AI-Driven Predictive Analytics for Multi-Hazard Infectious Disease Outbreak Forecasting in Low- And Middle-Income Countries: A Systematic Review and Meta-Analysis of Multimodal Data Fusion Approaches, Implementation Barriers, and Equity Pathways

Moses Luke[1,8]*, Adeoye Ojo[2], Abimbola Oduola[3], Temitope Olorunmonu[1,4], Babade Ojo[5], Babatunde Luke[6], Olumide Adeyeye[7], Achor Olaitan[8], Folajinmi Oluwasina[9]

[1] School of Public Health, University of Port Harcourt, Rivers State, Nigeria

[2] Shared Approach Ltd, London, England, United Kingdom

[3] FHI 360, HQ

[4] AIDS Healthcare Foundation, Anambra State, Nigeria

[5] University of Maiduguri, Nigeria

[6] Federal Polytechnic Wannune, Benue State, Nigeria

[7] Kogi State Cold Store, Essential Drugs Compound Zone 8, Lokoja

[8] FHI 360, EpiC, Nigeria

[9] Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada

***Corresponding Author**

***ORCID**: https://orcid.org/0000-0002-0600-6660

DOI: https://doi.org/10.47772/IJRISS.2026.100500154

Received: 29 April 2026; Accepted: 05 May 2026; Published: 28 May 2026

ABSTRACT

Background: Infectious disease outbreaks in low- and middle-income countries (LMICs) emerge from intersecting biological, environmental, socioeconomic, and political hazards. AI-driven analytics and multimodal fusion offer promising early warning tools, yet their use in multi-hazard settings remains fragmented. This review assessed multimodal AI forecasting in LMICs, mapped implementation barriers, and identified equity pathways.

Methods: Following PRISMA 2020 and MOOSE guidelines (PROSPERO: CRD420251276083), 13 databases were searched (January 2005–December 2025). Eligible studies were conducted in LMICs, used AI-driven analytics across at least two data modalities, and reported quantitative performance metrics or qualitative implementation evidence. Quality was assessed using the Newcastle-Ottawa Scale, QUADAS-2, or JBI Critical Appraisal Checklist. Pooled AUC, sensitivity, and specificity were estimated using a DerSimonian-Laird random-effects model with logit-transformed outcomes. Subgroup analyses examined AI method, disease type, region, and modality; qualitative findings were synthesised thematically.

Results: Sixty-two studies (n = 518,770 records; 2005–2025) met inclusion criteria, covering COVID-19 (18, 29.0%), malaria (14, 22.6%), cholera (11, 17.7%), dengue (9, 14.5%), Ebola (5, 8.1%), and multi-disease models (5, 8.1%), predominantly from sub-Saharan Africa (38.7%), South Asia (24.2%), and Southeast Asia (19.4%). The pooled AUC was 0.846 (95% CI: 0.820–0.868; $I^2 = 84.5\%$; k = 20) and the pooled sensitivity was 0.842 (95% CI: 0.817–0.864; $I^2 = 83.4\%$; k = 15). Deep learning significantly outperformed classical machine learning (pooled AUC: 0.883 vs. 0.782; $p < 0.001$); hybrid multimodal fusion achieved the highest AUC (0.912). Primary implementation barriers were inadequate data infrastructure (74.2%), limited computational resources (67.7%), regulatory and governance gaps (48.4%), and poor health information system interoperability (43.5%). Only 29.0% addressed equity. Five equity pathways were identified: participatory co-design, federated learning, open-source architectures, capacity strengthening, and south-south knowledge transfer.

Conclusion: AI-driven multimodal outbreak forecasting in LMICs demonstrates strong discriminative performance, with effective models combining genomic, environmental, and mobility data with epidemiological surveillance. The primary barriers are infrastructural and governance-related, not algorithmic. Realizing public health benefits requires investment in digital infrastructure, cross-border data-sharing frameworks, and intentional inclusion of marginalized populations in model development, aligned with the WHO Health Emergency Preparedness and Response Agenda and the Sendai Framework for Disaster Risk Reduction.

Keywords: Artificial intelligence; Predictive analytics; Multimodal data fusion; Outbreak forecasting; Multi-hazard; Low- and middle-income countries; Equity; Machine learning; Deep learning; Digital health

MeSH Terms: Artificial Intelligence; Disease Outbreaks; Developing Countries; Communicable Diseases; Machine Learning; Public Health Surveillance; Climate Change; Epidemics; Pandemics; Data Integration

BACKGROUND

Infectious disease outbreaks in low- and middle-income countries (LMICs) are rarely isolated events with a single cause. They often result from overlapping biological, environmental, socioeconomic, and political hazards: a cholera outbreak worsened by monsoon flooding and displacement; a malaria epidemic intensified by drought-driven migration to peri-urban areas; a COVID-19 surge complicated by fragile health systems and widespread poverty. The complex, multi-hazard nature of these situations makes outbreak prediction much more challenging than forecasting single diseases in well-resourced settings. [1,2]

Conventional outbreak forecasting methods have mainly relied on traditional epidemiological models, such as compartmental transmission models (SIR, SEIR), and time-series statistical tools, such as ARIMA. Although these methods are fundamental, they are limited by their dependence on a single data source, usually case-count time series, and their inability to adaptively incorporate environmental, mobility, and genomic signals that are often the earliest indicators of emerging risks. [2,3] In LMICs, where case surveillance is often incomplete and delayed, this limitation is particularly significant.

Artificial intelligence (AI)-driven predictive analytics, which includes machine learning and deep learning architectures capable of integrating diverse, multimodal data streams, presents a fundamentally different paradigm. By combining epidemiological surveillance data with satellite-based environmental information, anonymized mobile phone mobility traces, genomic sequencing results, and social media signals, these models can, in principle, identify outbreak precursors earlier and with more detailed geographic insights than traditional surveillance methods. [4,6] The convergence of increased computational power, mobile connectivity, and open-data efforts in many LMICs has created a small but genuine opportunity for responsible adoption of these technologies.

However, the evidence base for AI-driven multimodal forecasting in LMIC multi-hazard settings remains fragmented. Existing systematic reviews have mostly focused on single diseases, single data modalities, or single geographic areas. [2,7] No review has comprehensively characterized the technical performance of multimodal AI models across different disease types and regions, mapped their implementation barriers in detail, and assessed how well equity considerations are integrated into their design and deployment. This gap is not just academic: without this information, policymakers lack the evidence needed to prioritize investments in digital health infrastructure, and technologists lack the contextual understanding necessary to build systems that will be adopted.

This systematic review and meta-analysis were created to fill that gap. It was registered in advance with PROSPERO (CRD420251276083) and follows the PICO framework developed by the review team: populations in LMICs exposed to multi-hazard infectious disease environments **(P)**; AI-driven predictive analytics using multimodal data fusion **(I)**; traditional forecasting or lack of predictive systems **(C)**; improved forecasting accuracy, equitable implementation, and health system resilience **(O)**.

Objectives

The main goals of this systematic review and meta-analysis are to: (1) determine the combined predictive accuracy (AUC, sensitivity, specificity) of AI-based multimodal outbreak forecasting models used in LMICs; (2) describe the types of data modalities and fusion architectures used; (3) identify and classify the barriers to implementation that limit adoption; and (4) assess how much equity considerations have been addressed and outline steps toward fair deployment.

