqbal U, Li YC. Artificial intelligence in clinical risk prediction: promise, performance and the path forward? *BMJ health care Inf*. 2025;32(1):e101707. <https://doi.org/10.1136/bmjhci-2025-101707>.
19. Odeyemi CS, Olaniyan OM, Omodunbi BA, Samuel IB, Soladoye AA, Olawade DB. Hybridized artificial intelligence system for reducing neonatal mortality in Nigeria. *Int J Med Informatics*. 2026;206:106162. <https://doi.org/10.1016/j.ijmedinf.2025.106162>.
20. Park JS, Yoo R, Lim SY, Kim D, Chun MK, Han J, Lee JY, Choi SJ, Oh SH, Lee JS, Lee J. Development of a machine learning-based prediction model for serious bacterial infections in febrile young infants. *BMJ paediatrics open*. 2025;9(1):e003548. <https://doi.org/10.1136/bmjpo-2025-003548>.
21. Schinkel M, Boerman AW, Bennis FC, Minderhoud TC, Lie M, Peters-Sengers H, Holleman F, Schade RP, de Jonge R, Wiersinga WJ, Nanayakkara PWB. Diagnostic stewardship for blood cultures in the

- emergency department: A multicenter validation and prospective evaluation of a machine learning prediction tool. *EBioMedicine*. 2022;82:104176. <https://doi.org/10.1016/j.ebiom.2022.104176>.
22. Sharrow D, Hug L, You D, Alkema L, Black R, Cousens S, Croft T, Gaigbe-Togbe V, Gerland P, Guillot M, Hill K, Masquelier B, Mathers C, Pedersen J, Strong KL, Suzuki E, Wakefield J, Walker N, UN Inter-agency Group for Child Mortality Estimation and its Technical Advisory Group. Global, regional, and national trends in under-5 mortality between 1990 and 2019 with scenario-based projections until 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. *Lancet Global health*. 2022;10(2):e195–206. [https://doi.org/10.1016/S2214-109X\(21\)00515-5](https://doi.org/10.1016/S2214-109X(21)00515-5).
23. Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say I don't know. *npj Digit Med*. 2021;4(1):134. <https://doi.org/10.1038/s41746-021-00504-6>.
24. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, Scales N, Tanwani A, Cole-Lewis H, Pfohl S, Payne P, Seneviratne M, Gamble P, Kelly C, Babiker A, Schärli N, Chowdhery A, Mansfield P, Demner-Fushman D, Agüera Y, Arcas B, Natarajan V. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172–80. <https://doi.org/10.1038/s41586-023-06291-2>.
25. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56. <https://doi.org/10.1038/s41591-018-0300-7>.
26. Ukwaja KN, Talabi AA, Aina OB. Pre-hospital care seeking behaviour for childhood acute respiratory infections in south-western Nigeria. *Int Health*. 2012;4(4):289–94. <https://doi.org/10.1016/J.INHE.2012.09.001>.
27. UNICEF. (2026, March 17). *Progress in reducing child deaths slows as 4.9 million children under five die in 2024* [Press release]. Retrieved July 8, 2026, from <https://www.unicef.org/press-releases/progress-reducing-child-deaths-slows-49-million-children-under-five-die-2024>
28. UNICEF. (2026, March). *Under-five mortality*. UNICEF Data. Retrieved June 8, 2026, from <https://data.unicef.org/topic/child-survival/under-five-mortality/>
29. United Nations Inter-agency Group for Child Mortality Estimation. (2026). *Levels and trends in child mortality: Report 2025*. UNICEF. Retrieved July 8, 2026, from <https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2025/>
30. Van den Bruel A, Thompson MJ, Haj-Hassan T, Stevens R, Moll H, Lakharpaul M, Mant D. Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review. *BMJ*. 2011;342:d3082. <https://doi.org/10.1136/bmj.d3082>.
31. Waldock WJ, Guni A, Darzi A, Ashrafian H. Performance of predictive AI-based clinical decision support systems across clinical domains: A systematic review and meta-analysis. *PLOS Digit health*. 2026;5(3):e0001310. <https://doi.org/10.1371/journal.pdig.0001310>.
32. World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization. Accessed May 25, 2026 <https://www.who.int/publications/i/item/9789240029200>