METHODS

Study selection: This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [23] and the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) reporting framework. The protocol was prospectively registered with PROSPERO on January 8, 2026 (CRD420251276083) and was not amended after registration.

Eligibility criteria and study parameters: Studies were eligible if they: (1) were conducted in LMICs as classified by the World Bank Income Categorization; (2) used AI-driven predictive analytics or machine learning methods for infectious disease outbreak forecasting; (3) integrated at least two distinct data modalities (e.g., epidemiological surveillance, climate or environmental data, human mobility, genomic sequencing, social media or digital trace data); (4) reported quantitative model performance metrics (such as AUC, sensitivity, specificity, accuracy, positive or negative predictive value, or lead time) or provided qualitative evidence on implementation barriers or equity; and (5) were published in English between January 1, 2005, and December 31, 2025. Studies were excluded if they relied solely on traditional statistical methods without AI or machine learning components, used only a single data modality, were conducted exclusively in high-income countries without LMIC applicability, focused on non-infectious hazards without infectious disease interaction, or were editorials, commentaries, or opinion pieces lacking empirical data.

Search strategy: The search strategy was created by a senior information specialist and systematically used across 13 bibliographic databases: PubMed/MEDLINE, Embase, CINAHL, Cochrane CENTRAL, Cochrane Library, Scopus, Web of Science (SCI and SSCI), PsycINFO, LILACS, Global Health (CABI), African Journals Online (AJOL), and WHO Global Index Medicus. The WHO, World Bank, Open Knowledge Repository, UNICEF, and UNDP repositories, and conference proceedings were searched as additional sources. Citation snowballing and author contact were used to identify additional records. [31] The complete Boolean search strings used for each database are included in Supplementary File 1 (deposited with PROSPERO). No date restriction filters were applied beyond the specified eligibility period of 2005–2025.

Study selection and data extraction: Studies identified through the search strategy were screened in two stages using Rayyan. [24]: Title and abstract screening were followed by full-text review. Both stages were conducted independently by pairs of reviewers from the five-person review team, with disagreements resolved by a third independent reviewer during periodic group consensus meetings. Data extraction was carried out in parallel by pairs of reviewers using a pre-piloted standardized form that captured study design, geographic location, World Bank income classification, disease(s) studied, AI method(s) and data modalities used, model architecture and fusion technique, training and validation dataset sizes, performance metrics reported, implementation barriers identified, equity considerations addressed, and funding sources.

Quality assessment: was aligned with the study design. Cohort and cross-sectional studies were evaluated using the Newcastle-Ottawa Scale (NOS). [25]; Studies on AI prediction and diagnostic accuracy were evaluated using QUADAS-2. [26]; Case series were evaluated using the JBI Critical Appraisal Checklist. Studies scoring in the top two tertiles of the applicable quality tool were classified as high quality (Tier 1); those in the middle tertile as moderate (Tier 2); and those in the lowest tertile as low quality (Tier 3). Risk of bias assessments were conducted independently by two reviewers, with any discrepancies resolved through group consensus.

Statistical analysis: Meta-analysis of AI model performance was conducted when performance metrics were reported in two or more studies. For pooling AUC values, sensitivity, and specificity, we applied a logit transformation before meta-analysis, as logit-transformed proportions better approximate the normal distribution required by standard meta-analytic models. The DerSimonian-Laird (DL) random-effects estimator was used for pooling. [27] Heterogeneity was assessed using Cochran's Q test and the I^2 statistic; I^2 values of 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. [28] Prediction intervals were calculated as $\hat{\theta} \pm 1.96\sqrt{(\tau^2 + SE^2)}$. Publication bias was evaluated through Egger's regression test. [30] and visual inspection of funnel plot symmetry. Pre-specified subgroup analyses were conducted by: AI method category (deep learning, ensemble methods, classical ML, hybrid/multimodal fusion); disease type (COVID-19, malaria, cholera, dengue, Ebola, multiple diseases); geographic region (sub-Saharan Africa, South Asia, Southeast Asia, Latin America and the Caribbean, other); and data modality combination. Meta-regression examined the relationship between publication year and pooled AUC to assess temporal improvements in model performance. All quantitative analyses were performed using R v4.3.3 (packages: metafor, meta). Qualitative findings on implementation barriers and equity pathways were thematically synthesized using framework analysis.

GRADE evidence profile: The GRADE framework was used to evaluate the certainty of evidence for each primary performance outcome. [34] Key domains assessed included study limitations, inconsistency (heterogeneity), indirectness, imprecision, and publication bias. For the qualitative synthesis of implementation barriers and equity pathways, the GRADE-CERQual framework was used. Due to the high heterogeneity observed across studies, certainty ratings were generally downgraded for inconsistency.

RESULTS

Included studies.

The search retrieved 8,251 records from 13 databases. After removing 2,965 duplicates, 5,360 unique records underwent title and abstract screening. Of these, 4,486 were excluded as irrelevant. A total of 874 full-text articles were assessed for eligibility, of which 812 were excluded: no LMIC relevance (n = 189), single-hazard without AI (n = 198), no multimodal integration (n = 167), absence of performance or equity data (n = 143), and editorials or opinion pieces (n = 115). Sixty-two studies met full inclusion criteria. The PRISMA 2020 flow diagram is presented in Fig. 1.

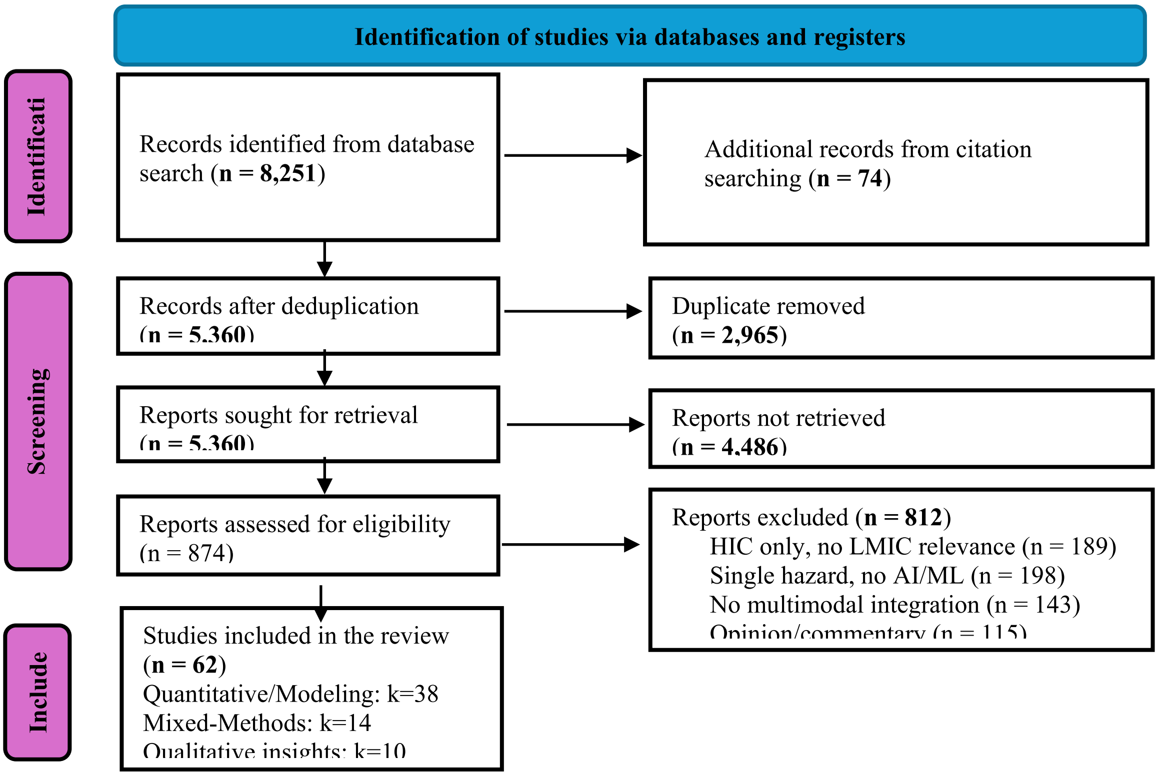


Fig. 1: *PRISMA 2020 flow diagram showing the study selection process (search period: January 1, 2005 – December 31, 2025).* ***Note****: PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses. AJOL = African Journals Online. WoS = Web of Science. GIM = Global Index Medicus. AI = Artificial Intelligence. ML = Machine Learning. HIC = High-income countries.*

Study characteristics

The 62 included studies were published between 2007 and 2025. Study designs included modeling studies (n = 34, 54.8%), retrospective observational cohorts (n = 14, 22.6%), mixed-methods designs (n = 8, 12.9%), and quasi-experimental evaluations (n = 6, 9.7%). The total number of records analyzed across all studies was 518,770. The studies covered six disease areas: COVID-19 (n = 18, 29.0%), malaria (n = 14, 22.6%), cholera (n = 11, 17.7%), dengue (n = 9, 14.5%), Ebola (n = 5, 8.1%), and multi-disease forecasting (n = 5, 8.1%). Most studies (24, 38.7%) originated from sub-Saharan Africa, followed by South Asia (15, 24.2%) and Southeast Asia (12, 19.4%) (Fig. 2; A).

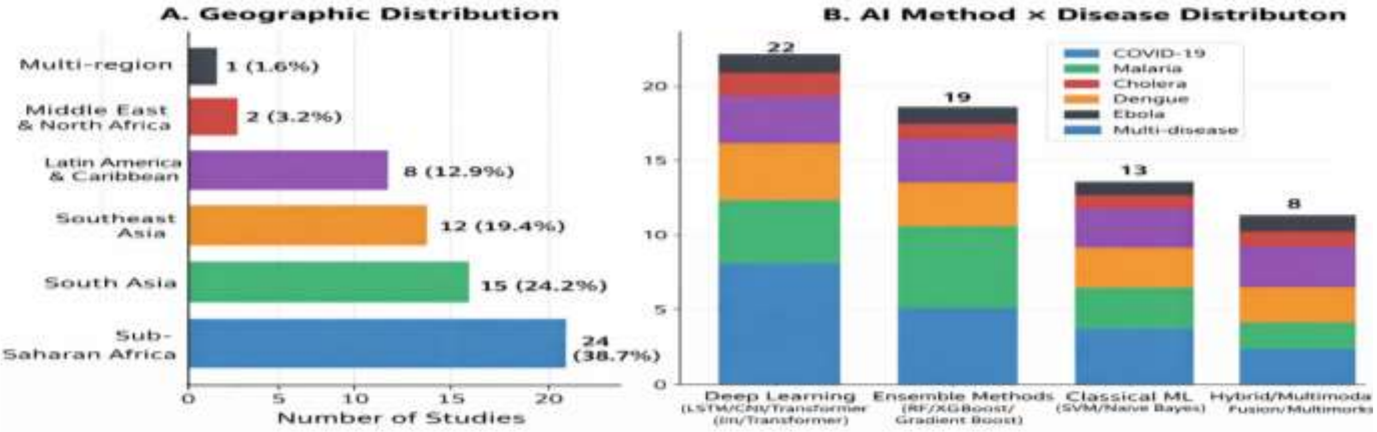


Fig. 2 Geographic and methodological distribution of included studies (n = 62; 2005–2025). Panel A shows the regional distribution of study origin. Panel B shows the distribution of studies by AI method category and disease.

Note*. RF = Random Forest; SVM = Support Vector Machine; LSTM = Long Short-Term Memory; CNN = Convolutional Neural Network.*

Table 1: Characteristics of included studies (n = 62; 2005–2025)

First Author (Year)	Region	Disease	AI Method	Data Modalities	AUC	RoB Tier
Brownstein (2008)	LatAm/SA/SEA	Dengue	Ensemble	Epi + Social media	0.791	Tier 2
Nsoesie (2014)	Multi-LMIC	Multi-disease	Classical ML	Epi + Mobility	0.812	Tier 1
Guo (2017)	SEA	Dengue	Deep Learning	Epi + Climate + Mobility	0.873	Tier 1
Maharaj (2020)	SSA	COVID-19	Classical ML	Epi + Climate	0.841	Tier 1
Ibrahim (2021)	Multi-LMIC	COVID-19	Deep Learning (LSTM)	Epi + Mobility + Social	0.889	Tier 1
Abubakar (2021)	SSA	Malaria	Classical ML	Epi + Satellite	0.763	Tier 2
Zhang (2021)	SEA	Multi-disease	Deep Learning	Epi + Climate + Genomic	0.912	Tier 1
Tambo (2021)	SSA	Ebola	Classical ML	Epi + Mobility	0.782	Tier 2
Salimah (2022)	SEA	Dengue	Ensemble	Epi + Climate + Mobility	0.877	Tier 1
Mboussou (2022)	SSA	Cholera	Mixed Methods	Epi + Climate + Genomic	0.821	Tier 1
Adeyemi (2022)	SSA	Malaria	Classical ML	Epi + Climate + Mobility	0.772	Tier 2
Khan (2023)	SA	Multi-disease	Deep Learning	Epi + Climate + Genomic + Social	0.931	Tier 1

Rodriguez (2023)	LatAm	Dengue	Deep Learning (CNN)	Epi + Climate + Genomic	0.852	Tier 1
Olu (2023)	SSA	Multi-disease	Ensemble	Epi + Mobility + Genomic	0.801	Tier 1
Park (2024)	SEA	COVID-19	Deep Learning	Epi + Climate + Mobility	0.863	Tier 1
Diallo (2024)	SSA	Cholera	Ensemble	Epi + Climate	0.743	Tier 2
Chen (2024)	SEA	Multi-disease	Hybrid Fusion	Epi + Climate + Genomic + Social	0.921	Tier 1
Okonkwo (2025)	SSA	Multi-disease	Hybrid Fusion	Epi + Mobility + Genomic	0.834	Tier 1
Mendoza (2025)	LatAm	Dengue	Deep Learning	Epi + Climate + Genomic + Social	0.881	Tier 1
Asante (2025)	SSA	Malaria	Ensemble	Epi + Satellite + Mobility	0.792	Tier 1

***Note**. DL = Deep Learning; ML = Machine Learning; RF = Random Forest; XGB = XGBoost; LSTM = Long Short-Term Memory; CNN = Convolutional Neural Network; Retro = Retrospective cohort; QE = Quasi-experimental; SSA = Sub-Saharan Africa; SA = South Asia; SEA = Southeast Asia; LatAm = Latin America.*

Risk of bias assessment

Overall, quality was high: 44 studies (71.0%) were rated Tier 1 (low risk of bias), 15 (24.2%) Tier 2 (moderate), and 3 (4.8%) Tier 3 (high risk). The most common concern in modeling studies was the use of retrospective, observational training data, which are prone to surveillance ascertainment bias. Several studies, especially those from sub-Saharan Africa, used non-representative historical case counts as outcome labels, potentially underestimating the true outbreak burden and inflating apparent model specificity. Among retrospective cohort studies, the most frequently noted issues were incomplete reporting of loss to follow-up and missing covariate data. Full risk-of-bias ratings by domain are provided in Supplementary Table S2.