33. World Health Organization. (2023). *Children: Improving survival and well-being*. Accessed May 22, 2026. <https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality>
34. World Health Organization. (2026, May 1). *Child mortality (under 5 years)*. Retrieved July 8, 2026, from <https://www.who.int/news-room/fact-sheets/detail/child-mortality-under-5-years>

Figures

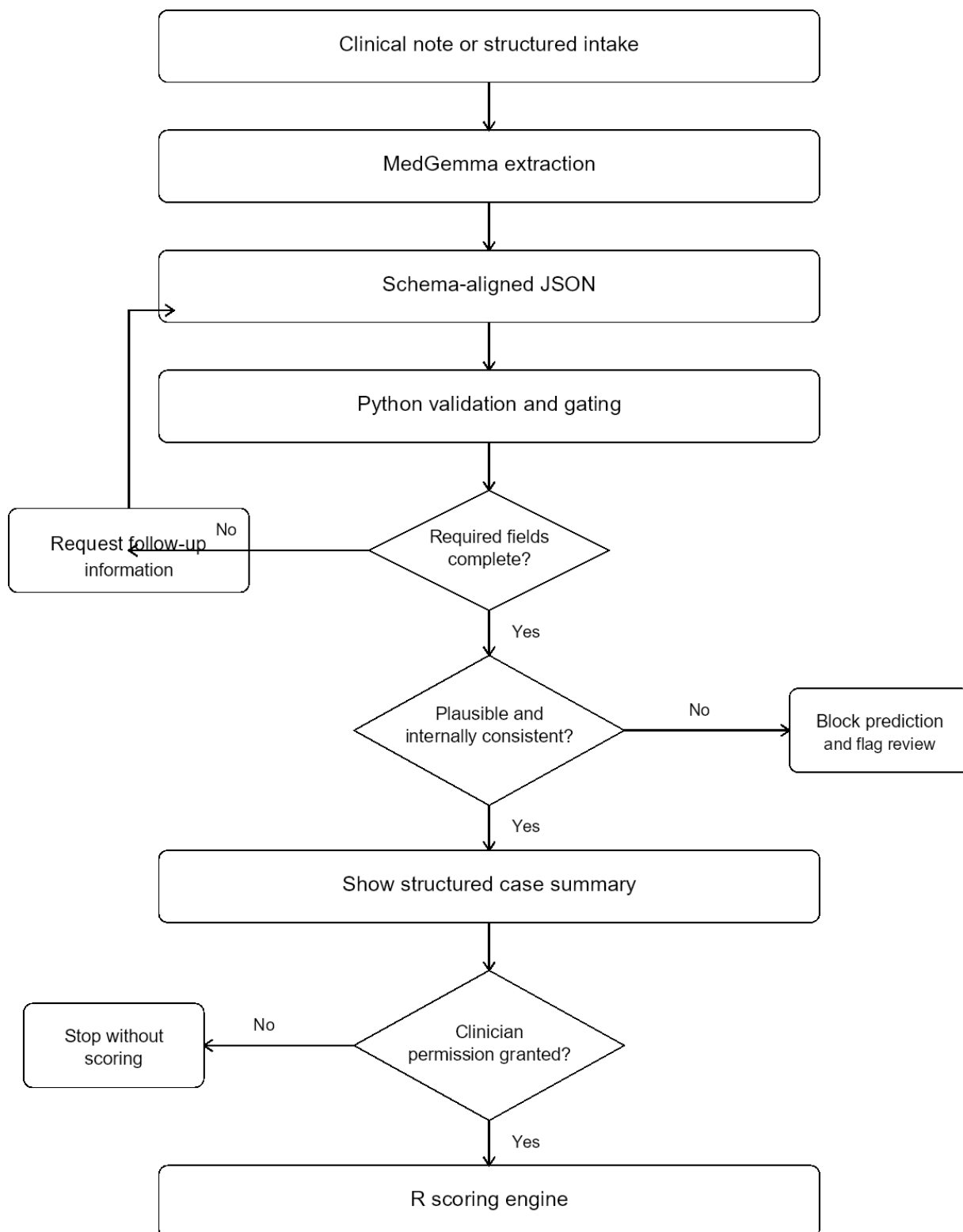


Figure 1

System Architecture