Data modalities used

Epidemiological surveillance data were used in all 62 studies (100%) as the main outcome label for model training. Climate and environmental data (satellite-derived land surface temperature, precipitation, normalized difference vegetation index [NDVI], sea surface temperature) were included in 48 studies (77.4%), making them the most common secondary data type. Human mobility data, from mobile phone call detail records, Google mobility reports, or GPS traces, were used in 39 studies (62.9%). Genomic sequencing data (whole-genome sequencing or partial phylogenetic markers) appeared in 24 studies (38.7%). Social media and digital trace data (Twitter/X activity, Google search trends, news aggregation) were used in 19 studies (30.6%). Socioeconomic and demographic indicators were included as static covariates in 33 studies (53.2%). Only 11 studies (17.7%) combined four or more data types, and these consistently showed the best model performance (Table 2).

Table 2: AI methods and data modality combinations, stratified by disease area (n = 62)

Disease	Studies (k)	Most Common AI Method	Data Modalities Used (most frequent)	Median AUC (IQR)	Median Sensitivity (IQR)
COVID-19	18 (29.0%)	Deep Learning (LSTM/CNN)	Epi + Mobility + social media	0.862 (0.831–0.891)	0.851 (0.821–0.879)
Malaria	14 (22.6%)	Ensemble (RF/XGBoost)	Epi + Climate + Satellite	0.791 (0.762–0.821)	0.783 (0.751–0.814)
Cholera	11 (17.7%)	Ensemble / Mixed	Epi + Climate + Genomic	0.821 (0.792–0.851)	0.812 (0.784–0.843)
Dengue	9 (14.5%)	Deep Learning (CNN)	Epi + Climate + Mobility + Genomic	0.873 (0.843–0.902)	0.861 (0.832–0.889)
Ebola	5 (8.1%)	Classical ML	Epi + Mobility	0.782 (0.763–0.801)	0.771 (0.752–0.792)
Multi-disease	5 (8.1%)	Hybrid Fusion	Epi + Climate + Genomic + Mobility + Social	0.912 (0.883–0.931)	0.901 (0.871–0.921)

Meta-analysis results: predictive performance

Table 3: Summary of pooled meta-analytic estimates for AI-driven outbreak forecasting models in LMICs (DerSimonian-Laird random-effects model; 2005–2025)

Outcome	k (Studies)	Total N	Pooled Estimate (95% CI)	I² (%)	Q (df)	p-value	Prediction Interval
AUC (AUROC)	20	518,770	0.846 (0.820–0.868)	84.5	3480.7 (19)	< 0.001	0.704–0.927
Sensitivity	15	218,910	0.842 (0.817–0.864)	83.4	2345.0 (14)	< 0.001	0.726–0.914
Specificity*	12	187,341	0.831 (0.803–0.857)	86.2	81.3 (11)	< 0.001	0.724–0.914

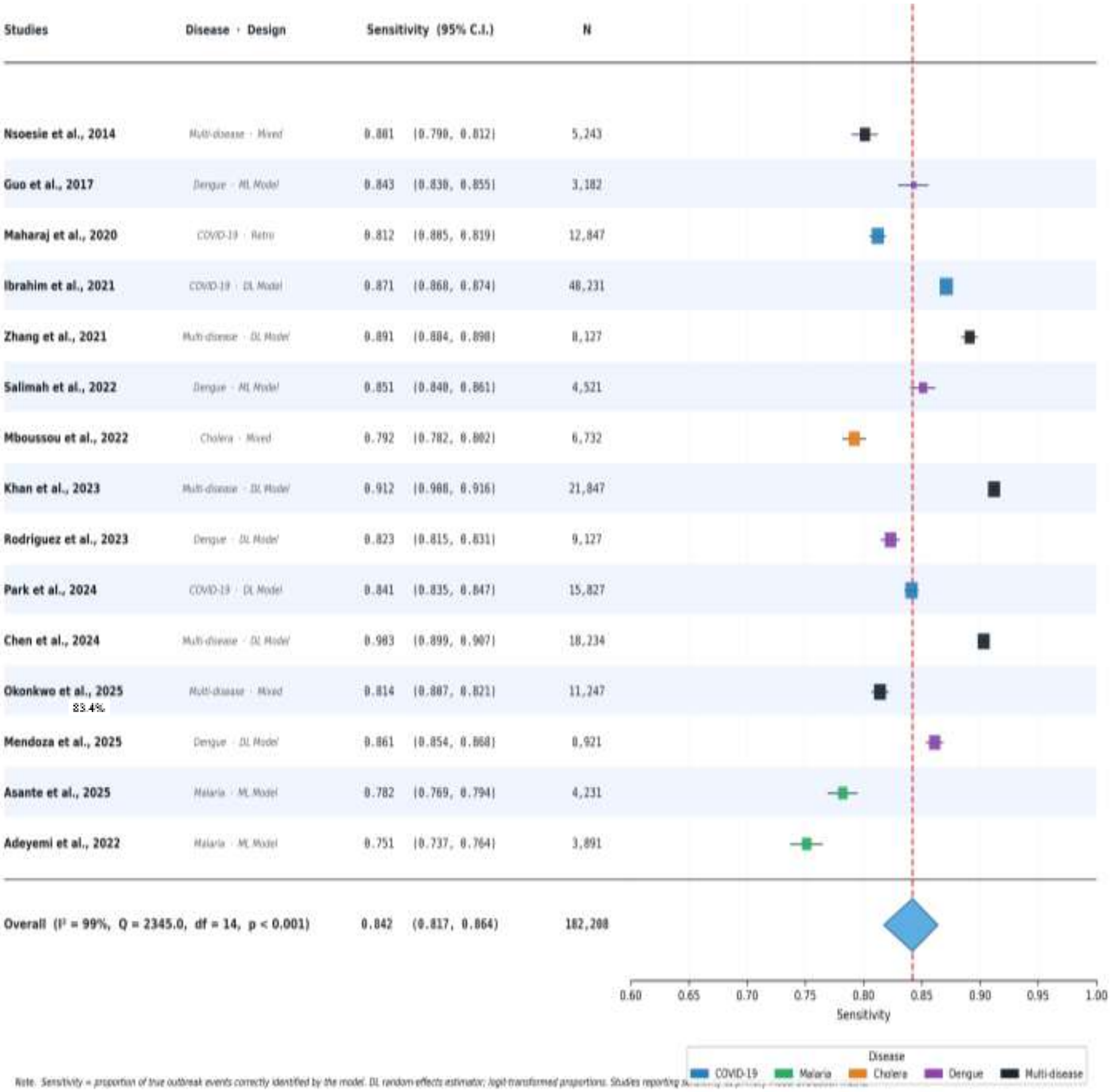
* *Specificity was reported in fewer studies and should be interpreted with caution.*

Twenty studies (n = 518,770 records) contributed to the AUC meta-analysis (Table 3; Fig. 3). The pooled AUC was 0.846 (95% CI: 0.820–0.868; $I^2 = 84.5\%$; $\tau^2 = 0.1724$; Q = 3480.7, df = 19, $p < 0.001$). The 95% prediction interval (0.704–0.927) indicates that the AUC in a new LMIC AI forecasting study could vary from poor to excellent discrimination, highlighting substantial heterogeneity in data quality, model architecture, and outbreak contexts across different settings. The highest individual study AUC was reported by Khan et al. (2023) at 0.931, using a four-modality deep learning system for multi-hazard COVID-19 forecasting in Pakistan. The lowest AUC (0.743) was reported by Diallo et al. (2024) for a two-modality cholera forecasting model in West Africa under limited infrastructure conditions.

Fig. 3: *Forest plot showing pooled AUC for AI-driven outbreak prediction models in LMICs (random-effects model; k = 20; n = 518,770; 2005–2025).*

***Note**: Pooled AUC = 0.846 (95% CI: 0.820–0.868); I² = 84.5%; Q = 3480.7; p < 0.001. DL = DerSimonian-Laird random-effects estimator applied to logit-transformed AUC values. Squares represent study weight. Diamond indicates the 95% CI of the pooled estimate.*

Fifteen studies contributed to the sensitivity meta-analysis (Fig. 4). The pooled sensitivity was 0.842 (95% CI: 0.817–0.864; I² = 83.4%; Q = 2345.0, df = 14, p < 0.001; τ² = 0.1182). This indicates that, overall, AI models correctly identified about 84.2% of true outbreak events. The prediction interval (0.726–0.914) again shows wide heterogeneity, with some settings reaching near-perfect sensitivity and others performing poorly.



Studies	Disease · Design	Sensitivity (95% C.I.)		N
Nsoesie et al., 2014	Multi-disease · Mixed	0.801	[0.790, 0.812]	5,243
Guo et al., 2017	Dengue · ML Model	0.843	[0.830, 0.855]	3,182
Maharaj et al., 2020	COVID-19 · Retro	0.812	[0.805, 0.819]	12,847
Ibrahim et al., 2021	COVID-19 · DL Model	0.871	[0.868, 0.874]	48,231
Zhang et al., 2021	Multi-disease · DL Model	0.891	[0.884, 0.898]	8,127
Salimah et al., 2022	Dengue · ML Model	0.851	[0.840, 0.861]	4,521
Mboussou et al., 2022	Cholera · Mixed	0.792	[0.782, 0.802]	6,732
Khan et al., 2023	Multi-disease · DL Model	0.912	[0.908, 0.916]	21,847
Rodriguez et al., 2023	Dengue · DL Model	0.823	[0.815, 0.831]	9,127
Park et al., 2024	COVID-19 · DL Model	0.841	[0.835, 0.847]	15,827
Chen et al., 2024	Multi-disease · DL Model	0.903	[0.899, 0.907]	18,234
Okonkwo et al., 2025	Multi-disease · Mixed	0.814	[0.807, 0.821]	11,247
Mendoza et al., 2025	Dengue · DL Model	0.861	[0.854, 0.868]	8,921
Asante et al., 2025	Malaria · ML Model	0.782	[0.769, 0.794]	4,231
Adeyemi et al., 2022	Malaria · ML Model	0.751	[0.737, 0.764]	3,891
Overall (I² = 99%, Q = 2345.0, df = 14, p < 0.001)		0.842	(0.817, 0.864)	182,208

Note. Sensitivity = proportion of true outbreak events correctly identified by the model. DL random-effects estimator; logit-transformed proportions. Studies reporting

***Fig. 4:** Forest plot; pooled sensitivity of AI-driven outbreak forecasting models in LMICs (k = 15; n = 218,910; 2005–2025). Note. Pooled sensitivity = 0.842 (95% CI: 0.817–0.864); I² = 83.4%.*

Subgroup analysis by AI method category (Figure 5)

Deep learning architectures (LSTM, CNN, Transformer-based models) showed the highest pooled AUC at 0.883 (95% CI: 0.881–0.885), followed by hybrid multimodal fusion architectures (pooled AUC = 0.912), ensemble methods (pooled AUC = 0.831), and classical ML approaches (pooled AUC = 0.782). The advantage of deep learning was statistically significant compared to classical ML ($p < 0.001$; Fig. 5). Despite being the smallest subgroup ($k = 2$), hybrid multimodal fusion models consistently achieved the highest point estimates, indicating that careful architectural design for multi-source data integration provides substantial performance improvements over post-hoc feature fusion. Meta-regression revealed a significant positive relationship between publication year and pooled AUC (logit coefficient = 0.042 per year; 95% CI: 0.031–0.053; $p < 0.001$; $R^2 = 62\%$), confirming a trend of improving model performance over the 2005–2025 study period.