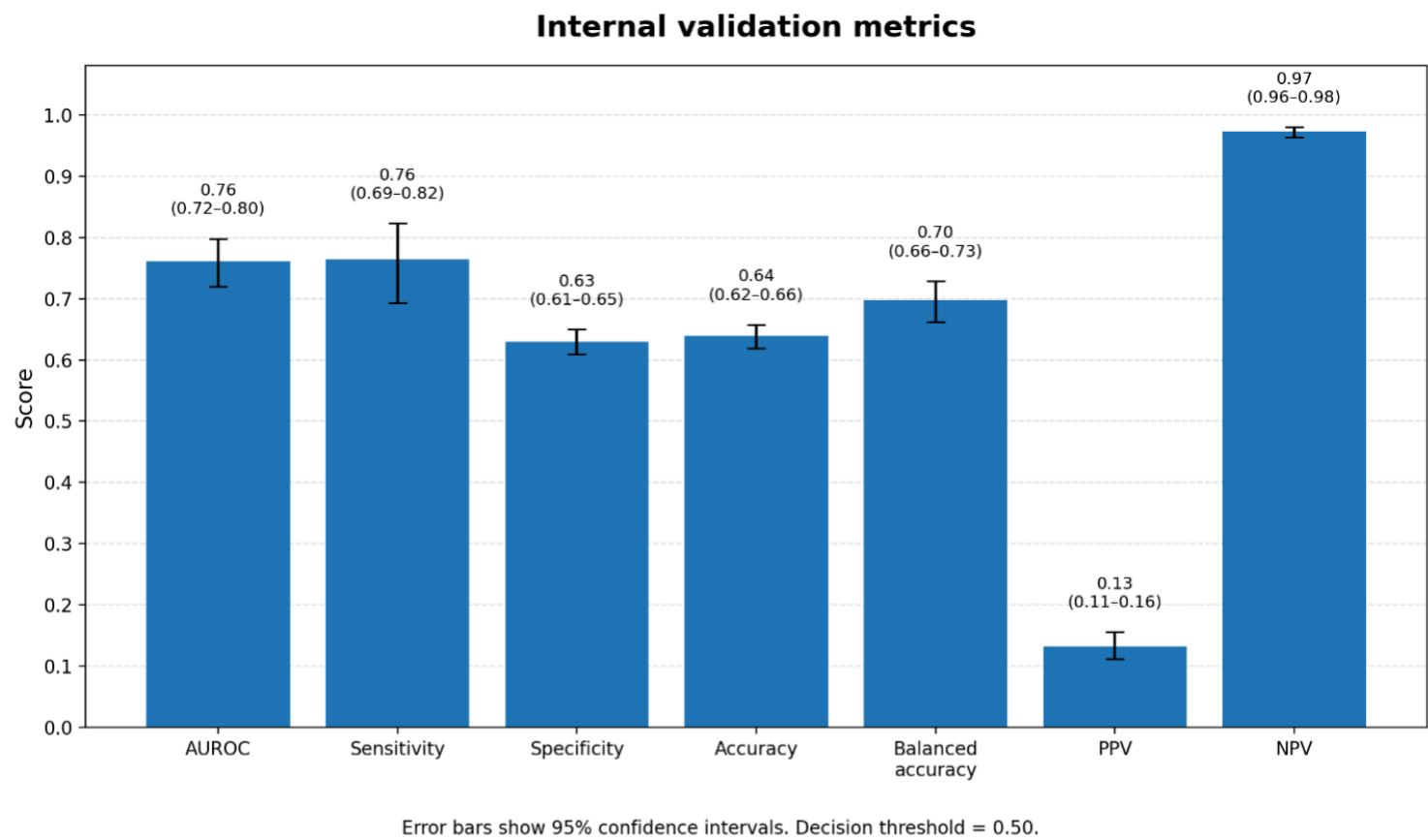


Figure 2

Internal Validation Performance Metrics

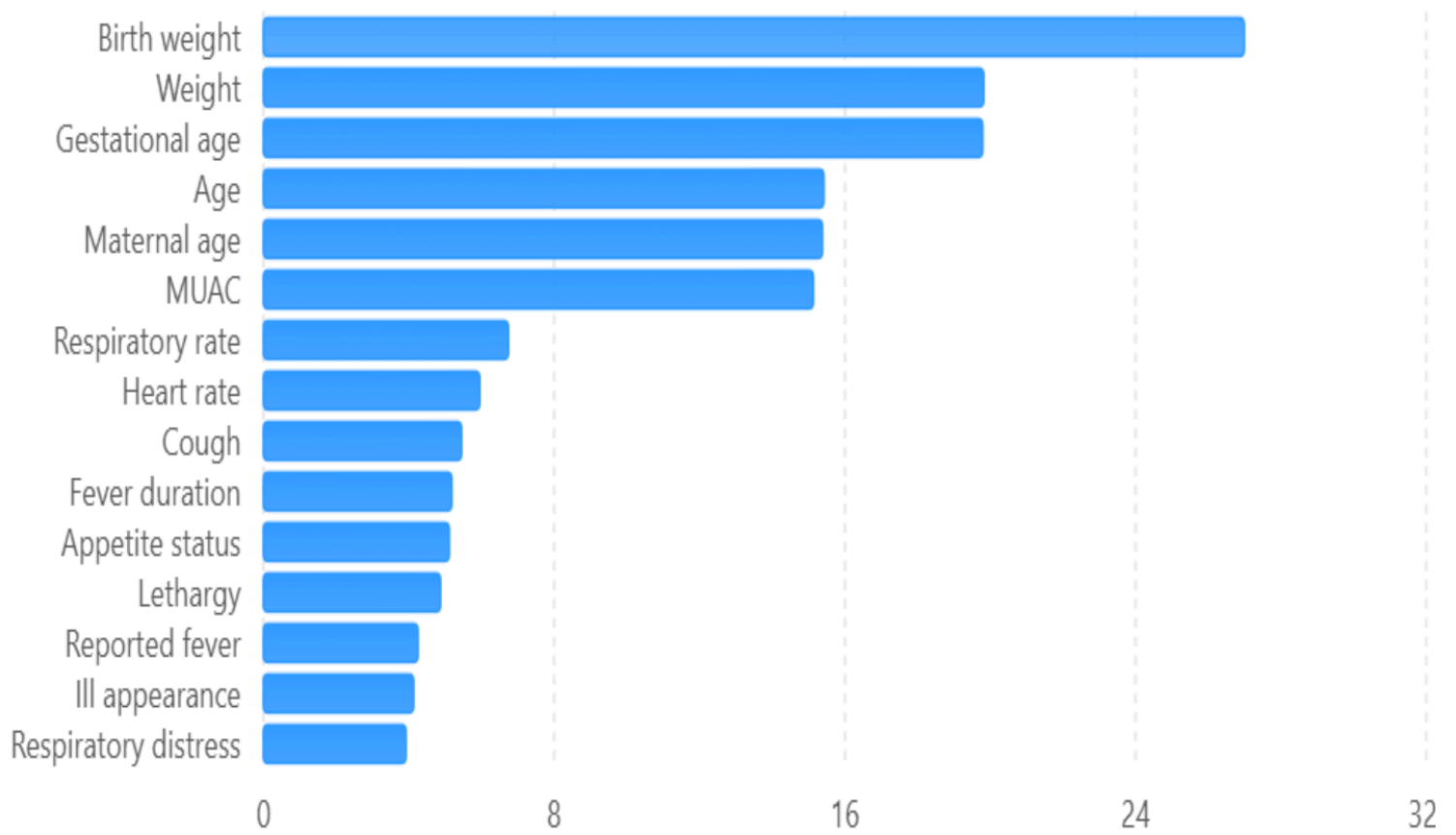


Figure 3

Feature importance ranking for the SBI prediction model (Top 15)