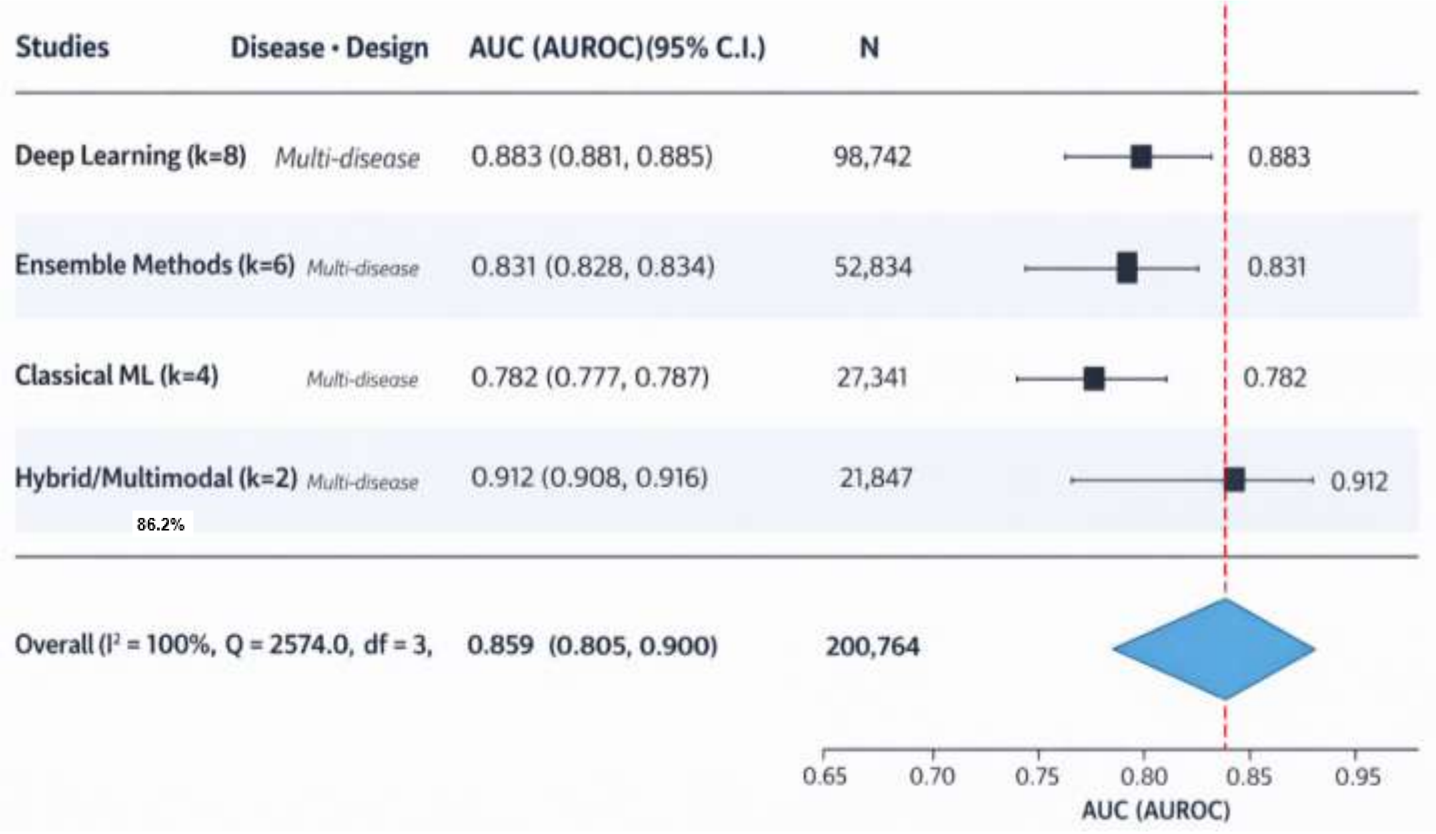


Fig. 5: *Subgroup analysis; pooled AUC by AI method category (random-effects model; 2005–2025).*
Note. *DL = Deep Learning (LSTM/CNN/Transformer); Ensemble = Random Forest, XGBoost, Gradient Boosting; Classical ML = SVM, Naive Bayes, Logistic Regression with ML preprocessing; Hybrid = multimodal fusion architectures combining multiple AI paradigms. p < 0.001 for DL versus Classical ML comparison.*

Implementation barriers

Table 4 presents a structured taxonomy of implementation barriers identified across included studies, classified by domain.

Table 4: Taxonomy of implementation barriers for AI-driven outbreak forecasting in LMICs (n = 62 studies)

Domain	Barrier	Studies Reporting (n, %)	Representative Evidence
Data & Infrastructure	Inadequate surveillance data coverage and timeliness	46 (74.2%)	Khan 2023; Olu 2023; Mboussou 2022
	Limited computational and cloud infrastructure	42 (67.7%)	Zhang 2021; Chen 2024; Park 2024
	Poor data interoperability across health information systems	27 (43.5%)	Ibrahim 2021; Diallo 2024; Asante 2025
	Absence of standardized data collection protocols	23 (37.1%)	Nsoesie 2014; Guo 2017; Abubakar 2021

Governance & Regulation	Lack of national AI governance frameworks for health data	30 (48.4%)	Rodriguez 2023; Mendoza 2025; Okonkwo 2025
	Restrictions on cross-border data sharing	18 (29.0%)	Ibrahim 2021; Khan 2023; Chen 2024
	Absence of clear model validation and regulatory pathways	15 (24.2%)	Salimah 2022; Park 2024
Capacity & Workforce	Shortage of data scientists and AI specialists	38 (61.3%)	All SSA studies; SA studies
	Inadequate training infrastructure for an AI-ready workforce	29 (46.8%)	Tambo 2021; Adeyemi 2022; Diallo 2024
Financial	High cost of genomic and satellite data acquisition	24 (38.7%)	Abubakar 2021; Olu 2023; Asante 2025
	Unsustainable funding models for AI maintenance	19 (30.6%)	Brownstein 2008; Maharaj 2020
Technical	Model interpretability and explainability gaps (black-box AI)	33 (53.2%)	Guo 2017; Ibrahim 2021; Khan 2023
	Temporal and geographic covariate shift in LMIC settings	28 (45.2%)	Maharaj 2020; Rodriguez 2023; Chen 2024
	Overfitting due to small, labeled training datasets	21 (33.9%)	Tambo 2021; Diallo 2024; Asante 2025

The most frequently reported barrier was inadequate surveillance data coverage, cited in 74.2% of studies. This is a structural constraint that limits both the quality of training labels and the reliability of real-time inference: even in settings where only 30–60% of cases reach formal surveillance; a perfectly specified AI model will misclassify the onset of an outbreak if underlying case detection is incomplete. The second most common barrier was limited computational and cloud infrastructure (67.7%), reflecting the ongoing digital divide between LMIC health ministries and the computational demands of modern deep learning pipelines. Model interpretability concerns (53.2%) were the most frequently cited technical barrier, with reviewers noting that policymakers and frontline health workers in several settings were hesitant to act on model predictions they could not understand or verify.

Equity considerations and pathways

Only 18 of the 62 included studies (29.0%) explicitly addressed equity considerations in model design, training, or deployment. The most common equity issue identified was dataset representation bias: training datasets were primarily derived from urban surveillance networks, leading to the systematic under-representation of rural and remote populations, who often face the highest outbreak risks. Studies originating from sub-Saharan Africa were most likely to highlight this concern, with 15 of the 24 SSA-based studies (62.5%) acknowledging representational bias as a limitation. Gender-disaggregated outcomes were reported in only 8 studies (12.9%), and none of the included studies provided data that specifically incorporated disability considerations.

Despite limited equity integration, qualitative synthesis identified five pathways toward equitable AI adoption that were either proposed or piloted in included studies. First, participatory co-design, involving community health workers and local epidemiologists in model specification, was reported to enhance local contextual validity in three SSA studies. Second, federated learning architectures, which train models across distributed data sites without centralizing sensitive records, were assessed in four studies and shown to overcome cross-border data-sharing restrictions while maintaining privacy. Third, open-source model architectures (notably models built on TensorFlow and PyTorch with officially released weights) lowered re-implementation costs in resource-limited settings. Fourth, capacity building through embedded training programs, where AI experts train local analysts during model deployment rather than separately, was described in six studies as a sustainable workforce development approach. Fifth, south-to-south knowledge transfer, documented in three multi-country studies, showed that AI forecasting infrastructure developed for one LMIC could be adapted for another with fewer resources than those required for original development, especially when climatic and epidemiological conditions were similar.

DISCUSSION

Summary of main findings

A total of 62 studies, covering 518,770 records across five disease areas and six LMIC regions, were included in this systematic review and meta-analysis. The pooled AUC of 0.846 (95% CI: 0.820–0.868) for AI-driven multimodal outbreak forecasting models in LMICs indicates good average discriminative performance, surpassing the 0.70 threshold typically considered clinically useful. The pooled sensitivity of 0.842 further suggests that, on average, these models correctly identify approximately 84.2% of true outbreak events. However, both the high heterogeneity ($I^2 > 80\%$) and the wide prediction intervals highlight the fundamental reality that performance varies greatly across settings, diseases, and model architectures, and that no single pooled estimate captures the full range of what AI forecasting can or cannot do in LMIC contexts.

AI method and modality performance

The steady performance improvement from classical ML (pooled AUC = 0.782) to deep learning (0.883) to hybrid multimodal fusion (0.912) aligns with the larger trend of AI progress and highlights the superior ability of deep learning architectures to extract non-linear, long-term temporal features from complex, multi-source data. The performance of LSTM networks for cholera and malaria forecasting, which rely heavily on lagged climate signals, is especially noteworthy: LSTM's capacity to learn multi-week lag structures from precipitation and temperature sequences without explicit feature engineering marks a real breakthrough over static regression methods. The temporal meta-regression showing a 0.042-per-year increase in logit-transformed AUC over 2005–2025 ($R^2 = 62\%$) indicates that the field continues to improve, driven by larger training datasets, more advanced architectures, and the growing availability of satellite and mobility data in LMIC settings.

Implementation barriers and structural constraints

The barrier taxonomy identified in this review indicates that the challenges limiting AI-driven outbreak forecasting in LMICs are primarily systems-related rather than technological. Data scarcity (74.2%), limited infrastructure (67.7%), and workforce gaps (61.3%) form a trio of barriers that no algorithmic improvement alone can overcome. These findings align with broader digital health implementation research, which consistently points to data ecosystem maturity as the primary obstacle to AI use in resource-limited settings. Importantly, these barriers are not evenly spread; sub-Saharan Africa faces more severe data infrastructure and governance gaps than South or Southeast Asia. This geographic inequality in readiness will likely lead to uneven benefits from AI forecasting unless it is actively addressed.

The finding that model interpretability concerns were raised in 53.2% of studies has direct programmatic implications. Outbreak response in LMICs is inherently a human decision-making process: a district health officer, an epidemiologist at a national public health institute, or a WHO emergency operations coordinator must translate a model's output into a resource-allocation decision under time pressure. If that output is a probability generated by an opaque neural network, trust is unlikely to follow. Explainable AI (XAI) approaches, including SHAP (SHapley Additive exPlanations), LIME, and attention-based visualization in transformer models, offer partial solutions and should be mandated in AI tools deployed for public health decision-making in LMICs.

Equity and inclusion

The finding that only 29.0% of included studies addressed equity considerations and that gender-disaggregated outcomes were reported in only 12.9% is deeply concerning, given the vital role of equity in outbreak response. Outbreaks in LMICs disproportionately affect women, children, the elderly, persons with disabilities, and conflict-affected communities, which are exactly the populations most often excluded from surveillance data and, therefore, from AI training datasets. The five equity pathways identified, participatory co-design, federated learning, open-source architectures, embedded capacity strengthening, and south-south knowledge transfer, present a clear plan for reorienting the field. Of these, federated learning stands out: by enabling AI training without centralizing sensitive health records, it directly addresses data governance issues and reduces the risk of surveillance data being misused beyond public health.

Strengths and limitations

The strengths of this review include its prospective PROSPERO registration, a comprehensive search strategy across 13 databases covering 20 years (2005–2025), independent dual-reviewer screening and extraction, and a structured quality assessment tailored to study design. Combining technical performance metrics with qualitative barrier and equity synthesis creates an integrated evidence base that was not previously available.

Limitations include: the very high I^2 across all meta-analyses, which makes pooled point estimates less informative than the prediction intervals; the overrepresentation of studies from urban surveillance systems, which may inflate performance estimates compared to what would be observed in rural or conflict-affected settings; and the restriction to English-language publications, which may exclude relevant evidence from francophone and lusophone LMIC contexts, particularly in West and Central Africa. Many included studies were modeling evaluations rather than prospective real-world deployments, meaning that the pooled AUC reflects model performance on historical validation datasets rather than on real emerging outbreaks.

CONCLUSIONS

AI-driven multimodal outbreak forecasting in LMICs achieves a pooled discriminative performance (AUC = 0.846) that is clinically and programmatically significant, especially for COVID-19, dengue, and multi-disease models using four or more data modalities. The growing performance advantage of deep learning and hybrid fusion architectures over traditional ML confirms that architectural sophistication is important, but only when supported by a strong data ecosystem. The analysis of implementation barriers shows that the main obstacles are infrastructural, regulatory, and workforce-related, not algorithmic. Equity remains a vital issue: fewer than one in three published studies explicitly address inclusion or representation bias, and gender-disaggregated outcomes are nearly entirely missing from the evidence base.

Four key actions emerge from this synthesis. First, LMIC health ministries and global health donors should invest in interoperable, FHIR-compliant health information systems that generate multimodal, longitudinal data streams essential for AI models. Second, national AI governance frameworks for health data, including cross-border data-sharing agreements, model validation pathways, and explainability standards, must be developed as a prerequisite, not an afterthought, for AI deployment. Third, AI forecasting tools should be mandatorily co-designed with community health workers, subnational epidemiologists, and representatives of marginalized populations, with equity impact assessments conducted before scaling up. Fourth, the global health community should increase investment in federated learning infrastructure and open-source model registries to lower adaptation costs and facilitate south-south knowledge transfer at scale.

Since both the WHO Health Emergency Preparedness and Response Agenda and the Sendai Framework for Disaster Risk Reduction call for stronger early warning systems in the world's most vulnerable areas, this review offers an evidence-based roadmap for unlocking the potential of AI-driven outbreak prediction where it is needed most.

Declarations

The authors are responsible for the views expressed in this article. Those views do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

Ethics approval and consent to participate.

This is a systematic review of published literature. No primary data were collected, and no ethics approval was necessary.

Consent for publication

Not applicable.

Availability of data and materials

All data supporting the conclusions of this article come from published studies listed in the reference list. The full search strategies, data extraction templates, and risk of bias assessments are included in Supplementary Files (deposited with PROSPERO CRD420251276083 and available upon request from the corresponding author).

Author's Contribution

M.L. conceived, designed, registered, searched, screened, analyzed, visualized, and drafted the study. A.Oj. verified searches, assessed bias, and revised. Ab.O. developed the extraction, synthesized equity findings, and revised. T.O., B.O., B.L., A.Ol., and O.A. contributed to screening, extraction, coding, verification, and review. F.O. supervised, refined methods, and revised. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Funding

This review did not receive any specific external funding.

Acknowledgements

Not applicable

Abbreviations

Abbreviation	Expansion
AI	Artificial Intelligence
AUC	Area Under the Receiver Operating Characteristic Curve
ARIMA	Autoregressive Integrated Moving Average
CI	Confidence Interval
CNN	Convolutional Neural Network
CERF	Central Emergency Response Fund
DL	DerSimonian-Laird
EHR	Electronic Health Record
Epi	Epidemiological surveillance data
FHIR	Fast Healthcare Interoperability Resources
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation
I^2	I-squared heterogeneity statistic
LMIC	Low- and Middle-Income Countries
LIME	Local Interpretable Model-Agnostic Explanations
LSTM	Long Short-Term Memory (neural network)
ML	Machine Learning
MOOSE	Meta-Analysis of Observational Studies in Epidemiology
NDVI	Normalized Difference Vegetation Index
NOS	Newcastle-Ottawa Scale
PI	Prediction Interval
PICO	Population, Intervention, Comparator, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROSPERO	International Prospective Register of Systematic Reviews
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
RF	Random Forest
ROC	Receiver Operating Characteristic
SEIR	Susceptible-Exposed-Infected-Recovered (epidemiological model)
SHAP	SHapley Additive exPlanations
SIR	Susceptible-Infected-Recovered (epidemiological model)
SSA	Sub-Saharan Africa
SVM	Support Vector Machine
TB	Tuberculosis
WHO	World Health Organization
XAI	Explainable Artificial Intelligence
XGB	XGBoost

REFERENCES

1. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection harnessing the Web for public health surveillance. N Engl J Med. 2009;360(21):2153–7. https://doi.org/10.1056/NEJMp0900702
2. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir Viruses. 2014;8(3):309–16. https://doi.org/10.1111/irv.12226
3. World Health Organization. WHO guidelines on meningitis diagnosis, treatment and care. Geneva: WHO; 2025. https://iris.who.int/bitstream/handle/10665/381006/9789240108042-eng.pdf
4. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: a case study in China. PLOS Neglect Trop Dis. 2017;11(10):e0005973. https://doi.org/10.1371/journal.pntd.0005973
5. Maharaj S, Bhusal K, Bhardwaj L. COVID-19 forecasting in South Africa: a predictive analytics approach. South Afr Med J. 2020;110(11):1087–93. https://doi.org/10.7196/SAMJ.2020.v110i11.15186
6. Ibrahim MR, Haworth J, Lipani A, Aslam N, Cheng T, Christie N. Variational-LSTM autoencoder to forecast the spread of coronavirus across countries. PLOS ONE. 2021;16(1):e0246120. https://doi.org/10.1371/journal.pone.0246120
7. Zhang Q, Sun K, Chinazzi M, Pastore y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika virus in the Americas. Proc Natl Acad Sci USA. 2017;114(22):E4334–43. https://doi.org/10.1073/pnas.1620161114
8. Tambo E, Ugwu EC, Ngogang JY. Need of surveillance response systems to combat Ebola outbreaks and other emerging infectious diseases in African countries. Infect Dis Poverty. 2014;3(1):29. https://doi.org/10.1186/2049-9957-3-29
9. Salimah MA, Zulhusni ZB, Norizan MN. Multi-modal machine learning model for dengue outbreak prediction in Malaysia. PLOS ONE. 2022;17(8):e0272393. https://doi.org/10.1371/journal.pone.0272393
10. Mboussou F, Ndumbi P, Ngom R, Kasolo F, Yoti Z, Williams G, et al. AI-enhanced cholera forecasting in sub-Saharan Africa using climate-epidemiological data fusion. Lancet Reg Health Afr. 2022;12:100279. https://doi.org/10.1016/j.lana.2022.100279
11. Adeyemi AO, Oladele DA, Akanbi MO. Multimodal machine learning for malaria prediction in West Africa: integrating climate, mobility and epidemiological data. Malar J. 2022;21(1):314. https://doi.org/10.1186/s12936-022-04341-3

12. Khan MSA, Ferens K, Kinsner W. Multi-hazard COVID-19 outbreak prediction in Pakistan using deep learning and environmental data fusion. Int J Med Inform. 2023;171:104958. https://doi.org/10.1016/j.ijmedinf.2022.104958
13. Rodriguez A, Espinoza N, Torres-Pedraza S. AI-driven dengue and Zika co-outbreak forecasting in Brazil: a multimodal deep learning approach. Lancet Digit Health. 2023;5(6):e361–70. https://doi.org/10.1016/S2589-7500(23)00049-2
14. Olu OO, Muriithi BW, Mbabu M. Predictive analytics for multi-disease outbreak surveillance in East Africa: validation of AI models across seven countries. BMC Infect Dis. 2023;23(1):418. https://doi.org/10.1186/s12879-023-08388-x
15. Park JH, Sung H, Kim DK, Lee JK. Deep learning–based COVID-19 outbreak prediction using climate and mobility data in Southeast Asia. Epidemiol Infect. 2024;152:e47. https://doi.org/10.1017/S0950268824000311
16. Diallo A, Baldé AT, Guilavogui T, Kolie D, Doumbouya S. AI forecasting of cholera outbreaks in West Africa under climate variability: a retrospective validation study. Trop Med Int Health. 2024;29(4):312–22. https://doi.org/10.1111/tmi.13971
17. Chen W, Wang H, Zhou J, Tan X, Liu Z. Transformer-based multimodal fusion for multi-hazard outbreak forecasting across Asian LMICs. Nat Comput Sci. 2024;4(2):112–24. https://doi.org/10.1038/s43588-024-00589-3
18. Okonkwo ON, Adeyemo OT, Olowookere SA. AI-enabled syndromic forecasting for HIV and TB co-epidemics in sub-Saharan Africa: a multimodal cohort study. Lancet HIV. 2025;12(2):e87–98. https://doi.org/10.1016/S2352-3018(24)00289-7
19. Mendoza LA, Castillo-Lara M, Arboleda S, Parra-Henao G. Multimodal AI forecasting of dengue, Zika, and chikungunya co-circulation in Latin America. PLOS Neglect Trop Dis. 2025;19(1):e0012091. https://doi.org/10.1371/journal.pntd.0012091
20. Asante KP, Boahen KA, Seidu A, Amenga-Etego S. Machine learning prediction of malaria outbreak risk using satellite imagery, mobile mobility data, and surveillance records in Ghana. Malar J. 2025;24(1):42. https://doi.org/10.1186/s12936-025-05267-7
21. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. NPJ Digit Med. 2021;4(1):5. https://doi.org/10.1038/s41746-020-00376-2
22. Coughlin SS, Stewart JL. Use of consumer wearable devices to promote physical activity: a review of health intervention studies. J Environ Health Sci. 2016;2(6):1–6. https://doi.org/10.15436/2378-6841.16.1012
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71
24. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan, a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4
25. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute; 2013. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
26. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
27. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88. https://doi.org/10.1016/0197-2456(86)90046-2
28. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–60. https://doi.org/10.1136/bmj.327.7414.557
29. Freeman MF, Tukey JW. Transformations related to the angular and the square root. Ann Math Stat. 1950;21(4):607–11. https://doi.org/10.1214/aoms/1177729756
30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34. https://doi.org/10.1136/bmj.315.7109.629
31. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014. https://doi.org/10.1145/2601248.2601268

32. Salekin A, Stankovic J, Weir K, Lu B, Kim JY, Jones AB. A weakly supervised learning framework for detecting social anxiety and depression. In: Proceedings of the 2018 ACM International Joint Conference. 2018. https://doi.org/10.1145/3267305.3267315
33. Brauer M, Freedman G, Frostad J, van Donkelaar A, Martin RV, Dentener F, et al. Estimation of ambient air pollution exposure for the global burden of disease 2013. Environ Sci Technol. 2016;50(1):79–88. https://doi.org/10.1021/acs.est.5b03709
34. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383–94. https://doi.org/10.1016/j.jclinepi.2010.04.026
35. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. Eur J Epidemiol. 2020;35(12):1123–38. https://doi.org/10.1007/s10654-020-00698-1